



TUGAS AKHIR - KS 141501

SEGMENTASI SUPPLIER MENGGUNAKAN METODE K-MEANS *CLUSTERING* (STUDI KASUS : PTPN X PG MERITJAN)

SUPPLIER SEGMENTATION USING K-MEANS *CLUSTERING* (CASE STUDY : PTPN X PG MERITJAN)

**NABILAH SHOFIANI
NRP 5213 100 051**

**Dosen Pembimbing :
Edwin Riksakomara, S. Kom., M. T.**

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017**



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS 141501

**SEGMENTASI SUPPLIER MENGGUNAKAN METODE K-
MEANS *CLUSTERING*
(STUDI KASUS : PTPN X PG MERITJAN)**

NABILAH SHOFIANI
NRP 5213 100 051

Dosen Pembimbing :
Edwin Riksakomara, S. Kom., M. T.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017



ITS
Institut
Teknologi
Sepuluh Nopember

FINAL PROJECT - KS 141501

***SUPPLIER SEGMENTATION USING K-MEANS
CLUSTERING
(CASE STUDY : PTPN X PG MERITJAN)***

**NABILAH SHOFIANI
NRP 5213 100 051**

**Supervisor :
Edwin Riksakomara, S. Kom., M. T.**

**DEPARTMENT OF INFORMATION SYSTEM
Faculty of Information Technology
Sepuluh Nopember Institute of Technology
Surabaya 2017**

LEMBAR PENGESAHAN
SEGMENTASI SUPPLIER MENGGUNAKAN
METODE K-MEANS *CLUSTERING*
(STUDI KASUS : PTPN X PG MERITJAN)

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

NABILAH SHOFIANI
5213 100 051

Surabaya, Juli 2017

KEPALA
DEPARTEMEN SISTEM INFORMASI

Dr. Ir. Aris Tjahyanto, M. Kom.

NIP 19650310199102001

LEMBAR PERSETUJUAN
SEGMENTASI SUPPLIER MENGGUNAKAN METODE
K-MEANS CLUSTERING
(STUDI KASUS : PTPN X PG MERITJAN)

TUGAS AKHIR
Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Oleh :

NABILAH SHOFIANI
NRP.5213 100 051

Disetujui Tim Penguji : Tanggal Ujian : Juli 2017
Periode Wisuda : September 2017

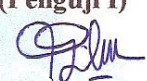
Edwin Riksakomara, S.Kom., M.T.


(Pembimbing I)

Wiwik Anggraeni, S.Si., M.Kom.


(Penguji I)

Renny Pradina, S.T., M.T.


(Penguji II)

SEGMENTASI SUPPLIER MENGGUNAKAN METODE K-MEANS CLUSTERING

(STUDI KASUS : PTPN X PG MERITJAN)

Nama Mahasiswa : Nabilah Shofiani
NRP : 5213 100 051
Departemen : Sistem Informasi FTIf-ITS
Pembimbing : Edwin Riksakomara, S.Kom., M.T.

ABSTRAK

PTPN X Meritjan adalah sebuah perusahaan milik negara di bidang agribisnis yang berfokus pada pengolahan tebu menjadi gula yang nantinya akan dipasarkan di seluruh pelosok Indonesia. Perusahaan ini memiliki jumlah supplier yang lebih dari 500 supplier yang terdiri dari para petani tebu. Perusahaan telah menerapkan SRM sebagai bukti bahwa mereka telah memahami pentingnya supplier untuk proses bisnisnya. Akan tetapi, PTPN X Meritjan merasa bahwa SRM yang mereka lakukan belum mendapatkan dampak yang signifikan. Hal ini disebabkan karena perusahaan melakukan segmentasi supplier berdasarkan daerah para supplier tersebut tinggal, sehingga segmentasi tersebut belum menggambarkan keadaan supplier dan perusahaan dalam dunia nyata.

Solusi untuk menangani masalah tersebut adalah dengan mengelompokkan dengan perilaku bisnis supplier yang dinamis. Supplier yang memiliki perilaku yang sama akan dikelompokkan menjadi satu. Perilaku bisnis supplier dapat dilihat melalui frekuensi pengiriman, rentan waktu pembelian dan nilai barang dalam bentuk uang yang dapat disebut dengan Analisis RFM. Penelitian ini menggunakan metode K-Means Clustering untuk memproses hasil analisis RFM, dimana metode menghasilkan pengelompokan supplier dimana seorang supplier hanya akan masuk dalam satu kelompok saja.

Luaran dari penelitian ini adalah berupa segmentasi supplier yang ditampilkan dalam visualisasi berbasis web yang diharapkan dapat membantu PTPN X Meritjan dalam melakukan segmentasi supplier untuk membuat strategi hubungan dengan tiap segmennya. Perusahaan memiliki derajat kepentingan yang berbeda untuk melakukan segmentasi, recency adalah 0.67, frequency 0.23 dan monetary 0.1. Segmentasi supplier yang dihasilkan terdiri dari 3 kelompok, dimana cluster 3 menjadi cluster yang terbaik berdasarkan kebutuhan perusahaan dengan nilai recency sebesar 156 hari, frequency pengiriman tebu sebanyak 4 kali dan nilai monetary sebesar Rp53,351,836.

Kata Kunci: segmentasi supplier, klastering, k-means, RFM analisis

**SUPPLIER SEGMENTATION USING K-MEANS
CLUSTERING
(CASE STUDY : PTPN X PG MERIJAN)**

Student Name : Nabilah Shofiani
NRP : 5213 100 051
Department : Sistem Informasi FTIf-ITS
Supervisor : Edwin Riksakomara, S.Kom., M.T.

PTPN X PG Meritjan is one of state-owned enterprised in agribusiness that focused on processing sugarcane to sugar that will be distributed in all over country, PTPN X PG Meritjan have more than 500 sugarcane supplier which is sugarcane farmer itself. The company already applied SRM as one of evidence that they already acknowledging the importance of supplier in their business process. However, PTPN X PG Meritjan believe that what they already did haven't give them a significant impact. This is caused by wrong supplier segmentation, they segmented the supplier by their residence and it can't describe supplier and company circumstances in real life.

Solution for this problem is to segment supplier by its dynamic business behavior. Supplier that have same behavior will be grouped into one group. Dynamic business behavