TESIS - KM 185401

# EXTREME GRADIENT BOOSTING UNTUK PENCARIAN PROTEIN YANG BERPENGARUH TERHADAP PRODUKSI INSULIN BERDASARKAN INTERAKSI PROTEIN-PROTEIN

MOH HAMIM ZAJULI AL FAROBY
NRP 0611 1850 010 010

DOSEN PEMBIMBING:
Prof. Dr. Mohammad Isa Irawan, M.T.
Prof. Dr. Ni Nyoman Tri P, M.Si.

PROGRAM MAGISTER
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

THESIS - KM 185401

# EXTREME GRADIENT BOOSTING TO SEARCHING INFLUENTIAL PROTEIN AGAINST INSULIN PRODUCTION BASED ON PROTEIN PROTEIN INTERACTION

MOH HAMIM ZAJULI AL FAROBY
NRP 0611 1850 010 010

SUPERVISORS:
Prof. Dr. Mohammad Isa Irawan, M.T.
Prof. Dr. Ni Nyoman Tri P, M.Si.

MASTER PROGRAM
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE AND DATA ANALYTICS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

**LEMBAR PENGESAHAN TESIS**

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar

**Magister Matematika (M.Mat)**

di Departemen Matematika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh:
**MOH. HAMIM ZAJULI AL FAROBY**
**NRP: 0611 1850 010 010**

Disetujui oleh:
Pembimbing:

1. Prof. Dr. Mohammad Isa Irawan, M.T. .....
   NIP. 19631225 198903 1 001

2. Prof. Dr. Ni Nyoman Tri Puspaningsih, M.Si. .....
   NIP. 19630615 198701 2 001

Penguji:

3. Dr. Dwi Ratna Sulistyaningrum, S.Si., M.T. .....
   NIP. 19690405 199403 2 003

4. Dr. Budi Setiyono, S.Si., M.T. .....
   NIP. 19720207 199702 1 001

5. Dr. Dieky Adzkiya, S.Si., M.Si. .....
   NIP. 19830517 200812 1 003

Kepala Departemen Matematika
Fakultas Sains dan Analitika Data

Subchan, S.Si., M.Sc., Ph.D.
NIP. 19710513 199702 1 001

# KATA PENGANTAR

Alhamdulillaahirobbil'aalamiin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan limpahan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan Tesis yang berjudul:

"EXTREME GRADIENT BOOSTING UNTUK PENCARIAN PROTEIN YANG BERPENGARUH TERHADAP PRODUKSI INSULIN BERDASARKAN INTERAKSI PROTEIN-PROTEIN"

sebagai salah satu syarat kelulusan Program Pasca Sarjana Departemen Matematika, Fakultas Sains dan Analitika Data Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Tesis ini dapat diselesaikan dengan baik berkat bantuan dan dukungan dari berbagai pihak, oleh karena itu, penulis menyampaikan ucapan terima kasih dan penghargaan kepada:

1. Kedua orang tua saya, Bapak Drs. Syahroni dan Ibu Dra. Fatimatu Zuhroh serta istri tercinta Mariatul Lutfia, S.M. dan adik-adik saya atas dukungan dan semangat yang telah diberikan.

2. Ketua Departemen Matematika ITS Bapak Subchan, S.Si., M.Sc., Ph.D. yang telah memberikan dukungan dan motivasi selama perkuliahan hingga terselesaikannya Tesis ini.

3. Kaprodi S2 Departemen Matematika Dr. Dieky Adzkiya, M.Si. yang telah memberikan arahan akademik selama penulis kuliah di Departemen Matematika FSAD-ITS.

4. Bapak Prof. Dr. Mohammad Isa Irawan, M.T. dan Ibu Prof. Dr. Ni Nyoman Tri Puspaningsih, M.Si. sebagai dosen pembimbing yang telah memberikan motivasi dan pengarahan dalam penyelesaian Tesis ini.

5. Bapak Dr. Budi Setiono, M.T., Dr. Dieky Adzkiya, M.Si. dan Ibu Dr. Dwi Ratna Sulistyaningrum, M.T. selaku dosen penguji.

6. Bapak Prof. Dr. Erna Apriliani, M.Si. sebagai dosen wali yang telah memberikan arahan akademik selama penulis kuliah di S2 Departemen Matematika FSAD-ITS.

7. Bapak dan Ibu dosen serta para staf Departemen Matematika ITS yang tidak dapat penulis sebutkan satu-persatu.

8. Teman-teman yang telah membantu dan memotivasi saya selama proses mengerjakan Tesis.

Penulis menyadari bahwa dalam penyusunan Tesis ini masih mempunyai banyak kekurangan. Kritik dan saran dari berbagai pihak yang bersifat membangun juga sangat diharapkan sebagai bahan perbaikan di masa yang akan datang.

Surabaya, 17 Agustus 2020

Penulis

# *EXTREME GRADIENT BOOSTING* UNTUK PENCARIAN PROTEIN YANG BERPENGARUH TERHADAP PRODUKSI INSULIN BERDASARKAN INTERAKSI PROTEIN-PROTEIN


| | | |
|---|---|---|
| Nama Mahasiswa | : | Moh Hamim Zajuli Al Faroby |
| NRP | : | 0611 1850 010 010 |
| Pembimbing | : | 1. Prof. Dr. Mohammad Isa Irawan, M.T. |
| | | 2. Prof. Dr. Ni Nyoman Tri P, M.Si. |



**Abstrak**

*Peran komputasi dalam bioinformatika sangat signifikan, analisis kebijakan. Protein sintesis insulin memiliki fungsi molekul yang sama yang mendukung peran yang berbeda tetapi saling mendukung, dan fungsi molekul dalam protein dapat mendukung dalam gen ontologi. Fungsi molekuler di GO memiliki bentuk grafik berarah, lebih dikenal sebagai Directed Acyclic Graph, dianalisis untuk mendapatkan bobot pada setiap node dengan menggunakan metode sentralitas pada grafik berarah, metode ini digunakan untuk mengekstraksi fitur dari setiap protein. Ada 11 fitur yang diambil dari hasil ekstraksi yang kemudian membentuk dataset. Dataset yang telah dibangun digunakan untuk klasifikasi dengan menggunakan metode Extreme Gradient Boosting, yang diklaim memiliki algoritma model yang lebih baik ketika dataset memiliki kompleksitas besar. Hasil klasifikasi menunjukkan akurasi dari model dengan dataset yang dibangun menggunakan metode Betweenness Centrality lebih baik dari pada dataset yang dibangun dari dua model metode centrality yang lain, yakni sebesar 74,56%. Hasil model prediksi dari klasifikasi dianalisis untuk membangun interaksi protein sintesis pada protein insulin, yang menghasilkan sebanyak 9 protein berpengaruh besar dan 18 protein berpengaruh sedang terhadap sintesis insulin. Protein yang ditemukan dalam model jaringan yang memiliki pengaruh besar, yakni IGF1R, INSR dan IGF1. Hasil tersebut sejalan dengan analisis pada Biomedical Text Mining pada aplikasi STRING-DB yang menunjukkan ke tiga protein tersebut berpengaruh besar terhadap Insulin.*



***Kata-kunci:*** *Extreme Gradient Boosting, Interaksi Protein-Protein, Insulin, Metode Sentralitas, dan Ontologi Gen*

# EXTREME GRADIENT BOOSTING TO SEARCHING INFLUENTIAL PROTEIN AGAINST INSULIN PRODUCTION BASED ON PROTEIN PROTEIN INTERACTION


| | | |
|---|---|---|
| Name | : | Moh Hamim Zajuli Al Faroby |
| NRP | : | 0611 1850 010 010 |
| Supervisors | : | 1. Prof. Dr. Mohammad Isa Irawan, M.T. |
| | | 2. Prof. Dr. Ni Nyoman Tri P, M.Si. |



## Abstract

*The role of computing in bioinformatics is very significant, policy analysis. Protein synthesis of insulin has the same molecular functions that support different but mutually supportive roles, and the function of molecules in proteins can support in ontology genes. The molecular function in GO has a form of directed graph, better known as Directed Acyclic Graph, analyzed to get weight at each node by using the centrality method on directed graph, this method is used to extract features from each protein. There are 11 features taken from the extraction which then form a dataset. The dataset was built for classification using the Extreme Gradient Boosting method, which is claimed to have a better algorithmic model when the datasets have large complexity. The classification results show that the accuracy of the model with the dataset that was built using the Betweenness Centrality method is better than the dataset that was built from the other two models of the centrality method, which is 74.56%. The results of the prediction model from the classification are analyzed to build the interaction of protein synthesis on insulin protein, which produces as many as 9 proteins having a large effect and 18 proteins having a moderate effect on insulin synthesis. Proteins found in tissue models that have a major influence, namely IGF1R, INSR and IGF1. These results are in line with the analysis of Biomedical Text Mining in STRING-DB applications which shows the three proteins have a major effect on Insulin.*



***Key-words:*** *Extreme Gradient Boosting, Protein-Protein Interactions, Insulin, Centrality Method, and Gene Ontology*

# DAFTAR ISI

# BAB 1
# PENDAHULUAN

Pada Bab ini, menjelaskan mengenai hal-hal yang mendasari ide dari topik penelitian. Selain itu, dirumuskan juga masalah-masalah yang berkaitan tentang topik penelitian beserta batasan-batasan masalahnya. Setiap penelitian memiliki tujuan dan juga bermaksud memberi manfaat kepada masyarakat luas baik itu berupa kontibusi keilmuan maupun manfaat langsung. Sehingga, pada bab ini juga dijelaskan maksud tujuan serta kebermanfaatan penelitian ini untuk kemudian hari.

## 1.1 Latar Belakang

Penyakit yang disebabkan oleh kelainan genetika menjadi suatu permasalahan serius di dunia kesehatan dari pada penyakit yang bukan dari kelainan genetika. Kelainan genetika sendiri disebabkan dari mutasi kode DNA (*deoxyribonucleic acid*) ketika transkripsi protein, asam amino yang tersusun tidak sesuai dengan fungsi yang diharapkan. Diabetes militus merupakan salah satu dari penyakit yang disebabkan oleh kelainan genetika. Diabetes militus terjadi ketika pankreas tidak cukup dalam memproduksi insulin atau tubuh pasien tidak mampu menggunakan fungsinya secara efektif. Insulin sendiri merupakan salah satu dari protein domain dalam organisme yang fungsinya mengubah glukosa menjadi energi. Diabetes militus terdiri dari 3 tipe, diabetes tipe-1, diabetes tipe-2 dan diabetes gestasional.(Defronzo, 1997)

Data dari *World Health Organization* (WHO) menyatakan bahwasannya kejadian penyakit tidak menular lebih besar dibandingkan penyakit menular, bahkan penyebab kematian yang diakibatkan penyakit tidak menular menjadi nomor satu di dunia yaitu 63,6%[1]. Secara global, diperkirakan 422 juta orang dewasa hidup dengan diabetes pada tahun 2014. Selama beberapa dekade terakhir, prevalensi diabetes meningkat lebih cepat di negara berpenghasilan rendah dan menengah daripada di negara berpenghasilan tinggi. Diabetes menyebabkan 1,5[2] juta kematian pada tahun 2012. Kandungan gula yang melebihi batas maksimum dari beban tubuh mengakibatkan bertambah 2,2 juta kematian. Empat puluh tiga persen (43%) dari 3,7[3] juta

[1]Departemen Kesehatan RI, (2018), *Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia: Hari Diabetes Sedunia*, Kementerian Kesehatan RI, Jakarta: Pusat Data dan Informasi, http://www.depkes.go.id/download.php?file=download/pusdatin/infodatin/hari-diabetes-sedunia-2018.pdf,[Sitasi: 10 Mei 2019].

[2]World Health Organization., (2016). *World Health Organization: Global Report on Diabetes.*, France: World Health Organization., http://www.who.int/diabetes/global-report/en/,[Sitasi: 14 Mei 2019].

[3]ibid

kematian yang disebabkan diabetes militus terjadi sebelum usia 70 tahun. Persentase kematian pada usia sebelum 70 tahun lebih tinggi di negara-negara berkembang dari pada di negara-negara maju.

Terapi kesehatan terhadap penderita diabetes militus melibatkan medis dengan menginjeksi insulin dari luar tubuh kedalam tubuh penderita. Hal ini tergolong menjadi permasalahan tersendiri, karena sumber insulin yang ada di departemen kesehatan tergolong mahal dan susah didapatkan. Selain itu, pasien juga merasa tidak nyaman karena harus datang ke klinik atau dokter yang mengakibatkan penurunan kepatuhan, terutama seseorang yang sedang bekerja. Setiap sesi terapi mingguan sekitar 5 jam dengan membutuhkan dokter profesional dibidang diabetes militus.(Shu dong, dkk, 2019) Maka dari pada itu dibutuhkan obat yang memudahkan pasien diabetes militus dalam memenuhi kebutuhan insulin di tubuhnya, mengingat pertumbuhan diabetes yang cepat dan juga kritis diseluruh dunia, salah satu alternatif merancang model global yang dapat mengidentifikasi protein baru yang terkait dengan penyakit diabetes militus. Obat-obatan baru yang dapat didisain dari hasil penelitian ini memberikan manfaat yang ekstra terhadap pasien terutama yang ketergantngan terhadap injeksi insulin, sehingga lebih ekonomis dan effisien.

Mengidentifikasi interaksi protein-protein merupakan salah satu kunci untuk mendapatkan protein baru yang terkait dengan suatu penyakit.(Kann, 2007) Terutama pada penyakit diabetes militus, identifikasi interaksi protein-protein dilakukan pada protein domain peyebab penyakit tersebut yaitu insulin. Pada penelitian yang dilakukan Bader dan Hogue, interaksi protein-protein berkorelasi dengan sifat fungsi protein dan jaringan interaksi protein yang sering digunakan untuk menemukan peran biologis potensi dari suatu protein dengan fungsi yang tidak diketahui.(Bader, dkk, 2003) Teknik komputasi maupun eksperimental telah digunakan untuk mengidentifikasi interaksi antar protein tersebut. Interaksi pada protein cenderung pada interaksi fisik yang terjadi dengan proses seluler yang sama dan mutasi pada gen. Interaksi secara fisik pada protein terjadi secara singkat untuk membentuk suatu ikatan sementara yang mengekspresikan fungsi biologis mereka di dalam sel.(Ideker dan Sharan, 2008) Akhir-akhir ini, teknik komputasi untuk mengidentifikasi interaksi protein-protein dipergiatkan. Diperkirakan protein domain menjadi sentral dari jaringan interaksi antar protein, sehingga beberapa metode centrality seperti *Betweenness Centrality*(BC), *Closeness Centrality*(CC), *Degree Centrality*(CC), *Eigen Vector Centrality*(EC), *Information Centrality*(IC), *Edge Clustering Coefficient Centrality*(NC) dan *Subgraph Centrality*(SC).(Zhong, dkk, 2018) Akan tetapi beberapa penelitian mengenalkan karakter biologi yang lain pada masalah prediksi interaksi protein-protein, seperti ekspresi gen, protein GO (*gene orthologous*), lokasi dari subseluler, dan property dari protein kompleks.Pada tesis ini dalam mengolah data protein menggunakan karakteristik pada GO yang meliputi *biological process, cellular component*, dan *molecular function*. Tiga komponen pada GO tersebut menjadi dasar bahwa antar protein memiliki interaksi

dengan mempunya kesamaan fungsional. Untuk menskor nilai hirarki pada setiap karakteristik GO, beberapa metode *centrality* digunakan untuk mempersiapkan dataset.

Metode komputasi menjadi salah satu alternatif untuk mengidentifikasi interaksi antar protein. Beberapa metode komputasi misalnya berdasarkan *data sequence*, melatih komputer pada basis data yang besar dan juga gabungan dari keduanya telah digunakan dalam memprediksi interaksi protein-protein.(Karlin, dkk, 2012) Metode yang popular sekarang ini untuk mengolah basis data yang besar adalah metode *machine learning*. Metode seperti Bayesian, *Probabilistic Decision Tree, Logistic Regression*, dan *Support Vector Machine* telah dikenalkan untuk menyelesaikan permasalahan memprediksi interaksi protein-protein dengan menggunakan properti dari protein untuk mengklasifikasikan data.(Pizzut, dkk, 2016) Pengklasifikasian menggunakan *Support Vector Machine* terhadap data interaksi protein-protein pada protein yang mempengaruhi penyakit diabetes militus menghasilkan akurasi sebesar 73,6%, dengan menggunakan sebanyak 2653 data latih.(Ranu Vyas, dkk, 2016) Sehingga, metode *machine learning* tergolong efektif untuk menyelesaikan permasalahan identifikasi pada dataset interaksi protein-protein.

Salah satu metode *Machine Learning* yang baru-baru ini dikenalkan pada tahun 2015 oleh Tianqi Chen yaitu *Extreme Gradient Boosting*. Model ini dapat mengukur lebih akurat dengan peningkatan dari versi *Gradient Boosted Machines* (GBM).(Chen, 2014) XGBoost digunakan dalam pengklasifikasian dan pembelajaran mesin untuk memberi hasil akurasi yang lebih baik dari pada metode tradisional yang sejenis dan metode modern kontemporer.(Chen dan Charlos, 2016) Metode XGboost memiliki akurasi lebih baik dibandingkan dengan *Support Vector Machine, Naive Bayes, Random Tree, Random Forest*,dan *Radial Basis Function Network* dalam mendeteksi protein essensial dengan dataset interaksi protein-protein.(Zhong, dkk, 2018) Sehingga hipotesis awal, metode *Extreme Gradient Boosting* memiliki performa baik untuk mendeteksi interaksi protein pada insulin yang terjadi pada penderita diabetes militus tipe-2.

Pada penelitian ini, digunakan metode *Extreme Gradient Boosting* sebagai pengklasifikasi data protein yang berinteraksi dengan insulin sebagai penyebab terjadi penyakit diabetes militus. Hasil dari skor setelah klasifikasi akan dibandingkan dengan hasil dari *Biomedicine Text Mining*. Sehingga nantinya dapat diketahui protein-protein yang berpengaruh terhadap kinerja produksi insulin sebagai protein inti dari penyebab terjadinya penyakit diabetes militus tipe-2. Serta harapannya ditemukan protein baru yang berpengaruh terhadap insulin sebagai calon obat atau terapi bagi penderita diabetes militus tipe-2.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang topik permasalahan pada penelitian ini, dapat dirumuskan suatu permasalahan yang menjadi fokus utama dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membangun dataset dengan mengkarakteristikkan interaksi

yang terjadi antar protein ?

2. Bagaimana mengklasifikasikan data interaksi protein-protein dengan menggunakan XGBoost?

3. Bagaimana menganalisa model XGBoost untuk memperoleh skor sebagai bobot interaksi?

## 1.3 Batasan Masalah

Pada penelitian ini dibuat batasan-batasan dalam meneliti, agar penelitian sesuai dengan yang diinginkan. Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Data penelitian merupakan data sekunder dari Protein Data Bank, Uniprot, EMBL dan NCBI.

2. Data ontologi gen yang digunakan adalah *molecular function*.

3. Hasil penelitian hingga terbentuknya jaringan interaksi rotein yang berpengaruh terhadap insulin.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang disebutkan diatas, tujuan dari penelitian ini adalah mengklasifikasikan protein yang mensintesis insulin dan data protein yang tidak mensintesis insulin menggunakan *Extreme Gradient Boosting*. Dataset interaksi protein dibangun berdasarkan model *Directed Acyclic Graph* pada ontologi gen. Model yang dihasilkan dari XGBoost dianalisa untuk mendapatkan skor interaksi antar protein yang digunakan untuk membangun jaringan interaksi protein-protein.

## 1.5 Manfaat Penelitian

Penelitian-penelitian yang dilakukan pastinya mempunya manfaat atau kontribusi keilmuan untuk masa depan. Pada penelitian ini, dalam menganalisa interaksi protein-protein pada insulin diharapkan dapat memberi manfaat yaitu:

1. Kontribusi peran matematika dalam bidang kesehatan khususnya penemuan kandidat obat-obatan baru.

2. Hasil dari analisis menggunaankan metode *Extreme Gradient Boosting* pada dataset interaksi protein-protein sebagai kontribusi keilmuan perkembangan metode *machine learning*.

3. Dapat mengetahui macam-macam protein yang mempengaruhi kinerja ataupun produksi insulin dalam tubuh.

4. Hasil dari penelitian ini sebagai acuan dasar pembuatan obat atau terapi kesehatan terhadap penderita diabetes militus tipe-2 yang dapat di manfaatkan bagi para dokter ataupun farmasis.

# BAB 2
# KAJIAN PUSTAKA DAN DASAR TEORI

Pada bab ini, dipaparkan mengenai penelitian-penelitian terdahulu yang berkaitan dengan topik penelitian. Selain itu, ditunjukkan beberapa teori-teori yang menjadi landasan penyelesaian masalah yang dikemukakan pada penelitian ini.

## 2.1 Penelitian-Penelitian Terkait

Pada sub bab ini, dipaparkan mengenai penelitian-penelitian yang berkaitan dengan topik penelitian.

(a). Pada penelitian yang dilakukan oleh Ranu Vyas dkk (2016) dalam makalah yang berjudul "*Building and Analysis of Protein-Protein Interaction Related to Diabetes Millitus Using Support Verctor Machine, Biomedical Text Mining and Network Analysis*" (Ranu Vyas, dkk, 2016), mereka meneliti dengan mengklasifikasikan data interaksi protein-protein pada penyakit diabetes militus dengan menggunakan *Support Vector Machine* yang menghasilkan akurasi pada 2653 data latih sebesar 73,6%. Sedangkan dengan menganalisa pada *biomedical text mining* menghasilkan akurasi sebesar 78.2%. Hal ini menunjukkan, identifikasi interaksi protein-protein dengan menggunakan metode *Machine Learning* menghasilkan akurasi yang bagus.

(b). Tahun 2016, pada artikel ilmiah yang dituliskan Tianqi Chen dan Carlos Guestrin yang berjudul "*XGBoost: A Scalable Tree Boosting System*"(Chen dan Charlos, 2016). Membuat model pohon *boosting* baru dengan regulasi yang berbeda dari *Gradient Tree Boosting* yang ada. Model tersebut berjalan 10 kali lebih cepat dengan pembagian pararel memori pada komputer.

(c). Penelitian Aditya Gupta, dkk., pada tahun 2016 yang berjudul "*Verifying the Value and Veracity of eXtreme Gradient Boosted Decision Trees on a Variety of datasets*"(Gupta, A., dkk, 2016). Membandingkan metode XGBoost dengan metode *machine learning* lain seperti SVM, KNN, *Random Forest, Logistic Regresion* dan lain-lain. Berdasarkan beberapa jenis dataset performa rata-rata XGBoost berada diatas metode *machine learning* yang lain.

(d). Pada penilitan yang dilakukan oleh Jiancheng Zhong dkk (2018) dalam makalah yang berjudul "*XGBFEMF: An XGBoost-Based Framework for Essential Protein Prediction*" (Zhong, dkk, 2018), Telah dilakukan analisa metode *centrality* dan juga perbandingan metode antara

XGBoots dengan metode *Machine Learning* yang lain, seperti *Support Vector Machine, Naïve Bayes, Random Tree, Random Forest* dll. Analisa yang didapatkan dengan dataset interaksi protein-protein dengan mengklasifikasikan protein essensial dan yang bukan essensial, metode XGBoost memiliki performa lebih bagus dari pada metode yang lain, pada hal ini dalam hal akurasi. Selain itu, XGBoost juga dapat digunakan sebagai pensekoran karakteristik dari ontologi protein. Pada penelitian tesis ini menggunakan data mentah dari GO protein Insulin yang dianalisis dengan metode *centrality*, sedangkan pada penelitian yang dilakukan zhong data berupa protein esensial yang berupak network interaksi dan sudah diinteraksikan, fitur pada dateset tesis ini juga berbeda. Pada penelitian tesis ini juga, model yang dihasilkan dari XGBoost dignakan untuk membangun interaksi dari protein dengan menerapkan kedekatan jarak dari rasio *odds* dari masing-masing protein Insulin.

(e). Tahun 2019, M. Syaifudin Usman, dkk., menuliskan artikel yang berjudul "*Identification of Significant Protein Associated with Diabetes Mellitus Using Network Analysis of Protein-Protein Interaction*"(Usman, M.S., dkk, 2019). Analisa antar protein dapat direpresentasikan dengan analisa jaringan menggunakan graf. Skor kedekatan antar protein yang berpengaruh terhadap diabetes militus dapat dijadikan bobot sebagai interaksi yang terjadi pada protein.

Penelitian-penelitian yang terkait diatas menunjukkan keterhubungan dengan tesis ini. Metode *Extreme Gradient Boosting* yang digunakan merupakan model Chen (Chen dan Charlos, 2016), yang menunjukkan hasil akurasi yang bagus digunakan pada dataset bioinformasi, khususnya PPI. Hipotesis awal, *Tree Boosting* yang dihasilkan saat melatih model dengan dataset dapat digunakan untuk penyekoran kesamaan fungsi antar protein. Sehingga untuk mendapatkan informasi skor kedekatan protein satu dengan yang lain dibutuhkan analisa pada pohon keputusannya. Dataset yang digunakan merupakan karakteristik ontologi fungsi molekul setiap gen yang mensintesis insulin dengan pemberian bobot menggunakan metode *centrality* seperti pada penelitian yang dilakukan Jianchen Zhong, dkk (Zhong, dkk, 2018). Analisis terhadap model pohon *boosting* untuk mendapatkan skor kesamaan antar protein digunakan untuk membangun jaringan antar protein.

## 2.2 Extreme Gradient Boosting

*Extreme Gradient Boosting* pertama kali diusulkan oleh Tianqi Chen pada tahun 2015. Beberapa tahun ini model XGBoots diterapkan secara luas disemua jenis bidang *data mining*. XGBoost merupakan pohon regresi yang memiliki aturan keputusan sama dengan pohon keputusan (*decision tree*). Dalam pohon regresi, simpul-simpul mewakili nilai-nilai untuk uji atribut dan simpul daun skor mewakili keputusan. XGBoost merupakan sebuah *scalable machine learning* untuk *boosting tree*. Faktor yang menjadi kelebihan XGBoost adalah skalabilitasnya di semua skenario. Sistem berjalan

sepuluh kali lebih cepat dari pada soluli populer yang ada pada suatu mesin dan dalam skala miliaran contoh dalam pengaturan memori yang terbatas. Skalabilitas XGBoost disebabkan oleh beberapa sistem penting dan optimasi algoritmik. Inovasi-inovasi ini meliputi: algoritma pembelajaran pohon baru untuk menangani data yang jarang, prosedur sketsa kuantitatif berbobot yang dibenarkan secara teoretis memungkinkan penanganan bobot sampel dalam pembelajaran pohon perkiraan. Komputasi paralel dan terdistribusi membuat pembelajaran lebih cepat yang memungkinkan eksplorasi model yang lebih cepat.

### 2.2.1 Fungsi Objektif

Diberikan dataset dengan $n$ sampel dan $m$ *fetures*, $\varphi = \{(x_i, y_i)\}(|\varphi| = n, x_i \in \mathbb{R}^m)$, model ansembel pohon menggunakan fungsi aditif $K$ untuk memprediksi *output*,

$$\hat{y}_i = \phi(x_i) = \sum_{i=1}^{K} f_k(x_i), f_k \in \mathcal{F} \tag{2.1}$$

dimana $\mathcal{F} = \{f(x) = w_q(x)\}$ untuk $q : \mathbb{R}^m \to T, w \in \mathbb{R}^T$ yang merupakan ruang dari pohon regresi. $q$ merupakan representasi dari struktur setiap pohon yang memetakan dataset ke index daun yang sesuai. $T$ adalah jumlah daun pada pohon. Setiap $f_k$ berkorespondensi dengan struktur tree yang independent $q$ dan bobot daun $w$. Seperti *decision tree*, setiap pohon regresi mempunyai skor yang kontinu pada setiap daunnya, $w_i$ merepresentasikan skor pada daun ke-$i$. Contoh, kita gunakan aturan *decision tree* pada pohon (pada $q$) untuk mengklasifikasikannya ke daun dan menghitung prediksi akhir dengan menjumlahkan keatas skor pada daun yang bersesuaian (pada $w$). Fungsi pembelajaran dari model yang digunakan yakni dengan meminimalkan fungsi objektif;(Chen dan Charlos, 2016)

$$\mathcal{L}(x) = \sum_{i=1} l(y_i, \hat{y}_i) + \sum_{k=1} \Omega(f_k) \tag{2.2}$$

dimana,

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \left\| w \right\|^2$$

$l$ merupakan fungsi kerugian yang dapat dibedakan antara prediksi $\hat{y}_i$ dan target $y_i$ atau bisa dikatakan fungsi tersebut harus diferensiabel, sedangkan $\Omega$ adalah resiko dari compleksitas model. Fungsi kerugian yang digunakan merupakan fungsi kerugian yang digunakan pada *logistic regression* atau biasa disebut $log(likelihood)$,

$$l(y_i, p) = y_i log(p) + (1 - y_i) log(1 - p) \tag{2.3}$$

dimana $p$ merupakan probabilitas yang nilainnya $\frac{e^{log(odds)}}{1+e^{log(odds)}}$, pada *logistik regresion* prediksi paling bagus yakni memaksimalkan persamaan $log(likelihood)$ tersebut. Sehingga untuk merepresentasikan fungsi kerugian

yang paling minimum adalah yang paling bagus yakni nilai paling bagus, bisa menggunakan,

$$l(y_i, p) = 1 - log(likelihood)$$

atau

$$l(y_i, p) = -[y_i log(p) + (1 - y_i) log(1 - p)]$$

persamaan ini yang nantinya akan digunakan untuk memprediksi $log(odds) = log\left(\frac{p}{1-p}\right)$ yakni dengan mentransformasikan probability $p$ terhadap $log(odds)$,

$$\begin{aligned}
l(y_i, p) &= -[y_i log(p) + (1 - y_i) log(1 - p)] \\
&= -y_i log(p) - log(1 - p) + y_i log(p - 1) \\
&= -y_i \left(log(p) - log(1 - p)\right) - log(1 - p) \\
&= -y_i.log(odds) - log(1 - p) \\
&= -y_i.log(odds) - log\left(1 - \frac{e^{log(odds)}}{1 + e^{log(odds)}}\right) \\
&= -y_i.log(odds) - log\left(\frac{1 + e^{log(odds)}}{1 + e^{log(odds)}} - \frac{e^{log(odds)}}{1 + e^{log(odds)}}\right) \\
&= -y_i.log(odds) - log\left(\frac{1}{1 + e^{log(odds)}}\right) \\
&= -y_i.log(odds) - \left(log(1) - log(1 + e^{log(odds)})\right) \\
&= -y_i.log(odds) + log\left(1 + e^{log(odds)}\right)
\end{aligned}$$

karena fungsi kerugian harus diferensiabel, lalu akan ditunjukkan bahwa fungsi kerugian diatas diferensiabel.

$$\frac{d}{d\left(log(odds)\right)} = -y_i log(odds) + log\left(1 + e^{log(odds)}\right) = -y_i + \frac{e^{log(odds)}}{1 + e^{log(odds)}} = -y_i + p$$

terbukti bahwa fungsi kerugian diferensiabel karena dapat memprediksi $log(odds)$ dan juga fungsi yang dapat memprediksi $p$.

### 2.2.2 Gradient Tree Boosting

Ensambel tree dari persamaan (2.2) mengandung fungsi sebagai parameter dan tidak dapat dioptimasikan dengan metode optimasi tradisional pada ruang Euclid. Sebagai gantinya, model tersebut di latih secara aditif. Diberikan $\hat{y}_i^{(t)}$ menjadi prediksi kejadian ke-$i$ pada iterasi ke-$t$, maka dibutuhkan $f_t$ untuk meminimalkan objektif berikut,

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i)) + \Omega(f_t)$$

Berarti $f_t$ ditambahkan secara *greedy* untuk meningkatkan model pada persamaan (2.2). Pendekatan perluasan Taylor orde kedua dapat digunakan untuk mengoptimasi dengan cepat objektif diatas.(Friedman, J, 2002)

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n} (l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t)$$

dimana, $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ yang merupakan orde pertama, turunan pertama dari suatu fungsi dikenal dengan *Gradient*, sehingga pada XGBoost menggunakan $g$ sebagai representasi turunan pertama (Taylor orde 1) dan $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}$ merupakan orde kedua dari statistika gradien pada fungsi kerugian yang dikenal dengan *Hessian*, pada XGBoost direpresentasikan dengan $h$ sebagai turunan keduanya. Dengan mengekspansi persamaan diatas didapatkan,

$$\begin{aligned}
\tilde{\mathcal{L}^{(t)}} = l\left(y_1, \hat{y}_1^0\right) + g_1 f_t(x_1) + \frac{1}{2} h_1 f_t^2(x_1) \\
+ l\left(y_2, \hat{y}_2^0\right) + g_2 f_t(x_2) + \frac{1}{2} h_2 f_t^2(x_2) \\
+ ... \\
+ l\left(y_n, \hat{y}_n^0\right) + g_n f_t(x_n) + \frac{1}{2} h_n f_t^2(x_n) + \Omega(f_t)
\end{aligned}$$

dari hasil ekspansi diatas, kondisi konstan dapat dihilangkan karena kondisi tersebut tidak mengefek pada nilai optimum pada output $f_t(x_i)$,sehingga didapatkan fungsi objektif sederhana pada step $t$,

$$\tilde{\mathcal{L}^{(t)}} = \sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t) \tag{2.4}$$

Didefenisikan $I_j = \{i | q(x_i) = j\}$ sebagai perumpamaan dari himpunan daun $j$. Persamaan 2.4 dapat dituliskan dengan memperluas $\Omega$ menjadi,

$$\tilde{\mathcal{L}^{(t)}} = \sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

$$= \sum_{j=1}^{T} \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \tag{2.5}$$

mencari nilai optimal dari suatu fungsi dapat menggunakan $f'(x) = 0$, sehingga didapatkan

$$\frac{d}{dw_j} \sum_{j=1}^{T} \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T = 0$$

$$\sum_{i \in I_j} g_i + \left( \sum_{i \in I_j} h_i + \lambda \right) w_j = 0$$

sehingga dari penyelesaian diatas untuk memastikan struktur $q(x)$, dapat dengan menghitung bobot optimal $w_j^*$ dari daun $j$ oleh,

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{2.6}$$

dan menghitung nilai korespondensi optimal oleh,

$$\tilde{\mathcal{L}^{(t)}}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{2.7}$$

Persamaan 2.7 dapat digunakan sebagai fungsi penyekoran untuk mengukur kualitas dari struktur pohon $q$. Karena tidak memungkinkan untuk menyebutkan semua kemungkinan struktur pohon $q$. Algoritma *greedy* berawal dari daun tunggal dan secara iteratif menambahkan cabang ke pohon yang digunakan sebagai penggantinya. Asumsikan $I_L$ dan $I_R$ adalah himpunan sampel untuk sisi kiri dan kanan setelah dipisah. Diberikan $I = I_L \cup I_R$, lalu persamaan reduksi kerugian setelah pemisahan adalah,

$$\mathcal{L}_{split} = \frac{1}{2} \left[ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] \tag{2.8}$$

Persamaan yang digunakan pada fungsi kerugian yakni *loglikelihood* dan nilai dari prediksi $p$ sama dengan nilai $log(odds)$, karena persamaan $w_j^*$ menggunakan turunan pertama dan turunan kedua dari fungsi kerugian, maka didapatkan turunan pertama dari $l(y_i, p_i)$,

$$l(y_i, log(odds)) = -y_i log(odds) + log(1 + e^{log(odds)})$$

$$\frac{d}{dlog(odds)} l(y_i, log(odds)) = -y_i + \frac{e^{log(odds)}}{1 + e^{log(odds)}} = -(y_i - p_i)$$

dan sebagai turunan keduanya adalah,

$$\frac{d^2}{dlog(odds)^2} l(y_i, log(odds)) = \frac{e^{log(odds)}}{1 + e^{log(odds)}} \times \frac{1}{1 + e^{log(odds)}} = p_i(1 - p_i)$$

sehingga didapatkan prediksi skornya adalah,

$$w_j^* = -\frac{\sum_{i \in I_j} -(y_i - p_i)}{\sum_{i \in I_j} p_i(1 - p_i) + \lambda} \tag{2.9}$$

$$= \frac{\sum_{i \in I_j} (y_i - p_i)}{\sum_{i \in I_j} p_i(1 - p_i) + \lambda}$$

lalu persamaan korespondensi optimal dapat dituliskan sebagai,

$$\tilde{\mathcal{L}^{(t)}}(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum_{i\in I_j} -(y_i - p_i)^2}{\sum_{i\in I_j} p_i(1-p_i)+\lambda} + \gamma T$$

nilai $\frac{1}{2}$ dapat diabaikan, dan juga apabila nilai $\gamma = 0$ persamaan dari korespondensi optimal menjadi,

$$\tilde{\mathcal{L}^{(t)}}(q) = \frac{\sum_{i\in I_j}(y_i - p_i)^2}{\sum_{i\in I_j} p_i(1-p_i)+\lambda} \tag{2.10}$$

dari persamaan 2.9 dan 2.10, model dari prediksi XGBoost dibangun dengan menggunakan *boosting* pada pohon keputusan.

### 2.2.3 Algoritma XGBoost untuk Klasifikasi

Untuk lebih memahami algoritma XGBoost, disajikan langkah-langkah pembuatan model sebagai berikut. Misalkan diberikan data sebagai berikut seperti pada Tabel 2.1, dari data diatas akan diklasifikasikan orang yang

Tabel 2.1: Contoh sampel dataset

| No. | Umur | Warna Favorit | Suka Popcorn | Suka Film |
|---|---|---|---|---|
| 1 | 12 | Biru | Suka | Ya |
| 2 | 87 | Hijau | Suka | Ya |
| 3 | 44 | Biru | Tidak | Tidak |
| 4 | 19 | Merah | Suka | Tidak |
| 5 | 32 | Hijau | Tidak | Ya |
| 6 | 14 | Biru | Suka | ya |

suka menonton film dan yang tidak suka menonton film. Mula-mula dihitung terlebih dahulu probabilitas data sebagai pemisah kelas, dengan menghitung $log(odds)$, $odds$ adalah perbandingan nilai pada kelas positif dengan kelas negatif.

$$p = log(odds) = log(\frac{4}{2}) = 0.7$$

maka didapatkan pemisah dari kedua kelas pada probabilitas 0.7 seperti pada Gambar 2.1, nilai residual didapat dengan menghitung selisih dari $(y_i - p_i)$,

$$data1 = 1 - 0.7 = 0.3$$

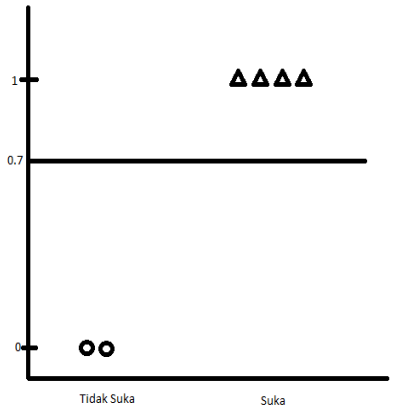
$$data2 = 1 - 0.7 = 0.3$$

$$data3 = 0 - 0.7 = -0.7$$

$$data4 = 0 - 0.7 = -0.7$$

$$data5 = 1 - 0.7 = 0.3$$

$$data6 = 1 - 0.7 = 0.3$$



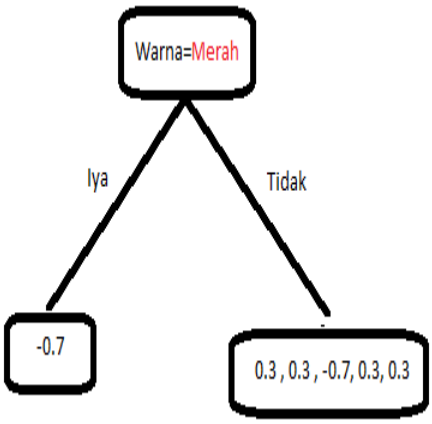

Gambar 2.1: Probabilitas mula-mula pada data

setelah itu menghitung nilai korespondensi dari prediksi awal, nilai $\lambda$ yang besar mengakibatkan nilai korespondensi menjadi lebih kecil, sehingga daun-daun pada pohon mudah terpangkas karena regulasi, pada contoh kali ini dimisalkan $\lambda = 0$,

$$\mathcal{L} = \frac{\sum_{i \in I_j} (y_i - p_i)^2}{\sum_{i \in I_j} p_i(1 - p_i) + \lambda}$$

$$\mathcal{L} = \frac{(0.3 + 0.3 - 0.7 - 0.7 + 0.3 + 0.3)^2}{6(0.7(1 - 0.7)) + 0} = 0.03$$

Selanjutnya membangun pohon klasifikasi atau biasa dikenal dengan CART (*Classification And Regression Tree*), berbeda dengan algoritma pada pohon keputusan, metode *boosting* yang digunakan pada XGBoost lebih mirip seperti *ensemble tree* layaknya *Random Forest*. Misalkan terpilih *threshold* root yakni $warna = merah$, maka terbentuklah pohon sebagai berikut,




Gambar 2.2: Pohon CART yang terbentuk dengan *threshold* $warna = merah$

Setelah data terpisah, maka hitung nilai korespondensi pada setiap daun

yang terbentuk,

$$\mathcal{L}_{left} = \frac{(0.3 + 0.3 - 0.7 + 0.3 + 0.3)^2}{5(0.7(1 - 0.7))} = 0.2381$$

$$\mathcal{L}_{Right} = \frac{(-0.7)^2}{0.7(1 - 0.7)} = 2.3333$$

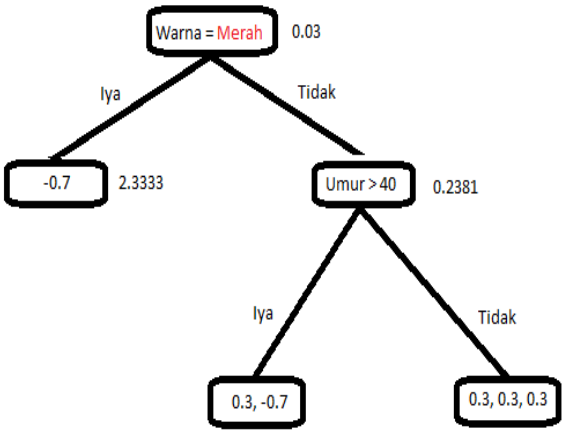
dari hasil penghitungan daun kanan dan kiri, maka didapatkan nilai *Gain* dari node tersebut adalah

$$\mathcal{L}_{spilt} = \mathcal{L}_{left} + \mathcal{L}_{Right} - root$$

$$\mathcal{L}_{spilt} = 0.2381 + 2.3333 - 0.03 = 2.5414$$

nilai *Gain* inilah yang menentukan nilai *threshold root* layak digunakan pada pohon CART, nilai *Gain* yang besar menunjukkan semakin layak nilai tersebut dijadikan *root*.

Selanjutnya dibentuk node dari daun yang memiliki data lebih banyak, sehingga terbentuk pohon seperti dibawah ini, sama seperti sebelumnya,




Gambar 2.3: Pohon yang terbentuk dari node yang memiliki data tidak tunggal

hitung nilai korespondensi pada daun-daun hasil pemisahan *node*.

$$\mathcal{L}_{left} = \frac{(0.3 - 0.7)^2}{2(0.7(1 - 0.7))} = 0.3809$$

$$\mathcal{L}_{Right} = \frac{(0.3 + 0.3 + 0.3)^2}{3(0.7(1 - 0.7))} = 1.2857$$

dari hasil penghitungan daun kanan dan kiri, maka didapatkan nilai *Gain* dari node tersebut adalah

$$\mathcal{L}_{spilt} = \mathcal{L}_{left} + \mathcal{L}_{Right} - root$$

$$\mathcal{L}_{spilt} = 0.3809 + 1.2857 - 0.2381 = 1.4285$$

pada nilai *Threshold* terjadi pengulangan perhitungan untuk mendapatkan nilai *Gain* maksimum, pada contoh diatas nilai $umur > 40$ merupakan nilai *Gain* maksimum, jika diubah maka nilai dari *Gain* akan lebih kecil.

Selanjutnya dilakukan penyetopan pembuatan cabang karena dibatasi. Pembatasan bisa dari kedalaman pohon atau juga dikarenakan regulasi nilai yang mengakibatkan pemangkasan seperti $\gamma$. Jika $Gain - \gamma = +$, maka tidak terjadi pemangkasan pada *node* pohon. Apabila $Gain - \gamma = -$, maka *node* tersebut akan dipangkas.

Setelah membangun pohon CART, lalu menghitung nilai output dari klasifikasinya. Penghitungan nilai output dilakukan pada setiap node yang bertindak sebagai daun pada pohon CART.

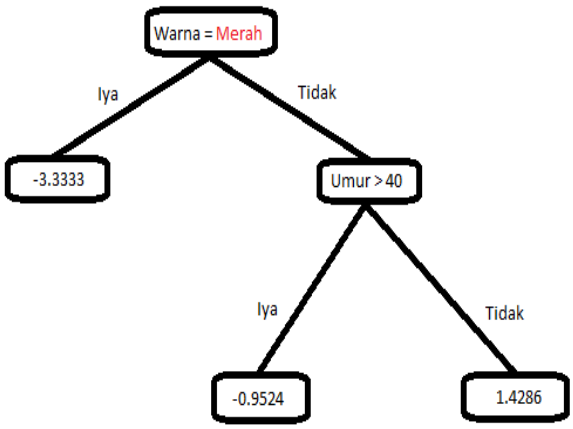
$$w_j^* = \frac{\sum_{i \in I_j}(y_i - p_i)}{\sum_{i \in I_j} p_i(1-p_i) + \lambda}$$

$$w_1^* = \frac{0.3 - 0.7}{0.7(1-0.7) + 0.7(1-0.7) + 0} = -0.9524$$

$$w_2^* = \frac{0.3 + 0.3 + 0.3}{0.7(1-0.7) + 0.7(1-0.7) + 0.7(1-0.7) + 0} = 1.4286$$

$$w_3^* = \frac{-0.7}{0.7(1-0.7) + 0} = -3.3333$$

sehingga pada node yang menjadi daun terdapat nilai output sebagai klasifikasi dari data seperti pada gambar dibawah ini,




Gambar 2.4: Pohon CART yang terbentuk pada $n = 1$

Pohon baru dibentuk setelah menghitung nilai prediksi yang baru, nilai prediksi yang baru memiliki formula sebagai berikut,

$$prediksi_i = log(odds) + (learning_rate)(nilaioutputdatapadapohon1)$$

$learning_rate$ merupakan rasio dimana model dapat mempelajari data berdasarkan pembelajaran terus-menerus dilakukan. Pada contoh kali ini dimisalkan $learning_rate = 0.3$, nilai ini digunakan sebagai nilai ketetapan pada XGBoost.

$$prediksi_1 = 0.03 + 0.3(1.4286) = 0.4586$$

$$prediksi_2 = 0.03 + 0.3(-0.9524) = -0.2557$$

$$prediksi_3 = 0.03 + 0.3(-3.3333) = -0.2557$$

$$prediksi_4 = 0.03 + 0.3(-3.3333) = -0.9699$$

$$prediksi_5 = 0.03 + 0.3(1.4286) = 0.4586$$

$$prediksi_6 = 0.03 + 0.3(1.4286) = 0.4586$$

nilai-nilai prediksi diatas masih dalam bentuk $log(odds)$, sehingga dibutuhkan pengkonversian kedalam bentuk probabilitas. Fungsi logistik yang digunakan adalah,

$$p_i = \frac{e^{log(odds)}}{1 + e^{log(odds)}}$$

maka didapatkan,

$$p_{1,5,6} = \frac{e^{0.4586}}{1 + e^{0.4586}} = 0.6$$

$$p_{2,3} = \frac{e^{-0.2557}}{1 + e^{-0.2557}} = 0.4$$

$$p_4 = \frac{e^{-0.9699}}{1 + e^{-0.9699}} = 0.27$$

nilai probabilitas tersebut digunakan sebagai probabilitas pada pembentukan pohon berikutnya, tahapan yang dilakukan sama seperti yang dilakukan tadi. Pembentukan pohon CART dilakukan terus menerus kecuali dibatasi oleh *user*. Nilai ketetapan dari banyaknya pohon CART yang dibangun pada XGBoost sebanyak 100 pohon.

### 2.3 Biomedicine Text Mining

*Text mining* merupakan informasi dan pengetahuan dalam jumlah besar dengan memproses ekstraksi pola pada teks seperti dokumen, jurnal, kutipanteks, buku, dll. *Text mining* merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu. *Biomedicine* merupakan disiplin ilmu dalam hal *bio science* dan teknologi untuk memecahkan masalah kesehatan global, seperti penyakit menular, kanker, kelainan dan kesejahteraan manusia. Sehingga *Biomedicine Text Mining* dapat diartikan sebagai penambangan teks dalam dokumen yang berhubungan dengan keilmuan dan teknologi untuk memecahkan masalah kesehatan.

Penambangan teks dalam hal *bio science* dan teknologinya berdasarkan pada sumber utama jurnal dan dataset bioinformasi. Database terkemuka yang menyediakan artikel dan dataset informasi biologis seperti *National Center For Biotechnology Information* (www.ncbi.nlm.nih.gov) dan *European Molecular Biology Laboratory.* Pada fitur NCBI menyediakan PubMed ( *Publication Medicine* ) dan PubChem (*Publication Chemistry*) yang disana terdapat sekitar 29 juta sitasi untuk literatur kesehatan. Data-data yang diperoleh dari sumbet text tersebut di tambang untuk mendapatkan komponen kesehatan yang telah diuji pada jurnal penelitian tentang kesehatan.

Penambangan teks pada data besar seperti NCBI maupun EMBL sudah banyak tersedia di berbagai web. Khususnya untuk interaksi antar protein seperti pada STRING (https://string-db.org/) dan InTact (https://www.ebi.ac.uk/intact). STRING dan Intact adalah database interaksi protein-protein yang diketahui dan diprediksi. Interaksi meliputi asosiasi langsung (fisik) dan tidak langsung (fungsional), interaksi ini berasal dari prediksi komputasi, dari transfer pengetahuan antara organisme dan dari interaksi yang dikumpulkan dari database (primer) lainnya. Database STRING saat ini mencakup 25.584.628[4] protein dari 5090 organisme sedangkan Intact saat ini mencakup 102.508[5] protein dan 585.731 interaksi dari berbagai organisme.

### 2.4 Protein

Protein merupakan salah satu komponen utama dalam sistem metabolism tubuh. Protein sendiri merupakan senyawa organik kompleks berbentuk polimer dari monomer asam amino yang dihubungkan satu sama lain dengan ikatan peptida. Senyawa kompleks protein terdiri dari molekul-molekul seperti karbon, hidrogen, oksigen, nitrogen serta kadang-kadang ada yang mengandung sulfur dan fosfor. Pada sistem organ manusia, protein kebanyakan berperan sebagai enzim atau subunit enzim, dan beberapa yang lain berperan dalam strukturl atau mekanis, seperti protein yang membentuk batang dan sendi sitoskeleton. Protein sendiri tergolong kedalam molekul yang sangat besar atau makrobiopolimer yang tersusun atas monomer asam amino, jenis asam amino pada protein ada 20 yang masing-masing terdiri atas seguah gugus karboksil, sebuah gugus amino dan rantai samping (grup 'R'). Rantai samping inilah yang akan mempengaruhi ciri-ciri dari keseluruhan protein.

#### 2.4.1 Asam Amino

Penyusun utama dari protein adalah asam amino. Terdapat 20 jenis asam amino, berdasarkan struktur molekulnya dapat diklasifikasikan menjadi,(Belitz, 2009)

(a). Asam amino non polar, dengan tanpa muatan rantai samping (grup 'R'), seperti, *glycine, alanine, valine, leucine, isoleucine, proline,*

[4]STRING Database – Statistic. 2019. *Functional Protein Association Network.* STRING Consortium 2019 [online]. https://string-db.org/. diakses 28 Juli 2019.

[5]IntAct - Statistics. 2019. *IntAct Molecular Interaction Database.* EMBL-EBI [online]. https://www.ebi.ac.uk/intact. diakses 28 Juli 2019

*phenylalanine, tryptophan* dan *methionine*.

(b). Asam amino polar, dengan tanpa muatan rantai samping, seperti *serine, threonine, cysteine, tyrosine, asparagine* dan *glutamine*.

(c). Asam amino bermuatan pada rantai samping, seperti *aspartic acid, glutamic acid, histidine, lysine* dan *arginine*.

Berdasarkan nutrisi pada protein, asam amino dapat diklasifikasikan menjadi,

(a). Asam amino essensial, merupakan asam amino yang diperlukan mahluk hidup akan tetapi tidak dapat memproduksi sendiri atau selalu kekurangan asam amino yang bersangkutan. Istilah essensial hanya untuk organisme heterotrof. Contoh dari asam amino essensial seperti *valine, leucine, isoleucine, phenylalanine, tryptophan, methionine, threonine, histidine, lysine* dan *arginine*.

(b). Asam amino non-essensial, merupakan jenis asam amino yang diperlukan oleh organisme dan setiap organisme mampu menghasilkan sendiri protein tersebut atau dengan kata lain tidak akan kekurangan asam amino tersebut. Contoh dari asam amino non-essensial seperti *glycine, alanine, proline, serine, cysteine, tyrosine, asparagine, glutamine, aspartic acid* dan *glutamic acid*.

### 2.4.2 Insulin

Insulin adalah hormon anabolik yang berefek sangat kuat terhadap metabolisme. Pada tubuh manusia insulin diproduksi di pankreas. Fungsi kerja insulin pada metabolisme yakni mengatur kadar gula dalam darah dengan mengubahnya menjadi energi atau glikogen. Insulin dapat berikatan dengan reseptor protein lain sangat diatur dan spesifik. Mendefinisikan langkah-langkah kunci yang mengarah pada kekhususan pensinyalan insulin menghadirkan tantangan besar bagi penelitian biokimia, tetapi hasilnya harus menawarkan pendekatan terapeutik baru untuk pengobatan pasien yang menderita keadaan resistan terhadap insulin, termasuk diabetes tipe-2.

Ada beberapa kondisi yang mengefek kepada metabolisme jika kerja insulin terganggu,

1. Resistensi Insulin, gangguan ini ketika sel otot, lemak dan hati tidak dapat menggunakan fungsi insulin dengan baik. Sehingga mengakibatkan kerja pankreas menjadi ekstra keras untuk menghasilkan insulin lebih banyak agar glukosa dapat diubah menjadi energi. Jika tidak ditanganai maka dapat menyebabkan diabetes.

2. Diabetes Mellitus, yakni kondisi dimana kadar gula dalam darah sangat tinggi akibatnya tubuh tidak mampu mengubah glukosa menjadi energi. Kelebihan muatan gula dalam darah mengakibatkan transportasi dalam darah terganggu, dikarenakan molekul gula glukosa yang besar.

3. Simdrom Metabolik, yakni keadaan dimana dapat meningkatkan resiko penyakit jantung dan masalah kesehatan lainnya, atau dengan kata lain keadaan dimana terjadi komplikasi.

4. Sindrom Ovarium Polikistik (PCOS), yaitu suatu keadaan kerja ovarium terganggu. PCOS mengakibatkan beberapa kadar hormon dalam tubuh menjadi *abnormal*.

(Virkamaki, A., dkk, 1999)

## 2.5 Diabetes Militus

Diabetes mellitus adalah penyakit kronis yang disebabkan oleh defisiensi bawaan dan / atau didapat dalam produksi insulin oleh pankreas, atau oleh ketidakefektifan insulin yang diproduksi. Kekurangan semacam itu menghasilkan peningkatan konsentrasi glukosa dalam darah, yang pada gilirannya merusak banyak sistem tubuh, khususnya pembuluh darah dan saraf.

Ada tiga jenis Diabetes Militus[6] , yaitu:

(a). Diabetes tipe-1 atau biasanya dikenal dengan *Insulin-Dependent* yaitu pankreas gagal menghasilkan insulin yang penting dalam metabolisme tubuh organisme. Penyebab pasti diabetes tipe 1 tidak diketahui. Secara umum disepakati bahwa diabetes tipe 1 adalah hasil dari interaksi yang kompleks antara gen dan faktor lingkungan, meskipun tidak ada faktor risiko lingkungan spesifik yang terbukti menyebabkan sejumlah besar kasus. Bentuk tipe ini sering dijumpai pada anak-anak dan remaja.

(b). Diabetes tipe-2 atau biasanya disebut dengan *Non-Insulin-Dependent* yaitu DM yang berasal dari ketidak mampuan tubuh untuk merespon dengan baik terhadap aksi insulin yang di produksi dengan baik. Dengan kata lain, tipe ini adanya gangguan fungsi dari insulin sehingga dia tidak bisa bekerja dengan baik dalam mengatur gula dalam darah. Diabetes tipe ini umum terjadi sekitar 90% dari jumlah kasus diabetes di dunia. Tipe-2 ini sering terjadi pada orang dewasa akat tetapi laporan WHO menyebutkan akhir-akhir ini kejadian pada remaja meningkat. Risiko diabetes tipe 2 ditentukan oleh interaksi faktor genetik dan metabolisme. Etnisitas, riwayat keluarga diabetes, dan diabetes gestasional sebelumnya digabungkan dengan usia yang lebih tua, kelebihan berat badan dan obesitas, diet yang tidak sehat, aktivitas fisik dan merokok meningkatkan risiko terjadinya tipe ini.

(c). Diabetes Gestasional merupakan gangguan sementara yang terjadi sekitar *trimester* kedua kehamilan dan menghilang setelah seorang wanita melahirkan. Tetapi wanita yang memiliki diabetes gestasional

[6]World Health Organization., (2016). *World Health Organization: Global Report on Diabetes.*, France: World Health Organization., http://www.who.int/diabetes/global-report/en/,[Sitasi: 14 Mei 2019].

harus dipantau ketat setelah melahirkan, karena mereka lebih mungkin berpotensi terkena diabetes tipe-2 di kemudian hari.

Ketika diabetes tidak dikelola dengan baik dapat menyebabkan komplikasi, keadaan seperti itu berkembang dan mengancam kesehatan dan membahayakan kehidupan. Komplikasi akut merupakan kontributor signifikan terhadap mortalitas, biaya dan kualitas hidup yang buruk. Glukosa darah yang *abnormal* tinggi dapat memiliki dampak yang mengancam jiwa jika memicu kondisi seperti *diabetic ketoacidosis* (DKA) pada tipe 1 dan 2, dan *koma hyperosmolar* pada tipe 2. Glukosa darah yang *abnormal* rendah dapat terjadi pada semua jenis diabetes dan dapat mengakibatkan kejang atau kehilangan kesadaran. Ini mungkin terjadi setelah melewatkan makan atau berolahraga lebih dari biasanya, atau jika dosis obat anti-diabetes terlalu tinggi.

## 2.6 Interaksi Protein-Protein

Interaksi protein-protein atau biasa disebut *Protein-Protein Interaction* (PPI) merupakan kondisi interaksi fisik antara dua protein (secara biner). Interaksi tersebut merupakan hasil dari biokimia yang didorong oleh gaya elektrostatik. Interaksi biner dari dua protein dapat digunakan untuk mengkontruksi model jaringan interaksi protein. Sehingga jaringan interaksi protein dapat menangani proses biologi, yang meliputi interaksi antar sel, metabolisme dan perkembanga.(Srinivasa, R,V., dkk, 2014)

Jenis-jenis interaksi protein dapat dibedakan sebagai berikut,

1. Homo-Oligomer dan Hetero-Oligomer, Homo-oligomer adalah makromolekul kompleks yang hanya terdiri dari satu jenis subunit protein. Hetero-oligomer merupakan Subunit protein yang berbeda berinteraksi untuk mengendalikan beberapa fungsi seluler.

2. Interaksi Stabil dan Sementara, Interaksi yang stabil melibatkan protein yang berinteraksi untuk waktu yang lama, mengambil bagian dari kompleks permanen sebagai subunit untuk menjalankan peran struktural atau fungsional. Di sisi lain, protein dapat berinteraksi secara singkat dan dengan cara yang dapat dibalikkan dengan protein lain hanya dalam konteks seluler tertentu seperti jenis sel , tahap siklus sel , faktor eksternal, keberadaan protein pengikat lainnya, dll. seperti yang terjadi pada sebagian besar protein.

3. Interaksi Ikatan Kovalen dan Non-Kovalen, Interaksi kovalen adalah interaksi yang paling kuat yang dibentuk oleh ikatan disulfida atau pembagian elektron . Meskipun jarang, interaksi ini sangat menentukan dalam beberapa modifikasi posttranslasional. Ikatan non-kovalen biasanya terbentuk selama interaksi sementara dengan kombinasi ikatan yang lebih lemah, seperti ikatan hidrogen , interaksi ion, gaya Van der Waals atau ikatan hidrofobik.

Penelitian-penelitian sebelumnya menyarankan penggunaan gen ontologi (GO) untuk memvalidasi nilai dari PPI dengan mengukur kesamaan yang ada

pada protein.(Jain, S., dkk, 2010) Pada tesis ini untuk menganalisa interaksi protein-protein menggunakan karakteristik dari gen ontologi. Interaksi antar protein terjadi karena memiliki kesamaan fungsional pada protein, maka kesamaan gen ontologi pada setiap protein dapat menunjukkan bahwasannya protein yang memiliki proses, struktur atau fungsi pada tingkat molekul sel dapat berinteraksi dengan protein lain yang memiliki ciri yang sama.

### 2.6.1 Gene Ontology

Gen ontologi atau *Gene Ontology* (GO) adalah tempat penyimpanan ontologi biologis serta keterangan dari gen maupun produk dari gen. Kesamaan fungsional antar protein dapat diukur dengan kesamaan semantiknya, fungsi yang mengembalikan nilai numerik menunjukkan kedekatan antara ontologi protein pada keterangan protein tersebut. Sehingga interaksi dari protein ditafsirkan sebagai asosiasi fungsional yang kuat antara dua protein yang berinteraksi, hal ini dapat diukur dengan menggunakan kesamaan simantik. Kesamaan simantik yang tinggi akan menunjukkan bahwa protein-protein tersebut berinteraksi dari pada yang memiliki kesamaan rendah.(Wang, J., dkk, 2010) Penyusun utama dari gen ontologi yaitu ada 3 seperti pada Gambar 2.5,

a. *Biological Process* yaitu serangkaian peristiwa molekuler yang diakui dari proses awal sampai akhir yang jelas serta berkaitan dengan fungsi unit hidup terintegrasi (misal. sel, jaringan, organ, dan organisme).

b. *Cellular Component* yaitu mendefenikan level struktur subseluler dan makromolekul kompleks, bagian dari sel ataupun lingkungan extraseluler berada.

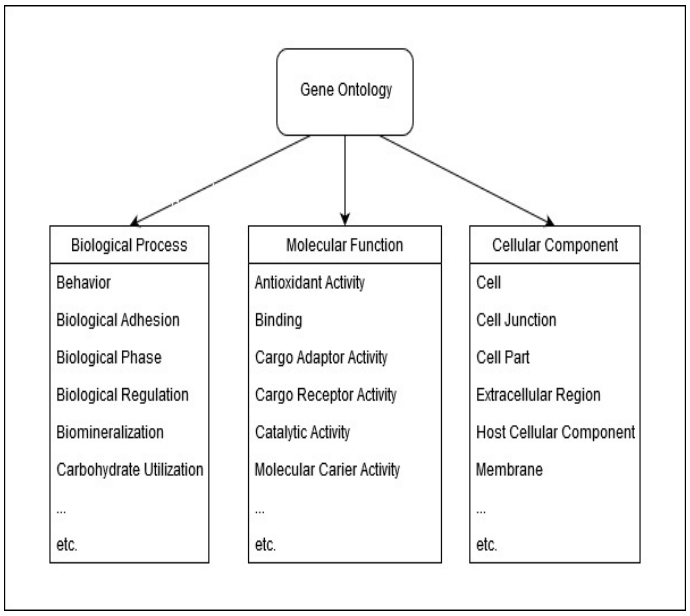
c. *Molecular Function* yakni karakteristik dari aktivitas-aktivitas dari produk gen pada level molekul, seperti *binding* atau *catalysis*.

Struktur gen ontologi berupa hirarki yang menjelaskan prilaku dari protein tersebut pada salah satu dari 3 ontologi utama gen. Anotasi pada gen ontologi dikodekan dan juga dapat dilihat struktur hirarkinya berupa *Directed Acyclic Graph* (DAG), dimana setiap *node* menyimbolkan *term* seperti kode dan konsep yang menjelaskan fungsi dari *feature*, sedangkan setiap *edge* menyetakan hubungan dari *term*. DAG pada gen ontologi seperti pada Gambar 2.6[7] dapat ditransformasikan sebagai bentuk graf berarah. Sehingga beberapa teori graf dapat digunakan untuk mendapatkan bobot dari setiap daun pada graf tersebut. Metode yang dapat digunakan dalam teori graf yaitu metode *centrality*.

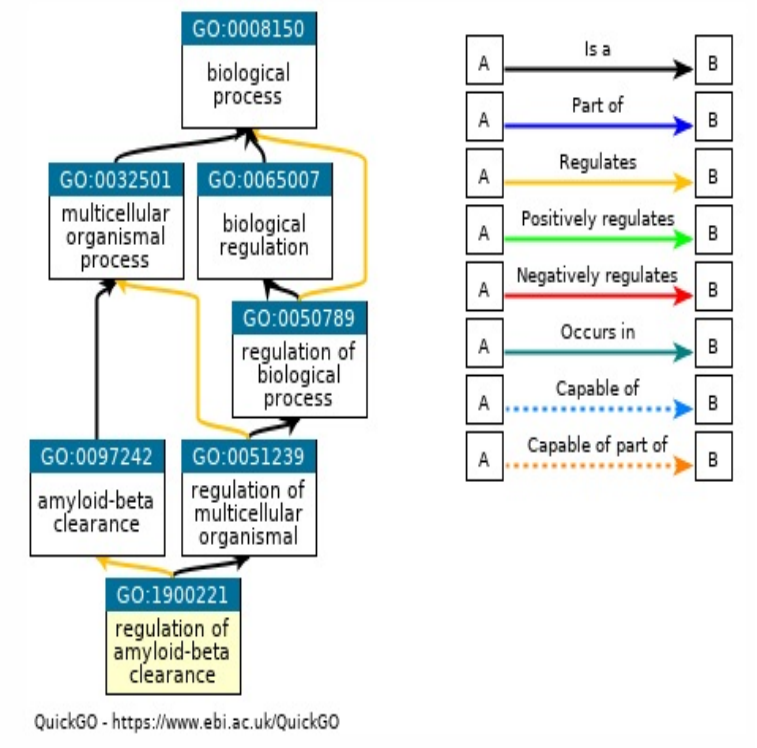
## 2.7 Centrality

Metode *centrality* merupakan salah satu metode graf yang memberikan bobot pada *node* dari graf. Metode *centrality* yang sering digunakan

[7]Quick GO - Gene Ontology and GO Annotation, [GO:1900221]Regulation of amyloid-beta clearance, `https://www.ebi.ac.uk/QuickGO/term/GO:1900221`

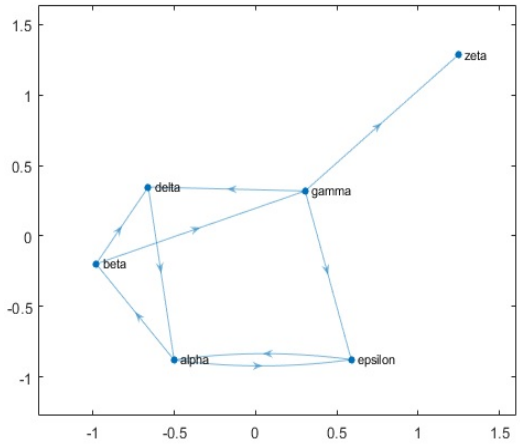



Gambar 2.5: Anotasi sebagai proses pada Gen Ontologi




Gambar 2.6: Contoh DAG serta *term* dan keterangan hubungan antara dua *term*

meliputi *Degree Centrality, Closeness Centrality, Betweenness Centrality* dan *Eigenvector Centrality*, dll. Metode-metode tersebut sering digunakan dalam masalah *network analysis*, untuk mendapatkan bobot dengan memperhatikan hubungan dari setiap *node* pada jaringan (graf).(Bogatti, S,P., 2005) Pada data GO, jaringan yang terjadi dapat diasumsikan sebagai graf berarah, yakni dalam konsep GO dikenal dengan DAG. Sehingga degan metode *centrality* dapat dihitung *measure* dari setiap proses yang ada pada DAG. Karena DAG merupakan graf berarah sehingga tidak semua metode *centrality* dapat diterapkan, misalnya *Eigenvector Centrality*. EC hanya bisa digunakan ketika graf yang akan diberi bobot termasuk dalam *undirected graf*. Contoh untuk graf berarah seperti pada Gambar 2.7, sehingga beberapa metode yang dapat digunakan dalam pemberian bobot seperti pada penjelasan dibawah ini.




Gambar 2.7: Contoh graf berarah

### 2.7.1 Closeness Centrality

Metode *closeness Centrality* (CC) merupakan metode yang menghitung jarak terbaik pada setiap *node*. Misalkan $d_{ij}$ menunjukkan jarak dari daun $i$ ke daun $j$. Bobot dari jarak diukur sebagai minimum dari hop yang diperlukan untuk berpindah dari $i$ ke $j$. Jarak rata-rata dari *node* $i$ ke yang lain deiberikan sebagai,

$$l_i = \frac{1}{n} \sum_j d_{ij}$$

Jika *node* memiliki $l_i$ kecil, serta dikatakan berdekatan dengan banyak *node* pada jaringan. Maka berakibat memberikan bobot yang disebut *closeness centrality* dari $i$,

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}$$

Sebagai contoh yang diberikan seperti pada Gambar 2.7, hasil bobot yang didapatkan dengan menggunakan *closeness centrality* dapat dilihat pada Tabel 2.2.

Tabel 2.2: Bobot yang dihasilkan dengan menggunakan *closeness centrality*

| No. | Node | Outcloseness | incloseness |
|---|---|---|---|
| 1 | Alpha | 0.11111 | 0.10667 |
| 2 | Beta | 0.125 | 0.08 |
| 3 | Gamma | 0.0125 | 0.071111 |
| 4 | Delta | 0.08333 | 0.091429 |
| 5 | Epsilon | 0.076923 | 0.10667 |
| 6 | Zeta | 0 | 0.071429 |

### 2.7.2 Betweenness Centrality

Metode *betweenness centrality* (BC) dari *node* $i$ dapat dikatakan sebagai seberapa sering *node* $i$ menemukan jalur terpendek antara dua *node* acak dari suatu jaringan. Misalkan, $g_{st}$ menjadi jumlah jarak terpendek antara $s$ dan $t$, lalu $n^i_{st}$ sebagai jumlah jarak terpendek antara $s$ dan $t$ yang melewati *node* $i$. Nilai BC dari $i$ dapat dinyatakan sebagai,

$$x_i = \sum_{s,t \in V} \frac{n^i_{st}}{g_s t}$$

dengan konvensi $\frac{n^i_{st}}{g_{st}} = 0$ jika keduanya bernilai 0. Dengan menggunakan contoh graf berarah pada Gambar 2.7, pembobotan dengan menggunakan metode *betweenness centrality* dapat dilihat pada Tabel 2.3.

Tabel 2.3: Bobot yang dihasilkan dengan menggunakan *betweenness centrality*

| No. | Node | Bobot |
|---|---|---|
| 1 | Alpha | 9 |
| 2 | Beta | 8 |
| 3 | Gamma | 5 |
| 4 | Delta | 2 |
| 5 | Epsilon | 1 |
| 6 | Zeta | 0 |

### 2.7.3 Page Rank Centrality

*Page Rank Centrality* (PRC) yaitu metode pembobotan pada graf berarah yang dihasilkan dari perjalanan acak pada suatu jaringan. Pada setiap *node* dalam graf, *node* berikutnya dipilih dengan probabilitas dari serangkaian *node* yang diteruskan pada *node* awal (pada kasus *undirected* graf dapat dikatakan tetangga). Jika suatu *node* tidak memiliki penerus maka *node* berikutnya dipilih dari semua *node* yang ada. Skor dari PRC merupakan waktu rata-rata yang dihabiskan pada setiap penelusuran secara acak. Jika *node* memiliki perulangan pada dirinya sendiri, maka ada kemungkinan algoritmanya melewati *vertex* tersebut, oleh karena itu perulangan pada diri sendiri dapat meningkatkan skor PRC.(Ivan, G. and Grolmusz, V., 2011)

Penghitungan skor dengan PRC sebagai berikut,

$$x_i = \alpha \sum_{j=1}^{n} A_{ji} \frac{x_j}{\sigma_j^+} + \beta$$

dimana $\sigma_j^+$ adalah *out-degree* dari *node* $j$. Namun, beberapa *node* memiliki $\sigma_j^+ = 0$ yang akan menyebabkan terjadi pembagian dengan nilai 0.Sehingga pada kasus ini, ditambahkan *vertex* yang merupakan perulangan dari $j$ ke $j$ sendiri, untuk menghasilkan $\sigma_j^+ = 1$. Sebagai catatan *node* tersebut tetap tidak memiliki kontribusi nilai terhadap *centrality* pada *node* yang lain.

Pada contoh graf berarah yang ditunjukkan pada Gambar 2.7, hasil perhitungan dengan *page rank centrality* dapat dilihat pada Tabel 2.4.

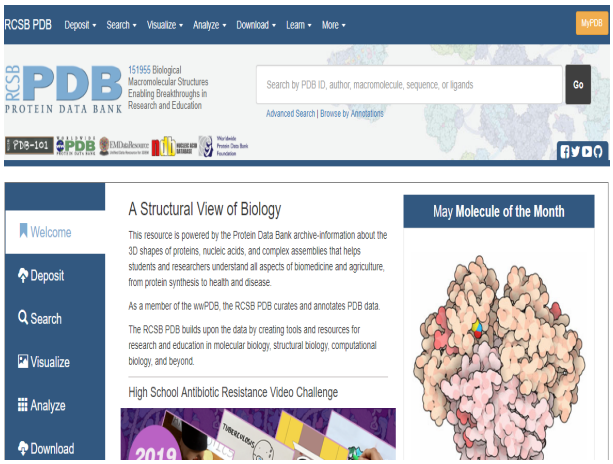
Tabel 2.4: Bobot yang dihasilkan dengan menggunakan *page rank centrality*

| No. | Node | Bobot |
|---|---|---|
| 1 | Alpha | 0.32098 |
| 2 | Beta | 0.17057 |
| 3 | Gamma | 0.10657 |
| 4 | Delta | 0.13678 |
| 5 | Epsilon | 0.20078 |
| 6 | Zeta | 0.06432 |

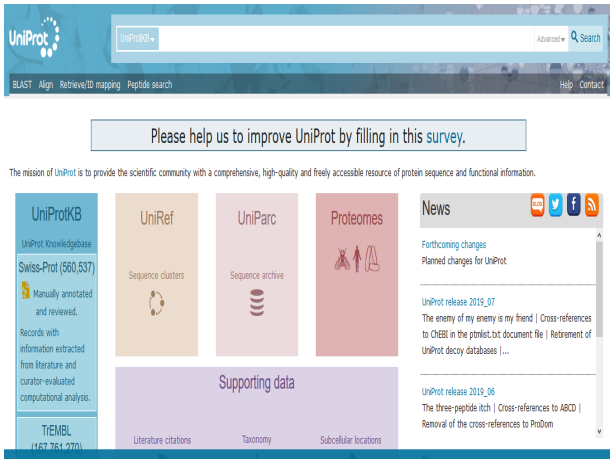
## 2.8 Database Bioinformatika

Data protein tersedia di beberapa database dunia, seperti Protein Data Bank (PDB)[8],Swiss-Prot (UniProt)[9], European Molecular Biology Laboratory (EMBL)[10] dan National Center of Biotechnology Information (NCBI)[11]. Data-data protein nantinya akan diambil dari database protein tersebut. Tujuannya agar saling melengkapi apabila pada salah satu database masih belum ada protein baru yang belum terdaftar. Halaman utama dari database protein seperti pada Gambar 2.8, Gambar 2.9 , Gambar 2.10 dan Gambar 2.11 dibawah ini.

[8]Protein Data Bank : http://www.rcsb.org/
[9]UniProt : https://www.uniprot.org/
[10]European Molecular Biology Laboratory : https://www.ebi.ac.uk/
[11]National Center Biotechnology Information : https://www.ncbi.nlm.nih.gov

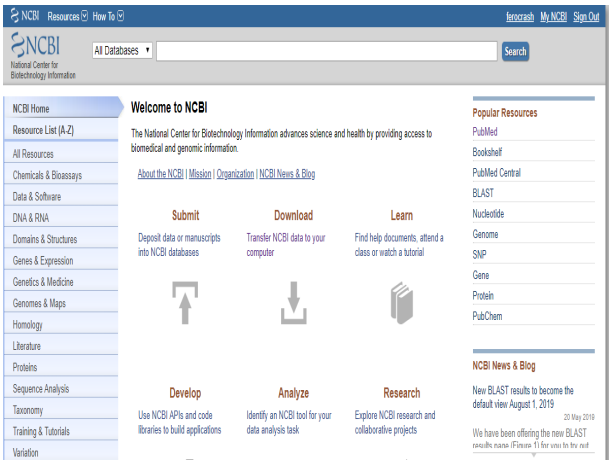



Gambar 2.8: Halaman depan web Protein Data Bank




Gambar 2.9: Halaman depan web UniProt


Gambar 2.10: Halaman depan web EMBL



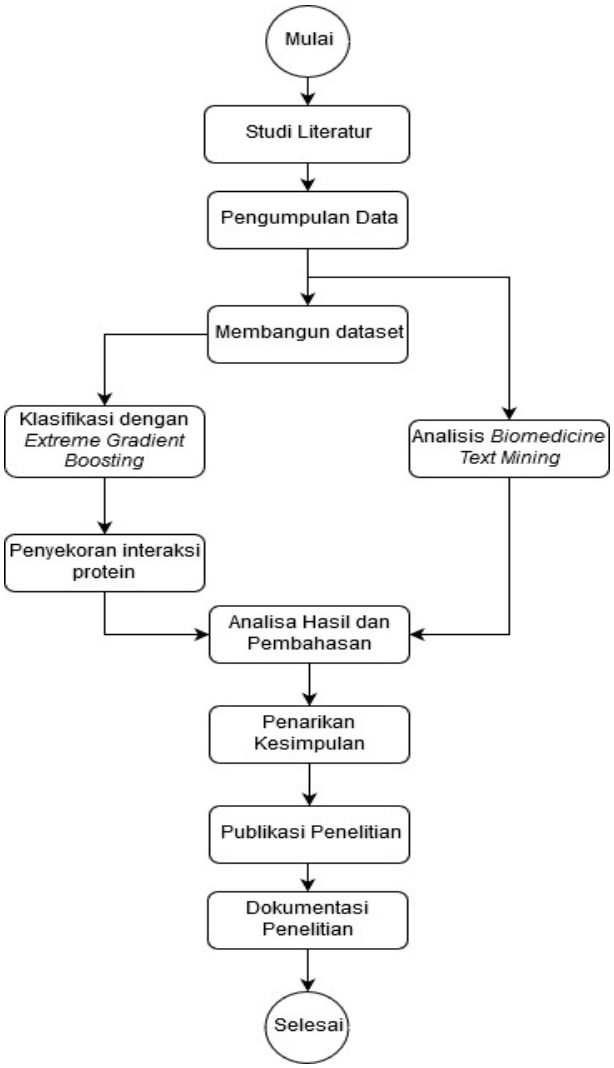

Gambar 2.11: Halaman depan dari web NCBI)

# BAB 3
# METODE PENELITIAN

Pada bab ini, dijelaskan tentang tahapan-tahapan penelitian yang dilakukan untuk menyelesaikan masalah yang telah dikemukakan pada rumusan masalah. Ditunjukkan pula jadwal penelitian untuk masing-masing tahapan penelitian tersebut.

## 3.1 Tahapan Penelitian

Penelitian ini meliputi beberapa tahapan-tahapan proses. Setiap proses dari tahapan-tahapan tersebut mempengaruhi dalam pengerjaan tesis ini. Langkah-langkah tersebut dapat dilihat pada Gambar 3.1 dibawah ini.




Gambar 3.1: Langkah-langkah metodologi dalam mengerjakan tesis

### 3.1.1 Studi Literatur

Pada tahap ini dilakukan studi literatur untuk mendukung pengerjaan penelitian ini dan pemahaman yang lebih mendalam mengenai metode *Extreme Gradient Boosting* dalam mengolah data interaksi protein-protein. Pemahaman akan metode dilakukan pada konsep matematika maupun konsep komputasinya. Selain mempelajari metode, mempelajari dataset yang akan dibangun juga diperlukan agar data yang digunakan tidak sembarangan. Literatur yang dipelajari dapat bersumber dari jurnal, buku, internet, maupun bimbingan dengan dosen pembimbing.

### 3.1.2 Pengumpulan Data

Tahap ini untuk mengumpulkan data dari objek penelitian, dalam hal ini pengumpulan data protein pada manusia yang mensintesis insulin dan protein yang tidak mensintesi insulin. Pengumpulan data bertujuan untuk membangun dataset sebelum diolah kedalam metode pengklasifikasian. Proses pengumpulan data dapat dilihat pada Gambar 3.3.

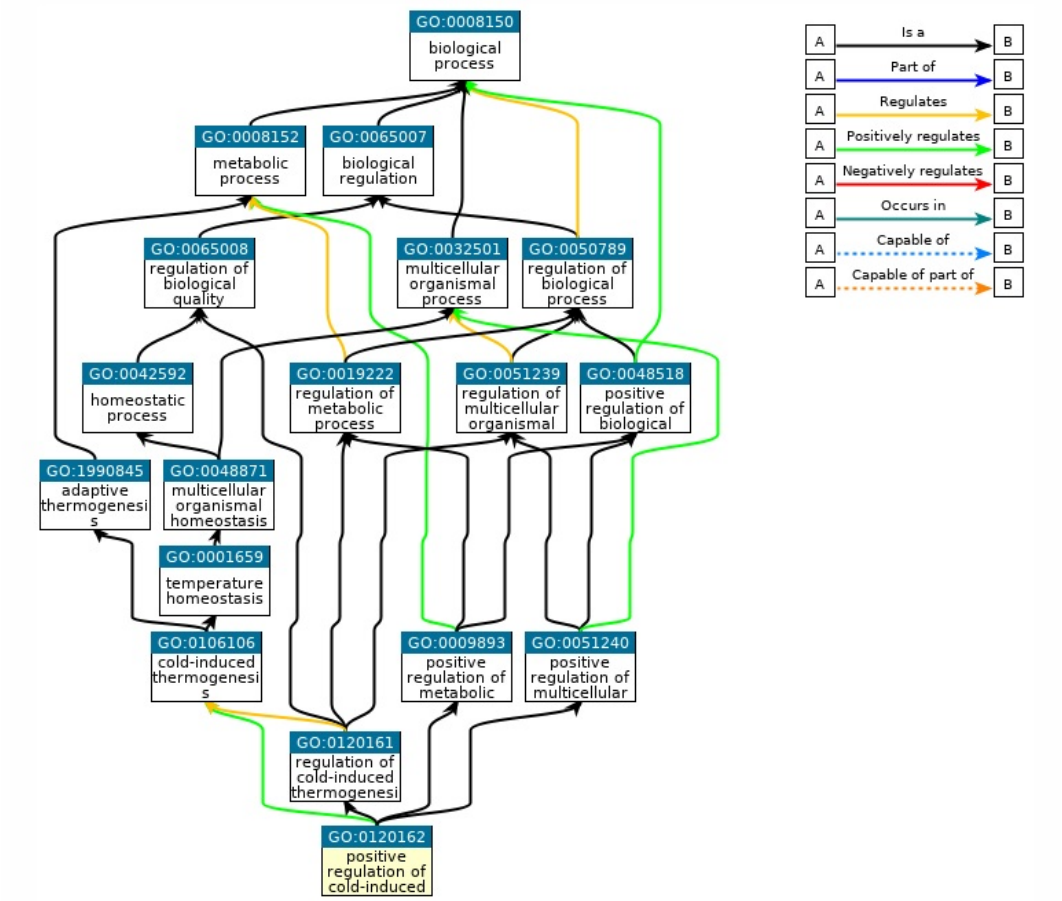
#### 3.1.2.1 Pengambilan data dari bank data

Pengambilan data protein dari data bank protein diambil dari dua sumber utama, yakni dari *Protein Data Bank* yang tersedia pada halaman https://www.rcsb.org/ dan pada UniProt yang tersedia pada halaman https://www.uniprot.org/. Pengumpulan data dengan mendownload protein dengan kata kunci "insulin". Atribut yang dibutuhkan pada saat pengumpulan data yaitu:

- *Gene Product* yaitu produk gen yang dihasilkan dari protein
- Protein ID atau Gene ID yaitu Nomer ID dari protein yang tertera pada database
- *Protein Function* yaitu keterangan dari fungsi protein tersebut
- *Gene Ontology* yaitu anotasi-anotasi yang terjadi pada protein, atau bisa dikatakan fenotip dari suatu Gen yang nantinya akan menjadi atribut utama pada pengklasifikasian. Anotasi dari gen ontologi di bedakan menjadi 3 jenis yaitu *Biological Process*, *Molecular Function*, dan *Cellular Component*.
- *Cross Reference* yaitu referensi keberadaan protein pada database protein apa saja.

Anotasi gen ontologi yang dijadikan data berupa teks seperti [GO:0120162] yang mengartikan kode fungsi *positive regulation of cold-induced thermogenesis*. Kode tersebut dapat menunjukkan hirarki dari fungsi protein tersebut berdasarkan fungsi utamanya seperti pada Gambar 3.2,

Anotasi GO juga disebut DAG, sehingga dibutuhkan pemberian bobot pada setiap daun pada graf tersebut untuk mendapatkan nilai pengaruhnya terhadap fungsi yang lain. Pemberian bobot dilakukan dengan menggunakan metode graf yaitu *centrality method* terhadap graf berarah.




Gambar 3.2: Anotasi dari GO:0120162 yang menunjukkan pengaruhnya terhadap kerja fungsi yang lain

#### 3.1.2.2 Data Pre-*Processing*

Proses Pre-*Processing* data digunakan untuk menyiapkan data sebelum dapat digunakan pada proses pembentukan dataset untuk klasifikasi. Proses ini bertujuan agar kontruksi dataset bisa seragam dan minim akan *noise* data. Proses pre-*processing* dilakukan dua kali yakni pada saat sebelum melakukan analisis menggunakan *centrality* dan pre-*processing* yang kedua sebelum data dilatih menggunakan algoritma XGBoost. Proses yang pertama meliputi,

1 Membersihkan deskripsi-deskripsi data yang tidak diperlukan, koleksi data pada file .xlsx disesuaikan dan dihilangkan kolom-kolom deskripsi yang tidak dibutuhkan pada proses selanjutnya. Setelah data dibersihkan, format data dirubah ke dalam bentuk .csv.

2 Mengkoleksi data GO *Molecular Function* dari data, pada proses data pada file .csv di dipilih kode GO yang merupakan *Molecular Function* dari protein tersebut.

3 Mendeskripsikan kode GO *Molecular Function* (Bentuk DAG). Proses ini menterjemahkan kode-kode GO yang belum bisa dimengerki ke dalam bentuk DAG, pada laman website www.quickgo.com.

4 Membangun DAG pada MATLAB. DAG hasil dari www.quickgo.com direplikasi kedalam perangkat lunak MATLAB untuk dapat diproses selanjutnya.

Proses pre-*processing* yang kedua dilakukan sebelum dataset dilatih menggunakan algoritma XGBoost, prosesnya meliputi.

1 Menambahkan label kelas pada dataset sebagai variabel $y$.

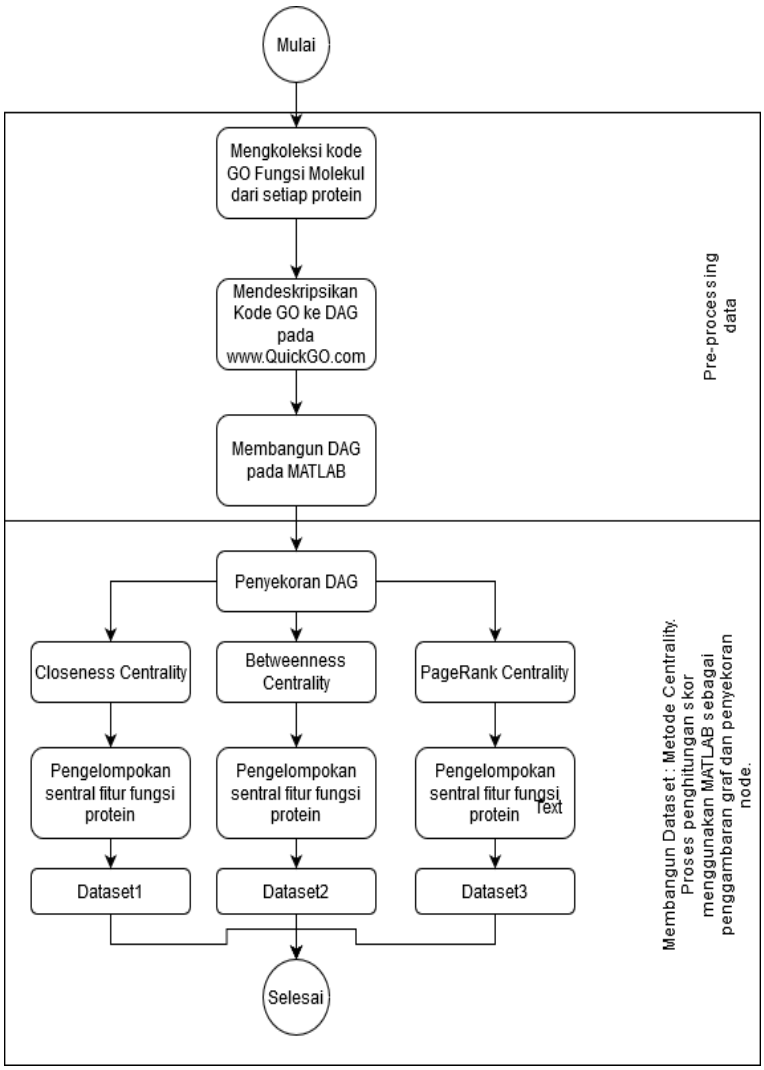
2 Menggabungkan *(merge)* data kelas protein insulin dengan protein yang bukan insulin.

3 Mengubah label *String* menjadi bentuk array biner 0 1 menggunakan *label_encoder*.

### 3.1.3 Membangun Datasets

Metode *centrality* digunakan untuk memberikan skor terhadap fungsi protein pada GO yang terhubung dengan fungsi yang lain yang direpresentasikan dalam bentuk graf berarah. Metode *centrality* yang digunakan yaitu BC, CC, dan PRC. Hasil-hasil penyekoran tersebut digunakan sebagai nilai fitur pengaruh suatu protein terhadap karakteristik ontologinya yang dijadikan datasets.

Karakteristik dari protein dilihat dari fungsi molekulnya dibagi menjadi 20 fungsi utama, yaitu *antioxidant activity, binding, cargo adaptor activity, cargo receptor activity, catalytic activity, molecular carrier activity, molecular function regulator, molecular sequestering activity, molecular transducer activity, negative regulation of molecular function, nutrient resivior activity,*




Gambar 3.3: Diagram alir proses pengumpulan data dan proses membangun dataset

*positive regulation of molecular function, protein folding chaperon, protein tag, regulation of molecular function, small molecular sensor activity, structural molecular activity, toxin activity, transcription regulator activity*, dan *transporter activity*. Setiap kode-kode GO yang berada pada setiap protein diterjemahkan dahulu dengan metode *centrality* untuk mengetahui pada fungsi utama mana dia berpengaruh. Alur dalam membangun dataset dapat dilihat pada Gambar 3.1.

Proses penyekoran dilakukan setelah kontruksi DAG pada setiap kode GO selesai. Metode *centrality* memberikan bobot pada setiap node dari graf yang merupakan skor sentral dari setiap node. Fungsi utama dari suatu protein dijadikan sentral utama dari GO, skor dari node tersebut digunakan untuk membangun dataset. Setiap metode *centrality* yang digunakan memiliki metode yang berbeda. Sehingga penyekoran untuk membangun dataset di lakukan sebanyak tiga kali, terhadap masing-masing dari metode yang digunakan.

### 3.1.4 Klasifikasi dengan XGBoost

Proses pengolahan data dimulai pada tahap ini. Data yang sudah siap digunakn akan dijadikan data latih dan data tes pada tahap klasifikasi ini. Metode klasifikasi yang digunakan adalah *Extreme Gradient Boosting* yang merupakan algoritma *boosting tree system*. Proses-proses yang ada pada tahap ini meliputi beberapa proses seperti, pembagian dataset, *cross validation*, proses latih XGBoost, proses tes model, optimasi model, dan analisis hasil model, yang dapat dilihat pada Gambar 3.4,

#### 3.1.4.1 Pembagian Data

Dataset yang telah disiapkan dengan format data .csv dibagi menjadi dua jenis data. Bagian data yang pertama sebanyak 80% dari total semua data yang dijadikan sebagai data latih. Sedangakan sisanya yakni 20% data yang lain digunakan sebagai data tes. Pembagian data menggunakan $train\_test\_spilt$ pada *Scikit Learn* dengan koefisien nilai acak $n = 9$.

#### 3.1.4.2 *Cross Validation*

Proses *cross validation* digunakan untuk mengacak komposisi data yang digunakan untuk proses pelatihan dan proses prediksi. Sehingga didapatkan variasi model dari data yang sama tetapi komposisi data berbeda pada data latih. Variasi data untuk *cross validation* dibuat sebanyak $10 - fold$, sehingga terdapat 10 jenis komposisi data yang berbeda pada dataset sama yang digunakan untuk melatih model XGBoost.

#### 3.1.4.3 Proses Latih XGBoost

Model desain dengan XGBoost merupakan modifikasi lanjut dari *gradient boosting*. Dasar dari metode ini merupakan mirip seperti *decision tree*. Perbedaan dari model sebelumnya terletak pada regulasi pada fungsi tujuannya. Pada GBM fungsi tujuannya hanya berupa fungsi kerugian saja, yakni dengan meminimalkan fungsi kerugian. Sedangkan pada XGBoost fungsi tujuannya berupa fungsi kerugian dan fungsi kompleksitas model. Melatih model klasifikasi XGBoost dengan data latih protein yang mensintesis insulin

dan bukan yang telah dibangun pada proses sebelumnya merupakan model yang akan digunakan untuk penyekoran selanjutnya.

Pada proses latih ini, algoritma XGBoost berjalan dengan membangun beberapa pohon keputusan sesuai dengan parameter *n_estimator* yang diinginkan. Proses atau algoritma XGboost pada pelatihan data menjadi model meliputi,

Input : Data $(x_i, y_i)_{i=1}^{n}$ dan fungsi kerugian yang diferensiable $L\left(y_i, f(x)\right)$.

Langkah-langkah:

1. Inisiasi model nilai awal konstan $F_o(x) = argmin_\gamma \sum_{i=1}^{n} L\left(y_i, f(x)\right)$ atau disebut nilai *odds*.

2. dari $n = 1$ ke $m$, sebagai *n_estimator* banyaknya pohon CART yang dibangun:

   a. hitung,
   $$r_{im} = -\alpha \left[\partial \frac{L\left(y_i, F(x)\right)}{\partial F(x_i)}\right]_{F(x)=f_{m-1}x}$$
   dimana $\alpha$ adalah *learning rate*.

   b. *fit* pohon CART ke nilai $r_{im}$ hingga $r_{jm}$, untuk $j = 1, 2, 3, ..., j_m$.

   c. untuk $j = 1, 2, 3, ..., j_m$ hitung,
   $$\gamma_{jm} = argmin_\gamma \sum_{x_i \in R_y} L\left(y_i, f_{m-1}(x_i) + \gamma\right)$$

   d. *update*
   $$F_m(x) = F_{m-1}(x) + \sum_{i=1}^{j_m} \gamma_m h_m(x_i)$$
   dimana $h_m(x_i)$ merupakan residual dari fungsi kerugian.

3. output $F_m(x)$

Pohon regresi untuk $h_m(x_i)$ memperkirakan sisa rata-rata pada setiap simpul terminal pohon, dalam meningkatkan gradien, komponen gradien rata-rata akan dihitung. Setiap node, ada faktor $\gamma$ yang $h_m(x_i)$ dikalikan yang menjelaskan perbedaan dampak dari masing-masing cabang perpecahan. *Gradient boosting* membantu memprediksi gradien optimal untuk model aditif, tidak seperti teknik *gradient descent* klasik yang mengurangi kesalahan dalam output pada setiap iterasi.

#### 3.1.4.4 Proses Tes Model

Proses tes atau prediksi merupakan proses untuk menguji model yang terbentuk dari data latih. Data tes yang digunakan merupakan data yang berbeda dari data latih. Data tes yang digunakan sebanyak 20% dari keseluruhan data. Hasil pengujian pada data tes sebagai data prediksi dapat dihitung nilai akurasi, presisi dan recall pada model pengklasifikasian dataset interaksi protein-protein. Nilai akurasi ini sebagai nilai acuan apakah model bagus dalam mengklasifikasikan atau tidak. Persamaan untuk mencari ketiga parameter uji tersebut sebagai berikut,

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Presisi = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

dimana TP menyatakan banyknya anggota dari kelas positif diklasifikasikan dengan benar, TN menyatakan banykanya anggota kelas negatif diklasifikasikan dengan benar, FP menyatakan jumlah anggota kelas positif diklasifikasikan dengan salah dan FN merupakan jumlah anggota kelas negatif yang diklasifikasikan dengan salah.

#### 3.1.4.5 Optimasi Model

Proses optimasi model dengan mengubah-ubah parameter-parameter XGBoost. Parameter-parameter yang dapat diubah untuk mendapatkan model terbaik yakni *n_estimator*, *learning_rate*, *max_depth*, $\gamma$, *subsample*, dan *colsample_bytree*. Parameter *n_estimator* adalah parameter yang digunakan untuk membangun berapa banyak pohon keputusan yang akan ditingkatkan. Pembuatan pohon keputusan yang semakin besar dapat menyebabkan model sangat sensitif terhadap fitur data, sehingga biasanya dapat terjadi keadaan *overfitting*. *learning_rate* adalah parameter yang digunakan untuk pembelajaran model XGBoost terhadap data latih. *max_depth* merupakan parameter untuk menentukan kedalaman dari setiap pohon keputusan yang terbentuk, kedalaman pohon keputusan menentukan peningkatan klasifikasi yang baik, semakin dalam pohon yang dibentuk akan membuat model lebih sensitif terhadap fitur data. Akan tetapi hal tersebut dapat membuat waktu komputasi menjadi lebih lama. $\gamma$ adalah parameter penentu atau regulasi dari pohon keputusan. Nilai $\gamma$ menjadi penentu bentuk pohon keputusan apakah node tersebut akan dipangkas atau tidak. *subsample* dan *colsample_bytree* merupakan parameter yang memiliki fungsi sama yakni membuat model latih *robust* terhadap noise. Pada *subsample* menjadi rasio dari baris dataset sedangkan untuk *colsample_bytree* sebagai rasio dari kolom dataset.

#### 3.1.4.6 Analisis Hasil Model

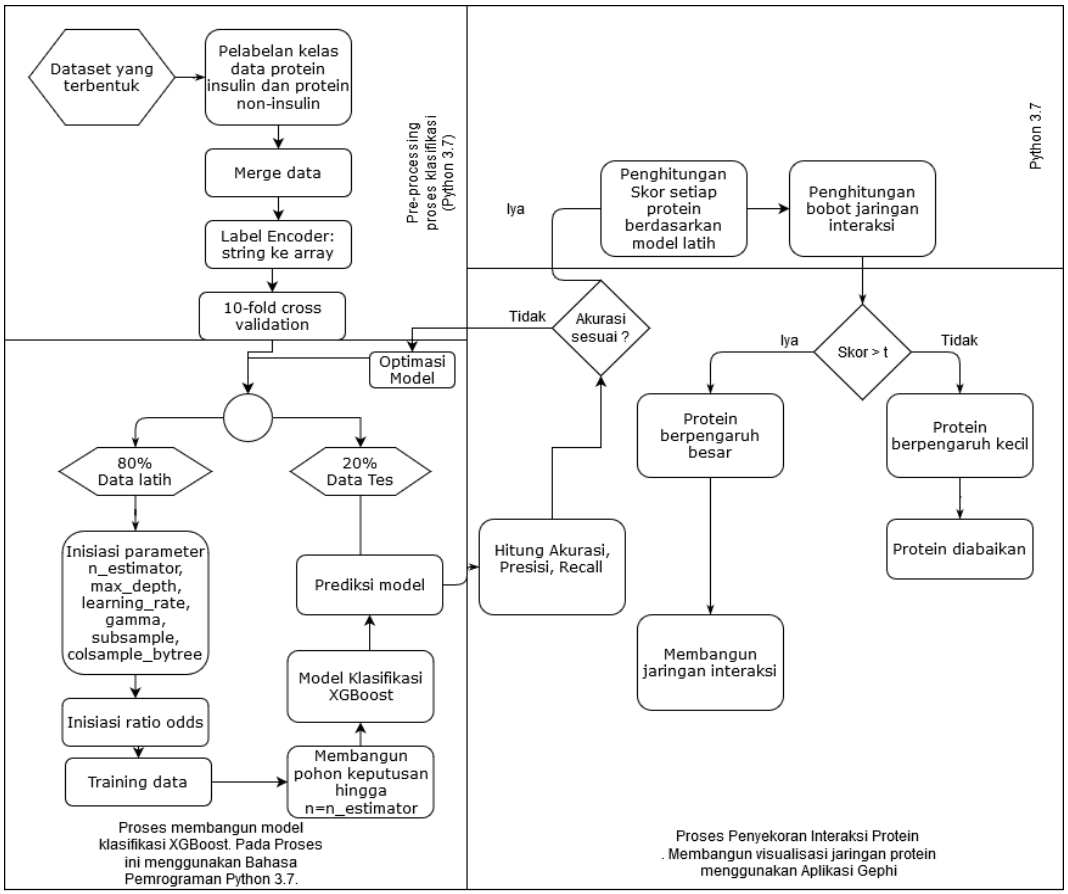
Proses analisis model meliputi proses pengamatan pada pohon CART, grafik fungsi kerugian (*error classification*), dan analisis terhadap fitur penting dari model XGBoost.

### 3.1.5 Perhitungan skor setiap protein

Penghitungan skor setiap protein berdasarkan model yang terbentuk nantinya berguna sebagai analisa interaksi yang terjadi pada setiap protein. Skor atau Bobot menggambarkan kedekatan antara protein satu dengan protein yang lainnya.

#### 3.1.5.1 Perhitungan skor interaksi protein

Perhitungan skor interaksi protein berdasarkan skor protein pada model latih XGBoost. Skor ini digunakan untuk membangun jaringan interaksi antar protein, sehingga dapat dianalisa protein yang berpengaruh terhadap insulin.

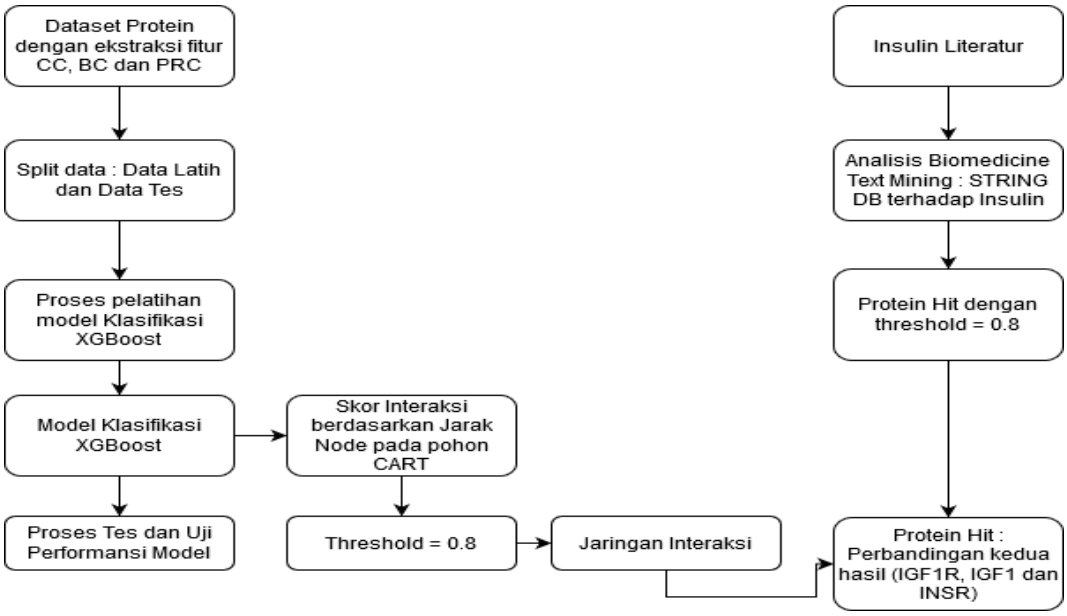



Gambar 3.4: Diagram alir proses klasifikasi pada XGBoost dan proses penyekoran untuk membangun jaringan interaksi

### 3.1.6 Biomedicine Text Mining Analysis

Selain mengklasifikasikan untuk mendapatkan protein yang berpengaruh pada insulin, juga dilakukan analisis *biomedicine text mining*. Analisis

ini dimaksudkan untuk mendapatkan protein-protein signifikan yang sudah diketahui keberadaannya dapat mempengaruhi insulin pada database PPI. *Text mining* dilakukan pada aplikasi web STRING. Setelah mendapatkan hasilnya, dibandingkan dengan hasil klasifikasi menggunakan XGBoost. Serangkaian proses pengolahan data tersusun seperti pada Gambar 3.5.




Gambar 3.5: Diagram proses pengolahan data

### 3.1.7 Analisis Hasil dan Pembahasan

Pada bagian ini, hasil klasifikasi XGBoost dan *text mining* dianalisis. Analisis yang dilakukan meliputi kemampuan pengklasifikasian, penyekoran setiap protein, pengaruh yang terjadi antar protein dan juga membandingkan hasil pada klasifikasi XGBoost dan *Biomedicine Text Mining*. Harapannya ditemukan interaksi protein yang berbeda yang berinteraksi kuat dengan insulin yang dihasilkan dengan metode XGboost dengan database pada STRING. Setelah dilakukan analisis baru dapat ditarik kesimpulan dari penelitian ini.

### 3.1.8 Penarikan Kesimpulan

Setelah mendapatkan pembahasan mengenai pengklasifikasian dan hasil *Biomedicine Text Mining* analisanya serta diskusi yang didapatkan, maka dapat ditarik berupa kesimpulan dari penelitian tesis.

### 3.1.9 Publikasi Penelitian

Publikasi penelitian bertujuan untuk memaparkan masil penelitian yang dilakukan pada tesis yang akan di *review* oleh pegiat akademisi yang lain. Publikasi penelitian yang dilakukan seperti seminar internasional terindex maupun jurnal internasional bereputasi.

### 3.1.10 Dokumentasi Penelitian

Dokumentasi bertujuan untuk mengarsipkan penelitian yang telah dilakukan. Dokumentasi dari buku Thesis ini menggunakan aplikasi LaTex dengan template thesis Matematika ITS.

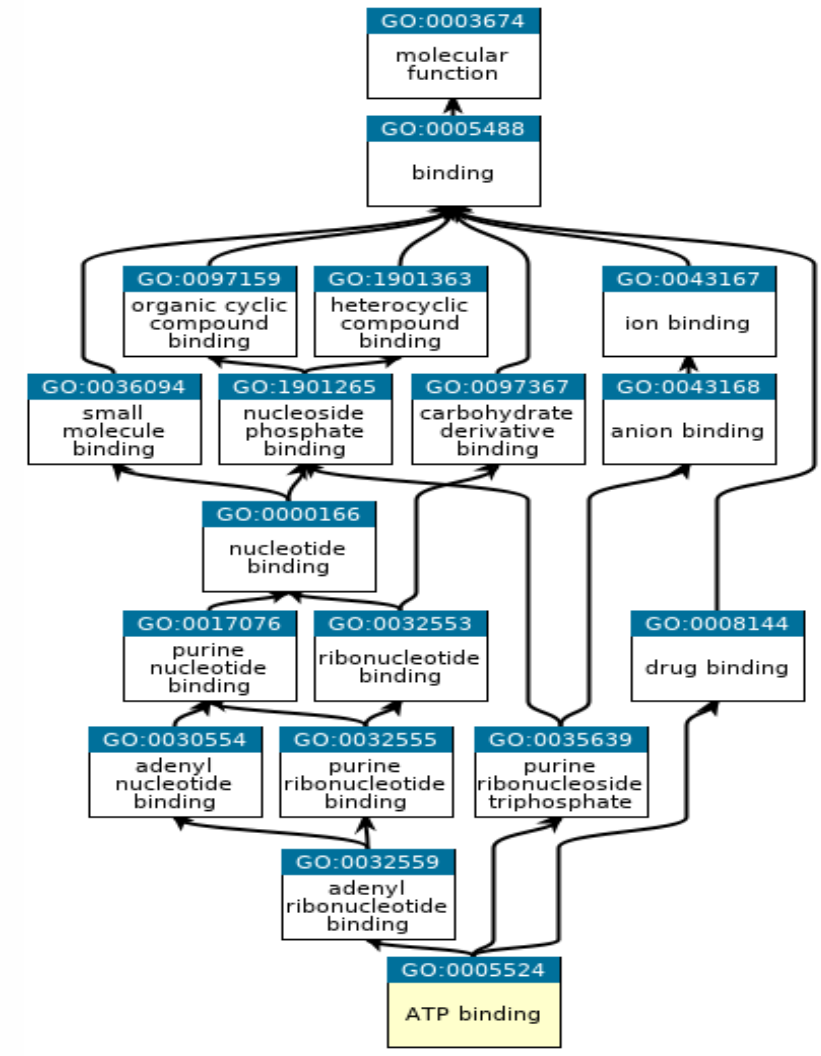
# BAB 4
# HASIL DAN PEMBAHASAN

Pada bagian ini dibahas mengenai hasil serta analisis model *Extreme Gradient Boosting* sebagai metode pengklasifikasian data protein-protein yang berinteraksi terhadap insulin, dataset yang digunakan pada XGBoost diolah menggunakan metode *Centrality* berdasarkan *Directed Acyclic Graph* (DAG) pada *Gene Ontology*. Hasil dari pengklasifikasian menggunakan XGBoost juga akan dianalisis melalui kinerja model (akurasi, presisi dan recall), analisis efisiensi model, pengaruh fitur terhadap model, dan analisis terhadap pohon CART pada *boosted tree*. Pada bagian ini juga dibahas analisis mengenai pembentukan interaksi antar protein yang mensintesis insulin berdasarkan model XGBoost yang telah terbentuk. Analisis interaksi ini berdasarkan skor prediksi dari pohon-pohon terhadap masing-masing data. Sehingga jika prediksi terhadap satu dengan yang lain mirip atau sangat mendekati maka interaksinya semakin kuat, apabila skor prediksi memiliki jarak yang berbeda jauh maka interaksi itu lemah atau bisa diabaikan. Pada bagian akhir juga dibahas perbandingan hasil interaksi protein insulin yang dibangun berdasarkan model XGBoost dengan hasil interaksi jaringan protein pada perangkat lunak STRING-DB berdasarkan penambangan teks interaksi protein untuk memvalidasi hasil yang telah dibangun.

## 4.1 Data Protein dan Gene Ontologi

Data protein yang diunduh dari uniprot dan PDB adalah data ontologi gen suatu protein tersebut. Data protein yang berpengaruh terhadap insulin sebanyak 1594 protein. Pada penelitian ini ontologi yang digunakan lebih spesifik terhadap fungsi molekul pada suatu protein saja, sehingga tidak semua data protein tersebut memiliki fungsi molekul terhadap insulin. Setelah dilakukan analisis terhadap kode GO *molecular function*, hanya sebanyak 1296 protein yang memiliki fungsi molekul terhadap insulin. Data pada kelas yang lain yakni mengambil protein yang tidak terkait dengan insulin (*non-insulin protein*), ada sebanyak 709 data yang digunakan sebagai dataset pada kelas *non-insulin*.

*Gene Ontology* sendiri merupakan karakteristik protein terhadap 3 sentral metabolisme, yakni proses biologi, fungsi molekul dan komponen seluler. Pada bank data, GO di simpan dalam bentuk kode GO dengan fungsi tertentu. Hirarki kode GO selalu terpusat atau tersentralisasi pada 3 inti metabolisme tadi. Sebagai contoh kode "GO:0005524", yang merupakan kode dari fungsi *"ATP Binding"*. Fungsi pengikat energi *(ATP binding)*, memiliki hirarki berupa *Directed Acyclic Graph* (DAG), seperti pada Gambar 4.1. *Ancestral Chart* selalu berujung pada fungsi molekul dari metablisme. Dari bentuk DAG seperti itu, DAG dianalisis untuk mendapatkan bobot pada setiap simpul graf

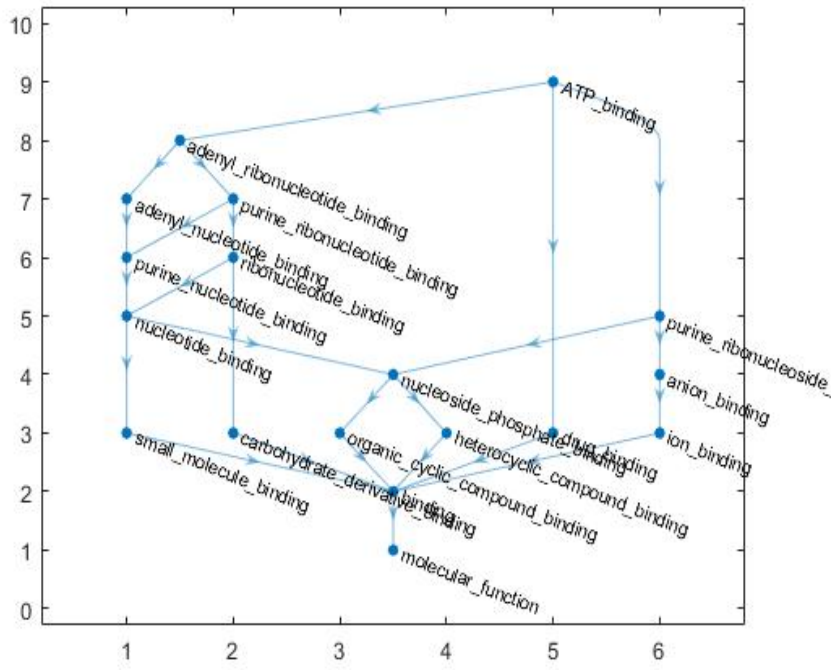



Gambar 4.1: *Ancestor chart* dari suatu GO yang menyatakan fungsi dari *ATP Binding*

untuk mengetahui nilai sentral dari fungsi molekul tersebut.

## 4.2 Hasil Ekstraksi Fitur

Proses penerjemahan DAG pada GO, diawali dengan membuat graf berarah yang sesuai dengan fungsi-fungsi yang berkorespondensi, seperti pada Gambar 4.2. Setelah itu, proses penghitungan skor terhadap masing-masing simpul pada graf dilakukan dengan metode *centrality*. Sebagai contoh, hasil dari penghitungan skor DAG fungsi molekul pada Gambar 4.1, yakni dengan kode GO "GO:0005524" sebagai fungsi *ATP Binding*. Analisis menggunakan tiga metode *centrality* memberikan skor seperti pada Gambar 4.3. Penghitungan bobot dari setiap simpul DAG menggunakan bantuan MATLAB. Dari hasil penyekoran pada Gambar 2.7 tersebut menunjukkan skor *Binding* dari fungsi molekul *ATP Binding* sebesar 0.026843 pada metode *closeness centrality*, 16 menggunakan metode *betweennees centrality*, dan 0.21222 dengan menggunakan metode *pagerank centrality*. Skor atau bobot ini yang dijadikan nilai dari fitur pada saat pembentukan dataset. Nilai dari suatu fitur pada protein menunjukkan pengaruh fungsi dari molekul protein tersebut terhadap fungsi yang lain (sebagai contoh fungsi utama protein *binding* pada Gambar 4.3). Semakin besar nilai dari suatu fitur pada suatu protein, menunjukkan protein tersebut memiliki fungsi yang besar dan juga sebaliknya.




Gambar 4.2: DAG dari fungsi GO *ATP Binding*

Fitur-fitur dari GO fungsi molekul yang digunakan untuk data-data protein tersebut terdiri dari 21 fitur yang

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.02 | 0 | 0.19804 |
| 'binding' | 0.026843 | 16 | 0.21222 |
| 'organic_cyclic_compound_binding' | 0.010381 | 1.6667 | 0.042812 |
| 'heterocyclic_compound_binding' | 0.010381 | 1.6667 | 0.042812 |
| 'ion_binding' | 0.0051903 | 2.6667 | 0.040924 |
| 'small_molecule_binding' | 0.0084775 | 6 | 0.049475 |
| 'nucleoside_phosphate_binding' | 0.012303 | 17.333 | 0.059124 |
| 'carbohydrate_derivative_binding' | 0.0055363 | 6 | 0.030143 |
| 'anion_binding' | 0.0046136 | 2.6667 | 0.027336 |
| 'nucleotide_binding' | 0.0095821 | 25 | 0.074801 |
| 'purine_nucleotide_binding' | 0.007909 | 14.167 | 0.052543 |
| 'ribonucleotide_binding' | 0.0051903 | 11.833 | 0.029306 |
| 'drug_binding' | 0.0034602 | 2 | 0.0227 |
| 'adenyl_nucleotide_binding' | 0.0046136 | 3.3333 | 0.027336 |
| 'purine_ribonucleotide_binding' | 0.0046136 | 11.667 | 0.027336 |
| 'purine_ribonucleoside_triphosphate' | 0.0034602 | 5 | 0.0227 |
| 'adenyl_ribonucleotide_binding' | 0.0034602 | 7 | 0.0227 |
| 'ATP_binding' | 0 | 0 | 0.017687 |

Gambar 4.3: Hasil pembobotan terhadap masing-masing daun pada graf dengan fungsi GO *ATP Binding* menggunakan ketiga metode *centrality*

merupakan fungsi utama dari molekul-molekul protein, yakni

| | | |
|---|---|---|
| $AA$ | : | *Antioxidant Activity* |
| $B$ | : | *Binding* |
| $CAA$ | : | *Cargo Adaptor Activity* |
| $CRA$ | : | *Cargo Receptor Activity* |
| $CC$ | : | *Catalytic Activity* |
| $MCA$ | : | *Molecular Carrier Activity* |
| $MFR$ | : | *Molecular Function Regulator* |
| $MSA$ | : | *Molecular Sequestering Activity* |
| $MTA$ | : | *Molecular Transducer Activity* |
| $NROMF$ | : | *Negative Regulation of Molecular Function* |
| $NRA$ | : | *Nutrient Resevior Activity* |
| $PROMS$ | : | *Positive Regulation of Molecular Function* |
| $PFC$ | : | *Protein Folding Chaperone* |
| $PT$ | : | *Protein Tag* |
| $ROMF$ | : | *Regulation of Molecular Function* |
| $SMSA$ | : | *Small Molecule Sensor Activity* |
| $SMA$ | : | *Structural Molecule Activity* |
| $TxA$ | : | *Toxin Activity* |
| $TRA$ | : | *Transcription Regulator Activity* |
| $TlRA$ | : | *Translation Regulator Activity* |
| $TA$ | : | *Transporter Activity* |

21 fitur yang digunakan untuk mengekstraksi fitur protein berdasarkan ontologi gen pada fungsi molekulnya. Berdasarkan data protein-protein yang mensintesis insulin. Hasil pengekstrakan kode GO pada penelitian ini, terdapat sebanyak 220 kode GO fungsi molekul protein yang terdapat pada protein-protein yang berhubungan dengan insulin. 220 kode tersebut setiap kodenya memiliki salah satu bahkan lebih dari ke-21 fitur protein diatas pada simpul DAGnya. Sehingga ketika protein memiliki fungsi fitur pada kode GO lebih dari satu, maka nilai fitur tersebut adalah jumlahan dari nilai skor dari beberapa GO tersebut. Sehingga semakin besar nilai fitur dari protein tersebut, dapat dikatakan protein yang memiliki fungsi yang menjadi fitur tersebut memiliki pengaruh sangat besar pada fungsi tertentu, begitupun sebaliknya. Karena, metode sentralitas pada graf tersebut, memberikan bobot yang menjadi sentral simpul dari graf.

### 4.3 Hasil Membangun Dataset dengan *Centrality*

Data yang digunakan untuk membentuk dataset sebanyak 1296 protein yang mensintesis insulin dan 709 yang bukan insulin, sehingga total dari seluruh data sebanyak 2005 data. Perbandingan data yang digunakan antara kelas positif dan negatif diusahakan tidak terlalu *inbalance*, karena jika terlalu tidak seimbang memungkinkan model yang terbentuk akan lebih dominan terhadap kelas data yang memiliki jumlah data lebih besar. Perbandingan data 2 : 1 masih termasuk dataset yang seimbang, sehingga dataset dengan 1296 positif dan 709 negatif dapat digunakan untuk proses selanjutnya. Semua data tersebut diambil dari bank data yakni Protein Data Bank (PDB) dan UniProt. Pada setiap protein memiliki fungsi sendiri-sendiri, fungsi yang sama antara protein menunjukkan adanya interaksi dari masing-masing protein tersebut. Hasil ekstraksi fitur tersebut yang diolah menggunakan metode *centrality* menghasilkan 3 dataset, yakni *data_ cc_ insulin.csv*, *data_bc_insulin.csv* dan *data_pr_insulin.csv*.

Dari analisis yang didapat pada dataset terdapat beberapa fitur yang memiliki nilai null (tidak memiliki nilai pada semua data), karena fungsinya dalam proses selanjutnya tidak ada dan hanya akan membuat dimensi dataset semakin besar, maka dataset dikompres dengan menghilangkan fitur yang tidak penting tersebut. Dari 21 fitur yang ada pada *GO molecular function* hanya 11 fitur yang memiliki nilai atau *feature score* yakni B, CRA, CC, MFR, MTA, NROMF, PROMF, ROMF, SMA, TRA, TA dan TA, seperti pada Gambar 4.4. Sedangkan 10 fitur lainnya yakni AA, CAA, MCA, MSA, NRA, PFC, PT, SMSA, TxA dan TlRA tidak memiliki nilai berarti atau bernilai kosong pada semua data. Sehingga 10 fitur tersebut dapat diabaikan, karena tidak berpengaruh atau berarti apapun dalam pembentukan model pada proses klasifikasi. Dataset tersebut disimpan dalam file *Comma Separated Values* (CSV) untuk mempermudah proses selanjutnya.

Visualisasi dataset ditunjukkan pada Gambar 4.4, visualisasi dataset terhadap 10 data teratas. Pada Gambar 4.4 dataset yang terbentuk dibangun menggunakan metode *Closeness Centrality*. Dimensi dataset yakni $X_{2005x11}$,

| index | b | cra | cc | mfr | ... | romf | sma | tra | ta |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | ... | | | | |
| 1 | 0.817346 | 0.0 | 0.108185 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |
| 2 | 0.455543 | 0.0 | 0.000000 | 0.013223 | ... | 0.013774 | 0.0 | 0.00000 | 0.0 |
| 3 | 1.379946 | 0.0 | 0.089289 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |
| 4 | 0.455543 | 0.0 | 0.000000 | 0.013223 | ... | 0.013774 | 0.0 | 0.00000 | 0.0 |
| 5 | 0.323143 | 0.0 | 0.038247 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |
| 6 | 0.841324 | 0.0 | 0.000000 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |
| 7 | 0.328672 | 0.0 | 0.000000 | 0.041446 | ... | 0.043548 | 0.0 | 0.00000 | 0.0 |
| 8 | 0.772972 | 0.0 | 0.147708 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |
| 9 | 0.655606 | 0.0 | 0.000000 | 0.000000 | ... | 0.000000 | 0.0 | 0.39815 | 0.0 |
| 10 | 0.602200 | 0.0 | 0.018896 | 0.000000 | ... | 0.000000 | 0.0 | 0.00000 | 0.0 |

Gambar 4.4: Dataset protein pensintesis insulin dengan ekstraksi fitur *closeness centrality* pada data 10 teratas

baris dataset berupa data protein-protein, sedangkan kolomnya sebagai fitur-fitur.

### 4.4 Analisis dan Hasil Model XGBoost

Model *Extreme Gradient Boosting* dibentuk dari dataset protein berdasarkan GO menggunakan metode *centrality* dengan data sebanyak 2005 data. Model XGBoost mengklasifikasikan data menjadi dua kelas yaitu kelas protein insulin dan kelas protein non-insulin. Sebelum dilakukan proses latih, dilakukan pre-*processing* untuk mempersiapkan dataset agar sesuai dengan format model pelatihan XGBoost. Label dataset diubah menjadi array 01 yang awalnya berupa *string*. Setelah itu data kedua kelas digabungkan sehingga menjadi dataset utuh yang memiliki dua kelas. Dari dataset utuh tersebut dibagi menjadi 2 yakni data latih dan data tes, pemisahan data menjadi dua jenis data dilakukan secara *random* dengan persentase data latih sebesar 80% dari banyaknya data, sedangkan data tes sebesar 20% dari banyaknya data. Data latih digunakan untuk membentuk model dari XGBoost, algoritma dari *ensamble tree* akan belajar dari kompleksitas dataset untuk membuat model. Untuk memvalidasi model yang telah dibangun, metode validasi dengan menggunakan *cross validation* dilakukan terhadap data latih sebanyak 10-*fold*, tujuannya untuk menguji performa dari model terhadap integritas pada data latih. Data tes digunakan untuk melihat bagaimana kinerja model terhadap data yang berbeda dengan data latih, atau bisa dikatakan bagaimana

model yang sudah terlatih dapat memprediksi hasil dengan benar dari suatu inputan yang belum dipelajari oleh model.

Dataset pada Gambar 4.4 merupakan dataset dari interaksi protein yang terbangun dari *molecular function* ontologi gen. Kolom data menyatakan fitur-fitur fungsi molekul data pada protein, sedangkan barisnya merupakan banyaknya data protein yang digunakan. Dalam persamaan yang terdapat pada algoritma XGBoost, data tersebut mula-mula di cari nilai $log(odds)$, yakni sebesar,

$$log(odds) = \frac{n_{kelaspositif}}{n}$$

$$log(odds) = \frac{1296}{2005} = 0.64638$$

nilai $log(odds)$ ini digunakan untuk menentukan probabilitas awal dari semua kelas, sehingga untuk kelas positif probabilitas awalnya adalah $1 - log(odds) = 0.35362$, sedangkan untuk kelas negatif $0 - log(odds) = -0.64638$. Nilai probabilitas awal ini digunakan untuk mendapatkan skor pada node pohon keputusan awal, untuk mendapatkan *gain* yang memiliki nilai maksimum. Sehingga didapatkan struktur pohon keputusan yang optimal. Untuk proses selanjutnya dari algoritma XGBoost, sama seperti pada contoh penyelesaian pada Sub Bab 2.2.3.

### 4.4.1 Struktur Model XGBoost

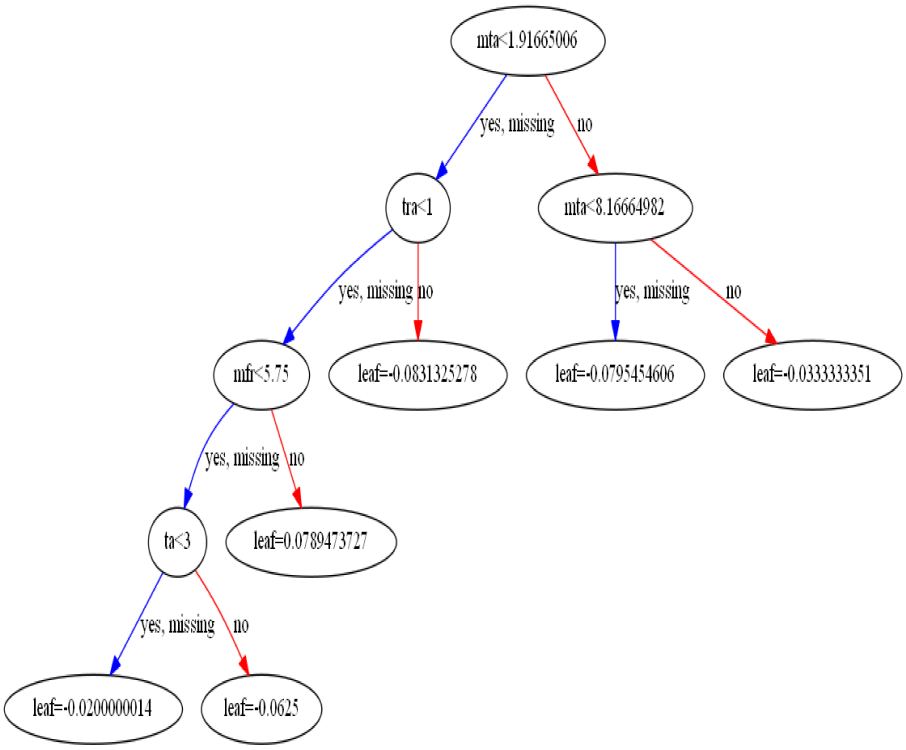
Struktur model dari *Extreme Gradient Boosting* merupakan peningkatan kualitas pohon keputusan pada setiap pembentukan pohon selanjutnya. Nilai korespondensi dari nilai prediksi merupakan jumlahan skor daun dari setiap pohon keputusan yang dikalikan dengan nilai $learning\_rate$.

$$\hat{y} = \alpha f_1(x_i) + \alpha f_2(x_i) + ... + \alpha f_m(x_i)$$

$$\hat{y} = \alpha \sum_{t=1}^{m} f_t(x_i)$$

fungsi $f_t(x_i)$ memetakan data $x_i$ terhadap posisi pada suatu daun di pohon keputusan ke $t$.

Pohon *Classification and Regression Tree* (CART) merupakan pohon-pohon keputusan yang terbentuk ketika pembentukan model, setiap pohon CART mewakili suatu pohon keputusan yang merupakan pengklasifikasi lemah. Sistem *boosting* / peningkatan menggabungkan pengklasifikasian yang lemah dari satu pohon keputusan dengan pohon keputusan yang lain. Setiap pembentukan pohon CART yang baru, klasifikasi yang terbentuk akan semakin meminimalkan fungsi kerugian yang dihasilkan dari klasifikasi. Pada penelitian ini dibangun sebanyak 1000 pohon CART yang merupakan model yang telah terbentuk dengan XGBoost. Kerena bahasa pemrograman memulai angka dari 0, maka pohon yang terbentuk antara $0 - 999$. Sebagai contoh untuk menggambarkan bentuk dari pohon CART yang dihasilkan dari model divisualisasikan seperti pada Gambar 4.5, Gambar 4.6, Gambar 4.7, Gambar 4.8, Gambar 4.9, dan Gambar 4.10.




Gambar 4.5: sub pohon CART ke-0 yang dibentuk oleh model klasifikasi XGBoost

Gambar 4.5 sebagai pohon CART pertama yang dibentuk. Pohon keputusan tersebut menunjukkan fitur *mta* sebagai *root*. Karena fitur *mta* memiliki kelompok besar yang mudah di pisahkan dan juga algoritma XGBoost mencari berulang-ulang struktur pohon yang memiliki fungsi kerugian paling kecil dengan menggunakan persamaan $\mathcal{L}_{split}$, yakni,

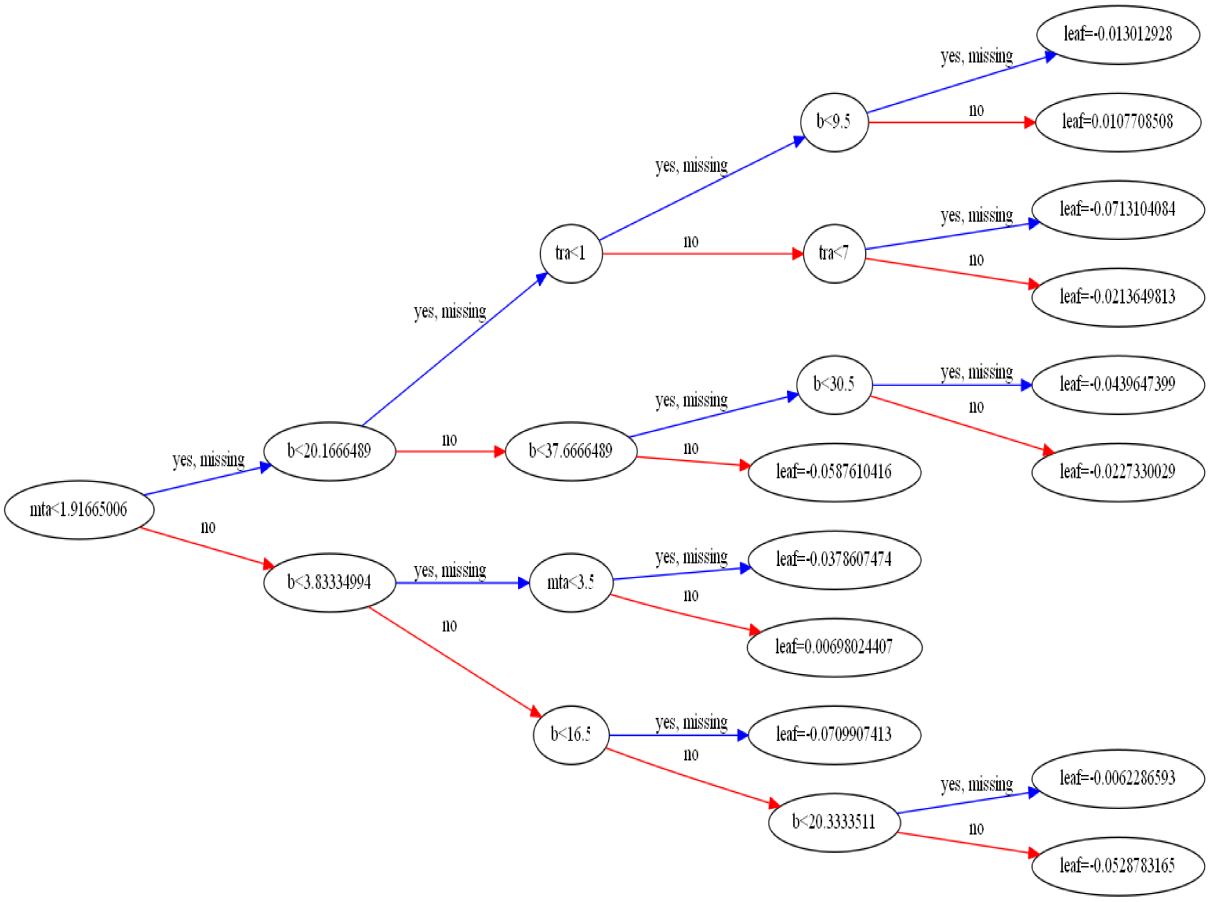
$$\mathcal{L}_{split} = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right]$$

*gain* terbesar dari suatu pohon keputusan memiliki fungsi kerugian yang kecil. Sehingga, algoritma XGBoost mencari terus menerus konstruksi pohon keputusan yang simpulnya memiliki *gain* terbesar dari simpul *root* hingga daun.

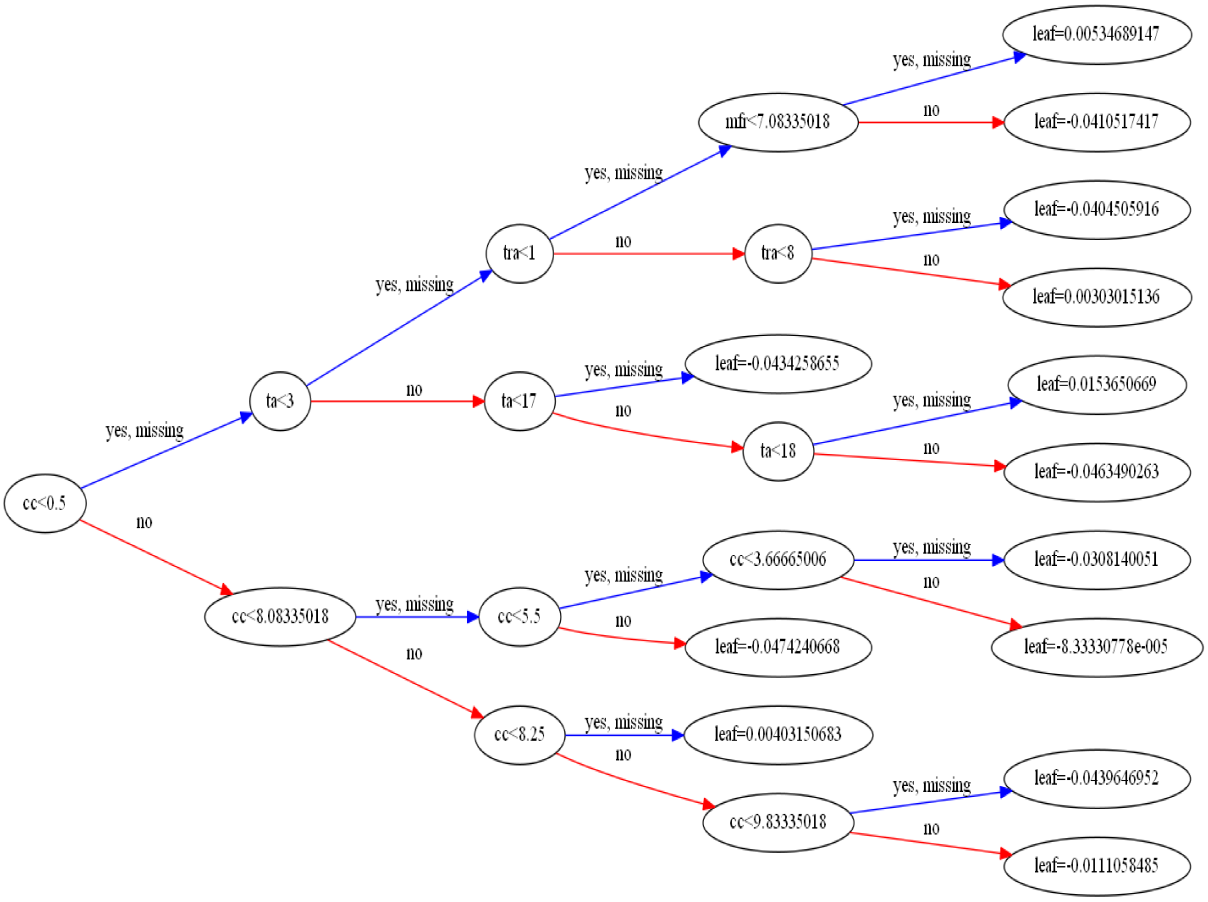
Pada pohon CART ke-9 pada Gambar 4.6, *root* dari pohon tersebut adalah fitur *Molecular Transducer Activity*. Fitur yang terdapat pada *root* dari suatu pohon mengartikan bahwa fitur tersebut memiliki pengaruh paling besar dalam pembentukan pohon pengklasifikasian. Sedangkan fitur yang berada pada *node* yang dekat dengan daun, pengaruhnya terhadap pembentukan pengklasifikasian kecil.

### 4.4.2 Parameter Optimum XGBoost

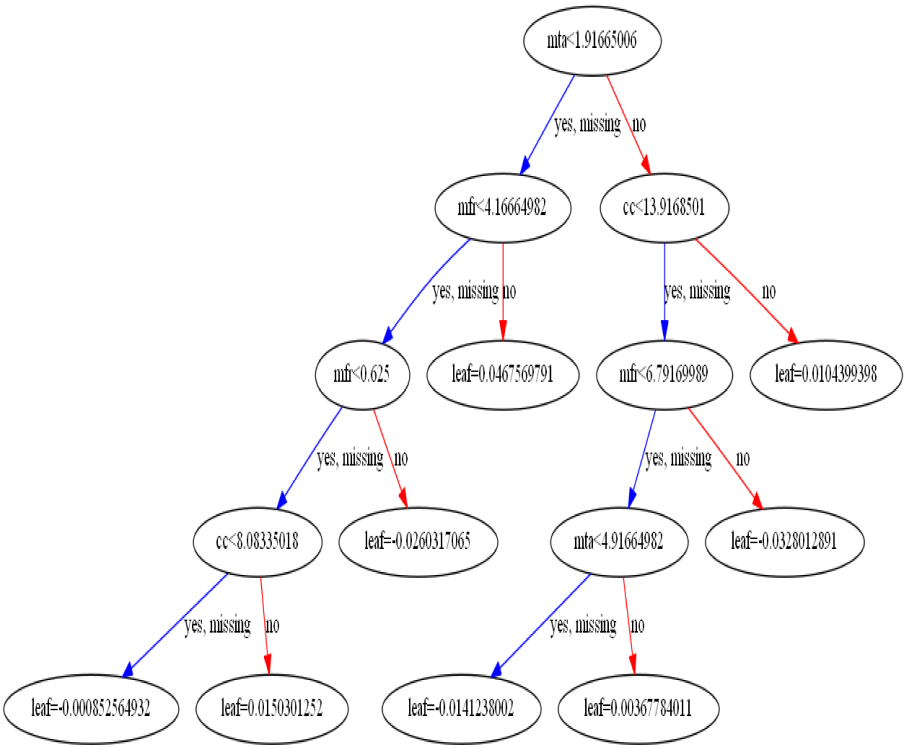
Model dari XGBoost didefinisikan sebagai jumlahan dari bobot daun pada pohon CART (*Classification and Regression Tree*). Model pada setiap sub



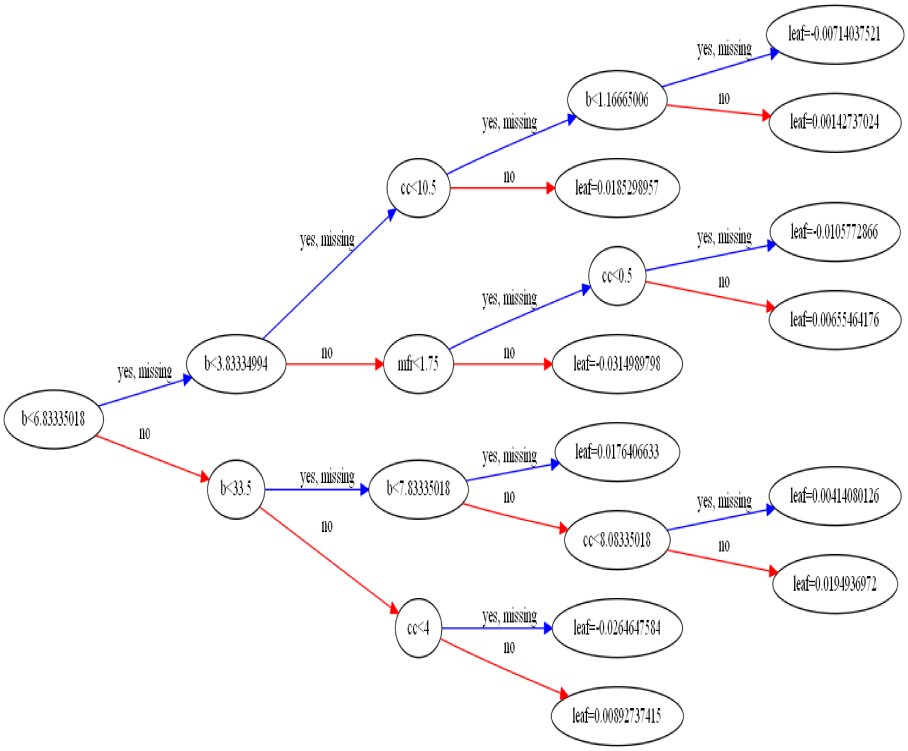

Gambar 4.6: sub pohon CART ke-9 yang dibentuk oleh model klasifikasi XGBoost



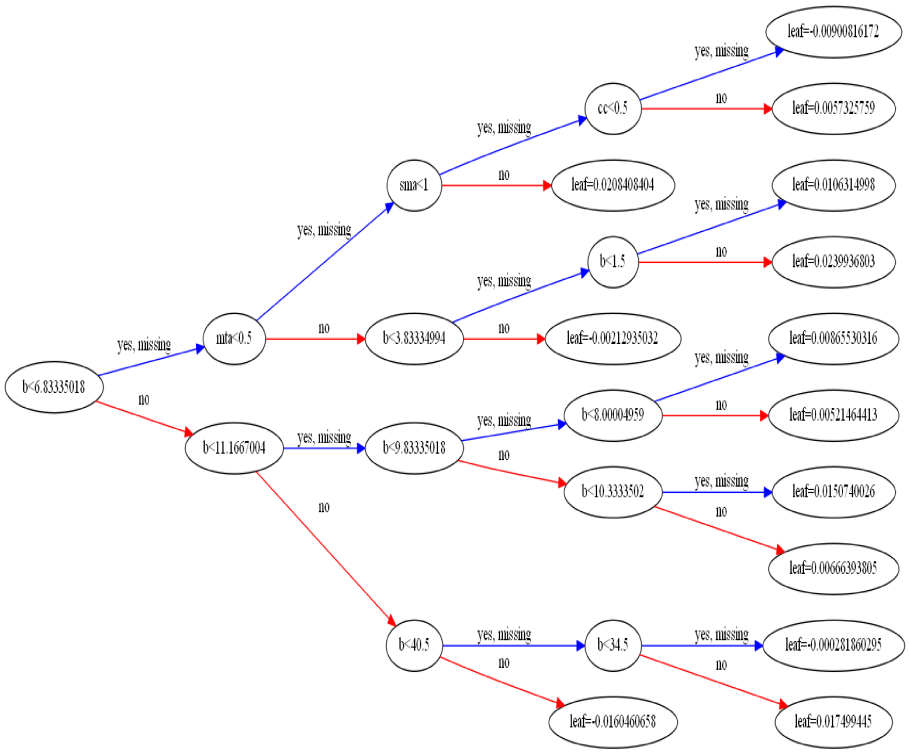

Gambar 4.7: sub pohon CART ke-49 yang dibentuk oleh model klasifikasi XGBoost




Gambar 4.8: sub pohon CART ke-99 yang dibentuk oleh model klasifikasi XGBoost




Gambar 4.9: sub pohon CART ke-499 yang dibentuk oleh model klasifikasi XGBoost




Gambar 4.10: sub pohon CART ke-999 yang dibentuk oleh model klasifikasi XGBoost

puhon CART merupakan pengklasifikasian yang jelek, sehingga XGBoost membangun model pengklasifikasian yang bagus dari klasifikasi pohon yang jelek.

Pada saat melatih model XGBoost, perlu memperhatikan pengaturan parameternya. Meskipun XGBoost sudah menentukan bentuk parameter standarnya, tetapi parameter standar belum tentu memenuhi hasil yang diinginkan. Parameter dari XGBoost ditentukan seperti pada Tabel 4.1, parameter-parameter ini ditentukan setalah melakukan beberapa kali pengujian terhadap efektifitas dan akurasi dari model XGBoost, kecuali parameter *n_estimaror* dan *max_depth*. Nilai parameter yang diuji ulang-ulan berdasarkan batas-batas yang berlaku pada algoritma. Sehingga didapatkan parameter optimum seperti pada Tabel 4.1.

Nilai $n_estimator$ diatas merupakan ujicoba awal pada pembentukan model, jika model yang terbentuk terlalu *overfitting* maka besar nilai dari $n_estimator$ bisa diperkecil untuk mendapatkan model yang lebih baik. Model yang baik tidak terjadi kondisi *overfitting* maupun kondisi *underfitting*, sehingga pada uji ROC luasan dibawah grafik tidak terlalu keatas dan juga tidak terlalu dekat dengan batas bawah.

Selain parameter model, juga terdapat parameter uji dari suatu model. Sebagai standar pengujian, parameter akurasi, presisi dan recall merupakan pengujian yang wajib untuk mengetahui kelayakan model yang terbentuk. Nilai akurasi digunakan untuk mengetahui kinerja model dalam

Tabel 4.1: Parameter-parameter yang digunakan pada model XGBoost

| **Parameter** | **Diskripsi** | **Nilai** |
|---|---|---|
| n_estimator | Banyaknya pohon CART yang dibangun dari data | 1000 |
| max_depth | Kedalaman maksimum pada setiap pohon | 5 |
| gamma | $[0, +\infty]$,mengatur banyaknya simpul pada pohon | 0 |
| learning_rate | $[0, 1]$ rasio model dalam mempelajari data | 0.05 |
| subsample | $(0, 1]$ rasio dari contoh pelatihan, membuat pelatihan *robust* terhadap *noise* | 0.8 |
| colsample_bytree | $(0, 1]$ rasio dari colom fitur ketika membentuk setiap pohon, membuat pelatihan lebih *robust* dari *noise* | 0.8 |

mengklasifikasikan setiap data dengan benar terhadap kelas sebenarnya. Nilai presisi digunakan untuk mengetahui tingkat ketepatan antara kelas asli dan terhadap kelas setelah diprediksi dengan model. Sedangkan parameter recal digunakan untuk mengetahui tingkat keberhasilan sistem dalam menemukan kembali informasi secara benar.

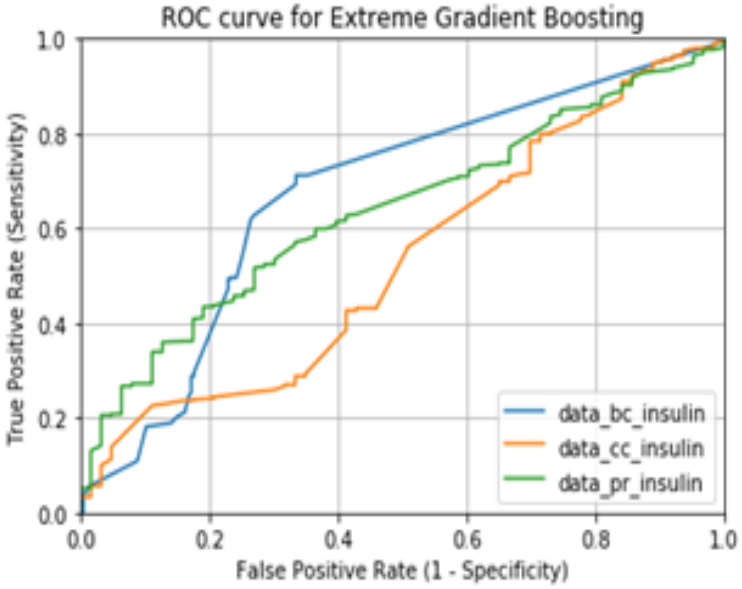
### 4.4.3 Perbandingan Kualitas Dataset

Dataset yang digunakan pada penelitian ini merupakan salah satu dari ketiga dataset yang dibangun. Dikarenakan data yang digunakan sama, akan tetapi dalam proses pemberian bobot di fitur proteinnya yang berbeda. Dataset yang digunakan merupakan data yang memiliki akurasi dan presisi paling bagus diantara yang lain, karena proses selanjutnya membutuhkan keakuratan model dalam memprediksi kedekatan data satu dengan data yang lain sebagai bentuk interaksi. Analisis dari hasil akurasi ketiga dataset juga dapat dijadikan sebagai referensi metode yang lebih bagus digunakan untuk mengekstraksi ciri fungsi protein berdasarkan gen ontologi yang berbentuk DAG. Hasil akurasi dari ketiga dataset dapat dilihat pada Tabel 4.2 yang menunjukkan akurasi yang dihasilkan dengan dataset yang dibangun oleh metode *Betweenness Centrality* lebih bagus dibandingkan dengan metode yang lain. Akurasi dataset yang dibangun dengan metode *Closeness Centrality* $data_bc_insulin$ pada proses *training* dengan validasi menghasilkan akurasi sebesar 73.48%, sedangkan pada proses prediksi atau tes sebesar 74.56%, serta memiliki presisi model sebesar 85.33%, yang mengartikan pengukuran data satu sama lain sangat dekat. Hasil dari pengujian dataset yang dibangun dengan metode *Betweenness Centrality* $data_cc_insulin$ pada proses *training* dengan validasi menghasilkan akurasi sebesar 73.30%, sedangkan pada proses prediksi atau tes sebesar 72.05%, serta memiliki presisi sebesar 83.32%. Hasil pengujian pada dataset yang dibangun menggunakan metode *Page Rank Centrality* $data_pr_insulin$ pada proses pelatihannya dapat memvalidasi dengan akurasi sebesar 72.55%, sedangkan pada proses prediksi atau tes sebesar 74.39%, serta memliki presisi sebesar 85.84%. Presisi dari semua dataset menunjukkan setiap data yang digunakan memiliki kedekatan yang besar. Hasil pengujian tersebut juga menunjukkan bahwa perbedaan yang

sangat kecil dari akurasi ketiga dataset, yang berarti menunjukkan metode *centrality* bagus digunakan untuk mengekstraksi ciri dari suatu GO fungsi molekul protein yang berupa DAG. Akan tetapi, untuk proses selajutnya pada penelitian ini hanya menggunakan dataset yang ekstraksi cirinya menggunakan metode *Betweenness Centrality* yang memiliki akurasi lebih bagus daripada dataset yang lain.

Tabel 4.2: Hasil pengukuran perbandingan dari ketiga dataset dengan klasifikasi XGBoost

| No | Dataset | Akurasi | | Presisi | Recall |
|---|---|---|---|---|---|
| | | Validasi | Test | | |
| 1 | $data_bc_insulin$ | 73.48% | 74.56% | 85.33% | 80.65% |
| 2 | $data_cc_insulin$ | 73.30% | 72.05% | 83.32% | 80.9% |
| 3 | $data_pr_insulin$ | 72.55% | 74.39% | 85.84% | 81.01% |



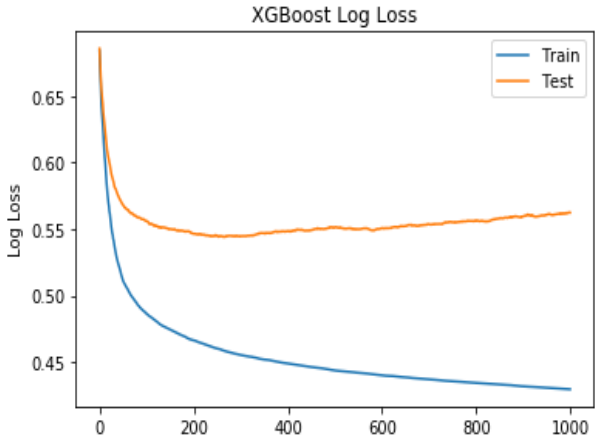

Gambar 4.11: Kurva ROC dari ketiga dataset terhadap model XGBoost

Hasil pemilihan dataset yang digunakan sebagai model interaksi diperkuat dengan menganalisis kurva ROC. Kurva ROC di Gambar 4.11 menunjukkan area di bawah kurva masing-masing model yang dibangun menggunakan model XGBoost. Nilai ROC diperoleh dari Area Under Curve (AUC), Semakin besar nilai AUC adalah model yang lebih baik (C. Marzban, 2004). Kurva $data_bc_insulin$ memiliki area terbesar daripada kurva lainnya atau memiliki nilai 0,6742, sedangkan $data_cc_insulin$ dan $data_pr_insulin$ masing-masing memiliki nilai 0,533 dan 0,6383. Titik-titik dalam grafik ROC menggambarkan semua kemungkinan TP dan FP jika kita menjalankan ambang batas dari bawah hingga atas.

### 4.4.4 Analisis Kinerja Model

Kinerja dari model yang dibangun dapat dilihat diri bagaimana model dapat meminimalkan *error* ketika mengklasifikasikan, karena fungsi objektif dari model XGBoost adalah meminimalkan fungsi kerugian dan fungsi kompleksitas mdel. Pada setiap model pohon CART yang dibentuk memiliki nilai kesalahan yang bervariasi, ketika pembentukan pohon pertama akan memiliki nilai kesalahan yang lebih besar dibandingkan pohon-pohon berikutnya, sedangkan seiring berjalannya iterasi pembentukan pohon sampai yang ditentukan, nilai *error* model akan semakin kecil hingga model memiliki *error* konstan. Pengujian kinerja terhadap XGBoost dalam membangun model pohon CART divisualisasikan pada Gambar 4.12. Grafik *XGBoost Log Loss* menunjukkan *error* yang terjadi pada saat pembentukan pohon ke-0 sampai pohon ke-1000. Pada grafik tersebut menunjukkan *error* pada proses pelatihan menghasilkan nilai yang semakin kecil setiap kali model membangun pohon-pohon baru, sehingga terbukti dalam memodelkan data XGBoost berusaha meminimalkan fungsi kerugian yang digunakan. Sedangkan pada proses prediksi menggunakan data tes, sama seperti pada saat proses pembelajaran (pelatihan) nilai *error* pada pembentukan pohon CART ke-0 memiliki nilai kesalahan lebih besar dari pada pohon CART yang dibangun lainnya. Hal ini wajar karena pembentukan pohon pertama merupakan pengklasifikasian yang lemah dengan pohon keputusan. Seiring bertambahnya pohon yang dibentuk, nilai *error* yang dihasilkan menjadi menurun. Akan tetapi, pada nilai pohon ke-n tertentu, nilai error pada grafik mulai mengalami kenaikan kembali, hal itu diwajarkan karena model berusaha menerima semua spesifikasi fitur yang terdapat pada dataset. Jika dilihat pada grafik, selisih *error* latih dan *error* tes tidak lebih dari 0.2. Sehingga model ini dengan menggunakan $n_estimator = 1000$ tidak terlalu overfitting terhadapa data. Dari grafik tersebut dapat dianalisis bahwa XGBoost mampu membuat model klasifikasi yang cukup baik dengan pembentukan pohon keputusan yang dapat memperkaya pembelajaran pada model. Nilai *error* prediksi yang besar bisa jadi dikarenakan kompleksitas model dalam pembelajarannya tersebut sehingga semaikin kompleks pohon CART yang terbentuk akan semakin besar pula daya selektif model terhadap prediksi data.

Pada Gambar 4.13 menunjukkan besaran kesalahan dalam mengklasifikasikan pada setiap pohon CART. Pada proses pelatihan model dengan memvalidasi hasil menunjukkan kesalahan klasifikasi pada data latih cenderung menurun pada setiap pembentukan pohon baru pada model. Sedangkan pada grafik data tes, kesalahan dalam mengklasifikasikan menunjukkan semakin besar pada setiap pembentukan pohon baru di model yang terbentuk pada saat pelatihan. Sama seperti analisis terhadap fungsi kerugian yang dihasilkan, kesalahan prediksi yang meningkat dikarenakan kompleksitas model. Seiring bertambahnya pohon yang digunakan galat yang terjadi pada proses latih semakin menurun sedangkan pada proses tes semakin meningkat. Keadaan ini dapat menyebabkan model yang *overfitting*, Model yang *overfitting* memiliki kerugian yang rendah selama

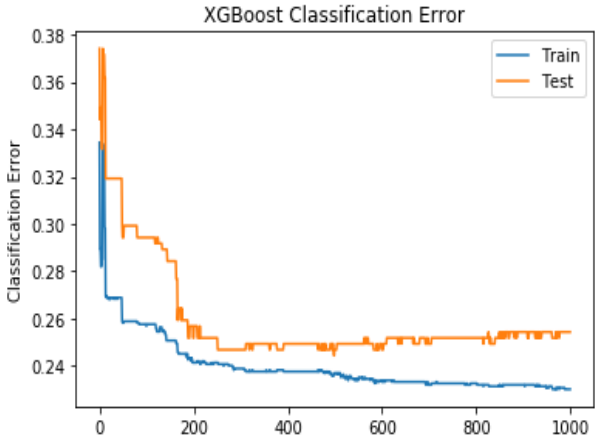



Gambar 4.12: Error pada fungsi kerugian terhadap masing-masing pohon CART, yang menunjukkan XGBoost membangun klasifikasi yang kuat dari pengklasifikasian lemah pada pohon sub-CART

pelatihan tetapi berfungsi dengan buruk saat memprediksi data baru. (Xue Ying, 2019) Model yang *overfitting* memiliki *bias* kecil dan varian yang besar. Akan tetapi, jika model terlalu sederhana memungkinkan model bersifat *underfitting* yakni model memiliki *bias* tinggi dan vasian yang rendah. Kedua keadaan ini bisa membuat model menjadi buruk, sehingga pada kebanyakan peneliti dua keadaan ini cenderung dihindari. Untuk mendapatkan model yang bagus dalam artian tidak mengalami *overfitting* maupun *underfitting*, maka pembentukan pohon CART dapat dibatasi pada keadaan galat yang setimbang yakni tidak terlalu kompleks dan tidak terlalu sederhana. Dari Gambar 4.13 dapat dianalisa bahwa model yang terbentuk tidak terlalu terjadi overfitting yakni memiliki selisisih sebesar 0.07345 eror antara *error* kelas data latih dan *error* kelas data tes, sehingga model sudah terbentuk dengan baik.

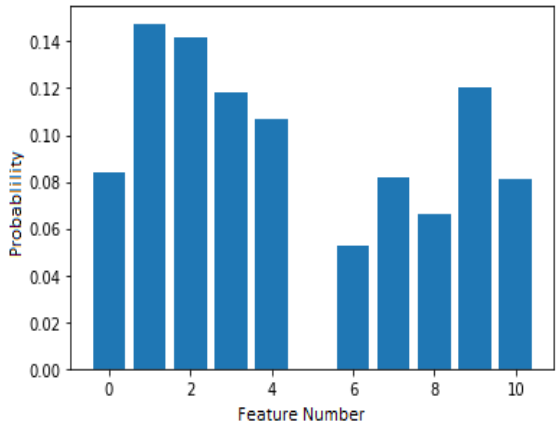
### 4.4.5 Analisis Pengaruh Fitur

Model dibangun berdasarkan nilai fitur-fitur yang dipelajari dari setiap data-data pada dataset. Fitur-fitur tersebut merupakan variabel *dependent* yang berperan penting dalam pembentukan model klasifikasi. Fitur-fitur yang dijadikan sebagai variabel memiliki pengaruh yang besar terhadap model yang terbentuk. Semakin besar keberagaman nilai pada suatu fitur semakin besar pula pengaruh yang diberikan terhadap pembentukan model. Jika nilai yang ada didalam fitur homogen maka fitur tersebut kurang terlalu berpengaruh terhadap pembentukan model klasifikasi dari XGBoost. Pada Gambar 4.14 menunjukkan probabilistik dari masing-masing fitur, total probabilistik dari semua fitur adalah 1. Maksudnya, jika probabilitas ke 11 nilai fitur tersebut dijumlahkan, nilai totalnya adalah 1. Pada Gambar 4.14 menunjukkan fitur ke-1 memiliki probabilistik lebih besar daripada fitur yang lain yakni sebesar 0.1476159. Nilai tersebut menyatakan fitur ke-1 lebih sering digunakan atau muncul (sebagai *node* pada pohon) sebagai pengagregat




Gambar 4.13: Nilai error pada data latih dan data test pada setiap sub pohon CART pada model pengklasifikasian XGBoost

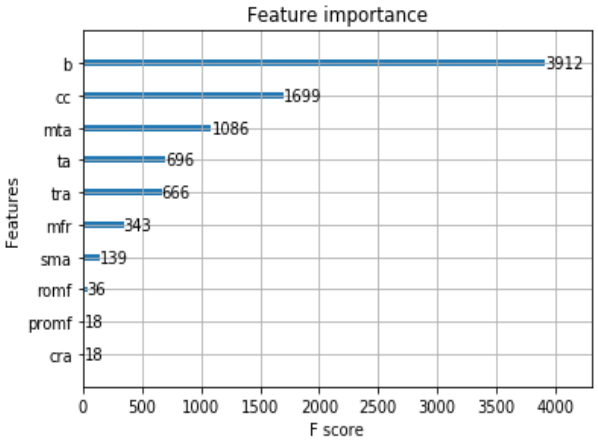
data pada pohon CART. Jika diurutkan, tiga probabilitas yang memiliki nilai terbesar yakni fitur ke-1, ke-2 dan ke-9, yang memiliki masing-masing nilai sebesar 0.1476159, 0.14203338, dan 0.1199996. Besaran dari probabilitas fitur ini digunakan sebagai seberapa sering fitur-fitur tersebut digunakan dalam pembentukan model XGBoost. Fitur ke-5 memiliki probabilitas 0.00, nilai tersebut mengartikan bahwa pada semua pohon CART yang terbentuk, kedua fitur tersebut tidak digunakan (sebagai *node* pada pohon) sebagai pengagregat data, sehingga fitur-fitur tersebut tidak memiliki pengaruh yang besar dalam pembentukan model. Keadaan seperti itu bisa disebabkan karena nilai yang ada dalam kedua fitur tersebut terlalu homogen.




Gambar 4.14: Fitur-fitur yang berpengaruh dalam membentuk klasifikasi model XGBoost

*F-score* atau *Feature Score* merupakan suatu nilai yang menunjukkan kedudukan fitur tersebut memiliki nilai yang sangat penting terhadap pembentukan model. Nilai *F-score* yang besar menunjukkan fitur tersebut

memiliki pengaruh yang signifikan dalam membentuk model klasifikasi. Pengaruh tersebut menjadikan fitur yang bernilai *F-score* tinggi lebih sering menjadi *root* atau pada level *node* yang dekat dengan *root* pada setiap pohon CART yang terbentuk. *F-score* yang kecil memiliki kemungkinan yang kecil pula sebagai agragator pada *root*, niali yang kecil akan lebih sering berada pada *branch tree* atau pada kedalaman pohon yang besar pula. Pada Gambar 4.15 menunjukkan *F-score* dari masing-masing fitur. Fitur *Binding* memiliki skor yang sangat besar dibandingkan yang lainnya, skor sebesar 3912 hampir $2.0x$ lipat dari fitur dengan skor tertinggi kedua, yang berarti *Binding* memiliki peranan lebih banyak menjadi *root* pada *boosting* pohon keputusan dalam membentuk model dari pengklasifikasian menggunakan metode XGBoost. Pada urutan ke-2 fitur yang memiliki peranan dalam membentuk model adalah *Catalytic Activity* dengan skor sebesar 1699. Fitur yang memiliki $F - score$ kecil, menjadi *node* pengagregat pada level yang lebih besar.




Gambar 4.15: Skor fitur yang berpengaruh terhaadap model pelatihan XGBoost

## 4.5 Membangun Interaksi Pada Setiap Protein

Interaksi protein ditentukan oleh kemiripan fungsi-fungsi pada dua atau lebih molekul protein. Pada bentuk model klasifikasi XGboost, kemiripan fungsi protein bisa diterjemahkan sebagai posisi daun pada setiap pohon CART. Jika suatu protein memiliki kemiripan fungsi yang sangat mirip maka posisi dari protein tersebut (daun) akan sangat dekat bahkan bisa jadi kedua protein memiliki nilai probabilistik pada daun yang sama. Sebaliknya, jika 2 protein tidak memiliki kemiripan fungsi satu sama lain maka posisi daun dipisahkan dengan tingkatan *node* yang jauh berbeda. Sehingga nilai probabilitas pada prediksi model XGBoost digunakan untuk menentukan kedekatan antara protein Insulin dengan protein yang lain.

### 4.5.1 Jarak pada Pohon Keputusan

Nilai probabilitas memiliki selisih yang sangat kecil, sehingga skor-skor tersebut perlu ditransformasi menjadi rasio *odds*,Rasio ini digunakan untuk

mengetahui dua *outcome* suatu variabel biner, yakni rasio dimana probabilitas terjadinya suatu kejadian dan tidak terjadi. Nilai ini sejalan dengan model prediksi dari XGBoost yang merupakan bentuk klasifikasi dari dua kelas, yakni kelas protein yang berinteraksi dan kelas protein yang tidak berinteraksi. Nilai *odds* didapatkan dengan menghitung,

$$O(y) = \frac{p(y)}{1 - p(y)}$$

dari hasil perhitungan nilai tersebut didapatkanlah sebaran nilai dari setiap probabilitas protein. Nilai dari *odds* tidak lagi berada pada selang $0 \leq p \leq 1$, sehingga bisa dilakukan pengukuran jarak dari setiap data.

Penghitungan jarak pada setiap data terhadap protein Insulin, dengan mendapatkan nilai selisih antara *odds* insulin dengan *odds* protein yang lain. Penghitungan jarak menggunakan metode jarak Manhattan, yakni dengan memutlakkan selisih nilai dari dua data,

$$d_{xy} = |o(x) - o(y)|$$

nilai selisih tersebut didapatkan untuk mengetahui jarak terdekat antar protein, yang berarti semakin kecil nilai jarak antar protein satu dengan yang lain maka kemiripan antara kedua protein tersebut sangat tinggi. Sebaliknya jika terdapat jarak yang sangat besar maka dua protein yang dihitung jaraknya tersebut memiliki kemiripan yang kecil.

Nilai rasio kemiripan digunakan untuk mengetahui seberapa besar dia mirip dengan fungsi molekul pada Insulin. Karena rasio identik dengan nilai persen atau pada selang $0 - 100$, maka dibutuhkan transformasi nilai jarak yang awalnya jarak terkecil menjadi mirip, pada rasio ini jarak terkecil ersebut diubah menjadi rasio terbesar, sehingga dapat merepresentasikan kemiripan fungsi molekul berdasarkan besarnya rasio persentase. Akan tetapi karena perhitungan jarak tersebut menghasilkan distribusi nilai yang random, maka untuk mendapatkan rasio yang setara susah didapatkan. Alternatifnya semua nilai jarak dinormalisasi terlebih dahulu sehingga data akan berada pada selang terbatas $[0, 1]$. Sehingga, perhitungan nilai rasio menjadi,

$$Rasio(y_i) = \left(1 - \left(\frac{y_i - min(y)}{max(y) - min(y)}\right)\right) * 100\%$$

### 4.5.2 Threshold Interaksi

Hasil dari penghitungan rasio kemiripan ini baru bisa dilakukan *Threshold* untuk protein-protein yang diprediksi berpengaruh atau memiliki kemiripan fungsi molekul dengan Insulin. Nilai *threshold* yang digunakan untuk rasio kemiripan protein sebesar 90%, jika nilai ratio diatas 90% maka protein tersebut memiliki pengaruh besar terhadap insulin dan berinteraksi, pada selang *threshold* $[60\% - 89\%]$, dinilai protein tersebut memiliki pengaruh sedang. Sedangkan jika nilai rasionya dibawah 60% protein tersebut tidak diinteraksikan atau bisa dikatakan memiliki pengaruh yang kecil terhadap produksi Insulin. Nilai *threshold* diatas berdasarkan batasan yang ada pada

STRING-DB, sehingga hasil mambangun interaksi dengan XGBoost dapat dibandingakan dengan analisis *Text Mining* interaksi antar protein pada STRING-DB.

### 4.5.3 Hasil Interaksi

Hasil dari *threshold* tersebut dijadikan sebagai bahan untuk membentuk jaringan interaksi pada Insulin. Jaringan interaksi protein Insulin yang terbentuk seperti pada Gambar 4.16. Jaringan interaksi pada insulin dibangun dengan menggunakan aplikasi graph yaitu Gephi 0.9.3. Data hasil perhitungan rasio dan *threshold* di masukkan kedalam Gephi untuk menghasilkan jaringan interaksi. Pada jaringan interaksi tersebut terdapat 1297 *node* dan 11538 *edge* yang merupakan representasi bahwa dua protein atau lebih memiliki interaksi. Jika *node* pada jaringan tersebut tidak memiliki *edge*, maka protein yang menjadi *node* tersebut tidak berinteraksi dengan manapun. Protein insulin dinyatakan dengan kode $P01308$, yang merupakan kode *entry* protein pada Protein Data Bank (PDB). Kode-kode yang memiliki pengaruh besar terhadap produksi insulin digambarkan dengan dominasi ukuran *node*nya. Semakin besar ukuran simpul (label) pada Gambar 4.16, maka pengaruhnya terhadap insulin juga akan semakin besar. Pada Gambar 4.16 dapat diketahui bahwa 9 protein yang berpengaruh sangat besar terhadap insulin, dan terdapat sekitar 18 protein yang memiliki pengaruh cukup besar terhadap insulin. Penjelasan lebih rinci terhadapa 9 protein yang berpengaruh sangat besar terhadap insulin dapat dilihat pada Tabel 4.3,dan protein yang berpengaruh cukup besar dapat dilihat pada Tabel 4.4,

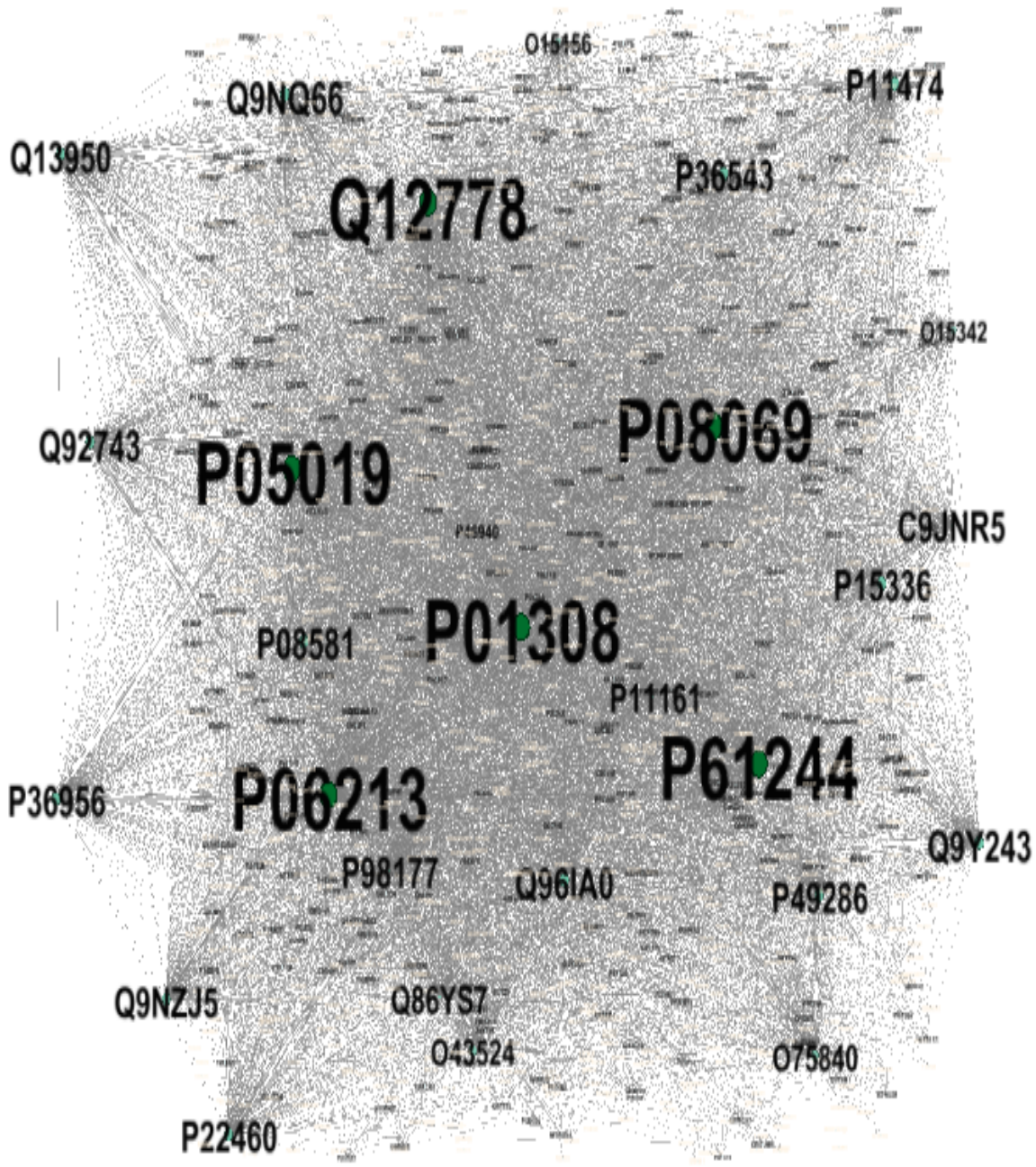
Dari Tabel 4.3 protein yang berpengaruh terhadap Insulin kebanyakan memiliki fungsi yang sama yakni pada transkripsi proteinnya. Hal itu membuktikan proses sintesis Insulin sebagai produk dari genotipe gen di bantu oleh protein lain pada saat transkripsi DNA menjadi asam amino calon Insulin. Selain faktor transkripsi sebagai pembentuk sintesis Insulin protein seperti INSR, IGFR1, dan IGF1 merupakan protein penting yang mempengaruhi hampir seluruh metabolisme yang mengaktifkan fungsi Insulin.

Tabel 4.3 dan Tabel 4.4 menunjukkan beberapa protein yang berpengaruh terhadap insulin, rasiorasio dari protein tersebut menunjukkan kedekatan atau kemiripan fungsi molekul dengan Insulin, sedangngkan *linkedge* menunjukkan protein tersebut berinteraksi dengan protein yang lain berdasarkan *threshold* yang digunakan.

## 4.6 Hasil Text Mining pada STRING-DB

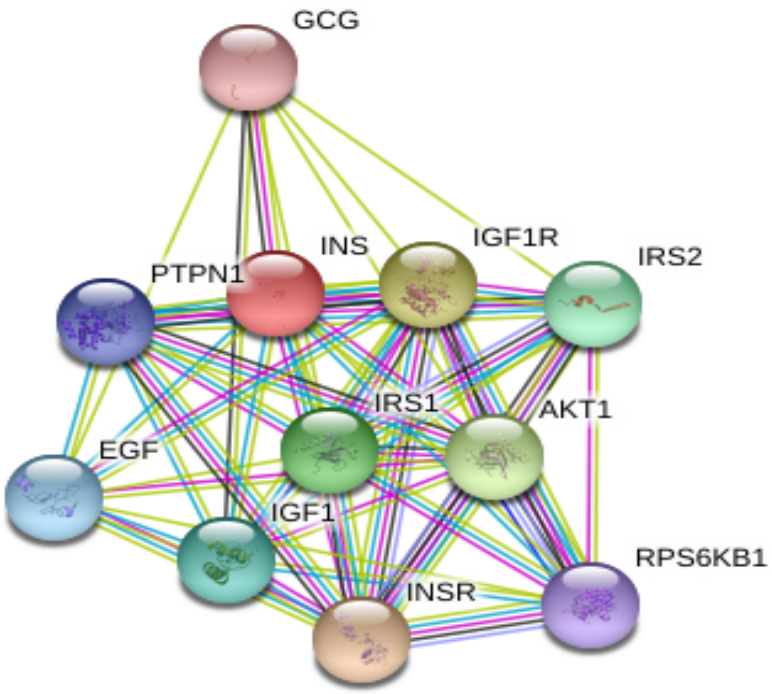
## 4.7 Perbandingan Hasil Interaksi XGBoost dengan *Text Mining* pada STRING Database

Penambangan teks merupakan proses mengeksplorasi data text yang dapat mengidentifikasi pola, konsep, topik, kata kunci sehingga menghasilkan informasi yang mudah diterjemahkan. Penambangan teks dalam penerjemahan interaksi proteinmenggunakan penambangan teks yang ada pada database STING-DB. Penggunaan perangkat lunak ini digunakan sebagai pembanding hasil yang didapatkan dari pembangunan interaksi protein yang




Gambar 4.16: Jaringan interaksi protein pada insulin yang terbentuk

dihasilkan dari model XGboost. Penambangan teks pada STRING-DB dapat diakses pada halaman https://string-db.org/, setelah masuk kehalaman utama masukkan kode protein insulin dengan kode INS atau langsung dengan kata kunci "Insulin", serta organisme "Human". Dengan menggunakan penambangan teks pada STRING-DB, didapatkan terdapat 11 protein yang berpengaruh besar terhapda insulin. Jaringan yang terbentuk seperti pada Gambar 4.18.




Gambar 4.17: Jaringan interaksi protein pada insulin yang terbentuk dari hasil penambangan teks pada STRING-DB

Dari hasil analisis tersebut, dengan membandingkan hasil dari jaringan yang terbentuk dari model XGBoost, didaptkan terdapat 3 protein yang sama yang memiliki interaksi terhadap insulin. Protein-protein tersebut adalah, INSR, IGF1R, IGF1. Akan tetapi interaksi pada STRING-DB tidak terkhusus pada protein pensintesis Insulin sebagai pemacu produktifitas insulin saja. Akan tetapi interaksi secara umum yang ada pada Insulin.

Tabel 4.5 menunjukkan hasil analisa dengan menggunakan *text mining* pada STRING-DB terhadap protein Insulin. Terdapat 16 protein yang memiliki interaksi kuat terhadap insulin, akan tetapi jika menggunakan *threshold=90%* maka hanya 11 protein yang memiliki pengaruh kuat terhadap insulin. Jika dibandingkan dengan hasil interaksi dengan model XGBoost, kedua hasil interaksi ini memiliki 3 protein yang sama-sama memiliki interaksi kuat terhadap Insulin, yakni INSR, IGFR1 dan IGF1. Sehingga bisa dimungkinkan model interaksi dengan mengunakan XGBoost serta

ekstraksi fitur fungsi molekul protein dengan menggunakan *centrality method* cukup bagus untuk mendeteksi protein yang memiliki pengaruh terhadap produksi/sintesis insulin.

Hasil dari membangun jaringan interaksi protein menggunakan XGBoost menghasilkan 9 protein yang berpengaruh besar dalam sintesis protein insulin. Dari segi biokimia, skema sintesis protein dapat dilihat seperti padaGambar 4.18 (D.L. Nelson dan M.M. Cox, 2004). Tiga protein yang teridentifikasi berpengaruh besar pada analisis jaringan protein XGboost dan String-DB, yakni INSR, IGF1R dan IGF1 menjadi protein-protein inti dalam metabolisme protein. IGF1R menjadi pintu pemicu utama dari tiga metabolisme insulin, yakni sintesis protein, motilitas sel dan poliferasi sel. Sedangakan IGF1 merupakan protein penstimulus IGF1R agar tetap bekerja stabil sebagaimana tugasnya menjadi pintu utama metabolisme. Selain ketiga protein tersebut, pada proses sintesis insulin juga terdapat protein-protein lain seperti IRS1, AKT, 4EBP1, dan lain-lain.(Jeremy M.B, dkk, 2015)

## 4.8 Analisis Biokimia

Pada unsur-unsur molekul protein, interaksi disebabkan adanya kontak fisik antara satu protein dengan protein yang lain. Kontak fisik ini lebih kedalam ikatan molekul dari protein tersebut. Ikatan yang dimaksudkan dalam interaksi protein merupakan ikatan Ionik yang merupakan ikatan molekul kuat. Ikatan dari molekul-molekul protein tersebut memiliki fungsi tertentu dan saling bergantung. Sehingga, salah satu protein tidak akan bisa aktif apabila protein yang seharusnya mendukung fungsi protein tersebut terganggu. Protein sendiri dibagi menjadi 7 jenis berdasarkan fungsinya dalam metabolisme,

1. Protein Hormonal, yang berfungsi sebagai bahan dasar pembentuk hormon. Hormon ini bertindak sebagai pembawa pesan kimia yang mengantarkan pesan melalui aliran darah. Setiap hormon ini akan memengaruhi satu sel tertentu di dalam tubuh yang dikenal sebagai sel target. Insulin sendiri merupakan protein jenis Hormonal yang merespon adanya kadar gula darah. Sehingga, protein-protein yang terlibat dalam sistem kerja Insulin, lebih kepada protein hormonal.

2. Protein Enzim (Katalis), pembentuk enzim. Enzim sendiri berfungsi untuk mendukung terjadinya reaksi kimia didalam tubuh.

3. Protein Struktural, yang merupakan jenis protein paling besar. Protein struktural berfungsi sebagai komponen penting yang membangun konstruksi tubuh dari tingkat sel. Contoh protein struktural yang paling umum adalah kolagen dan keratin. Protein jenis keratin adalah protein yang kuat dan berserat sehingga dapat membentuk struktur kulit, kuku, rambut, dan juga gigi. Sementara, protein struktural berbentuk kolagen berfungsi sebagai pembentuk tendon, tulang, otot, tulang rawan, dan juga kulit.

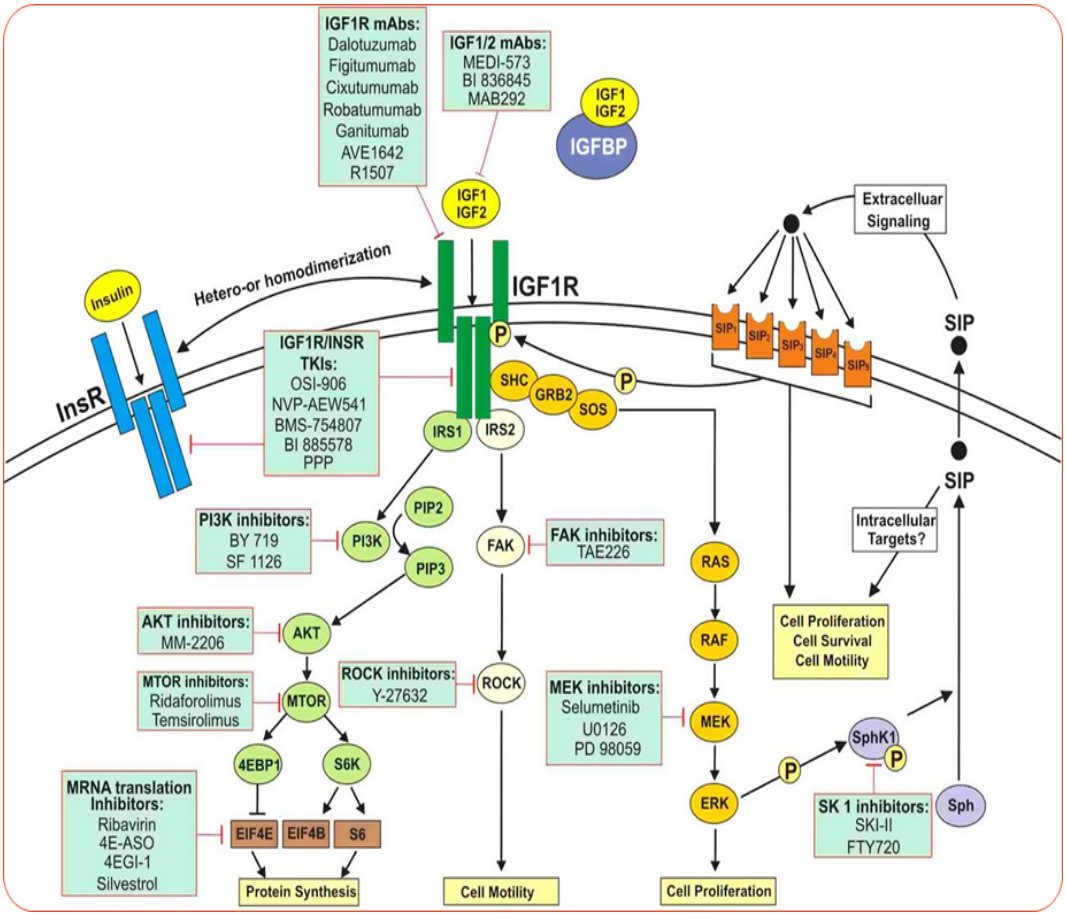
4. Protein Antibodi, berfungsi sebagai pelindung tubuh dari zat asing atau organisme asing yang memasuki tubuh. Protein-protein antibodi sebagai komponen pembentuk anti bodi dalam tubuh. Samakin banyak protein antibodi yang dihasilkan tubuh, maka pertahanan tubuh akan semakin optimal.

5. Protein Transport, yang berfungsi sebagai pengantar moleul-molekul dan zat yang dibutuhkan tubuh. Contohnya hemoglobin.

6. Protein Pengikat, berfungsi sebagai pengikat zat atau molekul yang diserap tubuh dan dibutuhkan dalam metabolisme.

7. Protein Penggerak, berfungsi sebagai pengatur kekuatan dan kecepatan jantung bergerak, serta mengatur otot ketika berkontraksi.

dari penjabaran jenis-jenis protein tersebut, sebagian besar protein yang berinteraksi dengan insulin merupakan protein hormonal. karena jenis dari Insulin sendiri merupakan hormon yang dikeluarkan oleh pankreas.




Gambar 4.18: Skema peranan protein-protein dalam metabolisme Insulin

Pada kajian biokimia, protein IGF1R, IGF1 dan INSR merupakan protein yang memiliki dampak besar dalam sentesis insulin. Pada Gambar 4.18

protein IGF1R menjadi pintu pemicu dari metabolisme insulin. *Insulin Growth Factor 1 Receptor* (IGF1R) merupakan reseptor utama dari pada pengaktifan Insulin. IGF1R diperkuat oleh protein IGF1 untuk mempertahankan fungsinya sehingga mengoptimalkan fungsi reseptor dari Insulin. Inti aktifasi insuli berada pada protein IGF1R, apabila protein ini terganggu fungsinya maka aktivasi Insulin juga akan terganggu. Selain itu, pada gambar Gambar 4.18, terdapat protein INSR yang memiliki peranan reseptor dari insulin juga yang memediasi aksi *pleiotropic* yang mengarahkan kepada proses fosforilasi. Ketiga protein ini berpengaruh sangat besar terhadap dampak fungsi insulin karena ketiganya ini protein reseptor utama Insulin.

Reseptor-reseptor insulin menjadi akar dari proses kerja *cascade* dari insulin. Kinerja dari IGF1R, IGF1 dan INSR mempengaruhi proses-proses inti dari insulin seperti, sintesis protein, motalitas sel, poliferasi sel dan survial sel. Apabila salah satu dari protein reseptor tidak aktif, maka fungsi yang ada dibawahnya juga akan tidak bekerja. sehingga insulin gagal untuk memerankan fungsinya dalam metabolisme tubuh manusia.

Tabel 4.3: Protein yang berpengaruh besar terhadap produksi insulin

| No | Kode | Nama Protein | Fungsi |
|---|---|---|---|
| 1 | P05019 | IGF1 | Faktor penumbuh insulin yang diisolasi dari plasma, secara struktural dan fungsional terkait dengan insulin serta memiliki aktivitas yang mendorong pertumbuhan yang jauh lebih tinggi. |
| 2 | Q12778 | FKHR | Faktor transkripsi yang merupakan target utama pensinyalan insulin dan mengatur homeostasis metabolik sebagai respons terhadap stres oksidatif. Mengikat elemen respons insulin (IRE) dengan urutan konsensus 5'-TT [G / A] TTTTG-3 'dan elemen ikatan keluarga Daf-16 terkait |
| 3 | P08069 | IGF1R | Reseptor tirosin kinase yang memediasi aksi *Insulin-like Growth Factor* 1 (IGF1). Ikatkan IGF1 dengan afinitas tinggi serta IGF2 dan insulin (INS) degan afinitas lebih rendah. IGF1R yang diaktifkan terlibat dalam pertumbuhan sel dan kontrol kelangsungan hidup. IGF1R sangat penting untuk transformasitumor dan kelangsungan hidup sel ganas. |
| 4 | P06213 | INSR | Reseptor tirosin kinase yang memediasi aksi pleiotropik insulin. Pengikatan insulin mengarah pada fosforilasi beberapa substrat intraseluler, termasuk substrat reseptor insulin (IRS1, 2, 3, 4), SHC, GAB1, CBL dan perantara pensinyalan sinyal lainnya. |
| 5 | P61244 | MAX | Regulator transkripsi. Membentuk kompleks protein pengikat DNA spesifik urutan dengan MYC atau MAD yang mengenali urutan inti 5'-CAC [GA] TG-3 '. |
| 6 | P36956 | SREBF1 | Diperlukan sebagai aktivator transkripsional untuk homeostasis lipid. Mengatur transkripsi gen reseptor LDL serta asam lemak dan pada tingkat yang lebih rendah jalur sintesis kolesterol (Dengan kesamaan). |
| 7 | Q13950 | AML3 | Faktor transkripsi yang terlibat dalam diferensiasi osteoblastik dan morfogenesis kerangka |
| 8 | Q9NZJ5 | PERK | Protein kinase penginderaan metabolik-stres yang memfosforilasi subunit alfa dari faktor inisiasi terjemahan eukariotik 2 (eIF-2-alpha / EIF2S1) pada 'Ser-52' selama respon protein tidak dilipat (UPR) dan sebagai respons terhadap ketersediaan asam amino yang rendah. |
| 9 | O15342 | ATP6H | Vakuolar ATPase bertanggung jawab untuk mengasamkan berbagai kompartemen intraseluler dalam sel eukariotik. |

Tabel 4.4: Protein-protein yang memiliki pengaruh cukup besar terhadap Insulin

| No | Kode | Nama Protein | Rasio | Link Edge |
|---|---|---|---|---|
| 1 | O75840 | KLF7 | 0.9868 | 280 |
| 2 | O15156 | ZBTB7B | 0.9825 | 198 |
| 3 | Q9Y243 | AKT3 | 0.9817 | 344 |
| 4 | P11474 | ESRRA | 0.9813 | 345 |
| 5 | O43524 | FOXO3 | 0.9806 | 273 |
| 6 | Q92743 | HTRA1 | 0.9799 | 348 |
| 7 | Q96IA0 | PTPRN | 0.9786 | 349 |
| 8 | P36543 | ATP6E | 0.9785 | 349 |
| 9 | Q9NQ66 | PLCB1 | 0.9778 | 365 |
| 10 | P08581 | MET | 0.9776 | 350 |
| 11 | P15336 | ATF2 | 0.9715 | 344 |
| 12 | P22460 | KCNA5 | 0.9698 | 338 |
| 13 | P49286 | MTNR1B | 0.9688 | 337 |
| 14 | P98177 | FOXO4 | 0.9655 | 344 |
| 15 | Q86YS7 | C2CD5 | 0.9654 | 247 |
| 16 | P46940 | IQGAP1 | 0.9649 | 107 |
| 17 | P11161 | EGR2 | 0.9603 | 319 |
| 18 | C9JNR5 | INS | 0.9603 | 340 |

Tabel 4.5: Hasil analisa *text mining* pada STRING-DB terhadap Insulin

| No | Nama Protein | Score |
|---|---|---|
| 1 | INSR | 0.999 |
| 2 | IGFR1 | 0.998 |
| 3 | IRS1 | 0.997 |
| 4 | IGF1 | 0.994 |
| 5 | IRS2 | 0.994 |
| 6 | EGF | 0.991 |
| 7 | PTPN1 | 0.991 |
| 8 | RPS6KB1 | 0.988 |
| 9 | EGFR | 0.967 |
| 10 | PDPK1 | 0.953 |
| 11 | IGFBP1 | 0.944 |

# BAB 5
# KESIMPULAN DAN SARAN

Dari analisis dan pembahasan yang sudah dilakukan pada bab sebelumnya, dapat diperoleh kesimpulan dan saran untuk pengembangan dan perbaikan penelitian selanjutnya.

## 5.1 Kesimpulan

*Molecular Function* merupakan salah satu dari tiga GO yang menterjemahkan fungsi dari suatu protein. Protein-protein yang memiliki kegiatan yang sama pada metabolisme menghasilkan interaksi antara protein yang memiliki fungsi sama. Data GO yang berupa *Directed Acyclic Graph* dapat diberikan label serta pembobotan pada setiap daun graf berarah menggunakan metode *centrality*. Metode *centrality* memberikan informasi penting terhadap fungsi universal dari suatu protein yang ditunjukkan sebagai root pada DAG. Metode *Betweenness Centrality* memiliki hasil akurasi lebih bagus sebesar 73.48% pada data latih dan 74.56% pada prediksi data menggunakan data tes dibandingkan dengan dua metode yang lain yakni *Closeness Centrality* dan *Page Rank Centrality*.

Metode XGBoost sebagai pengklasifikasi memiliki kemampuan mempelajari data dengan baik dengan memanfaatkan peningkatan kualitas pohon yang dibentuk, kebanyakan algoritma pembelajaran mesin ketika diberikan perilaku berlebih pada parameternya akan mengalami fluktuasi dan *ovverfitting*, sedangkan pada XGBoost model yang terbentuk hingga pohon yang terbentuk sebanyak 1000 pohon tidak terlalu mengalami *overfitting*. Model yang dihasilkan dipengaruhi oleh beberapa fitur yang dominan, seperti *binding* dan *catalysis activity*. Fitur-fitur yang dominan pengaruhnya pada model menjadikannya *root* pada kebanyakan pohon CART yang terbentuk.

Hasil dari model XGBoost tersebut dijadikan modal utama dalam membentuk interaksi antar protein pada insulin. Nilai *odd* pada model prediksi dapat dimanfaatkan untuk mengetahui seberapa dekat (dengan persamaan jarak)kesamaan fungsi dari suatu protein. Protein insulin sebagai pusat jaringan memiliki interaksi sangat besar terhadap 9 protein dan 18 protein yang berpengaruh cukup tinggi. Sehingga sebanyak 27 protein yang berpengaruh terhadap insulin berdasarkan model XGBoost. Sedangkan protein yang lain juga berinteraksi dengan insulin, akan tetapi tidak terlalu memiliki pengaruh besar terhadap insulin. 27 protein yang memiliki pengaruh besar tersebut memiliki beberapa kesamaan hasil ketika dibandingkan dengan hasil penambangan teks yang ada pada STRING-DB. Hasil perbandingan tersebut menunjukkan, interaksi protein-protein dapat dibangun dengan menggunakan fungsional protein yang ada pada gen ontologi, dengan memanfaatkan metode *centrality* pada teori graf sebagai ekstraksi

nilai fitur dan juga metode *Machine Learning* sebagai pembentukan model klasifikasi yang dapat dimanfaatkan untuk membangun interaksi protein.

## 5.2 Saran

Dari hasil penelitian diatas, peneliti mwmiliki saran terhadap pembaca maupun penelitian selanjutnya,

1 Penelitian selanjunya agar mengekstrak kesemua komponen pada GO yakni *Biological process, Cellular component* dan *Molecular function.*

2 Penelitian lanjutan dari hasil penelitian ini dapat dilakukan *Drug Virtual Screening* terhadap protein yang berpengaruh terhadap Insulin maupun terhadap Insulin sendiri untuk mendapatkan molekul kandidat obat terhadap penyakit yang berhubungan dengan Insulin.

3 *Molecular docking* juga dapat digunakan terhadap protein yang ditemukan memiliki pengaruh besar terhadap Insulin.

# DAFTAR PUSTAKA


Bader, G.D. Haque. C.W.V., (2003), *An Automated Method for Finding Molecular Complex in Large Protein Interaction Networks*, BMF Bioinformatics. Vol 4, No.2.

Belitz, H.D., Grosch, W., Schieberle, P., (2009), *Amino Acid, Peptides, Protein*, Springer: Food Chemistry.

Bogatti, S,P.,(2005), *Centrality and Network Flow*, Social Network, vol.27, p.55-71.

Chen, Tianqi., (2014). *Introduction to Boosted Trees.* University of Washington, Computer Science.,University of Washington 22 (2014).

Chen, Tianqi., and Charlos Guestrin., (2016). *XGBoost: A Scalable Tree Boosting System*,in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining,2016, pp. 785–794.

C. Marzban.,(2004), *The ROC Curve and the Area Under It as Performance Measure*, Weather and Forecasting. 19(6).

Defronzo,R.A, (1997), *Pathogenesis of type 2 diabetes: metabolic and molecular implication for identifying diabetes genes. Diabetes Care 36.*

D.L. Nelson., and M.M. Cox., (2004), *Lehninger Principles of Biochemistry*, Fourth Edition, W.H. Freeman & Company, New York.

Friedman, J., (2002), *Stochastic Gradient Boosting*, Computational statistic and data analysis, vol.38, no.4, pp.367-378.

Gupta, A., Gusain, K., Popli, B.,(2016), *Verifying the Value and Veracity of eXtreme Gradient Boosted Decision Trees on a Variety of Datasets*, ICIIS 11th International Conference on Industrial and Information System.

Ideker, T., Sharan, R.,(2008), *Protein Network in Disease*, Genome Res . Vol 18, pp.644-652.

Ivan, G., and Grolmusz, V.,(2011), *When the Web meets the cell: Using Personalized PageRank for Analyzed Protein Interaction Network*, Bioinformatics, vol.27, no.3, p.405-407.

Jain, S., and Bader, G.D.,(2010), *An Improved Method for Scoring Protein-Protein Interaction Using Semantic Similarity Within the Gene Ontology*, BMC Bioinformatics, vol.11, p.562.

J.M. Berg., J.L. Tymoczko., G.J. Gatto, and L. Stryer., (2015), *Biochemystry*, Eight Edition, W.H Freeman & Company, New York.

Kann,M.G., (2007), *Protein Interactions and disease: computational approaches to uncover the etiology of disease*, Brief Bioinformatics, Volume 8, pp. 333-346.

Karlin, D., Belshaw, R., (2012), *Detecting Remote Sequence Homology in Disordered Protein: Discovery of Conserved Motifs in the N-termini of Mononegavirales Phosphoprotein*, PLos One 7, e31719.

Pizzut, C., Rambo, S.E., (2016), *Algorithms and Tools for Protein-Protein Interaction Network Clustering, With A Special Focus on Population-based Stochastic Method*, Bioinformatics 30, pp. 1343-1352.

Ranu, Vyas., Sanket, Bapat., Esha, Jain., Mathukumarasamy, Karthikeyan., Sanjeev, Tambe., Baskar, D.K., (2016), *Building and Analysis of Protein-Protein Interaction Related to Diabetes Millitus Using Support Vector Machine, Biomedical Text Mining, and Network Analysis*, Computational Biology and Chemistry 65, pp.37-44.

Shu Dong., Hien Lau., Cody Chavarria., Michael, A., Alison Cimrel., John, PE. Sandra, E., Jack, L., James, N., dan Jonathan, RTL., (2019), *An Update Review on the Effects of Periodic Intensive Insulin Therapy*, 2019 International Conference on Health Research. https://doi.org/10.1016/j.curtheres.2019.04.003.

Srinivasa, R,V., Srinivas, K., Sujini, G,N., Kumar, G,N,S.,(2014), *Protein-Protein Interaction Detection: Method and Analysis*, International Journal of Proteomics, vol.2014(2014): 147648.10.1155/2014/147648 .

Usman, M.S., Kusuma, W.A., Afendi, F.M., and Heryanto, R.,(2019), *Identification of Significant Protein Associated with Diabetes Mellitus Using Network Analysis of Protein-Protein Interaction*, Computer Engineering and Application, vol.8, No.1.

Virkamaki, A., Kohjiro, U., Kahn, C.R., (1999), *Protein-Protein Interaction in Insulin Signaling and the Molecular Mechanism of Insulin Resistance*, The Journal of Clinic Investigation, Vol 103, No.7.

Wang, J., Zhou, X., Zhu, J., Zhou, C., and Guo, Z.,(2010), *Revealing and Avoiding Bias in Semantic Similarity Score for Protein Pairs*, BMC Bioinformatics, vol.11, p.290.

Xue Ying,(2019), *An Overview of Overfitting and its Solution*, IOP Conf. Series: Journal of Physic: Conf. Series 1168(2019)022022.

Zhong, J., Yusui, Sun., Wei, Pang., Minzhu, Xie., Jiahong, yang., and Xiwei, Tang.,(2018),*XGBFEMF:An XGBoost-Based Framework for Essential Protein Prediction*, Transaction On Nanobioscience, Vol. 17, No.3.

**LAMPIRAN 1**
**DATA PROTEIN**

A. Data protein yang berpengaruh terhadap insulin

| Entry | Entry name | Protein names | Gene names |
|---|---|---|---|
| P08069 | IGF1R_HUMAN | Insulin-like growth factor 1 receptor (EC 2.7.10.1) (Insulin-like growth factor I receptor) (IGF-I receptor) (CD antigen CD221) [Cleaved into: Insulin-like growth factor 1 receptor alpha chain; Insulin-like growth factor 1 receptor beta chain] | IGF1R |
| P51460 | INSL3_HUMAN | Insulin-like 3 (Leydig insulin-like peptide) (Ley-I-L) (Relaxin-like factor) [Cleaved into: Insulin-like 3 B chain; Insulin-like 3 A chain] | INSL3 RLF RLNL |
| P06213 | INSR_HUMAN | Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta] | INSR |
| P01308 | INS_HUMAN | Insulin [Cleaved into: Insulin B chain; Insulin A chain] | INS |
| P14616 | INSRR_HUMAN | Insulin receptor-related protein (IRR) (EC 2.7.10.1) (IR-related receptor) [Cleaved into: Insulin receptor-related protein alpha chain; Insulin receptor-related protein beta chain] | INSRR IRR |
| Q9UQB8 | BAIP2_HUMAN | Brain-specific angiogenesis inhibitor 1-associated protein 2 (BAI-associated protein 2) (BAI1-associated protein 2) (Protein BAP2) (Fas ligand-associated factor 3) (FLAF3) (Insulin receptor substrate p53/p58) (IRS-58) (IRSp53/58) (Insulin receptor substrate protein of 53 kDa) (IRSp53) (Insulin receptor substrate p53) | BAIAP2 |
| P01344 | IGF2_HUMAN | Insulin-like growth factor II (IGF-II) (Somatomedin-A) (T3M-11-derived growth factor) [Cleaved into: Insulin-like growth factor II; Insulin-like growth factor II Ala-25 Del; Preptin] | IGF2 PP1446 |
| P14735 | IDE_HUMAN | Insulin-degrading enzyme (EC 3.4.24.56) (Abeta-degrading protease) (Insulin protease) (Insulinase) (Insulysin) | IDE |
| P52945 | PDX1_HUMAN | Pancreas/duodenum homeobox protein 1 (PDX-1) (Glucose-sensitive factor) (GSF) (Insulin promoter factor 1) (IPF-1) (Insulin upstream factor 1) (IUF-1) (Islet/duodenum homeobox-1) (IDX-1) (Somatostatin-transactivating factor 1) (STF-1) | PDX1 IPF1 STF1 |
| P11717 | MPRI_HUMAN | Cation-independent mannose-6-phosphate receptor (CI Man-6-P receptor) (CI-MPR) (M6PR) (300 kDa mannose 6-phosphate receptor) (MPR 300) (Insulin-like | IGF2R MPRI |

| | | | |
|---|---|---|---|
| | | growth factor 2 receptor) (Insulin-like growth factor II receptor) (IGF-II receptor) (M6P/IGF2 receptor) (M6P/IGF2R) (CD antigen CD222) | |
| Q9UIQ6 | LCAP_HUMAN | Leucyl-cystinyl aminopeptidase (Cystinyl aminopeptidase) (EC 3.4.11.3) (Insulin-regulated membrane aminopeptidase) (Insulin-responsive aminopeptidase) (IRAP) (Oxytocinase) (OTase) (Placental leucine aminopeptidase) (P-LAP) [Cleaved into: Leucyl-cystinyl aminopeptidase, pregnancy serum form] | LNPEP OTASE |
| Q9Y5Q6 | INSL5_HUMAN | Insulin-like peptide INSL5 (Insulin-like peptide 5) [Cleaved into: Insulin-like peptide INSL5 B chain; Insulin-like peptide INSL5 A chain] | INSL5 UNQ156/PRO182 |
| P35568 | IRS1_HUMAN | Insulin receptor substrate 1 (IRS-1) | IRS1 |
| P05019 | IGF1_HUMAN | Insulin-like growth factor I (IGF-I) (Mechano growth factor) (MGF) (Somatomedin-C) | IGF1 IBP1 |
| O14654 | IRS4_HUMAN | Insulin receptor substrate 4 (IRS-4) (160 kDa phosphotyrosine protein) (py160) (Phosphoprotein of 160 kDa) (pp160) | IRS4 |
| P14672 | GLUT4_HUMAN | Solute carrier family 2, facilitated glucose transporter member 4 (Glucose transporter type 4, insulin-responsive) (GLUT-4) | SLC2A4 GLUT4 |
| P08833 | IBP1_HUMAN | Insulin-like growth factor-binding protein 1 (IBP-1) (IGF-binding protein 1) (IGFBP-1) (Placental protein 12) (PP12) | IGFBP1 IBP1 |
| Q13322 | GRB10_HUMAN | Growth factor receptor-bound protein 10 (GRB10 adapter protein) (Insulin receptor-binding protein Grb-IR) | GRB10 GRBIR KIAA0207 |
| P17936 | IBP3_HUMAN | Insulin-like growth factor-binding protein 3 (IBP-3) (IGF-binding protein 3) (IGFBP-3) | IGFBP3 IBP3 |
| P24593 | IBP5_HUMAN | Insulin-like growth factor-binding protein 5 (IBP-5) (IGF-binding protein 5) (IGFBP-5) | IGFBP5 IBP5 |
| P22692 | IBP4_HUMAN | Insulin-like growth factor-binding protein 4 (IBP-4) (IGF-binding protein 4) (IGFBP-4) | IGFBP4 IBP4 |
| P24592 | IBP6_HUMAN | Insulin-like growth factor-binding protein 6 (IBP-6) (IGF-binding protein 6) (IGFBP-6) | IGFBP6 IBP6 |
| P18065 | IBP2_HUMAN | Insulin-like growth factor-binding protein 2 (IBP-2) (IGF-binding protein 2) (IGFBP-2) | IGFBP2 BP2 IBP2 |
| P35858 | ALS_HUMAN | Insulin-like growth factor-binding protein complex acid labile subunit (ALS) | IGFALS ALS |
| Q9Y4H2 | IRS2_HUMAN | Insulin receptor substrate 2 (IRS-2) | IRS2 |
| Q16270 | IBP7_HUMAN | Insulin-like growth factor-binding protein 7 (IBP-7) (IGF-binding protein 7) (IGFBP-7) (IGFBP-rP1) (MAC25 protein) (PGI2-stimulating factor) (Prostacyclin-stimulating factor) (Tumor-derived adhesion factor) (TAF) | IGFBP7 MAC25 PSF |

| | | | |
|---|---|---|---|
| A0A024RAI2 | A0A024RAI2_HUMAN | Insulin-induced gene protein | INSIG2 hCG_1686419 |
| O19707 | O19707_HUMAN | MHC class II HLA-DQ-beta-1 (Fragment) | HLA-DQB1 |
| B7Z7W6 | B7Z7W6_HUMAN | cDNA FLJ53247, highly similar to Insulin-degrading enzyme | |
| C1PHA2 | C1PHA2_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | KIF5B-ALK |
| B4DKT5 | B4DKT5_HUMAN | cDNA FLJ53054, highly similar to Insulin-like growth factor 2 mRNA-binding protein 2 | |
| B6D4Y3 | B6D4Y3_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | ALK |
| B6EXY3 | B6EXY3_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | |
| V9HWC2 | V9HWC2_HUMAN | Epididymis secretory sperm binding protein Li 67p | HEL-S-67p |
| A0A024R0H1 | A0A024R0H1_HUMAN | Solute carrier family 16 (Monocarboxylic acid transporters), member 1, isoform CRA_b | SLC16A1 hCG_37455 |
| F5H6P3 | F5H6P3_HUMAN | Insulin-induced gene protein | INSIG1 |
| A0A3B3ITZ2 | A0A3B3ITZ2_HUMAN | Insulin-like peptide INSL6 | INSL6 |
| Q53Y25 | Q53Y25_HUMAN | Phosphotransferase (EC 2.7.1.-) | GCK hCG_1745191 tcag7.801 |
| A0A0C4DGZ4 | A0A0C4DGZ4_HUMAN | Insulin-induced gene protein | INSIG2 |
| D1MAM2 | D1MAM2_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | ALK |
| A0N9R6 | A0N9R6_HUMAN | Insulin-like growth factor type 2 receptor (Insulin-like growth factor type II receptor) (Fragment) | IGF2R igf2R |
| S4R3W1 | S4R3W1_HUMAN | Suppressor of cytokine-signaling 2 (Fragment) | SOCS2 |
| B3KTS7 | B3KTS7_HUMAN | cDNA FLJ38672 fis, clone HSYRA2000371, moderately similar to Insulin-like growth factor-binding protein 3 | |
| B7Z7I9 | B7Z7I9_HUMAN | cDNA FLJ57658, highly similar to Protein FAM3B | |
| H3BSX8 | H3BSX8_HUMAN | Insulin-like growth factor-binding protein complex acid labile subunit | IGFALS |
| B2RCP7 | B2RCP7_HUMAN | cDNA, FLJ96197, highly similar to Homo sapiens connective tissue growth factor (CTGF), mRNA | |
| X5DR71 | X5DR71_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) (Fragment) | NTRK1 |
| C9JMX4 | C9JMX4_HUMAN | Insulin-like growth factor-binding protein 3 (Fragment) | IGFBP3 |
| J7M2B1 | J7M2B1_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | EZR-ROS1 EZR-ROS1_E10;R34 |
| B8YNW1 | B8YNW1_HUMAN | Hepatocyte nuclear factor 1 alpha (Fragment) | HNF1alpha |
| C5H3H2 | C5H3H2_HUMAN | Chromosome 10 open reading frame 104, isoform CRA_a (Metabolic syndrome-associated protein) | C10orf104 hCG_1789694 |
| H0YMJ5 | H0YMJ5_HUMAN | Insulin-like growth factor 1 receptor (Fragment) | IGF1R |

| | | | |
|---|---|---|---|
| B5BU82 | B5BU82_HUMAN | Suppressor of cytokine signaling-2 | SOCS2 |
| E0YMJ8 | E0YMJ8_HUMAN | HNF1 beta A splice variant 3 (Hepatocyte nuclear factor 1-beta) | HNF1B |
| A9YLN6 | A9YLN6_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | SLC34A2/ROS1 fusion |
| A0A0S2Z559 | A0A0S2Z559_HUMAN | V-raf-1 murine leukemia viral oncogene-like protein 1 isoform 2 (Fragment) | RAF1 |
| B3KUJ4 | B3KUJ4_HUMAN | cDNA FLJ40030 fis, clone STOMA2009250, highly similar to Klotho | |
| A0A075EKM8 | A0A075EKM8_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | |
| A0N9R9 | A0N9R9_HUMAN | Insulin-like growth factor type II receptor (Fragment) | IGF2R |
| Q29922 | Q29922_HUMAN | HLA-DRB4 protein (Fragment) | HLA-DRB4 |
| Q30065 | Q30065_HUMAN | MHC class II HLA-DQ-alpha chain (Fragment) | HLA-DQA1 |
| B2RU28 | B2RU28_HUMAN | Relaxin 3 | RLN3 |
| M0QYY9 | M0QYY9_HUMAN | Insulin growth factor-like family member 4 | IGFL4 |
| C9JCQ0 | C9JCQ0_HUMAN | Receptor-type tyrosine-protein phosphatase-like N (Fragment) | PTPRN |
| Q6NT58 | Q6NT58_HUMAN | Leptin (Obesity factor) | LEP |
| A4D0Y8 | A4D0Y8_HUMAN | Leptin (Obesity factor) | LEP hCG_33000 tcag7.84 |
| Q75MT1 | Q75MT1_HUMAN | Uncharacterized protein GRB10 (Fragment) | GRB10 |
| E0YMI9 | E0YMI9_HUMAN | HNF1 alpha A splice variant 3 (Hepatocyte nuclear factor 1-alpha) | HNF1A |
| A8K3M3 | A8K3M3_HUMAN | Tyrosine-protein phosphatase non-receptor type (EC 3.1.3.48) | PTPN1 hCG_37134 |
| E9PNA7 | E9PNA7_HUMAN | Anoctamin | ANO1 |
| Q2PJC2 | Q2PJC2_HUMAN | Insulin receptor (Fragment) | |
| Q8TAY0 | Q8TAY0_HUMAN | Insulin-like growth factor binding protein, acid labile subunit | IGFALS |
| A0A0K0K1K3 | A0A0K0K1K3_HUMAN | Adenylate cyclase (EC 4.6.1.1) | HEL-S-172mP ADCY8 hCG_2008967 |
| F1D8T1 | F1D8T1_HUMAN | Hepatocyte nuclear factor 4 4 alpha variant 2 | NR2A1 |
| C9J392 | C9J392_HUMAN | Receptor-type tyrosine-protein phosphatase-like N (Fragment) | PTPRN |
| V9HWD0 | V9HWD0_HUMAN | Epididymis secretory protein Li 42 (Ras homolog gene family, member Q, isoform CRA_d) | HEL-S-42 RHOQ hCG_1995863 |
| A0A0S2Z4C6 | A0A0S2Z4C6_HUMAN | Serine/threonine-protein phosphatase (EC 3.1.3.16) (Fragment) | PPP3CA |
| A0A024RCF8 | A0A024RCF8_HUMAN | Synaptotagmin-like 4 (Granuphilin-a), isoform CRA_a | SYTL4 hCG_20193 |
| B6EXY4 | B6EXY4_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | |

| | | | |
|---|---|---|---|
| O19712 | O19712_HUMAN | MHC class II HLA-DQ-beta-1 (Fragment) | HLA-DQB1 |
| E7ENI6 | E7ENI6_HUMAN | Islet cell autoantigen 1 | ICA1 |
| Q9NW72 | Q9NW72_HUMAN | Anoctamin | |
| Q2MKP0 | Q2MKP0_HUMAN | Insulin receptor (Fragment) | |
| X5D991 | X5D991_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) (Fragment) | NTRK1 |
| C1PHA0 | C1PHA0_HUMAN | Tyrosine-protein kinase receptor (EC 2.7.10.1) | EML4-ALK variant 6 |
| C9IZC7 | C9IZC7_HUMAN | Islet cell autoantigen 1 (Fragment) | ICA1 |
| Q30079 | Q30079_HUMAN | MHC class II HLA-DQ-beta-1 (Fragment) | HLA-DQB1 |
| A0A024R7I3 | A0A024R7I3_HUMAN | RAB8A, member RAS oncogene family, isoform CRA_a | RAB8A hCG_36928 |
| A0A024RBB8 | A0A024RBB8_HUMAN | Oxysterol-binding protein | OSBPL8 hCG_27425 |
| B4DTR7 | B4DTR7_HUMAN | cDNA FLJ56779, highly similar to Insulin receptor | |
| O19705 | O19705_HUMAN | MHC class II HLA-DQ-alpha chain (Fragment) | HLA-DQA1 |
| M4PPE6 | M4PPE6_HUMAN | Phosphotransferase (EC 2.7.1.-) (Fragment) | GCK |
| B0FJM0 | B0FJM0_HUMAN | Insulin receptor (Fragment) | |
| A0A087WXV4 | A0A087WXV4_HUMAN | Hepatocyte nuclear factor 4-alpha | HNF4A |
| A0A3B3ISG5 | A0A3B3ISG5_HUMAN | Insulin-degrading enzyme | IDE |
| J7GXU7 | J7GXU7_HUMAN | Phosphoinositide-3-kinase regulatory subunit (Fragment) | PIK3R1 |
| A0N9R7 | A0N9R7_HUMAN | Insulin-like growth factor type II receptor (Fragment) | IGF2R |
| W0S0X4 | W0S0X4_HUMAN | Tyrosine-protein kinase (EC 2.7.10.2) | Pe1Fe3 |
| B4DPL6 | B4DPL6_HUMAN | cDNA FLJ53849, highly similar to Hepatocyte nuclear factor 1-beta | |

Dikarenakan data sangat banyak, maka hanya saya tampilkan 100 data saja pada lampiran buku tesis ini. Untuk lebih jelasnya saya lampirkan pada file Microsoft Exel (.xlxs) yang berada pada folder Lampiran dengan nama file $Uniprot\_Insulin\_Human\_Dataset.xlxs$.

B. Data protein yang tidak berpengaruh terhadap insulin

| Entry | Entry name | Protein names | Gene names |
|---|---|---|---|
| P01138 | NGF_HUMAN | Beta-nerve growth factor (Beta-NGF) | NGF NGFB |

| | | | |
|---|---|---|---|
| P04629 | NTRK1_HUMAN | High affinity nerve growth factor receptor (EC 2.7.10.1) (Neurotrophic tyrosine kinase receptor type 1) (TRK1-transforming tyrosine kinase protein) (Tropomyosin-related kinase A) (Tyrosine kinase receptor) (Tyrosine kinase receptor A) (Trk-A) (gp140trk) (p140-TrkA) | NTRK1 MTC TRK TRKA |
| P20783 | NTF3_HUMAN | Neurotrophin-3 (NT-3) (HDNF) (Nerve growth factor 2) (NGF-2) (Neurotrophic factor) | NTF3 |
| P08138 | TNR16_HUMAN | Tumor necrosis factor receptor superfamily member 16 (Gp80-LNGFR) (Low affinity neurotrophin receptor p75NTR) (Low-affinity nerve growth factor receptor) (NGF receptor) (p75 ICD) (CD antigen CD271) | NGFR TNFRSF16 |
| Q9UI33 | SCNBA_HUMAN | Sodium channel protein type 11 subunit alpha (Peripheral nerve sodium channel 5) (PN5) (Sensory neuron sodium channel 2) (Sodium channel protein type XI subunit alpha) (Voltage-gated sodium channel subunit alpha Nav1.9) (hNaN) | SCN11A SCN12A SNS2 |
| Q9Y5Y9 | SCNAA_HUMAN | Sodium channel protein type 10 subunit alpha (Peripheral nerve sodium channel 3) (PN3) (hPN3) (Sodium channel protein type X subunit alpha) (Voltage-gated sodium channel subunit alpha Nav1.8) | SCN10A |
| Q8ND25 | ZNRF1_HUMAN | E3 ubiquitin-protein ligase ZNRF1 (EC 2.3.2.27) (Nerve injury-induced gene 283 protein) (RING-type E3 ubiquitin transferase ZNRF1) (Zinc/RING finger protein 1) | ZNRF1 NIN283 |
| Q00994 | BEX3_HUMAN | Protein BEX3 (Brain-expressed X-linked protein 3) (Nerve growth factor receptor-associated protein 1) (Ovarian granulosa cell 13.0 kDa protein HGR74) (p75NTR-associated cell death executor) | BEX3 DXS6984E NADE NGFRAP1 |
| Q9NWD9 | BEX4_HUMAN | Protein BEX4 (BEX1-like protein 1) (Brain-expressed X-linked protein 4) (Nerve growth factor receptor-associated protein 3) | BEX4 BEXL1 NADE3 |
| P18146 | EGR1_HUMAN | Early growth response protein 1 (EGR-1) (AT225) (Nerve growth factor-induced protein A) (NGFI-A) (Transcription factor ETR103) (Transcription factor Zif268) (Zinc finger protein 225) (Zinc finger protein Krox-24) | EGR1 KROX24 ZNF225 |
| Q92982 | NINJ1_HUMAN | Ninjurin-1 (Nerve injury-induced protein 1) | NINJ1 |
| Q9NZG7 | NINJ2_HUMAN | Ninjurin-2 (Nerve injury-induced protein 2) | NINJ2 |
| O00458 | IFRD1_HUMAN | Interferon-related developmental regulator 1 (Nerve growth factor-inducible protein PC4) | IFRD1 |
| Q6UWH4 | GAK1B_HUMAN | Golgi-associated kinase 1B (Expressed in nerve and epithelium during development) (Protein FAM198B) | GASK1B C4orf18 ENED FAM198B AD021 |

| | | | |
|---|---|---|---|
| | | | UNQ2512/PRO6001 |
| P25189 | MYP0_HUMAN | Myelin protein P0 (Myelin peripheral protein) (MPP) (Myelin protein zero) | MPZ |
| P26367 | PAX6_HUMAN | Paired box protein Pax-6 (Aniridia type II protein) (Oculorhombin) | PAX6 AN2 |
| Q02962 | PAX2_HUMAN | Paired box protein Pax-2 | PAX2 |
| Q01453 | PMP22_HUMAN | Peripheral myelin protein 22 (PMP-22) (Growth arrest-specific protein 3) (GAS-3) | PMP22 GAS3 |
| Q9BXM0 | PRAX_HUMAN | Periaxin | PRX KIAA1620 |
| P11161 | EGR2_HUMAN | E3 SUMO-protein ligase EGR2 (EC 2.3.2.-) (AT591) (E3 SUMO-protein transferase ERG2) (Early growth response protein 2) (EGR-2) (Zinc finger protein Krox-20) | EGR2 KROX20 |
| P23560 | BDNF_HUMAN | Brain-derived neurotrophic factor (BDNF) (Abrineurin) [Cleaved into: BDNF precursor form (ProBDNF)] | BDNF |
| P34130 | NTF4_HUMAN | Neurotrophin-4 (NT-4) (Neurotrophin-5) (NT-5) (Neutrophic factor 4) | NTF4 NTF5 |
| P08034 | CXB1_HUMAN | Gap junction beta-1 protein (Connexin-32) (Cx32) (GAP junction 28 kDa liver protein) | GJB1 CX32 |
| Q9ULH0 | KDIS_HUMAN | Kinase D-interacting substrate of 220 kDa (Ankyrin repeat-rich membrane-spanning protein) | KIDINS220 ARMS KIAA1250 |
| Q96M96 | FGD4_HUMAN | FYVE, RhoGEF and PH domain-containing protein 4 (Actin filament-binding protein frabin) (FGD1-related F-actin-binding protein) (Zinc finger FYVE domain-containing protein 6) | FGD4 FRABP ZFYVE6 |
| Q86UL8 | MAGI2_HUMAN | Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2 (Atrophin-1-interacting protein 1) (AIP-1) (Atrophin-1-interacting protein A) (Membrane-associated guanylate kinase inverted 2) (MAGI-2) | MAGI2 ACVRINP1 AIP1 KIAA0705 |
| X6R3T8 | X6R3T8_HUMAN | SAP30-binding protein | SAP30BP |
| H7C5L4 | H7C5L4_HUMAN | Filamin-B (Fragment) | FLNB |
| U3KQR1 | U3KQR1_HUMAN | Adenylate cyclase type 3 (Fragment) | ADCY3 |
| Q6T2C9 | Q6T2C9_HUMAN | Brain-derived neurotrophic factor isoform 1 (Fragment) | BDNF |
| K7EJK1 | K7EJK1_HUMAN | Glial fibrillary acidic protein | GFAP |
| A0A0A0MQU1 | A0A0A0MQU1_HUMAN | Inverted formin-2 (Fragment) | INF2 |
| X6RKN2 | X6RKN2_HUMAN | Neurofascin (Fragment) | NFASC |
| A0A0C4DGV8 | A0A0C4DGV8_HUMAN | Semaphorin-3B | SEMA3B |

| | | | |
|---|---|---|---|
| E9PB39 | E9PB39_HUMAN | Rho guanine nucleotide exchange factor 10 | ARHGEF10 |
| M0R343 | M0R343_HUMAN | Melanoma-derived growth regulatory protein (Fragment) | MIA |
| E9PQR3 | E9PQR3_HUMAN | Ferritin | FTH1 |
| C9JVM2 | C9JVM2_HUMAN | Palmitoyltransferase (EC 2.3.1.225) (Fragment) | ZDHHC8 |
| K7EMP8 | K7EMP8_HUMAN | Glial fibrillary acidic protein | GFAP |
| J3KSH9 | J3KSH9_HUMAN | Integrin beta-4 (Fragment) | ITGB4 |
| M0R0X2 | M0R0X2_HUMAN | Neurotrophin-4 (Fragment) | NTF4 |
| K7EMA4 | K7EMA4_HUMAN | Sphingosine kinase 1 (Fragment) | SPHK1 |
| J3QKW0 | J3QKW0_HUMAN | SAP30-binding protein | SAP30BP |
| A0A024R0P1 | A0A024R0P1_HUMAN | HCG2043253, isoform CRA_a | hCG_2043253 |
| V9HW01 | V9HW01_HUMAN | Epididymis secretory protein Li 310 | HEL-S-310 |
| A0A024R6J9 | A0A024R6J9_HUMAN | Proline rich membrane anchor 1, isoform CRA_a | PRIMA1 hCG_1658529 |
| G5EA47 | G5EA47_HUMAN | Transmembrane protease serine 5 (Transmembrane protease, serine 5 (Spinesin), isoform CRA_a) | TMPRSS5 hCG_1730733 |
| A0A024R8K7 | A0A024R8K7_HUMAN | Integrin beta | ITGB4 hCG_27538 |
| A0A0B4J243 | A0A0B4J243_HUMAN | Protein FAM180B | FAM180B |
| K7EJ32 | K7EJ32_HUMAN | Sphingosine kinase 1 (Fragment) | SPHK1 |
| A2IDA2 | A2IDA2_HUMAN | Rhomboid 5 homolog 1 (Drosophila) (Fragment) | RHBDF1 Z69719.3-010 |
| D6W5D9 | D6W5D9_HUMAN | B-cell CLL/lymphoma 11A (Zinc finger protein), isoform CRA_b (BCL11A B-cell CLL/lymphoma 11A (Zinc finger protein) isoform 1) | BCL11A hCG_1986387 |
| A0A024R611 | A0A024R611_HUMAN | Coronin | CORO1A hCG_1770704 |
| Q8WYS3 | Q8WYS3_HUMAN | Inverted formin-2 | INF2 pp9484 |
| A0A0C4DGA1 | A0A0C4DGA1_HUMAN | Filamin-B (Fragment) | FLNB |
| F5H2M3 | F5H2M3_HUMAN | Transmembrane protease serine 5 | TMPRSS5 |
| C9JFD7 | C9JFD7_HUMAN | Serine/threonine-protein kinase Chk2 | CHEK2 |
| B2R789 | B2R789_HUMAN | START domain containing 13, isoform CRA_b (cDNA, FLJ93332, Homo sapiens START domain containing 13 (STARD13), transcriptvariant gamma, mRNA) | STARD13 hCG_32808 |
| B0LPE5 | B0LPE5_HUMAN | Non-specific serine/threonine protein kinase (EC 2.7.11.1) | AKT1 hCG_96740 |

| | | | |
|---|---|---|---|
| B3KUR8 | B3KUR8_HUMAN | cDNA FLJ40483 fis, clone TESTI2043758, highly similar to Neurofascin | |
| A8KAH9 | A8KAH9_HUMAN | RAP1A, member of RAS oncogene family (Ras-related protein Rap-1A) (cDNA FLJ75985, highly similar to Homo sapiens RAP1A, member of RAS oncogene family (RAP1A), transcript variant 2, mRNA) | RAP1A hCG_1818872 |
| K7EPT8 | K7EPT8_HUMAN | Glial fibrillary acidic protein (Fragment) | GFAP |
| F6KD01 | F6KD01_HUMAN | Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN (EC 3.1.3.16) (EC 3.1.3.48) (EC 3.1.3.67) (Phosphatase and tensin homolog) | PTEN |
| Q5SXY3 | Q5SXY3_HUMAN | Cholinergic receptor, nicotinic, beta 2 (Neuronal) (cDNA, FLJ94017, Homo sapiens cholinergic receptor, nicotinic, beta polypeptide 2(neuronal) (CHRNB2), mRNA) | CHRNB2 hCG_18732 |
| A0A087X1Y9 | A0A087X1Y9_HUMAN | Matrix metalloproteinase-28 (Fragment) | MMP28 |
| F5GXT6 | F5GXT6_HUMAN | Transmembrane protease serine 5 | TMPRSS5 |
| A0A024R3C5 | A0A024R3C5_HUMAN | Dopamine receptor D2, isoform CRA_c | DRD2 hCG_39593 |
| B8ZZ07 | B8ZZ07_HUMAN | Inactive rhomboid protein 1 (Rhomboid 5 homolog 1 (Drosophila)) (Fragment) | RHBDF1 Z69719.3-012 |
| A0A024R5S4 | A0A024R5S4_HUMAN | Ubiquitin specific peptidase 8, isoform CRA_a | USP8 hCG_38748 |
| F8WBS4 | F8WBS4_HUMAN | Inactive rhomboid protein 1 | RHBDF1 |
| J3QQL2 | J3QQL2_HUMAN | Integrin beta (Fragment) | ITGB4 |
| A0A0S2Z3Q4 | A0A0S2Z3Q4_HUMAN | V-crk sarcoma virus CT10 oncogene-like protein isoform 1 (Fragment) | CRK |
| A0A024R6Z0 | A0A024R6Z0_HUMAN | Dynein, cytoplasmic 1, light intermediate chain 2, isoform CRA_a | DYNC1LI2 hCG_28828 |
| M0QZQ0 | M0QZQ0_HUMAN | Uncharacterized protein | |
| M0R0V1 | M0R0V1_HUMAN | Pleckstrin homology domain-containing family A member 4 (Fragment) | PLEKHA4 |
| A0A024R8R0 | A0A024R8R0_HUMAN | SAP30 binding protein, isoform CRA_a | SAP30BP hCG_27541 |
| A0A0G2JLI8 | A0A0G2JLI8_HUMAN | Glial fibrillary acidic protein (Fragment) | GFAP |
| E9PPQ4 | E9PPQ4_HUMAN | Ferritin (Fragment) | FTH1 |
| E9PL19 | E9PL19_HUMAN | Poly(U)-binding-splicing factor PUF60 (Fragment) | PUF60 |
| J3QQJ0 | J3QQJ0_HUMAN | SAP30-binding protein (Fragment) | SAP30BP |
| F8W1F5 | F8W1F5_HUMAN | Formin-like protein 3 | FMNL3 |
| H7BZ30 | H7BZ30_HUMAN | Serine/threonine-protein kinase Chk2 (Fragment) | CHEK2 |

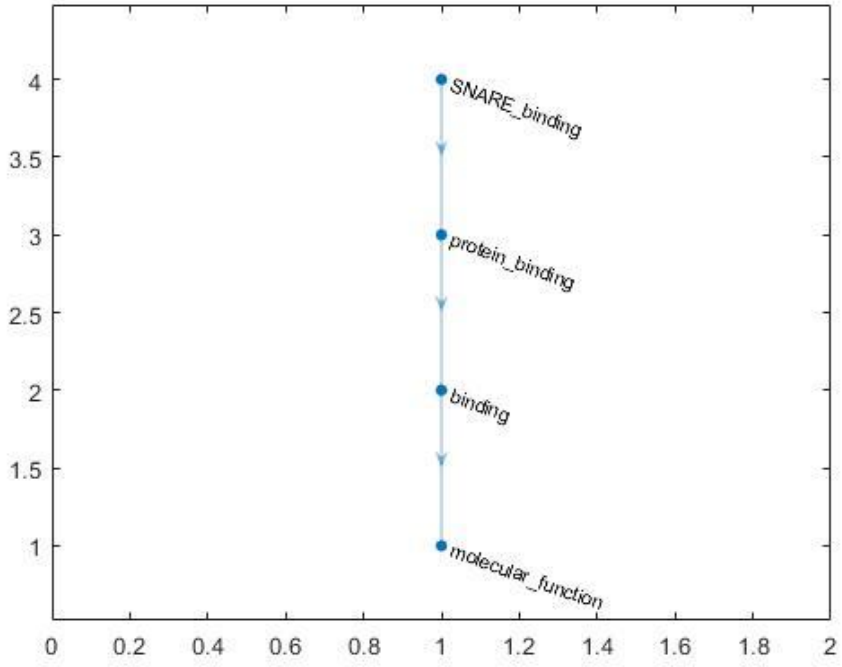
| | | | |
|---|---|---|---|
| A0A087WZE0 | A0A087WZE0_HUMAN | Semaphorin-3B (Fragment) | SEMA3B |
| B4DEK9 | B4DEK9_HUMAN | Semaphorin-3B (cDNA FLJ55830, highly similar to Semaphorin-3B) | SEMA3B |
| A0A024QZE4 | A0A024QZE4_HUMAN | Neurotrophin 5 (Neurotrophin 4/5), isoform CRA_a | NTF5 hCG_1749438 |
| E9PQW2 | E9PQW2_HUMAN | NADH-cytochrome b5 reductase 2 (Fragment) | CYB5R2 |
| D6RHX4 | D6RHX4_HUMAN | Neurofascin (Fragment) | NFASC |
| B7ZBF8 | B7ZBF8_HUMAN | Serine/threonine-protein kinase Chk2 (Fragment) | CHEK2 |
| Q53ZR5 | Q53ZR5_HUMAN | Sphingosine kinase 1 (Sphingosine kinase 1, isoform CRA_a) (cDNA FLJ78254) | SPHK1 hCG_30901 |
| Q5JP22 | Q5JP22_HUMAN | Epidermal growth factor-like protein 8 (Fragment) | EGFL8 |
| C0H5X0 | C0H5X0_HUMAN | MMP28 protein (Matrix metallopeptidase 28, isoform CRA_b) (Matrix metalloproteinase-28) | MMP28 hCG_1989249 |
| A8MU75 | A8MU75_HUMAN | Peripheral myelin protein 22 | PMP22 |
| B1B1J6 | B1B1J6_HUMAN | Chemokine-like protein TAFA-5 (Fragment) | TAFA5 |
| G3V192 | G3V192_HUMAN | Ferritin | FTH1 hCG_1776129 |
| V9HWC9 | V9HWC9_HUMAN | Superoxide dismutase [Cu-Zn] (EC 1.15.1.1) | HEL-S-44 |
| C9JM56 | C9JM56_HUMAN | Solute carrier family 25 member 48 | SLC25A48 |
| Q53QU7 | Q53QU7_HUMAN | Distal-less homeobox 2 (cDNA FLJ75693, highly similar to Homo sapiens distal-less homeo box 2 (DLX2), mRNA) | DLX2 hCG_16788 |
| A0A0E3SU01 | A0A0E3SU01_HUMAN | Brain-derived neurotrophic factor (BDNF) | BDNF |
| V9HWB4 | V9HWB4_HUMAN | Epididymis secretory sperm binding protein Li 89n | HEL-S-89n |
| J3QLM2 | J3QLM2_HUMAN | SAP30-binding protein (Fragment) | SAP30BP |

Dikarenakan data sangat banyak, maka hanya saya tampilkan 100 data saja pada lampiran buku tesis ini. Untuk lebih jelasnya penulis lampirkan pada file Microsoft Exel (.xlxs) yang berada di CD-Tesis pada folder Lampiran dengan nama file $Uniprot\_non\_Insulin\_Human\_Dataset.xlxs$.

**LAMPIRAN 2**
**KODE ONTOLOGI GEN DAN SKOR CENTRALITY**

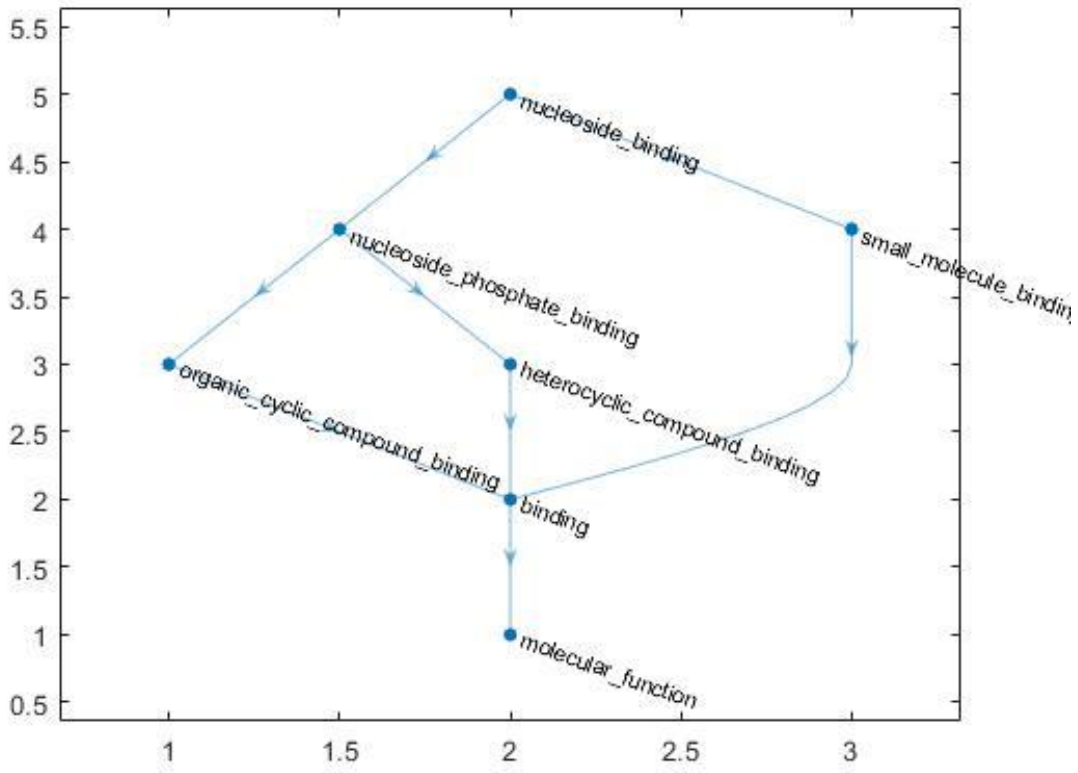
1. GO:0000149 = SNARE Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.16667 | 0 | 0.37018 |
| 'binding' | 0.14815 | 2 | 0.2988 |
| 'protein_binding' | 0.11111 | 2 | 0.21486 |
| 'SNARE_binding' | 0 | 0 | 0.11616 |



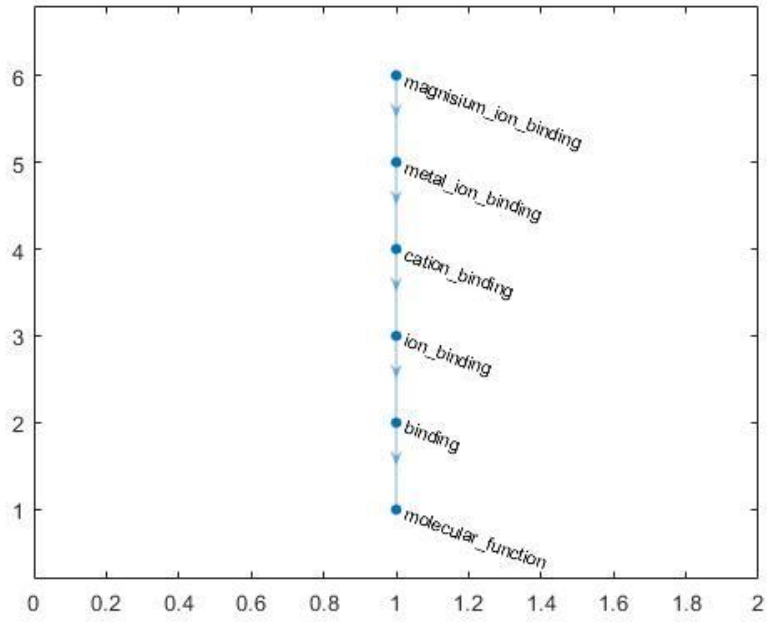

2. GO:0000166 = Nucleotide Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.076923 | 0 | 0.30228 |
| 'binding' | 0.099206 | 5 | 0.2872 |
| 'organic_cyclic_compound_binding' | 0.037037 | 1 | 0.093342 |
| 'heterocyclic_compound_binding' | 0.037037 | 1 | 0.093342 |
| 'nucleoside_phosphate_binding' | 0.027778 | 2 | 0.082847 |
| 'small_molecule_binding' | 0.027778 | 2 | 0.082847 |
| 'nucleotide_binding' | 0 | 0 | 0.058143 |



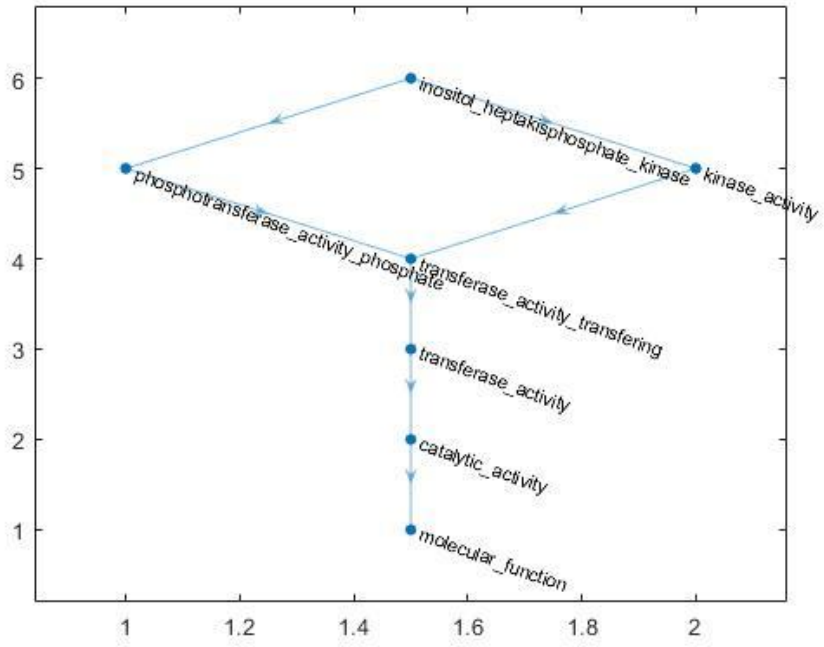

3. GO:0000287 = Magnesium Ion Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.066667 | 0 | 0.25218 |
| 'binding' | 0.064 | 4 | 0.22516 |
| 'ion_binding' | 0.06 | 6 | 0.1934 |
| 'cation_binding' | 0.053333 | 6 | 0.15618 |
| 'metal_ion_binding' | 0.04 | 4 | 0.11234 |
| 'magnisium_ion_binding' | 0 | 0 | 0.06073 |



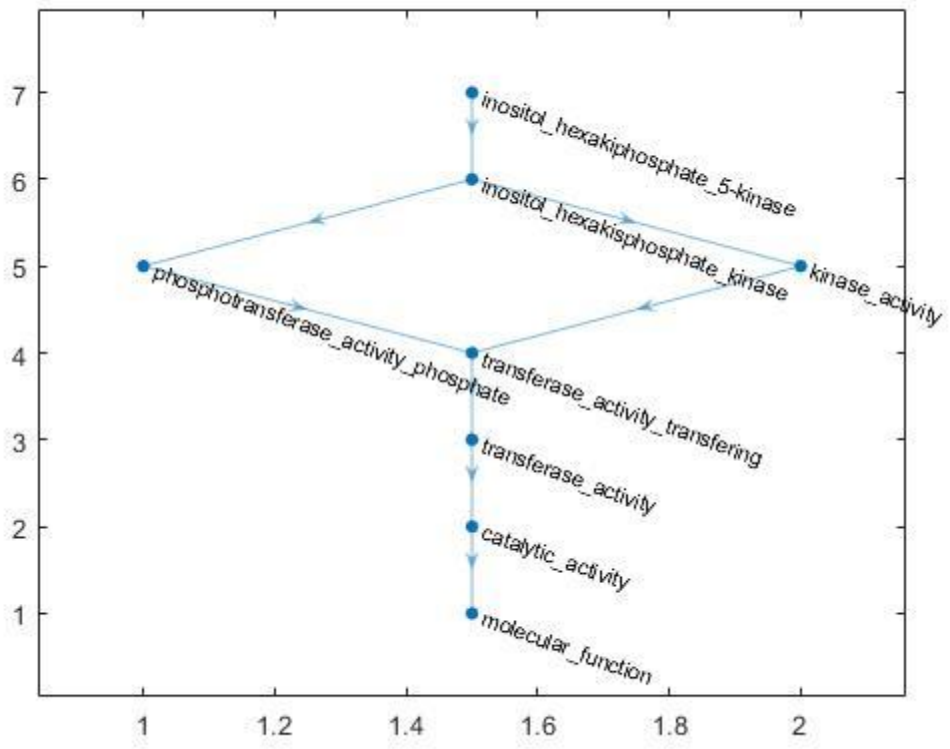

4. GO:0000829 = Inositol Heptakisphosphate Kinase Activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.052632 | 0 | 0.23171 |
| 'catalytic_activity' | 0.053419 | 5 | 0.21424 |
| 'transferase_activity' | 0.055556 | 8 | 0.19369 |
| 'transferase_activity_transfering' | 0.0625 | 9 | 0.16957 |
| 'phosphotransferase_activity_phosphate' | 0.027778 | 2 | 0.070615 |
| 'kinase_activity' | 0.027778 | 2 | 0.070615 |
| 'inositol_heptakisphosphate_kinase' | 0 | 0 | 0.049559 |



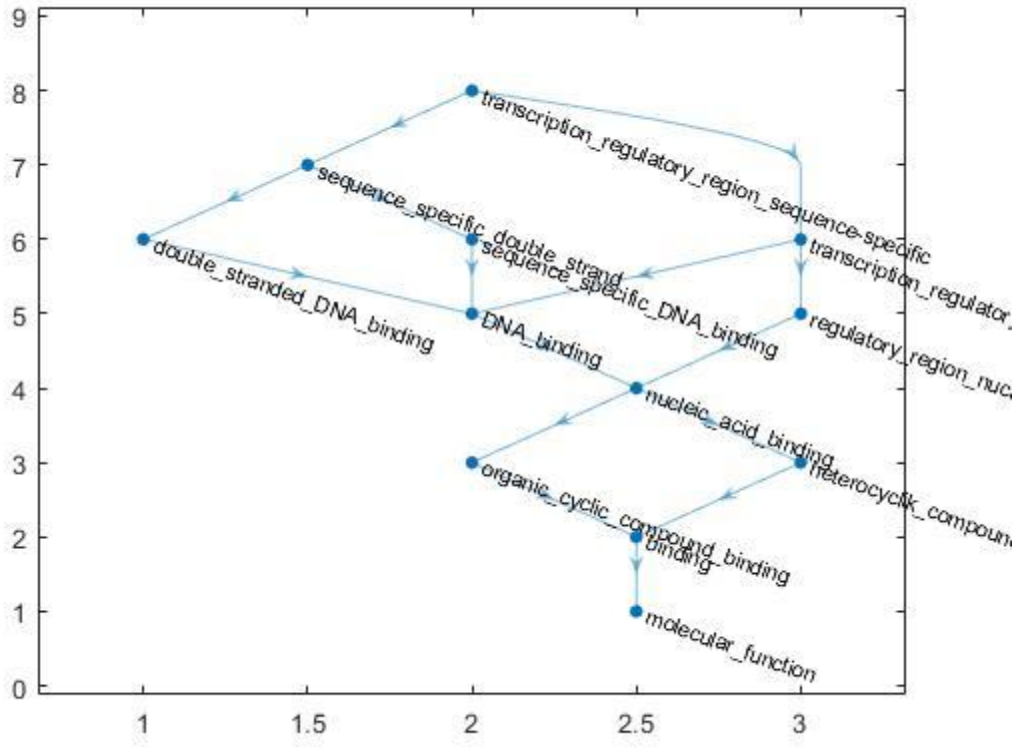
5. GO:0000832 = Inositol Hexakisphosphate 5-Kinase Activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.04 | 0 | 0.20442 |
| 'catalytic_activity' | 0.040816 | 6 | 0.19291 |
| 'transferase_activity' | 0.042517 | 10 | 0.17935 |
| 'transferase_activity_transfering' | 0.046647 | 12 | 0.16339 |
| 'phosphotransferase_activity_phosphate' | 0.027211 | 4 | 0.072293 |
| 'kinase_activity' | 0.027211 | 4 | 0.072293 |
| 'inositol_hexakisphosphate_kinase' | 0.020408 | 6 | 0.074873 |
| 'inositol_hexakiphosphate_5-kinase' | 0 | 0 | 0.040466 |



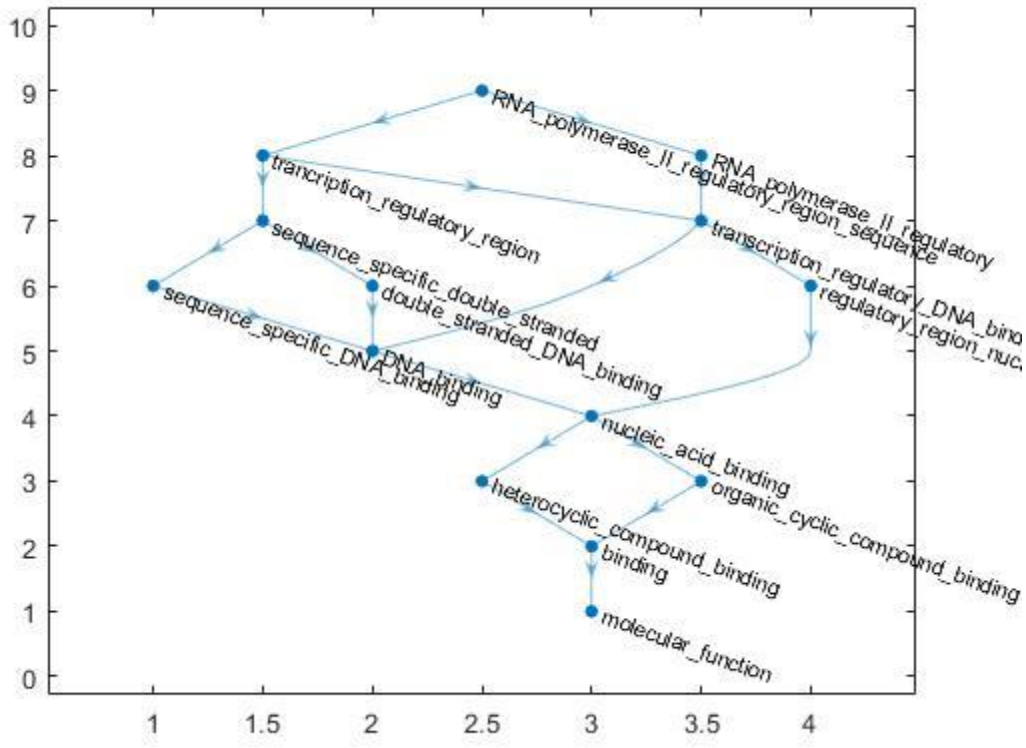

6. GO:0000976 = Transcription Regulatory Region Sequence-Specific DNA Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.023256 | 0 | 0.17401 |
| 'binding' | 0.025826 | 10 | 0.1755 |
| 'organic_cyclic_compound_binding' | 0.024042 | 8 | 0.088658 |
| 'heterocyclik_compound_binding' | 0.024042 | 8 | 0.088658 |
| 'nucleic_acid_binding' | 0.028926 | 28 | 0.15025 |
| 'regulatory_region_nucleic_acid_binding' | 0.011019 | 5 | 0.039878 |
| 'DNA_binding' | 0.029516 | 20 | 0.10768 |
| 'double_stranded_DNA_binding' | 0.011019 | 3 | 0.039878 |
| 'sequence_specific_DNA_binding' | 0.011019 | 3 | 0.039878 |
| 'transcription_regulatory_region_DNA' | 0.0082645 | 7 | 0.035391 |
| 'sequence_specific_double_strand' | 0.0082645 | 2 | 0.035391 |
| 'transcription_regulatory_region_sequence-specific' | 0 | 0 | 0.024833 |



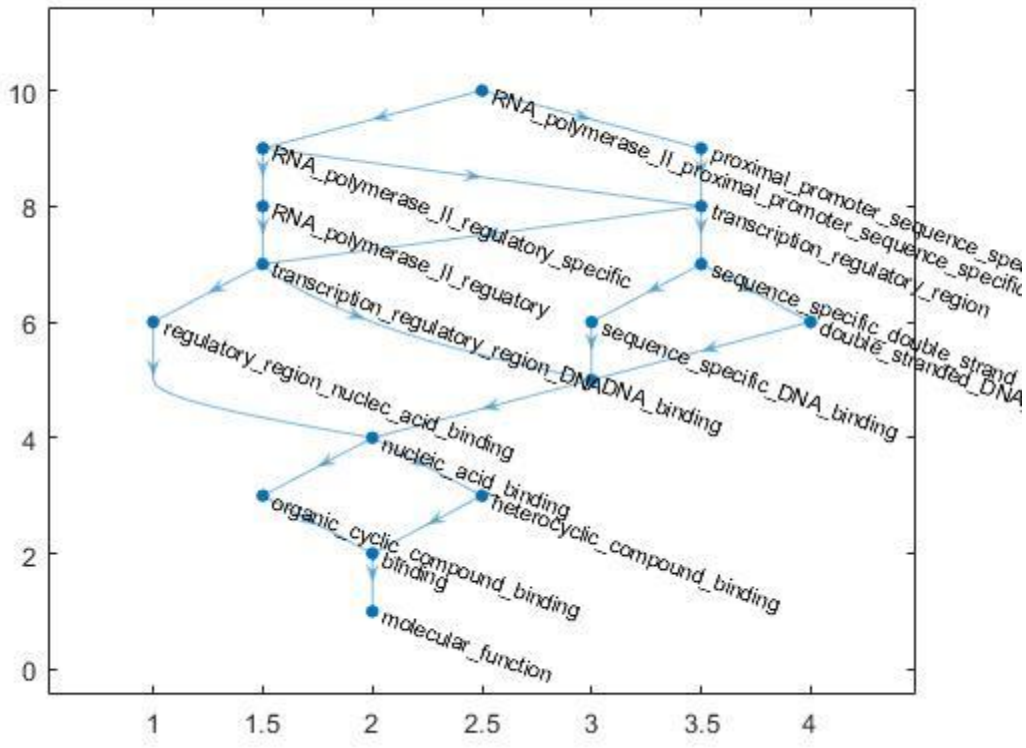

7. GO:0000977 = RNA polymerase II transcription regulatory region sequence-specific DNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.017857 | 0 | 0.15495 |
| 'binding' | 0.019816 | 12 | 0.15854 |
| 'heterocyclic_compound_binding' | 0.019088 | 10 | 0.081352 |
| 'organic_cyclic_compound_binding' | 0.019088 | 10 | 0.081352 |
| 'nucleic_acid_binding' | 0.022823 | 36 | 0.14401 |
| 'regulatory_region_nucleic_acid_binding' | 0.011834 | 10 | 0.044164 |
| 'DNA_binding' | 0.024162 | 25 | 0.10163 |
| 'transcription_regulatory_DNA_binding' | 0.013314 | 21 | 0.056624 |
| 'sequence_specific_DNA_binding' | 0.0088757 | 3 | 0.033828 |
| 'double_stranded_DNA_binding' | 0.0088757 | 3 | 0.033828 |
| 'sequence_specific_double_stranded' | 0.0078895 | 4 | 0.032286 |
| 'RNA_polymerase_II_regulatory' | 0.0059172 | 4 | 0.028659 |
| 'trancription_regulatory_region' | 0.0059172 | 7 | 0.028659 |
| 'RNA_polymerase_II_regulatory_region_sequence' | 0 | 0 | 0.020117 |



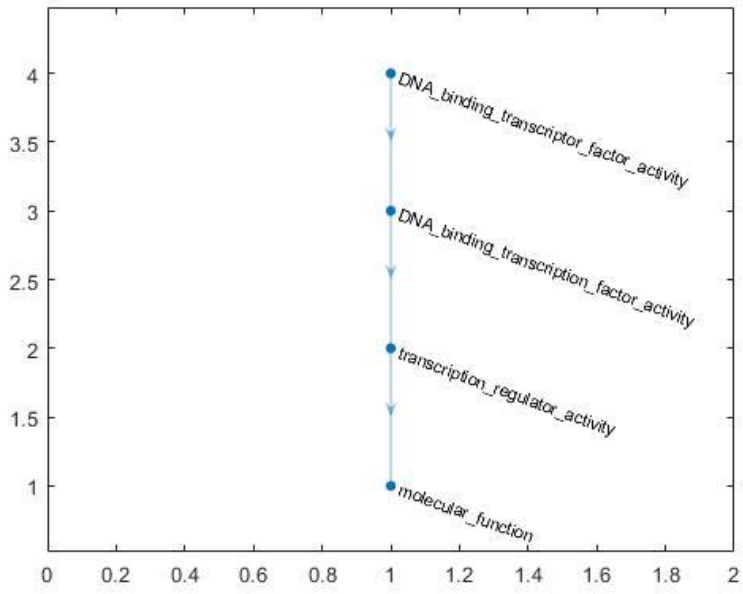

8. GO:0000978 = RNA polymerase II cis-regulatory region sequence-specific DNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.014085 | 0 | 0.13774 |
| 'binding' | 0.015556 | 14 | 0.14238 |
| 'organic_cyclic_compound_binding' | 0.015238 | 12 | 0.073917 |
| 'heterocyclic_compound_binding' | 0.015238 | 12 | 0.073917 |
| 'nucleic_acid_binding' | 0.017926 | 44 | 0.13462 |
| 'regulatory_region_nuclec_acid_binding' | 0.010667 | 15 | 0.041945 |
| 'DNA_binding' | 0.018947 | 30 | 0.096795 |
| 'transcription_regulatory_region_DNA' | 0.012346 | 35 | 0.059432 |
| 'sequence_specific_DNA_binding' | 0.008547 | 3 | 0.03227 |
| 'double_stranded_DNA_binding' | 0.008547 | 3 | 0.03227 |
| 'sequence_specific_double_strand' | 0.0088889 | 8 | 0.036662 |
| 'RNA_polymerase_II_reguatory' | 0.0059259 | 6.6667 | 0.026796 |
| 'transcription_regulatory_region' | 0.01 | 26.333 | 0.047004 |
| 'RNA_polymerase_II_regulatory_specific' | 0.0044444 | 8.3333 | 0.023783 |
| 'proximal_promoter_sequence_specific' | 0.0044444 | 4.6667 | 0.023783 |
| 'RNA_polymerase_II_proximal_promoter_sequence_specific_DNA_binding' | 0 | 0 | 0.016691 |



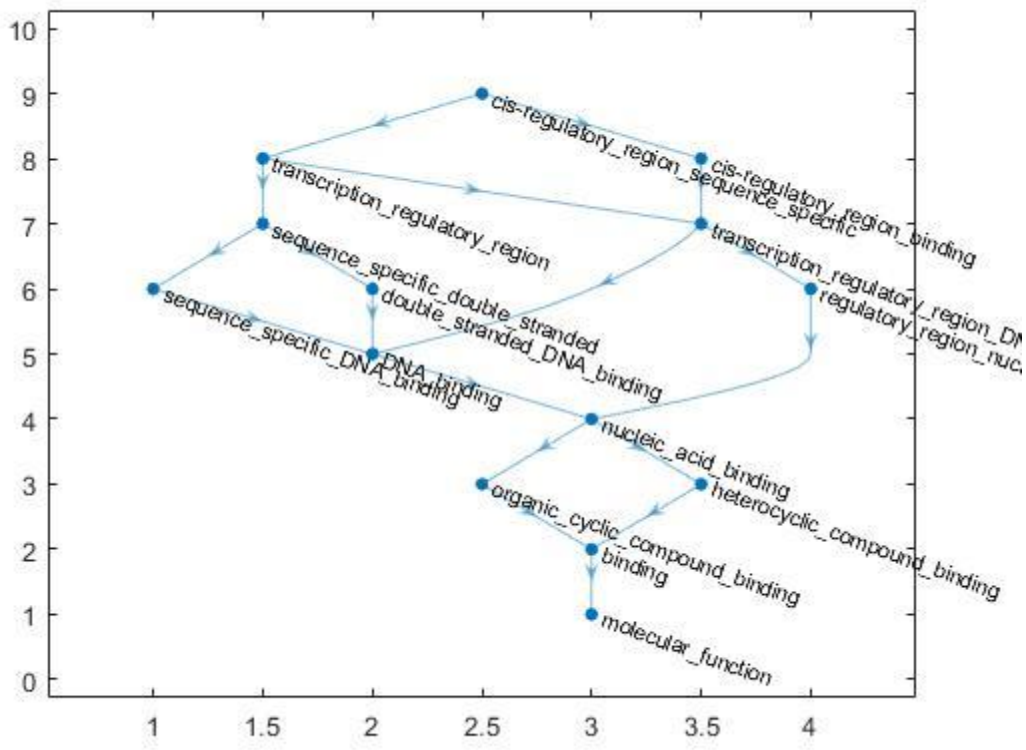

9. GO:0000981 = DNA-binding transcription factor activity, RNA polymerase II-specific

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.16667 | 0 | 0.37018 |
| 'transcription_regulator_activity' | 0.14815 | 2 | 0.2988 |
| 'DNA_binding_transcription_factor_activity' | 0.11111 | 2 | 0.21486 |
| 'DNA_binding_transcriptor_factor_activity' | 0 | 0 | 0.11616 |



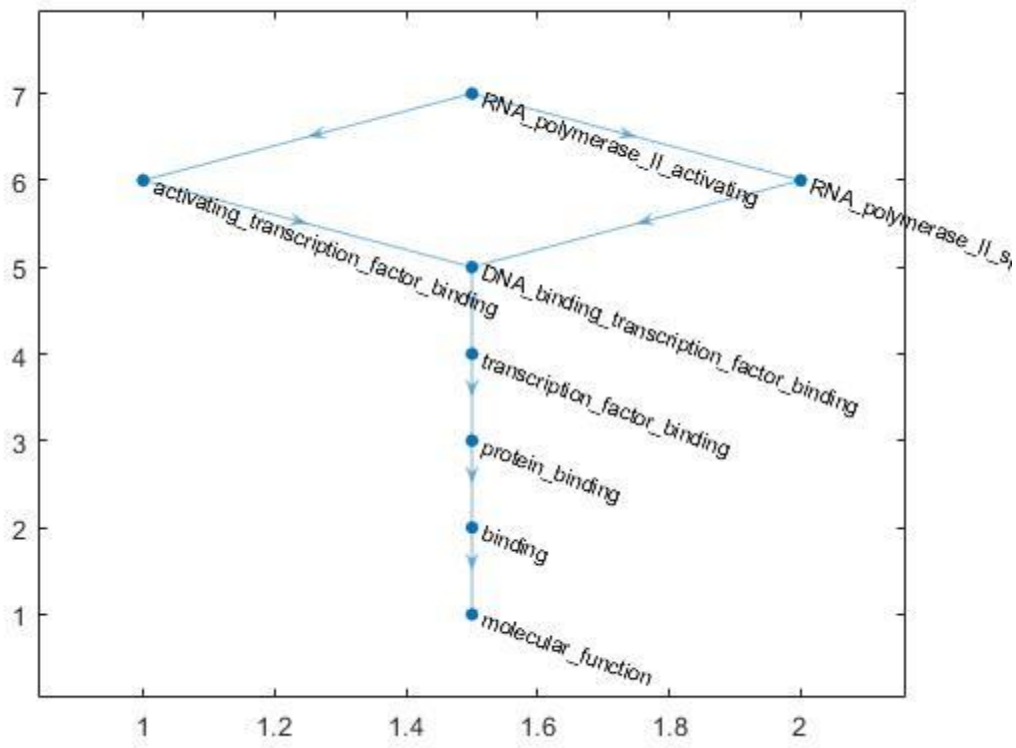
10. GO:0000987 = Cis-regulatory region sequence-specific DNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.017857 | 0 | 0.15495 |
| 'binding' | 0.019816 | 12 | 0.15854 |
| 'organic_cyclic_compound_binding' | 0.019088 | 10 | 0.081352 |
| 'heterocyclic_compound_binding' | 0.019088 | 10 | 0.081352 |
| 'nucleic_acid_binding' | 0.022823 | 36 | 0.14401 |
| 'regulatory_region_nucleic_acid_binding' | 0.011834 | 10 | 0.044164 |
| 'DNA_binding' | 0.024162 | 25 | 0.10163 |
| 'transcription_regulatory_region_DNA' | 0.013314 | 21 | 0.056624 |
| 'sequence_specific_DNA_binding' | 0.0088757 | 3 | 0.033828 |
| 'double_stranded_DNA_binding' | 0.0088757 | 3 | 0.033828 |
| 'sequence_specific_double_stranded' | 0.0078895 | 4 | 0.032286 |
| 'cis-regulatory_region_binding' | 0.0059172 | 4 | 0.028659 |
| 'transcription_regulatory_region' | 0.0059172 | 7 | 0.028659 |
| 'cis-regulatory_region_sequence_specific' | 0 | 0 | 0.020117 |



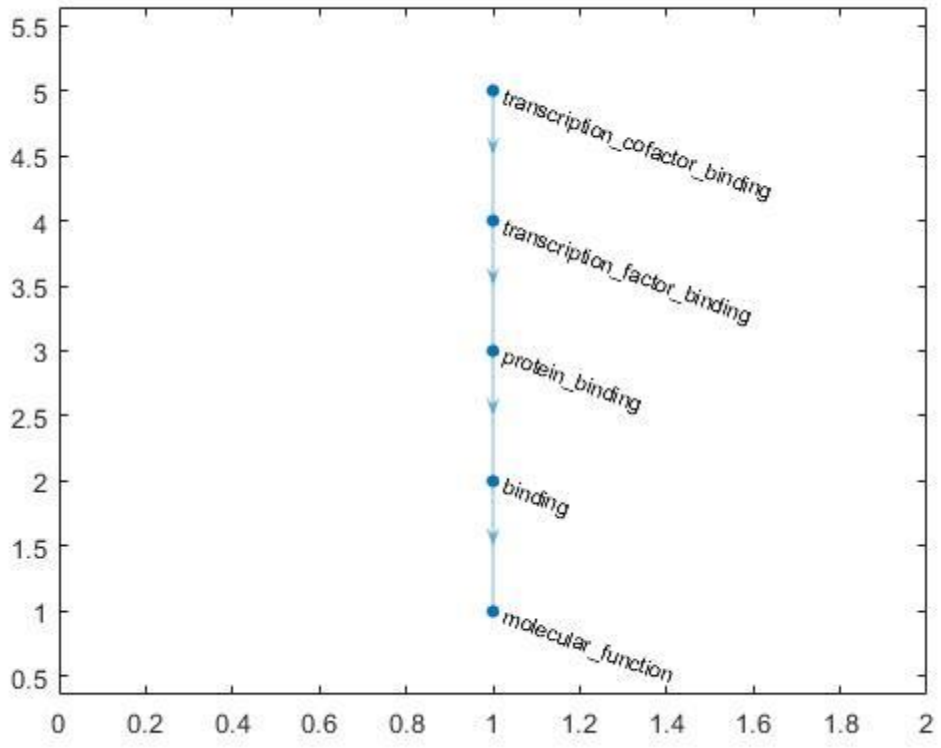
11. GO:0001102 = RNA polymerase II activating transcription factor binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.038462 | 0 | 0.19767 |
| 'binding' | 0.038668 | 6 | 0.18584 |
| 'protein_binding' | 0.039246 | 10 | 0.1719 |
| 'transcription_factor_binding' | 0.040816 | 12 | 0.15546 |
| 'DNA_binding_transcription_factor_binding' | 0.045918 | 12 | 0.13609 |
| 'activating_transcription_factor_binding' | 0.020408 | 2.5 | 0.056648 |
| 'RNA_polymerase_II_specific' | 0.020408 | 2.5 | 0.056648 |
| 'RNA_polymerase_II_activating' | 0 | 0 | 0.039749 |



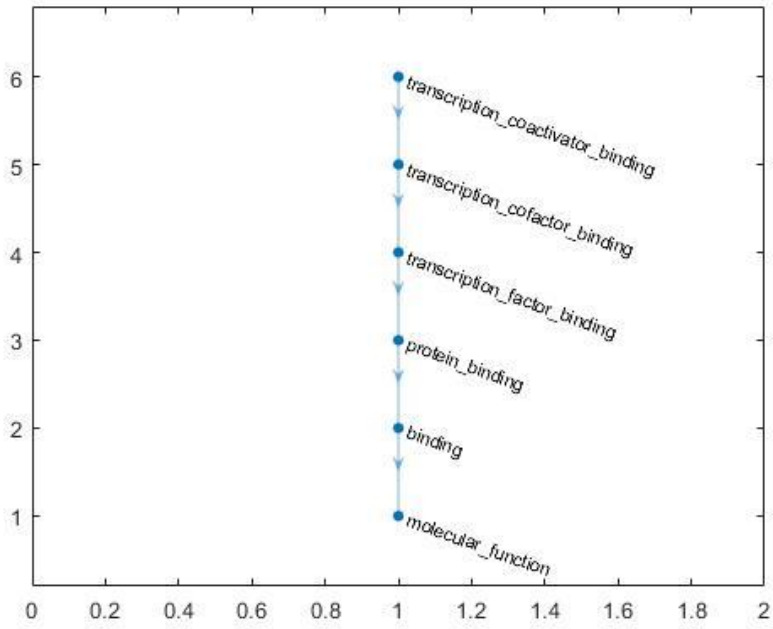

12. GO:0001221 = Transcription cofactor binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.1 | 0 | 0.30108 |
| 'binding' | 0.09375 | 3 | 0.25872 |
| 'protein_binding' | 0.083333 | 4 | 0.20885 |
| 'transcription_factor_binding' | 0.0625 | 3 | 0.15018 |
| 'transcription_cofactor_binding' | 0 | 0 | 0.081168 |



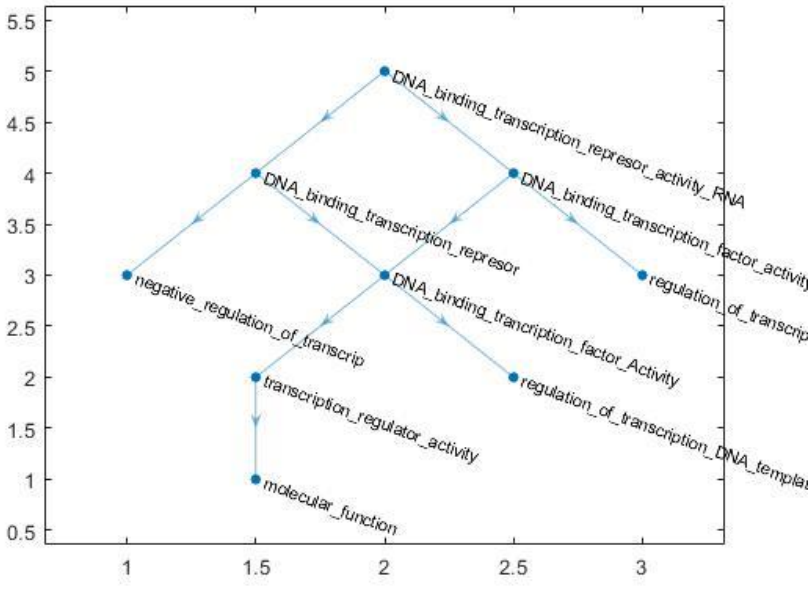

13. GO:0001223 = Transcription coactivator binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.066667 | 0 | 0.25218 |
| 'binding' | 0.064 | 4 | 0.22516 |
| 'protein_binding' | 0.06 | 6 | 0.1934 |
| 'transcription_factor_binding' | 0.053333 | 6 | 0.15618 |
| 'transcription_cofactor_binding' | 0.04 | 4 | 0.11234 |
| 'transcription_coactivator_binding' | 0 | 0 | 0.06073 |



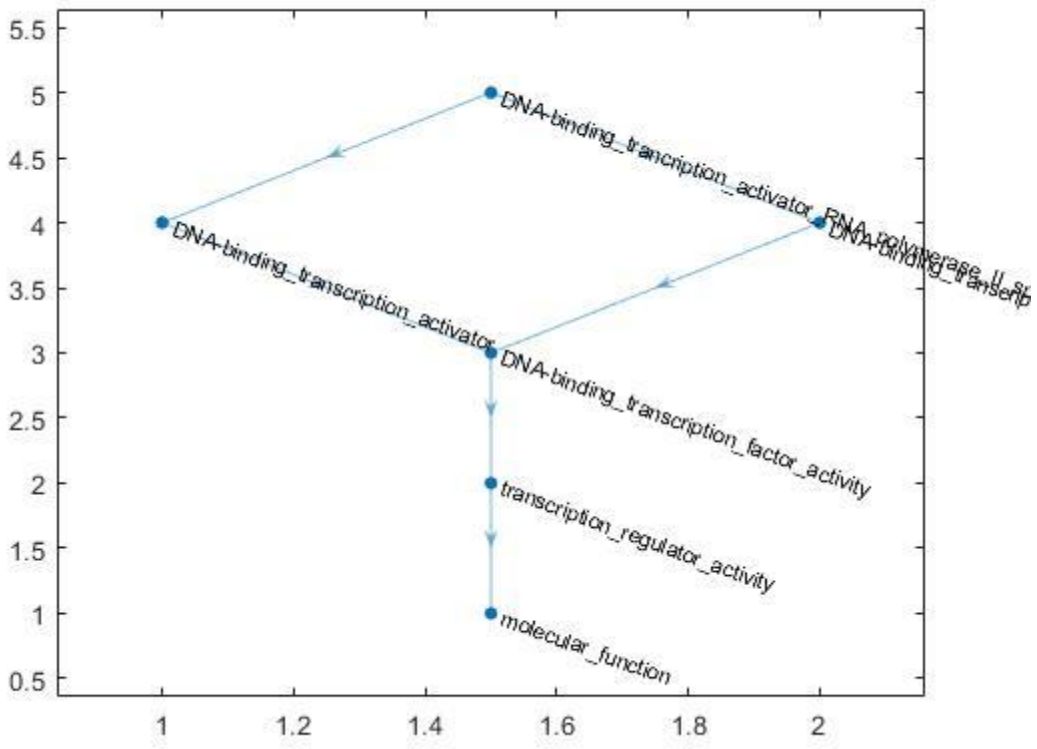

14. GO:0001227 = DNA-binding transcription repressor activity, RNA polymerase II-specific

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.030048 | 0 | 0.16769 |
| 'transcription_regulator_activity' | 0.03125 | 4 | 0.12276 |
| 'regulation_of_transcription_DNA_template' | 0.03125 | 0 | 0.12276 |
| 'DNA_binding_trancription_factor_Activity' | 0.035156 | 9 | 0.13991 |
| 'negative_regulation_of_transcrip' | 0.020833 | 0 | 0.1016 |
| 'regulation_of_transcription_by_RNA' | 0.020833 | 0 | 0.1016 |
| 'DNA_binding_transcription_represor' | 0.015625 | 3 | 0.090185 |
| 'DNA_binding_transcription_factor_activity_RNA_polymerase' | 0.015625 | 3 | 0.090185 |
| 'DNA_binding_transcription_represor_activity_RNA' | 0 | 0 | 0.063292 |



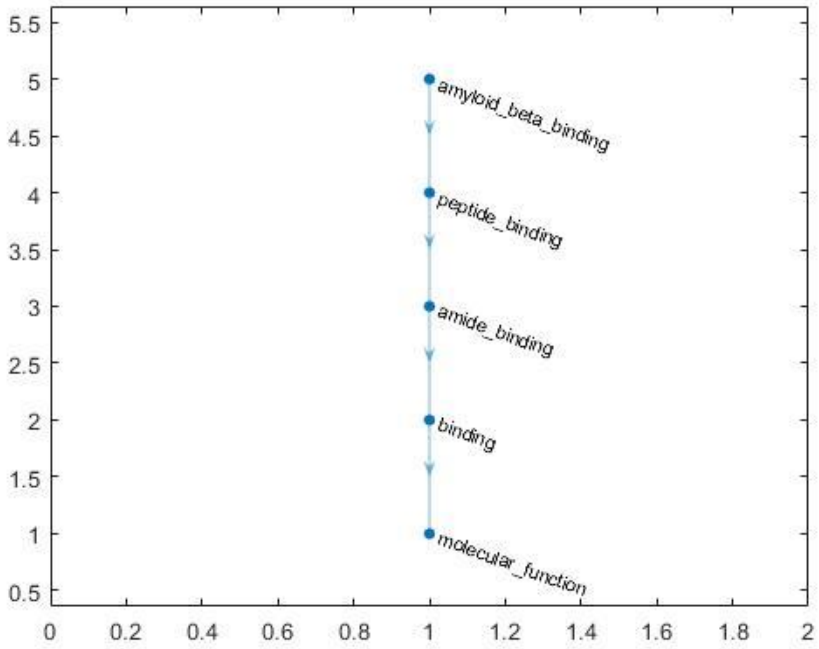
15. GO:0001228 = DNA-binding transcription activator activity, RNA polymerase II-specific

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.076923 | 0 | 0.27878 |
| 'transcription_regulator_activity' | 0.08 | 4 | 0.25215 |
| 'DNA-binding_transcription_factor_activity' | 0.09 | 6 | 0.22077 |
| 'DNA-binding_transcription_activator' | 0.04 | 1.5 | 0.091907 |
| 'DNA-binding_transcription_factor_activity_RNA_polymerase' | 0.04 | 1.5 | 0.091907 |
| 'DNA-binding_trancription_activator_RNA_polymerase_II_specific' | 0 | 0 | 0.064492 |



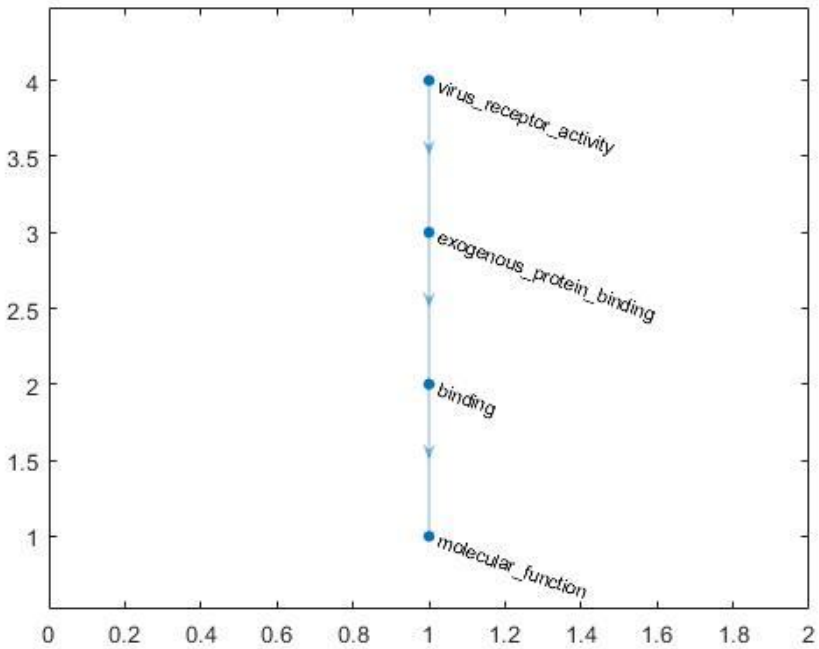

16. GO:0001540 = Amyloid-beta binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.1 | 0 | 0.30108 |
| 'binding' | 0.09375 | 3 | 0.25872 |
| 'amide_binding' | 0.083333 | 4 | 0.20885 |
| 'peptide_binding' | 0.0625 | 3 | 0.15018 |
| 'amyloid_beta_binding' | 0 | 0 | 0.081168 |



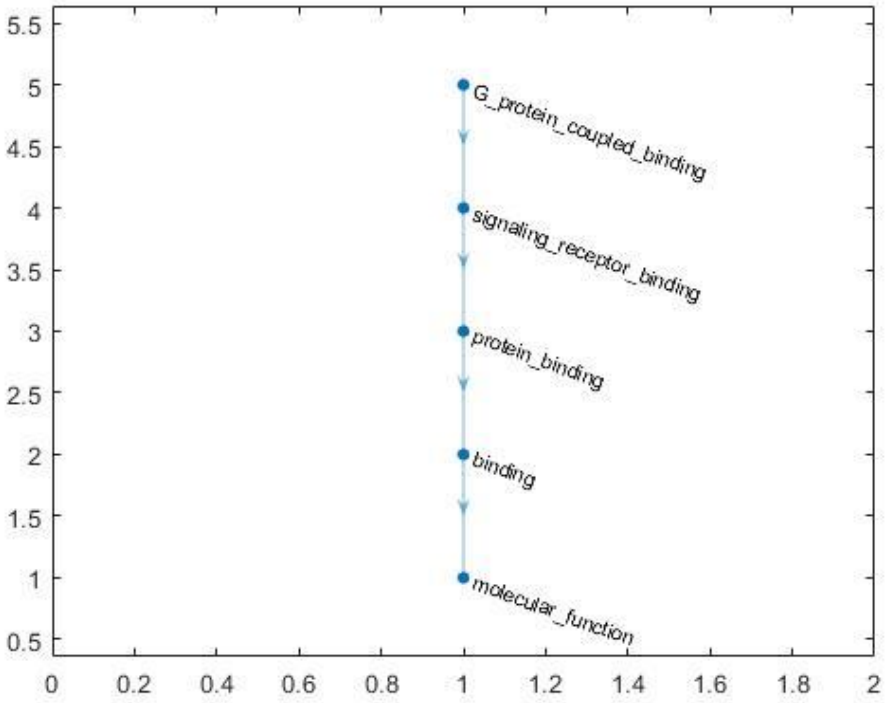

17. GO:0001618 = Virus receptor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.16667 | 0 | 0.37018 |
| 'binding' | 0.14815 | 2 | 0.2988 |
| 'exogenous_protein_binding' | 0.11111 | 2 | 0.21486 |
| 'virus_receptor_activity' | 0 | 0 | 0.11616 |



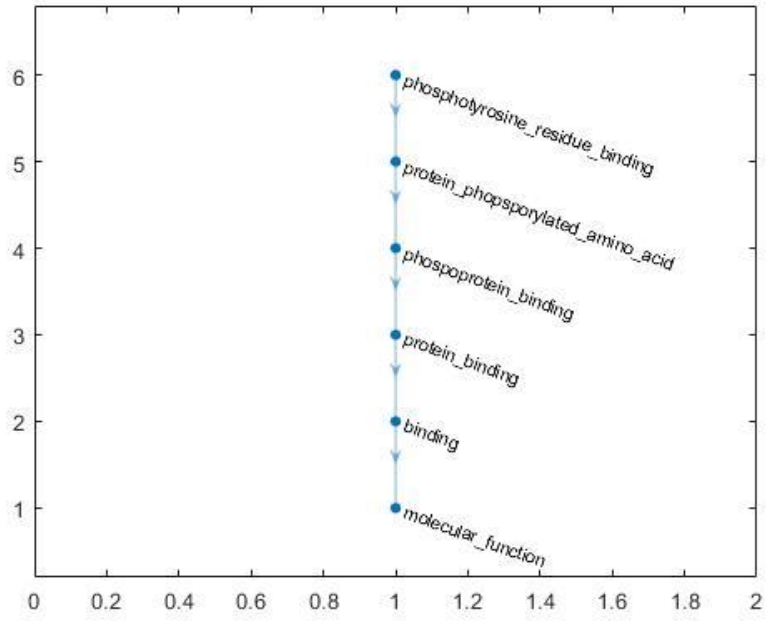
18. GO:0001664 = G protein-coupled receptor binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.1 | 0 | 0.30108 |
| 'binding' | 0.09375 | 3 | 0.25872 |
| 'protein_binding' | 0.083333 | 4 | 0.20885 |
| 'signaling_receptor_binding' | 0.0625 | 3 | 0.15018 |
| 'G_protein_coupled_binding' | 0 | 0 | 0.081168 |



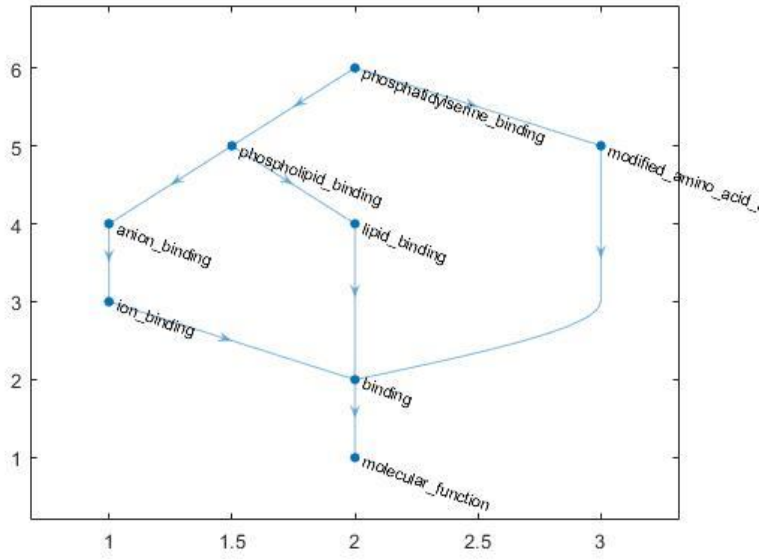

19. GO:0001784 = Phosphotyrosine residue binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.066667 | 0 | 0.25218 |
| 'binding' | 0.064 | 4 | 0.22516 |
| 'protein_binding' | 0.06 | 6 | 0.1934 |
| 'phospoprotein_binding' | 0.053333 | 6 | 0.15618 |
| 'protein_phopsporylated_amino_acid' | 0.04 | 4 | 0.11234 |
| 'phosphotyrosine_residue_binding' | 0 | 0 | 0.06073 |



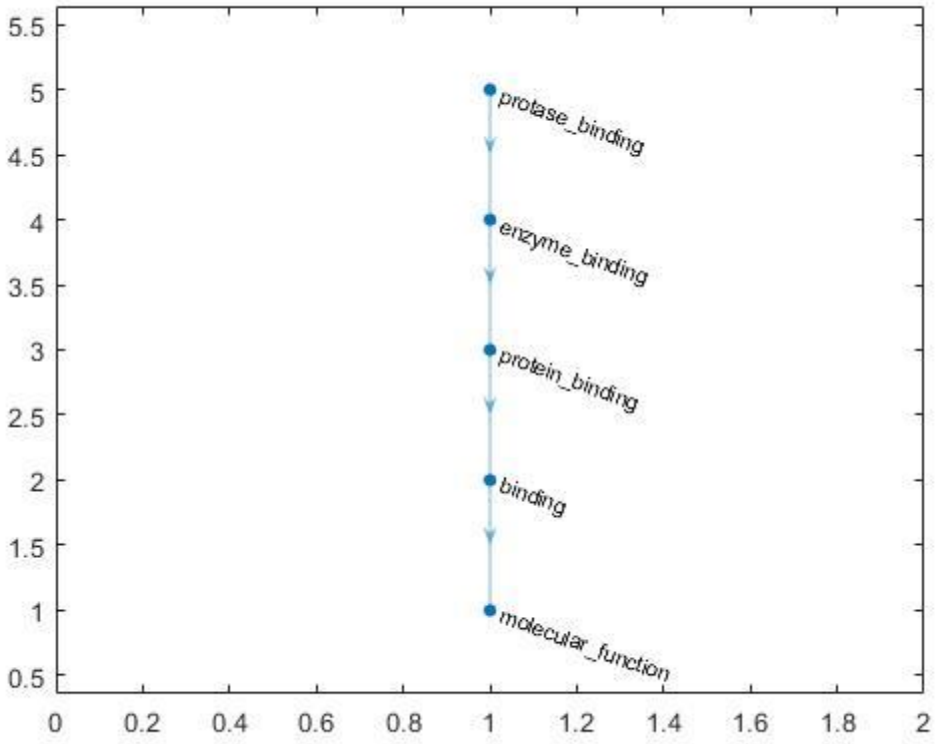

20. GO:0001786 = Phosphatidylserine binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.0625 | 0 | 0.27685 |
| 'binding' | 0.081633 | 6 | 0.26909 |
| 'ion_binding' | 0.030612 | 2 | 0.11391 |
| 'anion_binding' | 0.027211 | 2 | 0.077348 |
| 'lipid_binding' | 0.027211 | 2 | 0.077348 |
| 'modified_amino_acid_binding' | 0.020408 | 2 | 0.068646 |
| 'phospholipid_binding' | 0.020408 | 3 | 0.068646 |
| 'phosphatidylserine_binding' | 0 | 0 | 0.048167 |



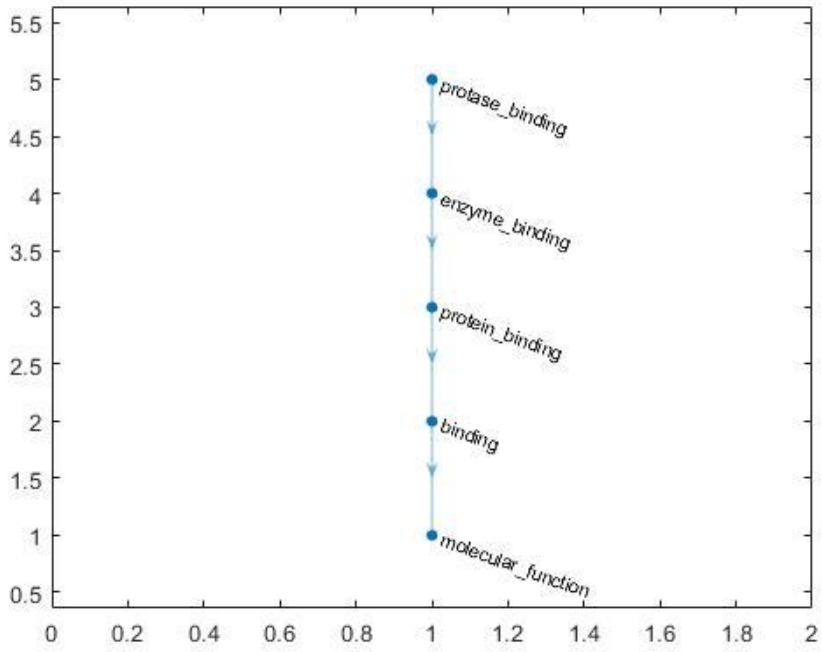
21. GO:0001968 = Fibronectin Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.16667 | 0 | 0.37018 |
| 'binding' | 0.14815 | 2 | 0.2988 |
| 'protein_binding' | 0.11111 | 2 | 0.21486 |
| 'fibronectin_binding' | 0 | 0 | 0.11616 |



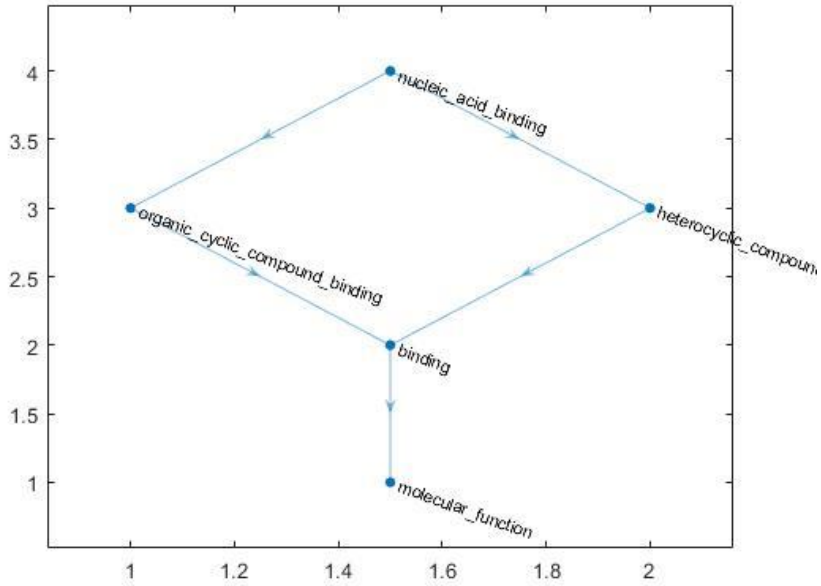

22. GO:0002020 = Protease binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.1 | 0 | 0.30108 |
| 'binding' | 0.09375 | 3 | 0.25872 |
| 'protein_binding' | 0.083333 | 4 | 0.20885 |
| 'enzyme_binding' | 0.0625 | 3 | 0.15018 |
| 'protase_binding' | 0 | 0 | 0.081168 |



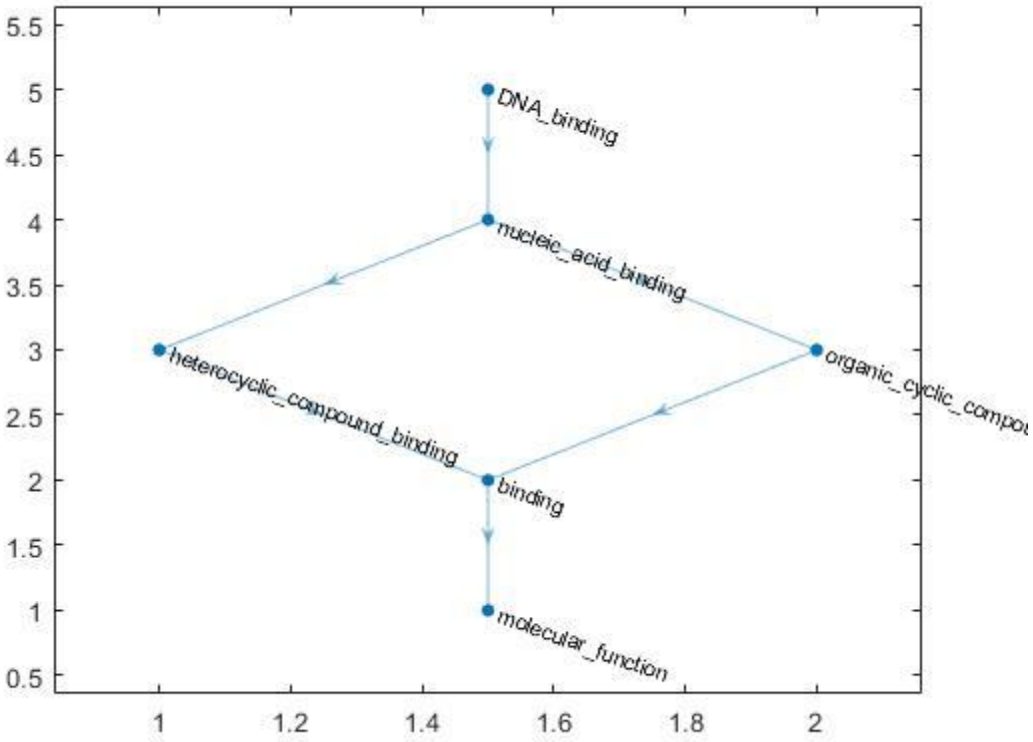

23. GO:0003676 = Nucleic acid binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.125 | 0 | 0.34958 |
| 'binding' | 0.14063 | 3 | 0.30612 |
| 'organic_cyclic_compound_binding' | 0.0625 | 1 | 0.12744 |
| 'heterocyclic_compound_binding' | 0.0625 | 1 | 0.12744 |
| 'nucleic_acid_binding' | 0 | 0 | 0.089419 |



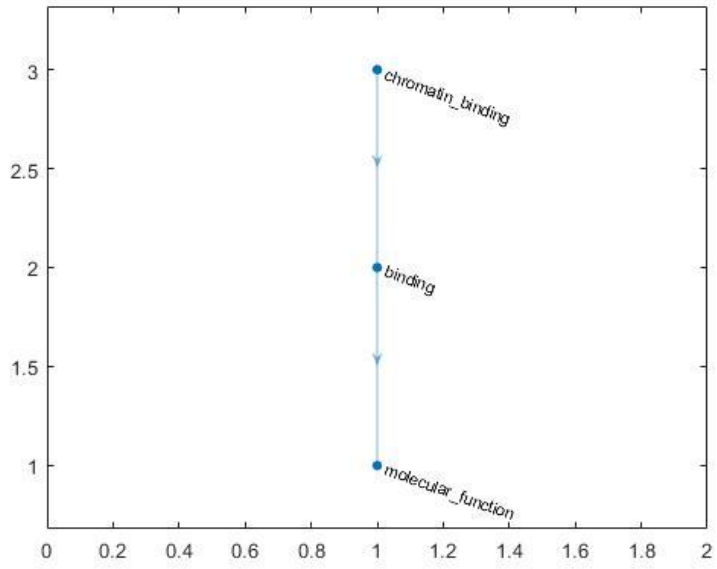
24. GO:0003677 = DNA Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.083333 | 0 | 0.29757 |
| 'binding' | 0.091429 | 4 | 0.2711 |
| 'heterocyclic_compound_binding' | 0.053333 | 2 | 0.11996 |
| 'organic_cyclic_compound_binding' | 0.053333 | 2 | 0.11996 |
| 'nucleic_acid_binding' | 0.04 | 4 | 0.12426 |
| 'DNA_binding' | 0 | 0 | 0.067148 |



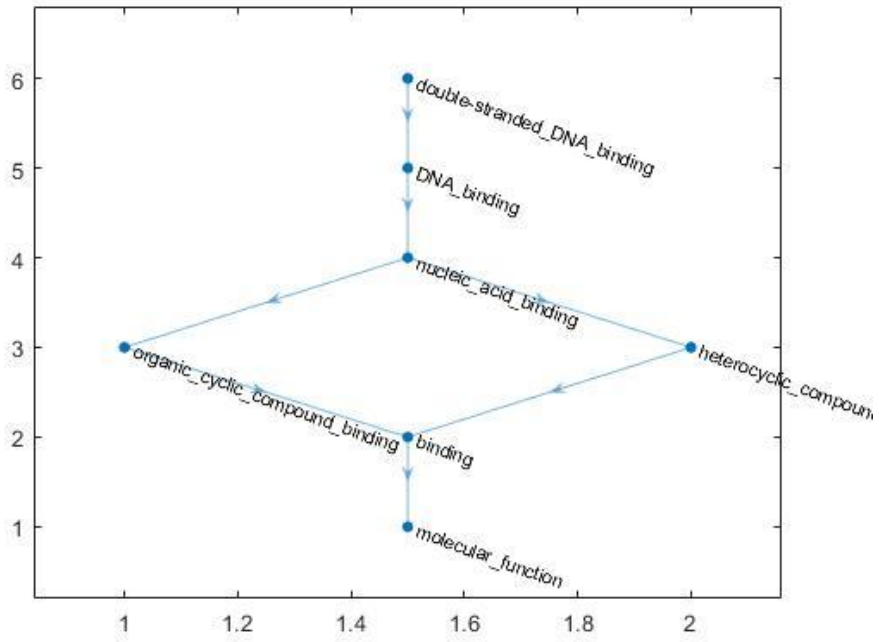

25. GO:0003682 = Chromatin binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.33333 | 0 | 0.47443 |
| 'binding' | 0.25 | 1 | 0.34117 |
| 'chromatin_binding' | 0 | 0 | 0.1844 |



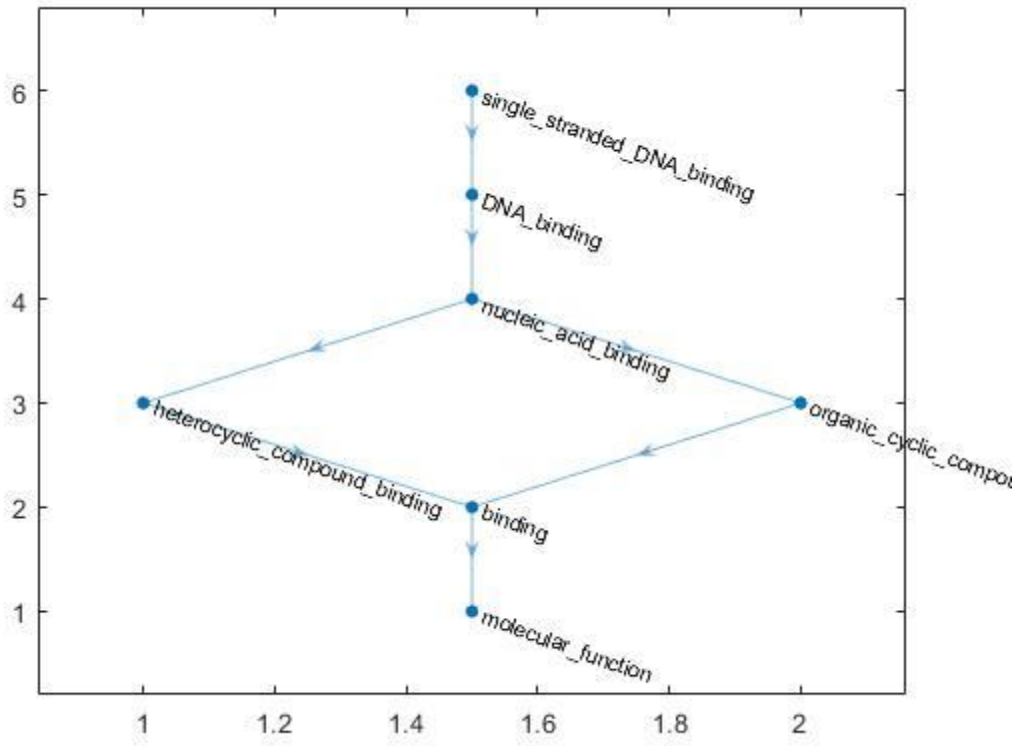

26. GO:0003690 = Double-stranded DNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'binding' | 0.063131 | 5 | 0.23935 |
| 'organic_cyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'heterocyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'nucleic_acid_binding' | 0.037037 | 8 | 0.1351 |
| 'DNA_binding' | 0.027778 | 5 | 0.097145 |
| 'double-stranded_DNA_binding' | 0 | 0 | 0.052517 |



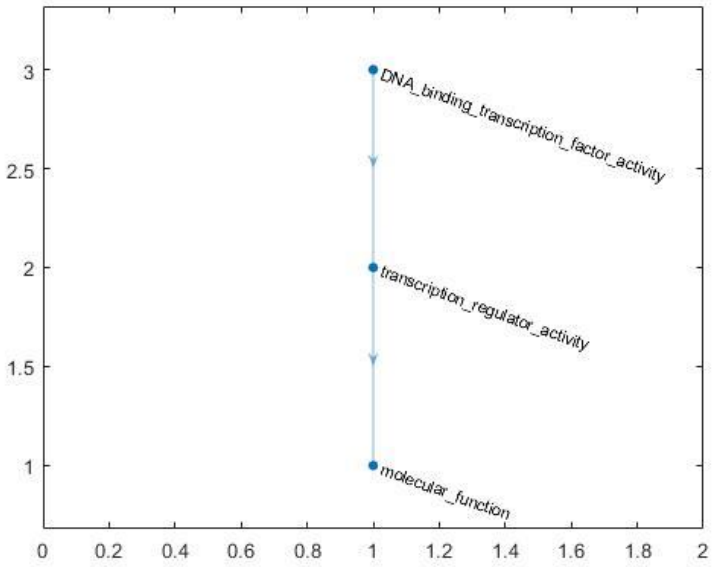
27. GO:0003697 = Single-stranded DNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'binding' | 0.063131 | 5 | 0.23935 |
| 'heterocyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'organic_cyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'nucleic_acid_binding' | 0.037037 | 8 | 0.1351 |
| 'DNA_binding' | 0.027778 | 5 | 0.097145 |
| 'single_stranded_DNA_binding' | 0 | 0 | 0.052517 |



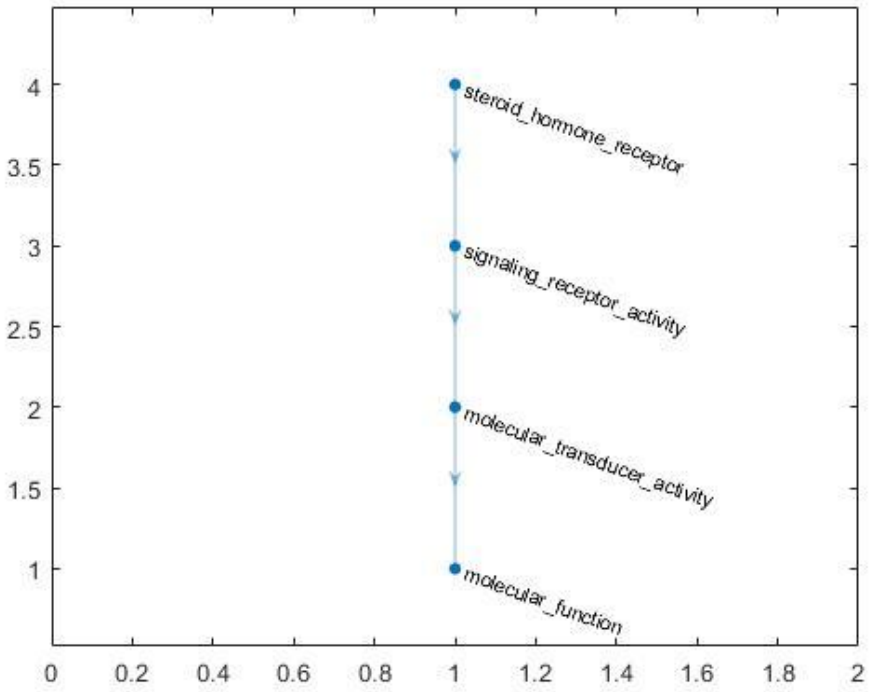

28. GO:0003700 = DNA-binding transcription factor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.33333 | 0 | 0.47443 |
| 'transcription_regulator_activity' | 0.25 | 1 | 0.34117 |
| 'DNA_binding_transcription_factor_activity' | 0 | 0 | 0.1844 |



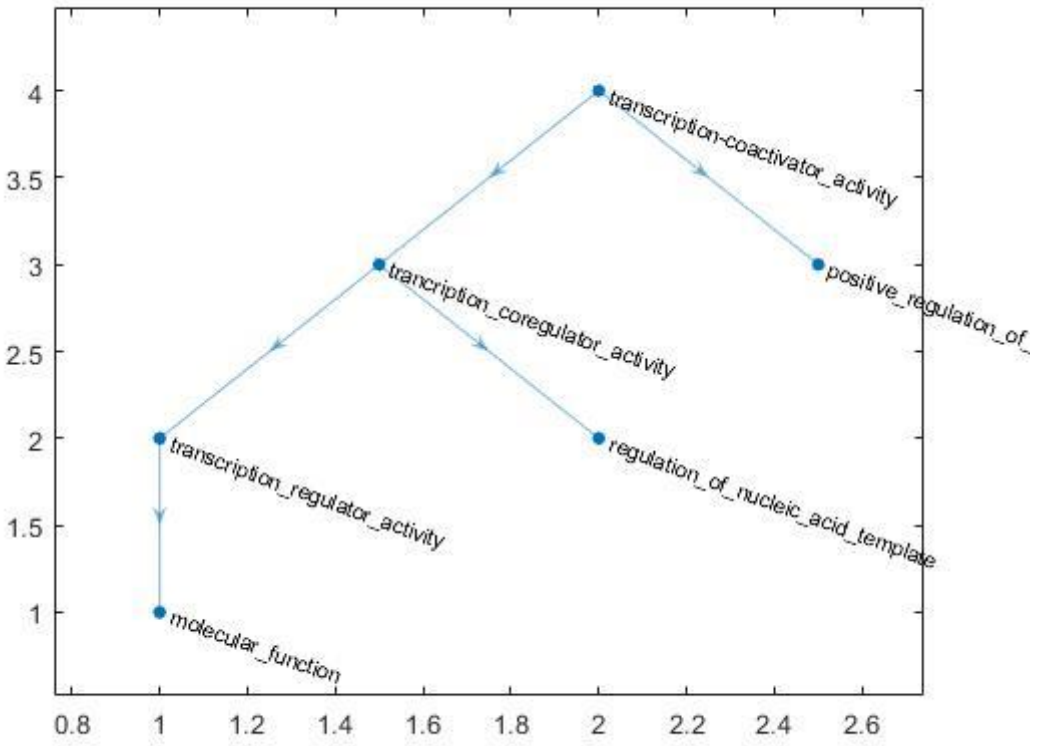

29. GO:0003707 = Steroid hormone receptor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.16667 | 0 | 0.37018 |
| 'molecular_transducer_activity' | 0.14815 | 2 | 0.2988 |
| 'signaling_receptor_activity' | 0.11111 | 2 | 0.21486 |
| 'steroid_hormone_receptor' | 0 | 0 | 0.11616 |



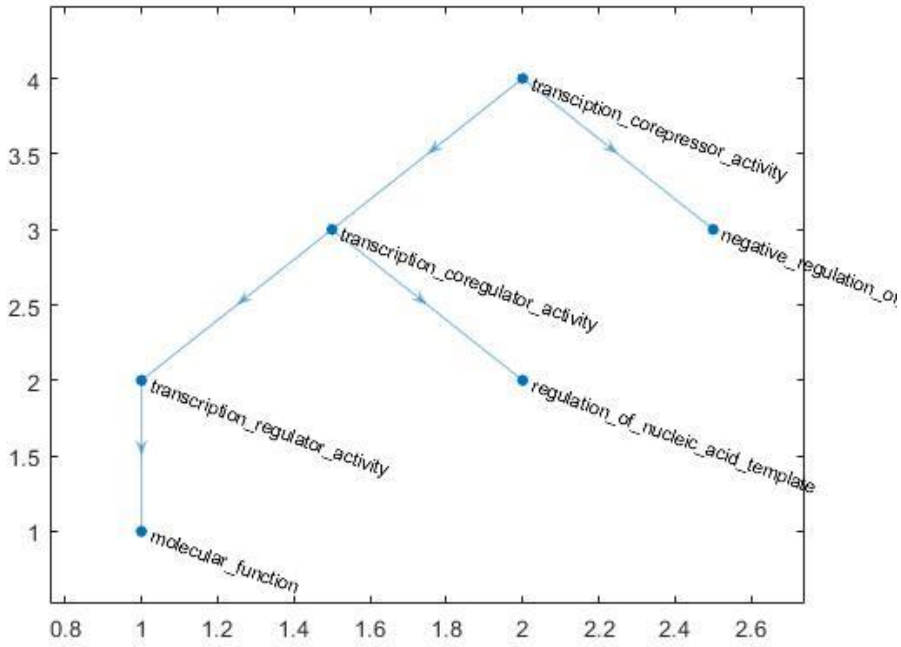
30. GO:0003713 = Transcription coactivator activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.06 | 0 | 0.2509 |
| 'transcription_regulator_activity' | 0.053333 | 2 | 0.17032 |
| 'regulation_of_nucleic_acid_template' | 0.053333 | 0 | 0.17032 |
| 'positive_regulation_of_nucleic' | 0.04 | 0 | 0.15118 |
| 'trancription_coregulator_activity' | 0.04 | 3 | 0.15118 |
| 'transcription-coactivator_activity' | 0 | 0 | 0.10609 |



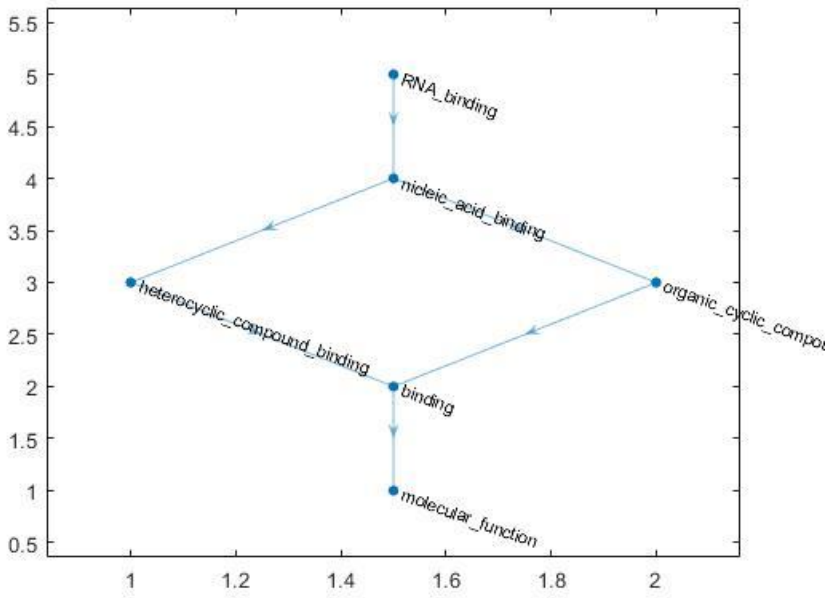

31. GO:0003714 = Transcription corepressor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.06 | 0 | 0.2509 |
| 'transcription_regulator_activity' | 0.053333 | 2 | 0.17032 |
| 'regulation_of_nucleic_acid_template' | 0.053333 | 0 | 0.17032 |
| 'transcription_coregulator_activity' | 0.04 | 3 | 0.15118 |
| 'negative_regulation_of_nucleic_acid' | 0.04 | 0 | 0.15118 |
| 'transciption_corepressor_activity' | 0 | 0 | 0.10609 |



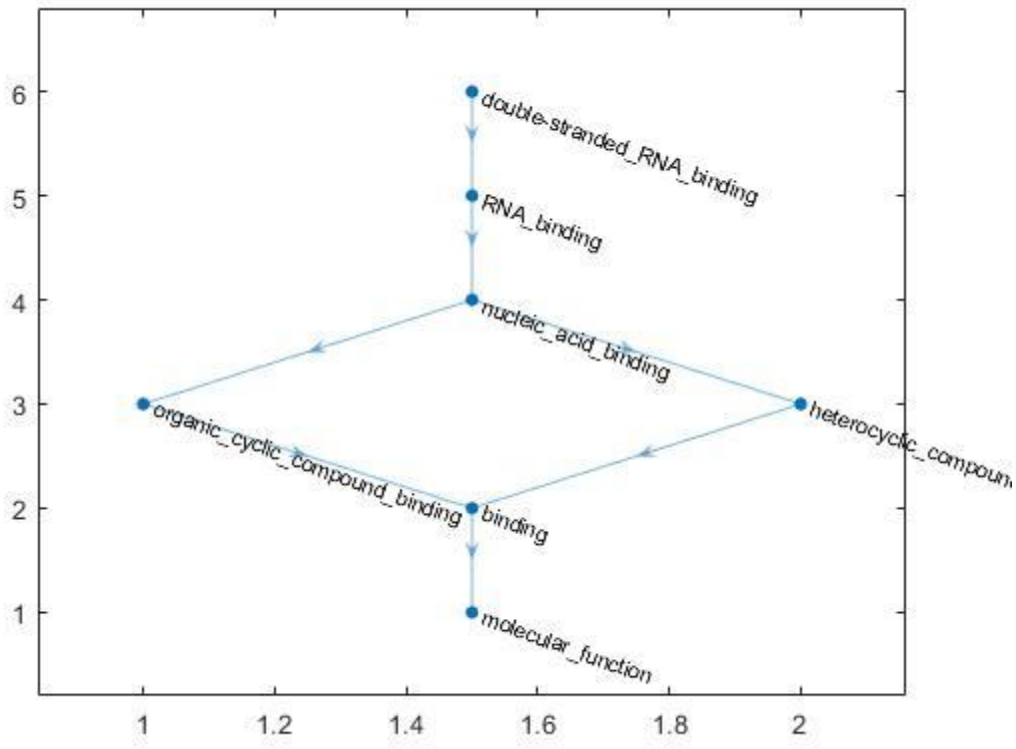

32. GO:0003723 = RNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.083333 | 0 | 0.29757 |
| 'binding' | 0.091429 | 4 | 0.2711 |
| 'heterocyclic_compound_binding' | 0.053333 | 2 | 0.11996 |
| 'organic_cyclic_compound_binding' | 0.053333 | 2 | 0.11996 |
| 'nicleic_acid_binding' | 0.04 | 4 | 0.12426 |
| 'RNA_binding' | 0 | 0 | 0.067148 |



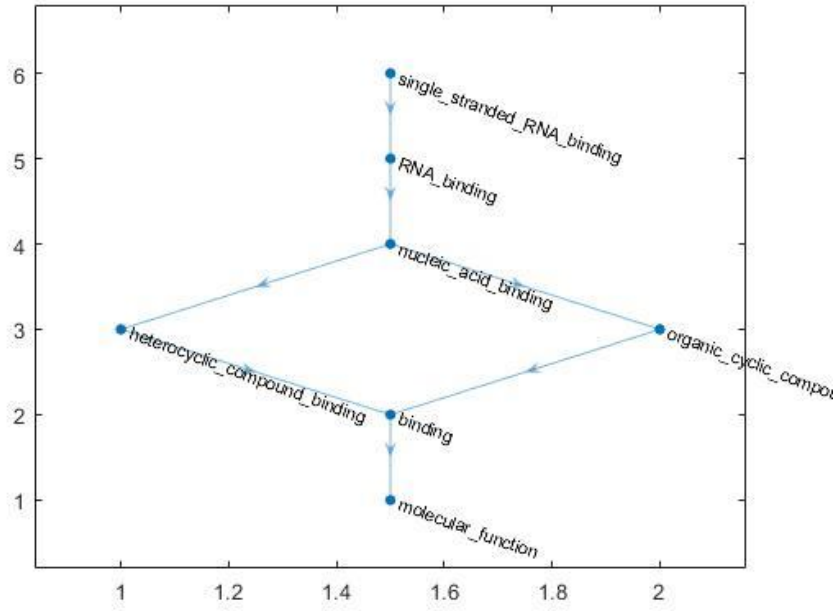
33. GO:0003725 = Double-stranded RNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'binding' | 0.063131 | 5 | 0.23935 |
| 'organic_cyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'heterocyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'nucleic_acid_binding' | 0.037037 | 8 | 0.1351 |
| 'RNA_binding' | 0.027778 | 5 | 0.097145 |
| 'double-stranded_RNA_binding' | 0 | 0 | 0.052517 |



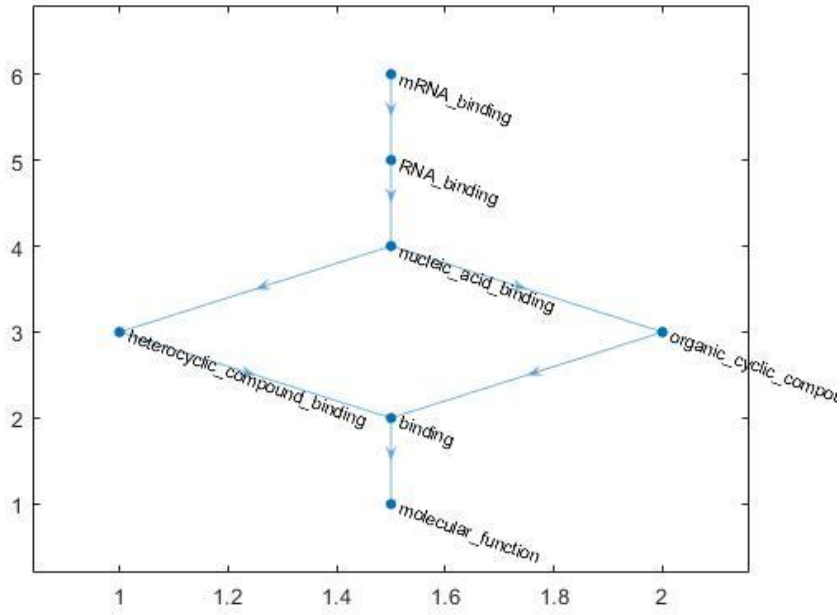

34. GO:0003727 = Single-stranded RNA binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'binding' | 0.063131 | 5 | 0.23935 |
| 'heterocyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'organic_cyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'nucleic_acid_binding' | 0.037037 | 8 | 0.1351 |
| 'RNA_binding' | 0.027778 | 5 | 0.097145 |
| 'single_stranded_RNA_binding' | 0 | 0 | 0.052517 |



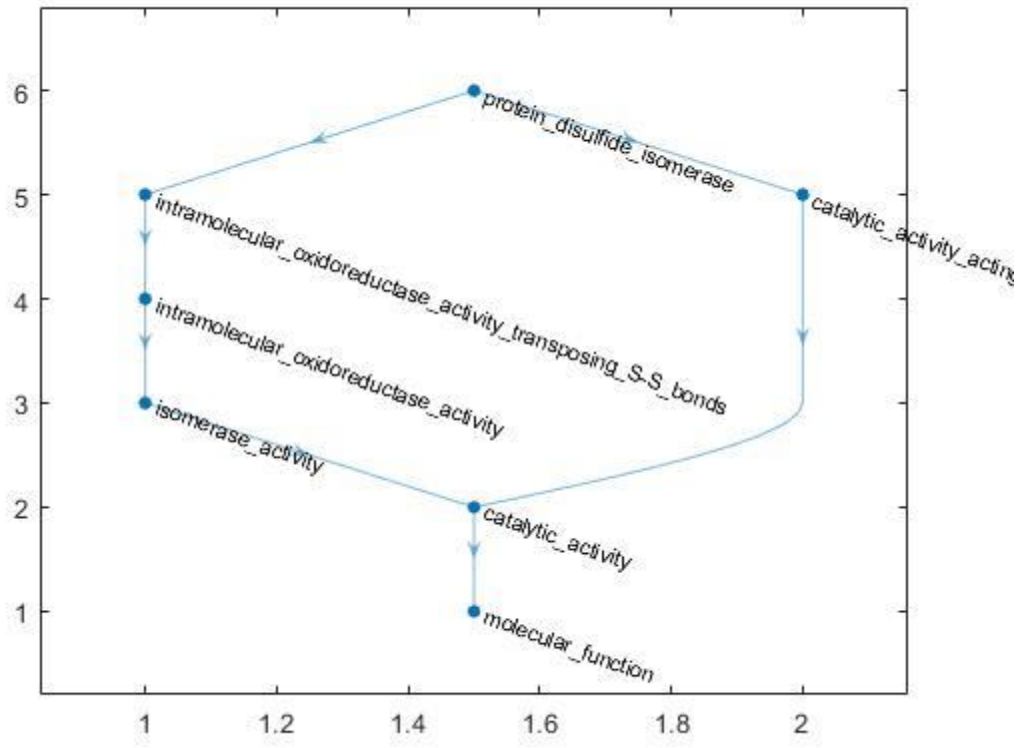

35. GO:0003729 = mRNA Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'binding' | 0.063131 | 5 | 0.23935 |
| 'heterocyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'organic_cyclic_compound_binding' | 0.041667 | 3 | 0.10993 |
| 'nucleic_acid_binding' | 0.037037 | 8 | 0.1351 |
| 'RNA_binding' | 0.027778 | 5 | 0.097145 |
| 'mRNA_binding' | 0 | 0 | 0.052517 |



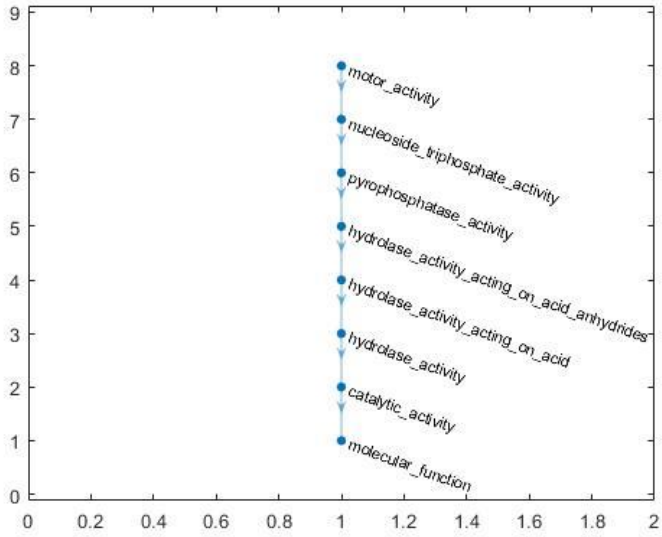

36. GO:0003756 = Protein disulfide isomerase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.066667 | 0 | 0.26722 |
| 'catalytic_activity' | 0.07716 | 5 | 0.25108 |
| 'isomerase_activity' | 0.041667 | 4 | 0.15519 |
| 'intramolecular_oxidoreductase_activity' | 0.037037 | 4 | 0.11912 |
| 'catalytic_activity_acting_on_a_protein' | 0.027778 | 2 | 0.07676 |
| 'intramolecular_oxidoreductase_activity_transposing_S-S_bonds' | 0.027778 | 2 | 0.07676 |
| 'protein_disulfide_isomerase' | 0 | 0 | 0.053872 |



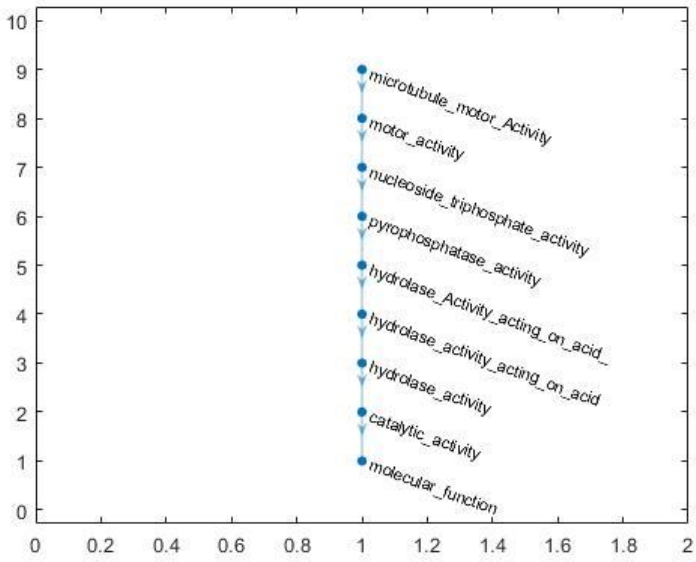

37. GO:0003774 = Motor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.035714 | 0 | 0.18764 |
| 'catalytic_activity' | 0.034985 | 6 | 0.17526 |
| 'hydrolase_activity' | 0.034014 | 10 | 0.16066 |
| 'hydrolase_activity_acting_on_acid' | 0.032653 | 12 | 0.14348 |
| 'hydrolase_activity_acting_on_acid_anhydrides' | 0.030612 | 12 | 0.12325 |
| 'pyrophosphatase_activity' | 0.027211 | 10 | 0.099478 |
| 'nucleoside_triphosphate_activity' | 0.020408 | 6 | 0.071551 |
| 'motor_activity' | 0 | 0 | 0.038681 |



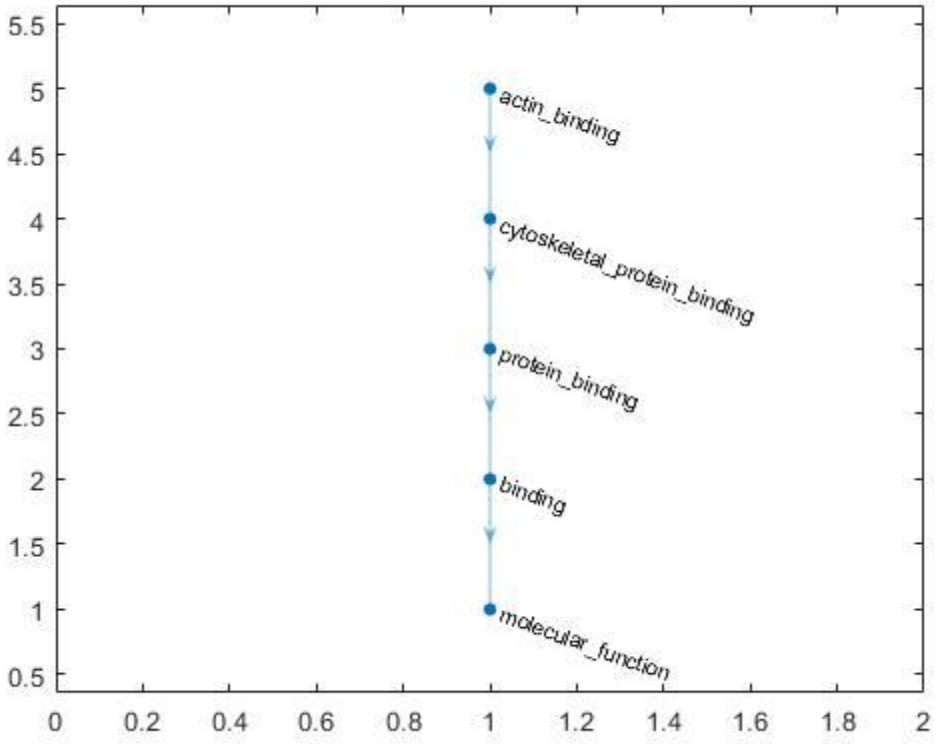

38. GO:0003777 = Microtubule motor activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.027778 | 0 | 0.16529 |
| 'catalytic_activity' | 0.027344 | 7 | 0.15656 |
| 'hydrolase_activity' | 0.026786 | 12 | 0.14626 |
| 'hydrolase_activity_acting_on_acid' | 0.026042 | 15 | 0.13411 |
| 'hydrolase_Activity_acting_on_acid_' | 0.025 | 16 | 0.11979 |
| 'pyrophosphatase_activity' | 0.023438 | 15 | 0.10292 |
| 'nucleoside_triphosphate_activity' | 0.020833 | 12 | 0.083073 |
| 'motor_activity' | 0.015625 | 7 | 0.059723 |
| 'microtubule_motor_Activity' | 0 | 0 | 0.032268 |



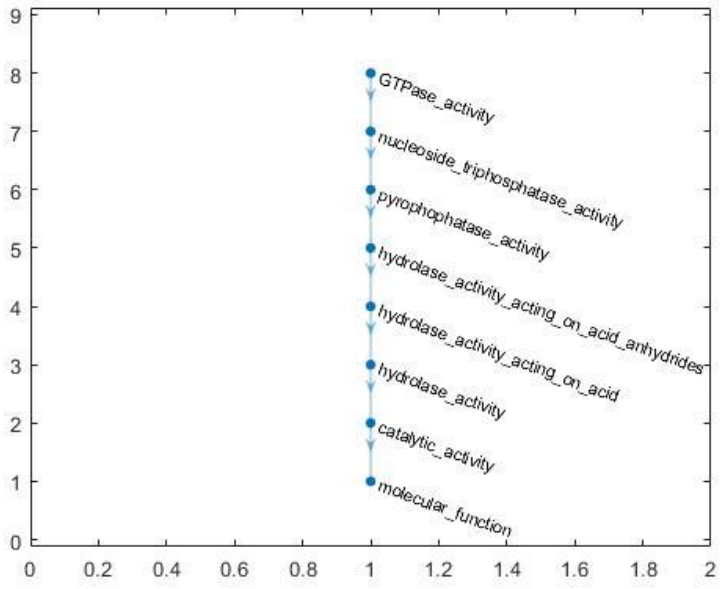
39. GO:0003779 = Actin Binding

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.1 | 0 | 0.30108 |
| 'binding' | 0.09375 | 3 | 0.25872 |
| 'protein_binding' | 0.083333 | 4 | 0.20885 |
| 'cytoskeletal_protein_binding' | 0.0625 | 3 | 0.15018 |
| 'actin_binding' | 0 | 0 | 0.081168 |



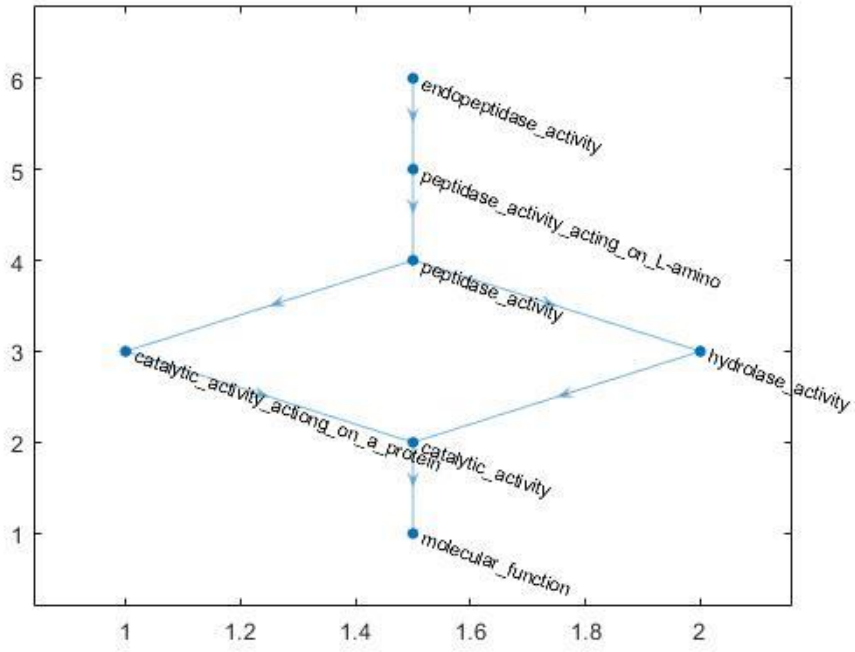

40. GO:0003924 = GTPaseactivity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.035714 | 0 | 0.18764 |
| 'catalytic_activity' | 0.034985 | 6 | 0.17526 |
| 'hydrolase_activity' | 0.034014 | 10 | 0.16066 |
| 'hydrolase_activity_acting_on_acid' | 0.032653 | 12 | 0.14348 |
| 'hydrolase_activity_acting_on_acid_anhydrides' | 0.030612 | 12 | 0.12325 |
| 'pyrophophatase_activity' | 0.027211 | 10 | 0.099478 |
| 'nucleoside_triphosphatase_activity' | 0.020408 | 6 | 0.071551 |
| 'GTPase_activity' | 0 | 0 | 0.038681 |



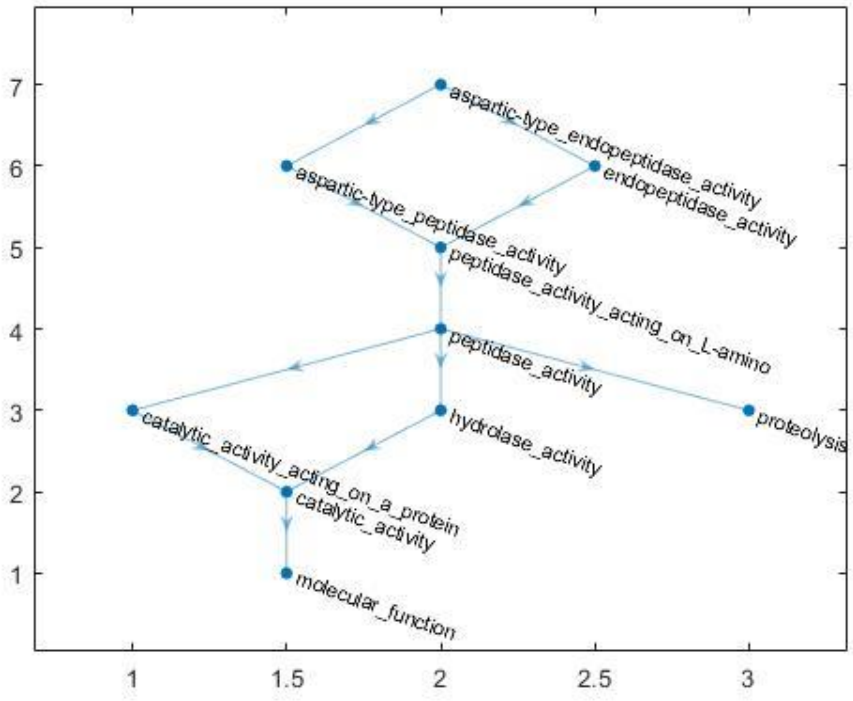

41. GO:0004175 = Endopeptidase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.058824 | 0 | 0.25603 |
| 'catalytic_activity' | 0.063131 | 5 | 0.23935 |
| 'catalytic_activity_actiong_on_a_protein' | 0.041667 | 3 | 0.10993 |
| 'hydrolase_activity' | 0.041667 | 3 | 0.10993 |
| 'peptidase_activity' | 0.037037 | 8 | 0.1351 |
| 'peptidase_activity_acting_on_L-amino' | 0.027778 | 5 | 0.097145 |
| 'endopeptidase_activity' | 0 | 0 | 0.052517 |



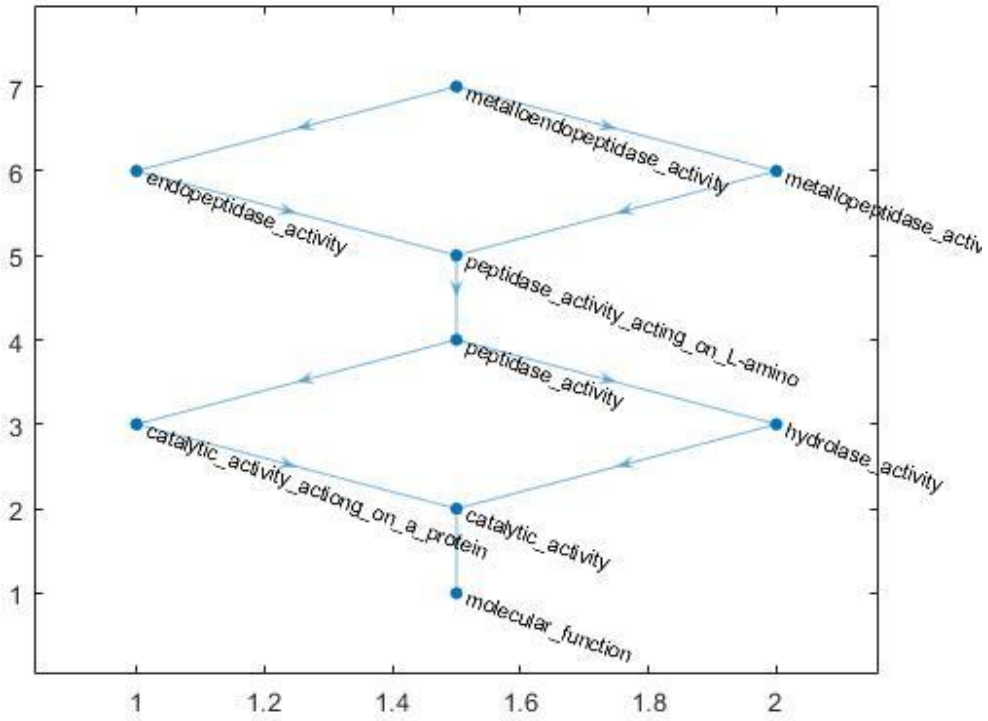
42. GO:0004190 = Aspartic-type endopeptidase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.028219 | 0 | 0.18141 |
| 'catalytic_activity' | 0.030247 | 7 | 0.16984 |
| 'catalytic_activity_acting_on_a_protein' | 0.023742 | 5 | 0.078126 |
| 'hydrolase_activity' | 0.023742 | 5 | 0.078126 |
| 'proteolysis' | 0.023742 | 0 | 0.078126 |
| 'peptidase_activity' | 0.024691 | 20 | 0.14488 |
| 'peptidase_activity_acting_on_L-amino' | 0.027778 | 18 | 0.12682 |
| 'aspartic-type_peptidase_activity' | 0.012346 | 3.5 | 0.052808 |
| 'endopeptidase_activity' | 0.012346 | 3.5 | 0.052808 |
| 'aspartic-type_endopeptidase_activity' | 0 | 0 | 0.037062 |



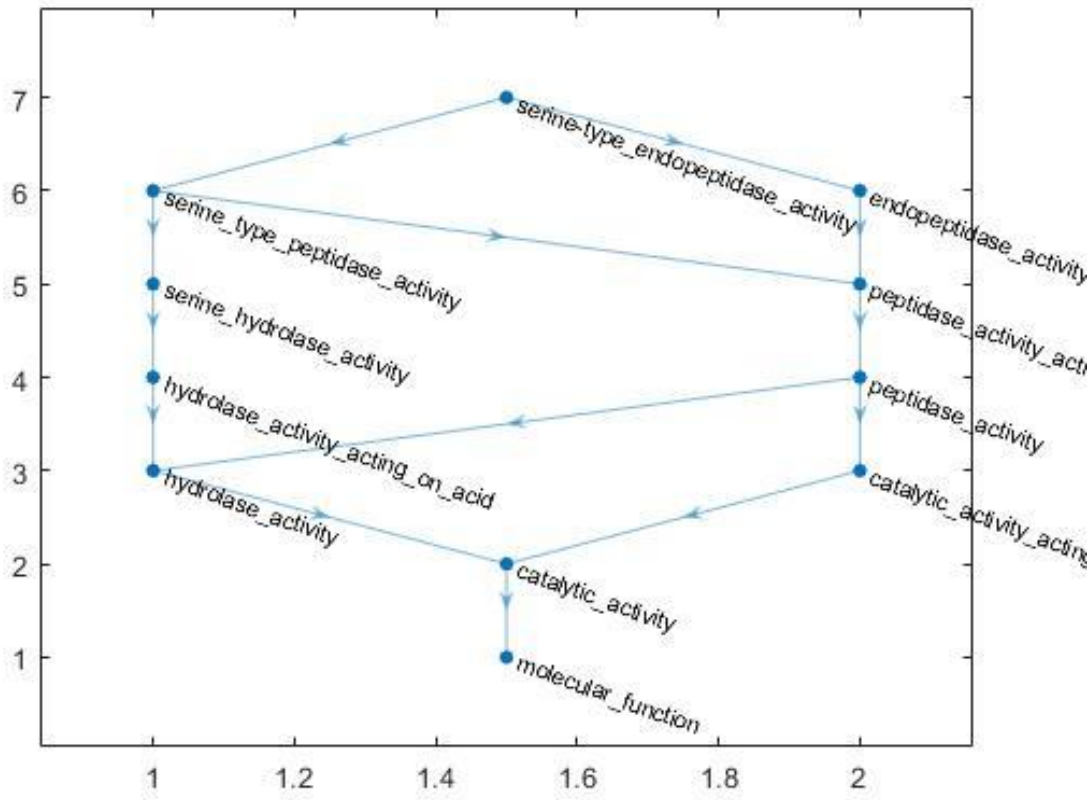

43. GO:0004222 = Metalloendopeptidase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.035714 | 0 | 0.2054 |
| 'catalytic_activity' | 0.038281 | 7 | 0.19926 |
| 'catalytic_activity_actiong_on_a_protein' | 0.030048 | 5 | 0.095991 |
| 'hydrolase_activity' | 0.030048 | 5 | 0.095991 |
| 'peptidase_activity' | 0.03125 | 16 | 0.14101 |
| 'peptidase_activity_acting_on_L-amino' | 0.035156 | 15 | 0.12347 |
| 'endopeptidase_activity' | 0.015625 | 3 | 0.051404 |
| 'metallopeptidase_activity' | 0.015625 | 3 | 0.051404 |
| 'metalloendopeptidase_activity' | 0 | 0 | 0.036068 |



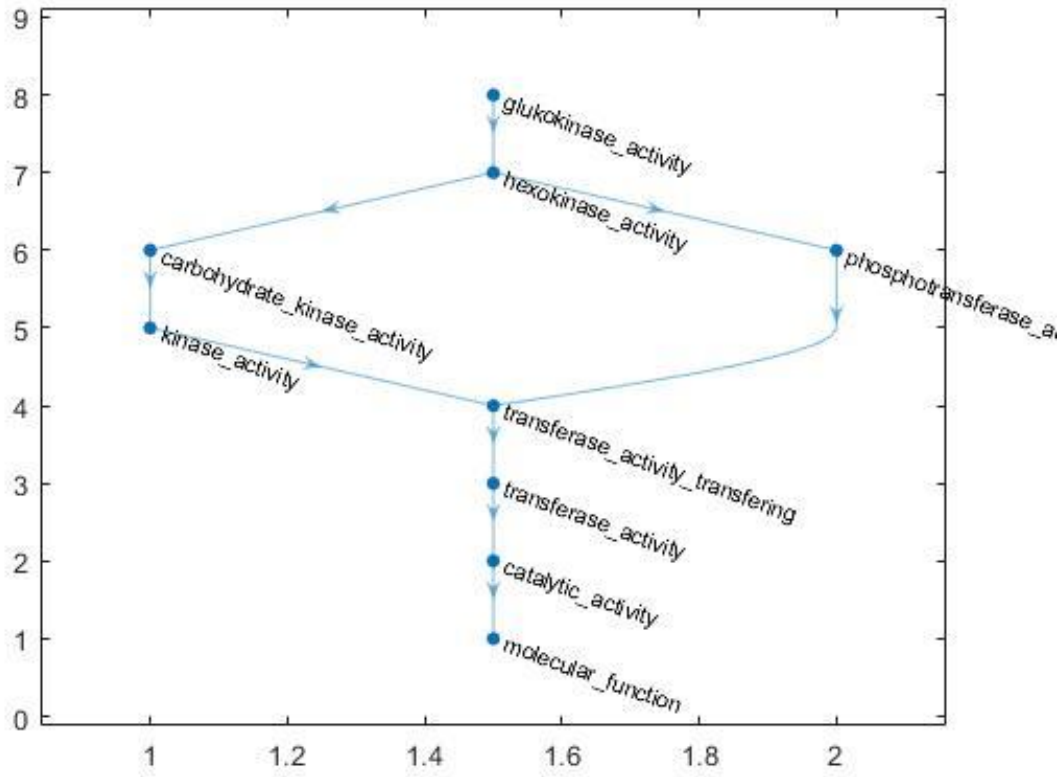

44. GO:0004252 = Serine-type endopeptidase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.028571 | 0 | 0.19745 |
| 'catalytic_activity' | 0.0324 | 9 | 0.19822 |
| 'hydrolase_activity' | 0.030625 | 9.5333 | 0.12858 |
| 'hydrolase_activity_acting_on_acid' | 0.015 | 4.9 | 0.068283 |
| 'catalytic_activity_acting_on_a_protein' | 0.019231 | 4.4667 | 0.070558 |
| 'serine_hydrolase_activity' | 0.013333 | 3.9 | 0.046369 |
| 'peptidase_activity' | 0.02 | 14.1 | 0.097996 |
| 'peptidase_activity_acting_on_L_amino' | 0.0225 | 13.1 | 0.081323 |
| 'serine_type_peptidase_activity' | 0.01 | 5.3667 | 0.04116 |
| 'endopeptidase_activity' | 0.01 | 2.6333 | 0.04116 |
| 'serine-type_endopeptidase_activity' | 0 | 0 | 0.028893 |



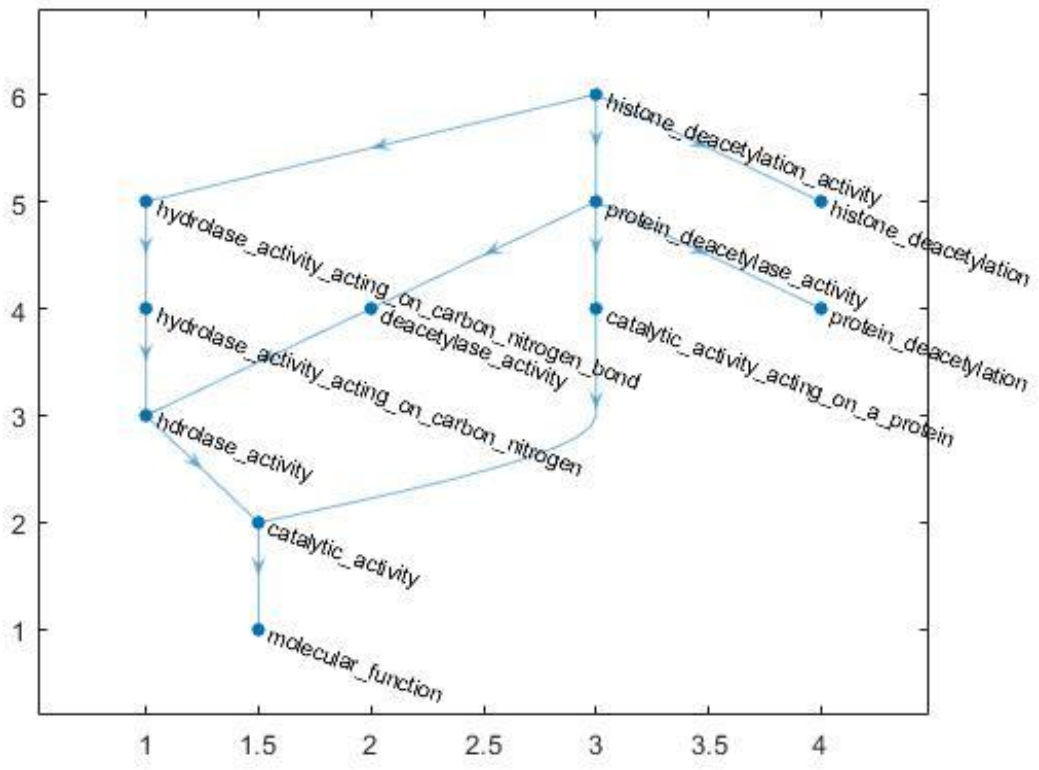

45. GO:0004340 = Glucokinase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.033333 | 0 | 0.18611 |
| 'catalytic_activity' | 0.034801 | 7 | 0.17859 |
| 'transferase_activity' | 0.0375 | 12 | 0.16977 |
| 'transferase_activity_transfering' | 0.043403 | 15 | 0.15945 |
| 'kinase_activity' | 0.023438 | 4 | 0.086201 |
| 'phosphotransferase_activity_alcohol_group' | 0.020833 | 8 | 0.061152 |
| 'carbohydrate_kinase_activity' | 0.020833 | 2 | 0.061152 |
| 'hexokinase_activity' | 0.015625 | 7 | 0.063342 |
| 'glukokinase_activity' | 0 | 0 | 0.034246 |



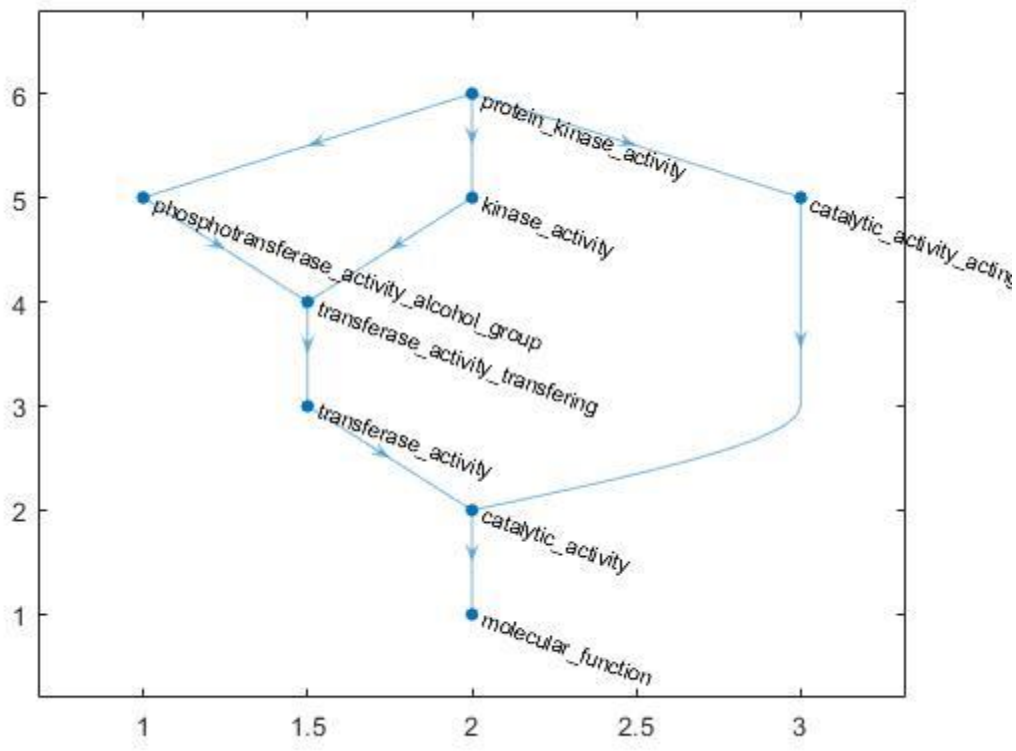

46. GO:0004407 = Histone deacetylase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.029091 | 0 | 0.21717 |
| 'catalytic_activity' | 0.035 | 7 | 0.21046 |
| 'hdrolase_activity' | 0.027778 | 6 | 0.1505 |
| 'hydrolase_activity_acting_on_carbon_nitrogen' | 0.013333 | 3.5 | 0.07995 |
| 'catalytic_activity_acting_on_a_protein' | 0.013333 | 4 | 0.052146 |
| 'deacetylase_activity' | 0.013333 | 1.5 | 0.052146 |
| 'hydrolase_activity_acting_on_carbon_nitrogen_bond' | 0.01 | 1.5 | 0.049078 |
| 'protein_deacetylation' | 0.013333 | 0 | 0.052146 |
| 'protein_deacetylase_activity' | 0.01 | 5.5 | 0.049078 |
| 'histone_deacetylation' | 0.01 | 0 | 0.049078 |
| 'histone_deacetylation_activity' | 0 | 0 | 0.038244 |



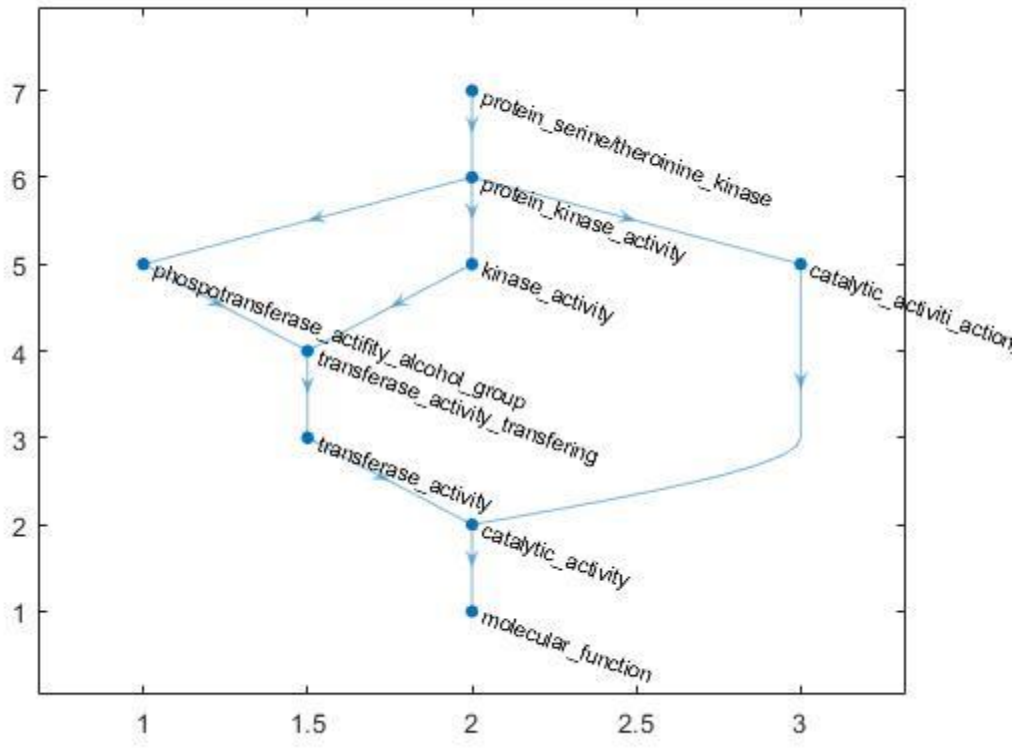

47. GO:0004672 = Protein kinase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.052632 | 0 | 0.24312 |
| 'catalytic_activity' | 0.061224 | 6 | 0.23363 |
| 'transferase_activity' | 0.040816 | 6 | 0.16518 |
| 'transferase_activity_transfering' | 0.045918 | 7 | 0.14186 |
| 'catalytic_activity_acting_on_a_protein' | 0.020408 | 2 | 0.057209 |
| 'phosphotransferase_activity_alcohol_group' | 0.020408 | 1 | 0.057209 |
| 'kinase_activity' | 0.020408 | 1 | 0.057209 |
| 'protein_kinase_activity' | 0 | 0 | 0.044576 |



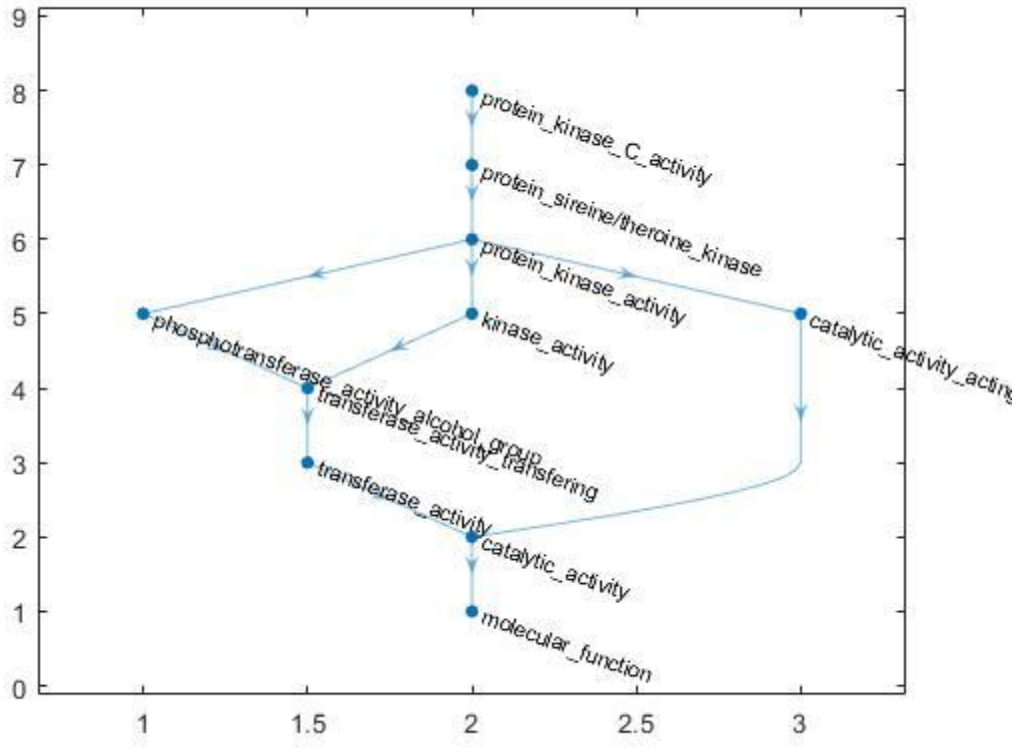

48. GO:0004674 = Protein serine/threonine kinase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.043478 | 0 | 0.2203 |
| 'catalytic_activity' | 0.051042 | 7 | 0.21512 |
| 'transferase_activity' | 0.032552 | 6 | 0.15185 |
| 'transferase_activity_transfering' | 0.035714 | 8 | 0.13459 |
| 'catalytic_activiti_action_on_a_protein' | 0.020833 | 4 | 0.057115 |
| 'phospotransferase_actifity_alcohol_group' | 0.020833 | 2 | 0.057115 |
| 'kinase_activity' | 0.020833 | 2 | 0.057115 |
| 'protein_kinase_activity' | 0.015625 | 7 | 0.069331 |
| 'protein_serine/theroinine_kinase' | 0 | 0 | 0.037472 |



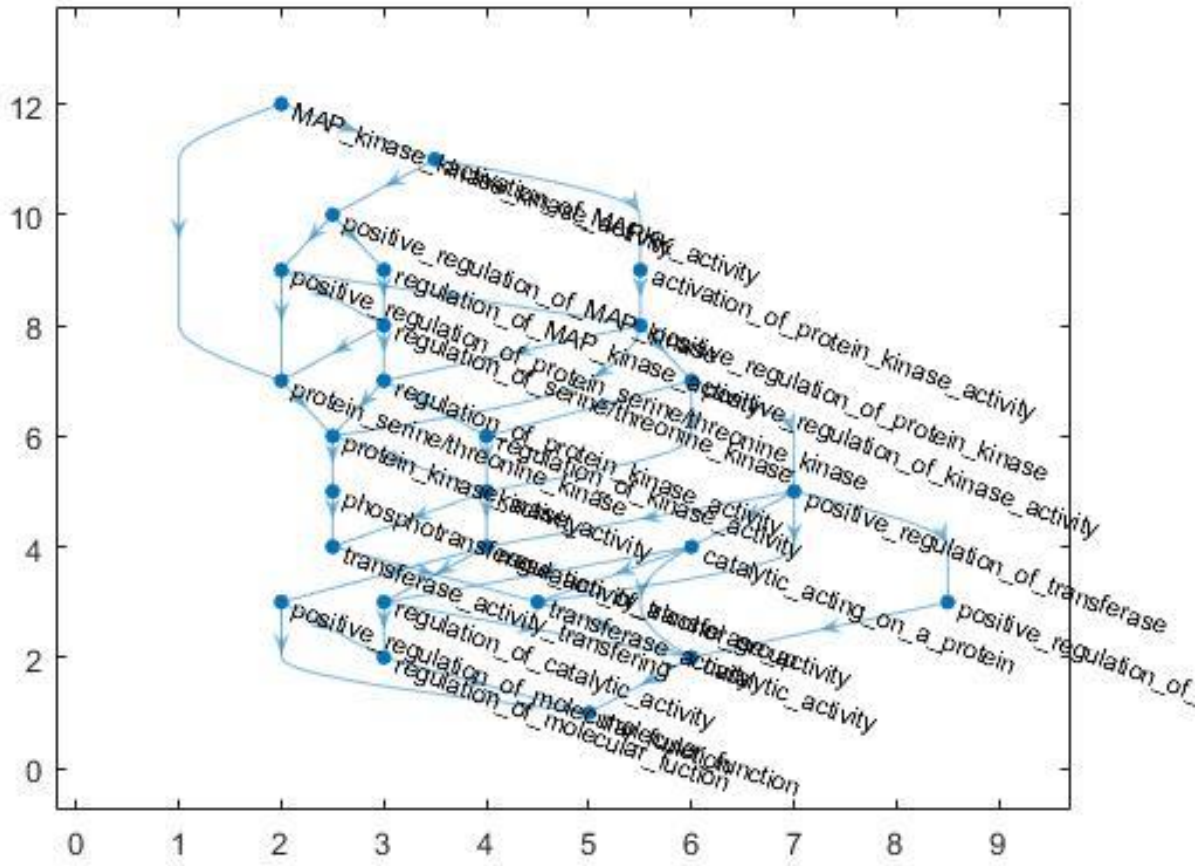

49. GO:0004697 = protein kinase C activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.035714 | 0 | 0.19937 |
| 'catalytic_activity' | 0.041585 | 8 | 0.19695 |
| 'transferase_activity' | 0.026144 | 6 | 0.13893 |
| 'transferase_activity_transfering' | 0.028058 | 9 | 0.12582 |
| 'catalytic_activity_acting_on_a_protein' | 0.018519 | 6 | 0.055232 |
| 'phosphotransferase_activity_alcohol_group' | 0.018519 | 3 | 0.055232 |
| 'kinase_activity' | 0.018519 | 3 | 0.055232 |
| 'protein_kinase_activity' | 0.016461 | 14 | 0.082185 |
| 'protein_sireine/theroine_kinase' | 0.012346 | 8 | 0.059096 |
| 'protein_kinase_C_activity' | 0 | 0 | 0.03195 |




50. GO:0004709 = MAP kinase kinase kinase activity

| Name | incloseness | betweenness | pagerank |
|---|---|---|---|
| 'molecular_function' | 0.010204 | 0 | 0.14188 |
| 'catalytic_activity' | 0.0094439 | 9.2071 | 0.1045 |
| 'regulation_of_molecular_fuction' | 0.0068762 | 0.5 | 0.036016 |
| 'transferase_activity' | 0.00864 | 12.419 | 0.076262 |
| 'positive_regulation_of_molecular_function' | 0.0064 | 19.701 | 0.028168 |
| 'regulation_of_catalytic_activity' | 0.0071138 | 13.742 | 0.03206 |
| 'transferase_activity_transfering' | 0.0075 | 15.85 | 0.068915 |
| 'positive_regulation_of_catalytic' | 0.0034133 | 3.569 | 0.013745 |
| 'regulation_of_transferase_activity' | 0.0075 | 56.275 | 0.041772 |
| 'catalytic_acting_on_a_protein' | 0.0034133 | 6.7365 | 0.013745 |
| 'phosphotransferase_activity_alcohol_group' | 0.0062452 | 8.7667 | 0.035854 |
| 'positive_regulation_of_transferase' | 0.0035636 | 39.462 | 0.015721 |
| 'kinase_activity' | 0.0081939 | 69.771 | 0.065935 |
| 'protein_kinase_activity' | 0.008 | 60.434 | 0.059897 |
| 'regulation_of_kinase_activity' | 0.0057143 | 9.3667 | 0.029144 |
| 'positive_regulation_of_kinase_activity' | 0.00384 | 50.783 | 0.01877 |
| 'protein_serine/threonine_kinase' | 0.00576 | 24.947 | 0.032615 |
| 'regulation_of_protein_kinase_activity' | 0.0056889 | 19.067 | 0.031605 |
| 'positive_regulation_of_protein_kinase' | 0.0044444 | 66.403 | 0.029545 |
| 'regulation_of_serine/threonine_kinase' | 0.0036364 | 16 | 0.030223 |
| 'positive_regulation_of_protein_serine/threonine_kinase' | 0.0024 | 20.917 | 0.017498 |
| 'regulation_of_MAP_kinase_activity' | 0.0024 | 2.0833 | 0.017498 |
| 'activation_of_protein_kinase_activity' | 0.0021333 | 25 | 0.016702 |
| 'positive_regulation_of_MAP_kinase' | 0.0021333 | 7 | 0.016702 |
| 'activation_of_MAPKK_activity' | 0.0016 | 12 | 0.014826 |
| 'MAP_kinase_kinase_kinase_activity' | 0 | 0 | 0.010406 |




Sengaja penulis hanya menyertakan lampiran dari kode GO dan hasil metode centrality hanya 50 kode saja. Total keseluruhan kode GO yang digunakan sebanyak 221 kode, sehingga untuk mengirit biaya dan kertas hanya disajikan 50 kode pada buku tesis. Untuk lebih lengkapnya dari keseluruhan kode dan hasil dari metode centrality dilampirkan pada file pada folder $DAG\_MF\_Skor$ dan folder $DAG\_Molecular\_function$.

**LAMPIRAN 3**
**DATASET**

A. Dataset *Betweenness Centrality*

| b | cra | cc | mfr | mta | nromf | promf | romf | sma | tra | ta | Kelas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 0 | 18.5667 | 6.1667 | 12.6 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 24.6667 | 0 | 0 | 6.1667 | 6.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 49.3333 | 0 | 18.5667 | 6.1677 | 12.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 49.6667 | 0 | 18.5473 | 6.1667 | 12.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 16.33337 | 0 | 7.6666 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 19.6667 | 0 | 0 | 17.1667 | 6.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 40 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | Insulin |
| 13 | 0 | 8.8333 | 0 | 4.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 7 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 48.1667 | 0 | 18.1667 | 6.1667 | 12.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 38 | 0 | 15.3333 | 3.3333 | 6.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 49 | 0 | 18.5473 | 6.1667 | 12.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 46.3333 | 0 | 18.3333 | 6.1667 | 12.6667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | Insulin |
| 43 | 0 | 8.3333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 40 | 0 | 3.3333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 43 | 0 | 6.6667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 37 | 0 | 12.667 | 0 | 6.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 15.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 4.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 18 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 30 | 0 | 16.7071 | 1.5 | 0 | 0 | 21.201 | 1 | 0 | 0 | 0 | Insulin |
| 21 | 0 | 3.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 6.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 25 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 4 | 0 | Insulin |
| 19 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 4.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 23 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | Insulin |
| 3 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 25 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 9.5 | 0 | 3.5 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 35 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 4.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 28 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.3334 | 0 | 0 | 12.3334 | 4.3334 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 13 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 10 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | Insulin |
| 10 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.3334 | 0 | 0 | 12.3334 | 4.3334 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.3334 | 0 | 0 | 12.3334 | 4.3334 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 7.5 | 2.5 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.3334 | 0 | 0 | 12.3334 | 4.3334 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 4.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5.3334 | 0 | 0 | 12.3334 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | Insulin |
| 16 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | Insulin |
| 20 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 4.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | Insulin |
| 21 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 8.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 1 | 0 | 0 | Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 25 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 19 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 5 | 0 | 0 | 7.5 | 2.5 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | Insulin |
| 21 | 0 | 6.5 | 1.5 | 0 | 1.5 | 0 | 0.5 | 0 | 0 | 0 | Insulin |
| 6 | 0 | 13.167 | 3.5 | 0 | 0 | 2.75 | 1.1667 | 0 | 0 | 0 | Insulin |
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 18 | 0 | 15.1667 | 0 | 2.8333 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 16 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 18 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 23 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | Insulin |
| 16 | 0 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 20.6667 | 0 | 0 | 6.1667 | 2.1667 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 34 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 44 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 11 | 0 | Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 3 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 | Insulin |
| 26 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |
| 12 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 6.6667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 0 | 8.1667 | 6.6667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 8.1667 | 6.6667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 10 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 20 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 12 | 0 | 8.1667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

Dataset yang dibangun dengan metode *Betweenness Centrality* berada pada file *Comma Separated Values (*.csv*)* dengan nama file *dataset_bc_insulin*. *csv* yang terlampir pada folder lampiran.

B. Dataset *Closeness Centrality*

| b | cra | cc | mfr | mta | nromf | promf | romf | sma | tra | ta | Kelas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.817346 | 0 | 0.108185 | 0 | 0.028638 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.455543 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.379946 | 0 | 0.089289 | 0 | 0.020135 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.455543 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.323143 | 0 | 0.038247 | 0 | 0.020135 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.841324 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.328672 | 0 | 0 | 0.041446 | 0.054436 | 0 | 0 | 0.043548 | 0 | 0 | 0 | insulin |
| 0.772972 | 0 | 0.147708 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.655606 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.39815 | 0 | insulin |
| 0.6022 | 0 | 0.018896 | 0 | 0.258503 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.194446 | 0 | 0.063131 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.106973 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.640808 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.313672 | 0 | 0 | 0.026446 | 0.034436 | 0 | 0 | 0.027548 | 0 | 0 | 0 | insulin |
| 0.317951 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.062329 | insulin |
| 0.3699 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.317951 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.46365 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.27615 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.27615 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.4243 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.27615 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.70887 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.090843 | 0 | 0.04913 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.063131 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.118272 | 0 | 0.034801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.156881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.157618 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.174993 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.156881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.354389 | 0 | 0.049385 | 0.008876 | 0 | 0 | 0.015276 | 0.027586 | 0 | 0 | 0 | insulin |
| 0.252153 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.077223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.169801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.152966 | 0 | 0.047852 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.120593 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.10052 | 0 | 0 | 0 | 0.14815 | 0 | 0 | 0 | 0 | 0.08 | 0 | insulin |
| 0.01962 | 0 | 0.034985 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.2419 | 0 | 0.0324 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.235405 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.061224 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.157618 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.08362 | 0 | 0.034985 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.061224 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.118272 | 0 | 0.034801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.109427 | 0 | 0 | 0 | 0.14815 | 0 | 0 | 0 | 0 | 0.25 | 0 | insulin |
| 0.09375 | 0 | 0.038281 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.137139 | 0 | 0.041585 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.10667 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.051042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.016296 | 0 | 0.007619 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.172136 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.306105 | 0 | 0.034985 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.157618 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.034985 | insulin |
| 0.2515 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.031154 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.267585 | 0 | 0.051042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.347422 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.268046 | 0 | 0.047852 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026446 | 0 | 0 | 0.026446 | 0.034436 | 0 | 0 | 0.027548 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.203177 | 0 | 0 | 0 | 0.14815 | 0 | 0 | 0 | 0 | 0.125399 | 0 | insulin |
| 0.10667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.25247 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.063131 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.10667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.137725 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.301552 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.268743 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.090843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.10667 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.293286 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.113566 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0 | insulin |
| 0.741377 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.034985 | insulin |
| 0.109427 | 0 | 0 | 0 | 0.14815 | 0 | 0 | 0 | 0 | 0.25 | 0 | insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.39005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026446 | 0 | 0 | 0.026446 | 0.034436 | 0 | 0 | 0.027548 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.10667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.118272 | 0 | 0.034801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.064 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.047852 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.210884 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.058894 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026446 | 0 | 0 | 0.026446 | 0.034436 | 0 | 0 | 0.027548 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.090843 | 0 | 0.051042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.009862 | 0 | 0 | 0.009862 | 0.01253 | 0 | 0 | 0.010144 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.156881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.276843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.026843 | 0 | 0.061224 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.728781 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.118272 | 0 | 0.034801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.106973 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.061224 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.106973 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0.047852 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.2515 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.82698 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.091429 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.2419 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.674931 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026446 | 0 | 0 | 0.026446 | 0.034436 | 0 | 0 | 0.027548 | 0 | 0 | insulin |
| 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.157618 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.169801 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026446 | 0 | 0 | 0.026446 | 0.017218 | 0 | 0 | 0.027548 | 0 | 0 | insulin |
| 0.156881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.14116 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.034985 | insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| 0.156881 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | insulin |
| . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | .. | . | . | . | . | . |
| 0.357851 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.20042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.44994 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.040989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.081633 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.120593 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.254993 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 1.024417 | 0 | 0 | 0.015 | 0.02 | 0 | 0 | 0.016 | 0 | 0 | 0 | non_insulin |
| 0.168504 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.058894 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0.015603 | 0.005154 | 0.25 | 0 | 0.005154 | 0.011613 | 0 | 0 | 0 | non_insulin |
| 0.166889 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.040989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.5559 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.07716 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.75993 | 0 | 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.34375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.28119 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0.015603 | 0.005154 | 0 | 0 | 0.005154 | 0.011613 | 0 | 0 | 0 | non_insulin |
| 0.753653 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.291281 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.073139 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.15775 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.120593 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.090843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.15716 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.66565 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |
| 0.2963 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.009795 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.54215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.7206 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.115609 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.090843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.132031 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.254993 | 0 | 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.07716 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.063131 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.013223 | 0 | 0 | 0.013223 | 0.017218 | 0 | 0 | 0.013774 | 0 | 0 | non_insulin |
| 1.379946 | 0 | 0.089289 | 0 | 0.020135 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.47622 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.17375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.128 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.092838 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.174993 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.064 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.15775 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.10667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2963 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2675 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.70911 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.17375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.063802 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2006 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.530096 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.376296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.806639 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.276843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.120593 | 0 | 0.072828 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.29215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.39815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.323143 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 1.235903 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |
| 0.397869 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.17375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.120593 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.254993 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 1.024417 | 0 | 0 | 0.015 | 0.02 | 0 | 0 | 0.016 | 0 | 0 | non_insulin |
| 0.168504 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.058894 | 0 | 0 | 0 | 0.25 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0.015603 | 0.005154 | 0.25 | 0 | 0.005154 | 0.011613 | 0 | 0 | non_insulin |
| 0.166889 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.040989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.5559 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.07716 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.75993 | 0 | 0.046296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.34375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.28119 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0.015603 | 0.005154 | 0 | 0 | 0.005154 | 0.011613 | 0 | 0 | non_insulin |
| 0.753653 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.10667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2963 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2675 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.70911 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.17375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.063802 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.2006 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.530096 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.09375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.376296 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.806639 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.276843 | 0 | 0.021786 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.14815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.026843 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.015379 | non_insulin |
| 0.120593 | 0 | 0.072828 | 0 | 0.009468 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.29215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |
| 0.39815 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | non_insulin |

Dataset yang dibangun dengan metode *Closeness Centrality* berada pada file *Comma Separated Values (*.csv*)* dengan nama file *dataset_cc_insulin. csv* yang terlampir pada folder lampiran.

C. Dataset *PageRank Centrality*

| b | cra | cc | mfr | mta | nromf | promf | romf | sma | tra | ta | Kelas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.1412 | 0 | 0.62914 | 0 | 0.21746 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.195474 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 3.43972 | 0.35091 | 0.49093 | 0 | 0.151898 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.195474 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.80982 | 0 | 0.27581 | 0 | 0.151898 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.9547 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.110918 | 0 | 0 | 0.172499 | 0.3065 | 0 | 0 | 0.173387 | 0 | 0 | 0 | Insulin |
| 2.08637 | 0 | 0.6364 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.33987 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.63997 | 0 | Insulin |
| 1.38028 | 0 | 0.13821 | 0 | 0.406732 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.49659 | 0 | 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.358344 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 1.94055 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.005448 | 0 | 0 | 0.110874 | 0.19863 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0.81776 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.33182 | Insulin |
| 1.00784 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.81776 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.26656 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.74912 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.74912 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.04792 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.74912 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.65989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.43738 | 0 | 0.30404 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.48332 | 0 | 0.17859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.49807 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.369169 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.51102 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.49807 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.18617 | 0 | 0.27962 | 0.038613 | 0 | 0 | 0.066781 | 0.117568 | 0 | 0 | 0 | Insulin |
| 0.7621 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.324784 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.51896 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.4774 | 0 | 0.2074 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.47094 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.52601 | 0 | 0 | 0 | 0.2988 | 0 | 0 | 0 | 0 | 0.25215 | 0 | Insulin |
| 0.19556 | 0 | 0.17526 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.55752 | 0 | 0.1746 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.71359 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.23363 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.369169 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.42072 | 0 | 0.17526 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.23363 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.48332 | 0 | 0.17859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.43714 | 0 | 0 | 0 | 0.2988 | 0 | 0 | 0 | 0 | 0.34117 | 0 | Insulin |
| 0.25872 | 0 | 0.19926 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.63517 | 0 | 0.19695 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.27961 | 0 | 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.21512 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.12921 | 0 | 0.060997 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.85053 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.11342 | 0 | 0.17526 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.369169 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17526 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.7426 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.18332 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.9066 | 0 | 0.21512 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.20709 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.90683 | 0 | 0.2074 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.199248 | 0 | 0 | 0.110874 | 0.19863 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.69586 | 0 | 0 | 0 | 0.2988 | 0 | 0 | 0 | 0 | 0.43452 | 0 | Insulin |
| 0.27961 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 1.23237 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.27961 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.46889 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.99204 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25215 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.76974 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.43738 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.27961 | 0 | 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 1.23237 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.41726 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25215 | 0 | Insulin |
| 1.85098 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17526 | Insulin |
| 0.43714 | 0 | 0 | 0 | 0.2988 | 0 | 0 | 0 | 0 | 0.34117 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.85632 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.19779 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.199248 | 0 | 0 | 0.110874 | 0.19863 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.27961 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.48332 | 0 | 0.17859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.22516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.2074 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.81611 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25136 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.199248 | 0 | 0 | 0.110874 | 0.19863 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.43738 | 0 | 0.21512 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.087416 | 0 | 0 | 0.049004 | 0.08749 | 0 | 0 | 0.049041 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.49807 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.55339 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.23363 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.64891 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.25215 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.48332 | 0 | 0.17859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.358344 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.23363 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.358344 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.2074 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.7426 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.76004 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.55752 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 1.80471 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.199248 | 0 | 0 | 0.110874 | 0.19863 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.369169 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.51896 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.199248 | 0 | 0 | 0.110874 | 0.099315 | 0 | 0 | 0.110566 | 0 | 0 | 0 | Insulin |
| 0.49807 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.47624 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | 0 | Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17526 | Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| . | . | . | . | . | . | . | . | . | . | . . |
| . | . | . | . | . | . | . | . | . | . | . . |
| . | . | . | . | . | . | . | . | . | . | . . |
| 0.43738 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.50323 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.40956 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |
| 0.5976 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.072011 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.11728 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.64552 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.19779 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.46673 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.43738 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.44732 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.76317 | 0 | 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25108 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | Non-Insulin |
| 3.43972 | 0.35091 | 0.49093 | 0 | 0.151898 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.31183 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51087 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.45032 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.38487 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51102 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.22516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.48388 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.27961 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.5976 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.76959 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.66492 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51087 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25601 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.72525 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.31283 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.79111 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2.39389 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.55339 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |
| 0.47094 | 0 | 0.3626 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.77611 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.63997 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.80982 | 0 | 0.34117 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 3.03757 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |
| 1.18174 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51087 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.18024 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.81029 | 0 | Non-Insulin |
| 0.52325 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.52982 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.76917 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.2988 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1384 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0.23245 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.69922 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.47305 | 0.28808 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0.25108 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | Non-Insulin |
| 0.26787 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.2351 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1384 | Non-Insulin |
| 0.4516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.17526 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.6543 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.55752 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1384 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.16884 | Non-Insulin |
| 0.22516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.78268 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0.19822 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0.92259 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.22516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.44118 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.01435 | 0 | Non-Insulin |
| 1.26914 | 0 | 0 | 0 | 0.57362 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.64849 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0.43757 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.76974 | 0 | 0.67079 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.19779 | 0 | 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.56142 | 0 | 0.67796 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.77101 | 0 | 0.55325 | 0 | 0 | 0 | 0.028168 | 0.036016 | 0 | 0 | Non-Insulin |
| 0.76917 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1384 | Non-Insulin |
| 1.27308 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.61376 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51102 | 0 | 0.14748 | 0 | 0.070369 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.86442 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.51087 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.97634 | 0 | 0.41207 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.48388 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.31513 | Non-Insulin |
| 1.07739 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.55095 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0.23245 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.713 | 0 | 0.12453 | 0.021431 | 0 | 0 | 0.020408 | 0.048283 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.51684 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 2.00445 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.89212 | 0 | Non-Insulin |
| 0.76974 | 0 | 0.25872 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.77616 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.14629 | Non-Insulin |
| 0.81624 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.41001 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0.23245 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0 | 0 | 0 | 0 | 0.23245 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.91427 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.5699 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17032 | 0 | Non-Insulin |
| 0.22516 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.55534 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.55752 | 0 | 0.12453 | 0.021431 | 0 | 0 | 0.020408 | 0.048283 | 0 | 0 | Non-Insulin |
| 1.28549 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.25215 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.22516 | 0 | 0.1591 | 0.041118 | 0 | 0.041118 | 0 | 0.086797 | 0 | 0 | 0 | Non-Insulin |
| 0.087416 | 0 | 0 | 0.049004 | 0.08749 | 0 | 0 | 0.049041 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.02147 | 0 | 0.6457 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.23935 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.82276 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.266 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.27961 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.38065 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.89212 | 0 | Non-Insulin |
| 0.8342 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.17526 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.55095 | 0 | Non-Insulin |
| 1.27433 | 0 | 0.47384 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 1.12385 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.099624 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Non-Insulin |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.59989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.49659 | 0 | 0.19926 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.21222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2988 | 0 | 0.12453 | 0.021431 | 0 | 0 | 0.020408 | 0.048283 | 0 | 0 | 0 | Non-Insulin |
| 0.75175 | 0 | 0.21512 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.48332 | 0 | 0.17859 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.93571 | 0 | 0.15656 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.657144 | 0 | 0 | 0.055437 | 0.099315 | 0 | 0 | 0.055283 | 0 | 0 | 0 | Non-Insulin |
| 0.19779 | 0 | 0.63683 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Non-Insulin |
| 0.2711 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.17526 | Non-Insulin |

Dataset yang dibangun dengan metode *PageRank Centrality* berada pada file *Comma Separated Values (*.csv*)* dengan nama file *dataset_pr_insulin*. *csv* yang terlampir pada folder lampiran.

**LAMPIRAN 4**
**CODE PROGRAMMING PYTHON**

- **Data_Processing.py**

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.feature_selection import SelectFromModel
from xgboost import XGBClassifier
from numpy import sort
from sklearn.metrics import accuracy_score
import math


def read_data(filename):
    data = pd.read_csv(filename,header=0)
    x = data.iloc[:, :-1]
    y = data.iloc[:, -1]
    return x, y


def normalization(x, method="standardization"):
    if method == "standardization":
        scaler = StandardScaler()
    else:
        scaler = MinMaxScaler()
    x_norm = scaler.fit_transform(x)
    x_norm_df = pd.DataFrame(data=x_norm,columns=x.columns.values)
    return x_norm_df, scaler


def transformation(scaler, x):
    return pd.DataFrame(scaler.transform(x), columns=x.columns.values)


def rmse_fun(y_true, y_pred):
    return np.sqrt(np.mean((y_true - y_pred) ** 2))


def mape_fun(y_true, y_pred):
    return np.mean(np.abs(1.0*(y_pred - y_true)/y_true))


def r2_fun(y_true, y_pred):
    return 1 - np.mean((y_true - y_pred) ** 2)/np.mean(1.0*(y_true - np.mean(y_true))**2)


def overall_score(y_true, y_pred):
    rmse = rmse_fun(y_true, y_pred)
    mape = mape_fun(y_true, y_pred)
    r2 = r2_fun(y_true, y_pred)
    return [rmse, mape, r2]

def feature_selection(model, X_train, y_train, X_test, y_test):
    thresholds = sort(model.feature_importances_)
    for thresh in thresholds:
```

```
    # select features using threshold
    selection = SelectFromModel(model, threshold=thresh, prefit=True)
    select_X_train = selection.transform(X_train)
    # train model
    selection_model = XGBClassifier()
    selection_model.fit(select_X_train, y_train)
    # eval model
    select_X_test = selection.transform(X_test)
    predictions = selection_model.predict(select_X_test)
    akurasi = accuracy_score(y_test, predictions)
    print("Thresh=%.3f, n=%d, akurasi: %.2f%%" % (thresh, select_X_train.shape[1],
                                  akurasi*100.0))

def pre_rec(y_true,y_pred):
  TP=0
  TN=0
  FP=0
  FN=0
  for i in range(len(y_true)):
    if (y_true[i]==0) & (y_pred[i]==0):
      TP=TP+1
    if (y_true[i]==0) & (y_pred[i]==1):
      FP=FP+1
    if (y_true[i]==1) & (y_pred[i]==0):
      FN=FN+1
    if (y_true[i]==1) & (y_pred[i]==1):
      TN=TN+1
  presisi=TP/(TP+FP)
  recall=TP/(TP+FN)
  return [presisi,recall]

def log_odds (y_prob):
  y_prob = y_prob[0:1295,0]
  odds = []
  for i in range(len(y_prob)):
    val = math.log(y_prob[i]/(1-y_prob[i]))
    odds.extend(val)
  return odds
```

- **Xgboost_main_tesis.ipynb**

```
from xgboost import XGBClassifier, plot_tree, plot_importance
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import Data_Processing as dp
from matplotlib import pyplot
import matplotlib.pyplot as plt
from sklearn.externals import joblib
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import LabelEncoder
import graphviz

file='D:\Tesis\Komputasi_XGBoost\data_bc_insulin.csv'
X, y=dp.read_data(file)
X.head()
```

| | b | cra | cc | mfr | mta | nromf | promf | romf | sma | tra | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 49.0000 | 0 | 18.56670 | 6.1667 | 12.6000 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **1** | 24.6667 | 0 | 0.00000 | 6.1667 | 6.1667 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **2** | 49.3333 | 0 | 18.56670 | 6.1677 | 12.1667 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **3** | 49.6667 | 0 | 18.54730 | 6.1667 | 12.1667 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **4** | 20.0000 | 0 | 16.33337 | 0.0000 | 7.6666 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |

```
y.head()
```

```
0    Insulin
1    Insulin
2    Insulin
3    Insulin
4    Insulin
Name: Kelas, dtype: object
```

```
label_encoder=LabelEncoder()
label_encoder=label_encoder.fit(y)
encoder_y=label_encoder.transform(y)
encoder_y
```

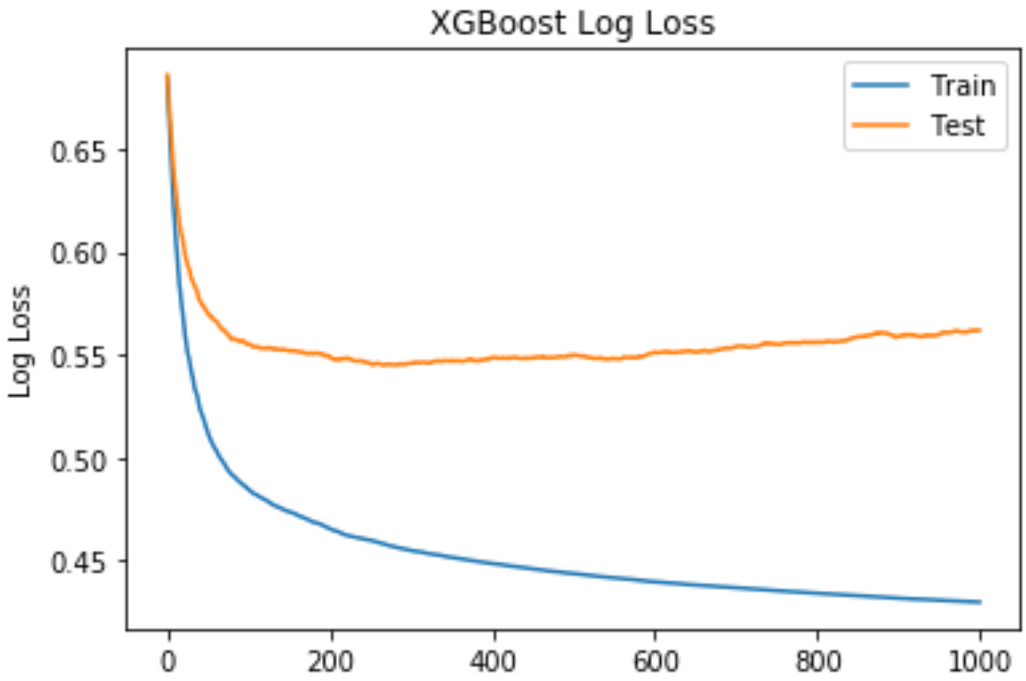
```
array([0, 0, 0, ..., 1, 1, 1])
```

```
seed =1
test_size = 0.2
X_train, X_test, y_train, y_test = train_test_split(X, encoder_y, test_size=test_size, random_state=seed)
X_train.head()
```

| | b | cra | cc | mfr | mta | nromf | promf | romf | sma | tra | ta |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **902** | 21.0000 | 0 | 0.0 | 0.0000 | 0.0000 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **679** | 9.0000 | 0 | 7.0 | 0.0000 | 0.0000 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **768** | 7.0000 | 0 | 0.0 | 0.0000 | 0.0000 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **385** | 5.6667 | 0 | 0.0 | 6.1667 | 2.1667 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |
| **1283** | 3.0000 | 0 | 0.0 | 0.0000 | 0.0000 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 |

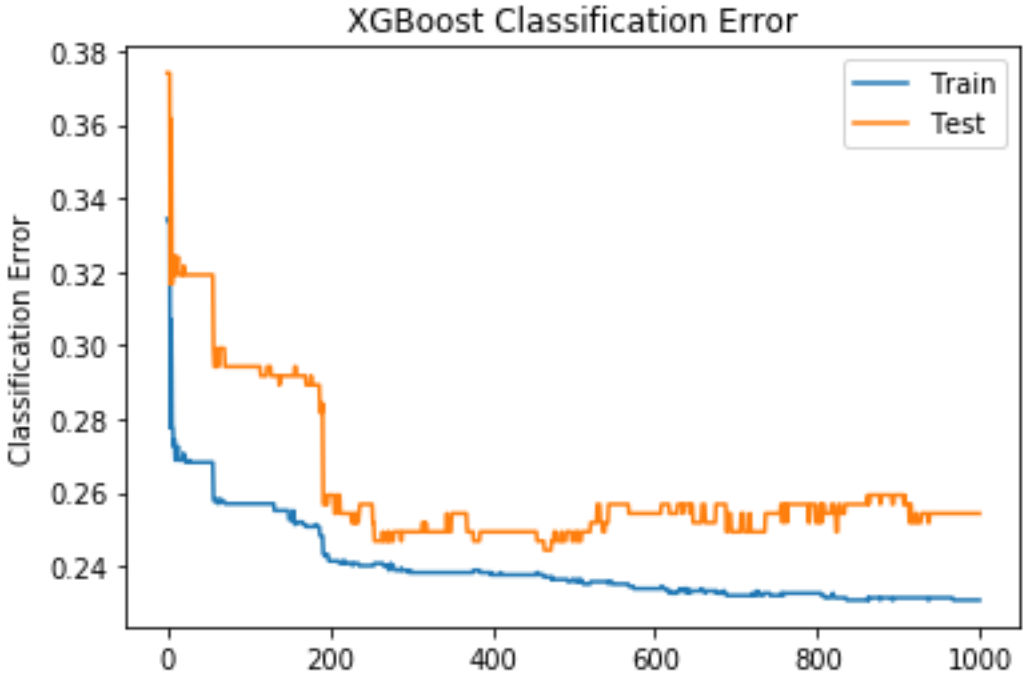
```
model = XGBClassifier(learning_rate=0.05, n_jobs=4,
            n_estimators=1000, max_depth=4,
            subsample=0.8, colsample_bytree=0.8,
            gamma=0, random_state=1)

eval_set = [(X_train, y_train), (X_test, y_test)]
model.fit(X_train, y_train, eval_metric=["error", "logloss"], eval_set=eval_set,
      verbose=True)

kfold = KFold(n_splits=10, random_state=7)
results = cross_val_score(model, X_train, y_train, cv=kfold)
print("Train_Accuracy: %.2f%% (%.2f%%)" % (results.mean()*100, results.std()*100))

y_pred = model.predict(X_test)
y_val = model.predict(X_train)
y_prob = model.predict_proba(X, validate_features=True)
predictions = [round(value) for value in y_pred]
```

```
accuracy = accuracy_score(y_test, predictions)
print("Test_Accuracy: %.2f%%" % (accuracy * 100.0))

TP=0
TN=0
FP=0
FN=0
for i in range(len(y_train)):
   if (y_train[i]==0) & (y_val[i]==0):
      TP=TP+1
   if (y_train[i]==0) & (y_val[i]==1):
      FP=FP+1
   if (y_train[i]==1) & (y_val[i]==0):
      FN=FN+1
   if (y_train[i]==1) & (y_val[i]==1):
      TN=TN+1
presisi=TP/(TP+FP)
recall=TP/(TP+FN)
print("Precision = %.2f%%" % (presisi*100.0))
print("Recall = %.2f%%" % (recall*100.0))

results = model.evals_result()
epochs = len(results['validation_0']['error'])
x_axis = range(0, epochs)

fig, ax = pyplot.subplots()
ax.plot(x_axis, results['validation_0']['logloss'], label='Train')
ax.plot(x_axis, results['validation_1']['logloss'], label='Test')
ax.legend()
pyplot.ylabel('Log Loss')
pyplot.title('XGBoost Log Loss')
pyplot.show()
```

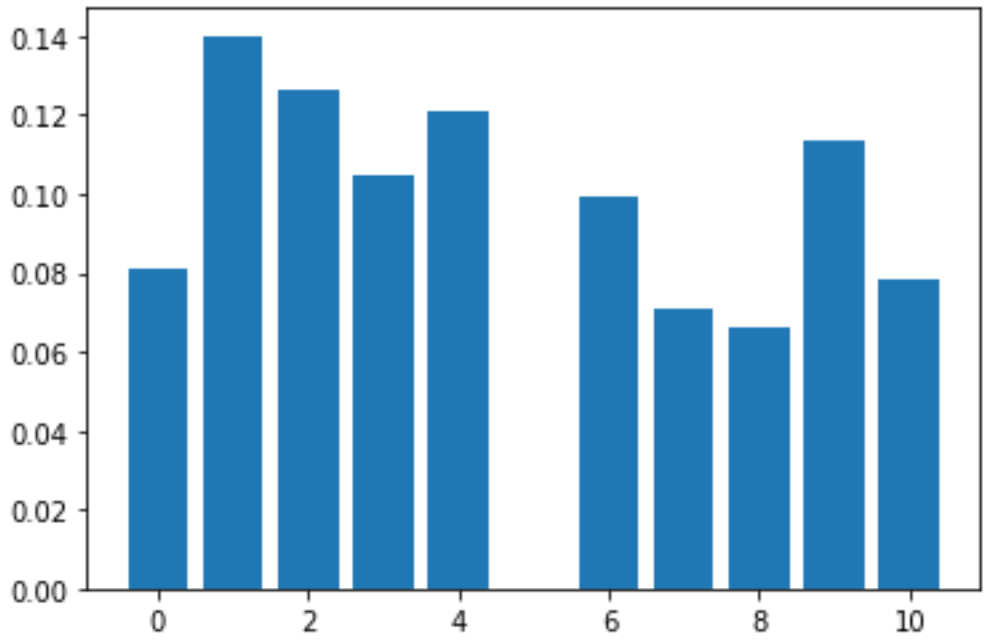



```
fig, ax = pyplot.subplots()
ax.plot(x_axis, results['validation_0']['error'], label='Train')
ax.plot(x_axis, results['validation_1']['error'], label='Test')
```

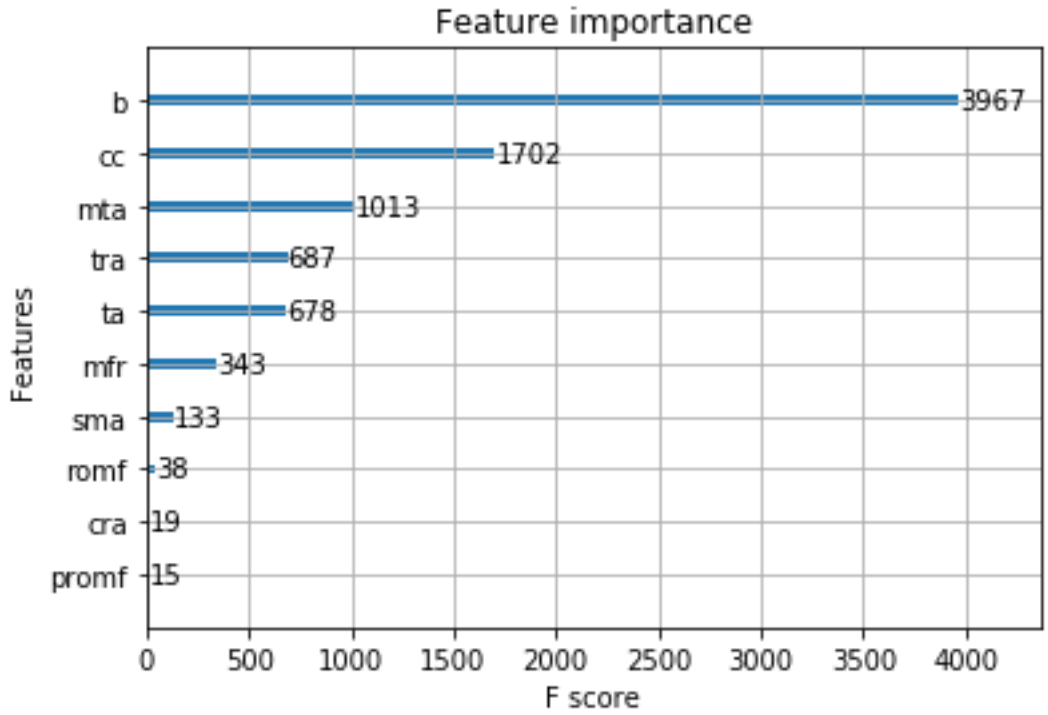
```
ax.legend()
pyplot.ylabel('Classification Error')
pyplot.title('XGBoost Classification Error')
pyplot.show()
```



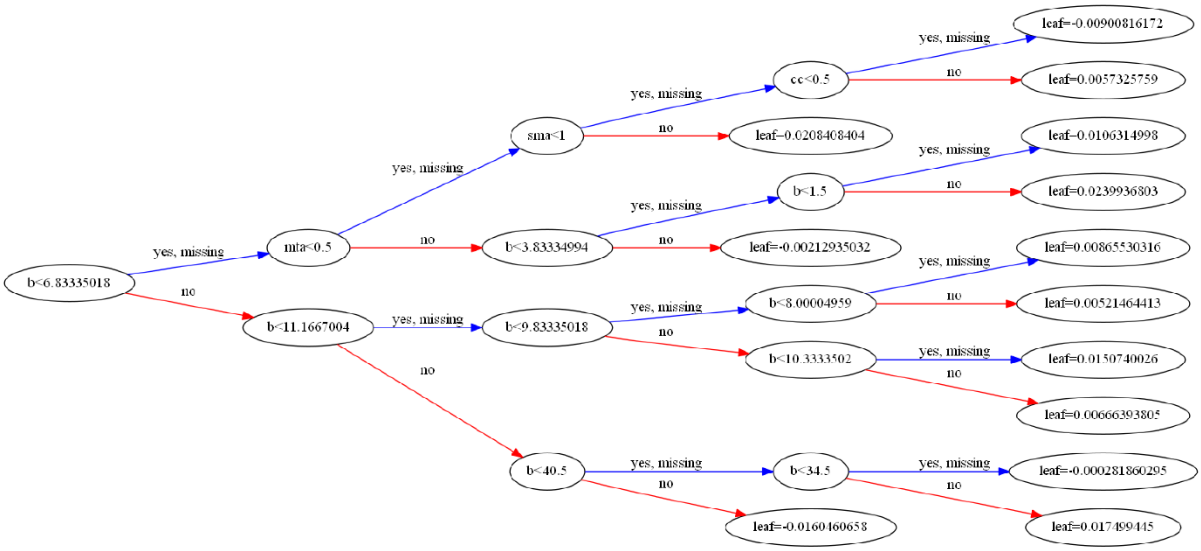

```
joblib.dump(model, 'model_xgboost_bc_insulin.pkl', compress=True)
print(model.feature_importances_)
pyplot.bar(range(len(model.feature_importances_)), model.feature_importances_)
pyplot.show()
```




```
plot_importance(model)
pyplot.show()
```




```
plot_tree(model, num_trees=998, rankdir="LR")
fig = plt.gcf()
fig.set_size_inches(180,50)
pyplot.show()
```



## BIODATA PENULIS


Nama : Moh. Hamim Zajuli Al Faroby
TTL : Banyuwangi, Juli 1995
Alamat : Desa Gladag, Kecamatan Rogojampi, Kabupaten Banyuwangi, Jawa Timur
Email : alfa.crash@gmail.com
No. Hp : 081331653603


Sebelum memulai pendidikan di Institut Teknologi Sepuluh Nopember, saya pernah mengemban pendidikan formal di MI Miftahul Ulum Mangir – Rogojampi lulusan 2007, Setelah itu melanjutkan ke tingkat sekolah menengah pertama di SMP Bustanul Makmur Genteng – Banyuwangi lulusan 2010 dan setelah itu di SMA Negeri 1 Glagah – Banyuwangi lulusan 2013, Setelah itu menempuh Sarjana di departemen Matematika, FMKSD ITS lulusan pada wisuda 117. Selain pendidikan formal saya juga pernah menempuh pendidikan non formal di pondok pesantren yaitu PP. Bustanul Makmur II tahun 2007-2010 dan juga PP. Sirojut Tholibien 2010-2013.

Penelitian yang saya minati ke bidang Bioinformatika, Tugas Akhir yang saya kerjakan ketika mendapatkan gelas Sarjana Sains berjudul "Identifikasi Jenis Kanker Darah (Leukemia) terhadap pengaruh parameter kernel *Support Vector Machine* dan ekstraksi ciri rantai Markov orde 2", dan diseminarkan pada *International Conference on Applied & Industrial Mathematics and Statistics 2019* di Pahang, Malaysia. Jurnal dari seminar tersebut telah terbit di IOP Publish *Journal of Physics*: Conference Series. Topik tesis yang sekarang saya ambil juga merupakan topik Bioinformatika yang mendapat hibah dana bantuan penelitian dari Kementerian Riset dan Teknologi Republik Indonesia yakni Penelitian Magister.

Pengalaman organisasi saya sejak duduk di SMP. Saat itu mengikuti Organisasi SC (*Student Council*) nama OSIS di sekolah saya dan juga pengurus Takmir di SMP. Saat beranjak di SMA organisasi yang pernah saya ikuti di Takmir masjid Al-Hurriyah sebagai kepala dan ketua Ekstrakulikuler Bulutangkis. Saat berada di ITS saya ikut organisasi dibidang dakwah yaitu Ibnu Muqlah sejak tahun 2014-2016 dan juga pada organisasi UKM IBC (ITS Badminton Community) sebagai staf kepelatihan anggota.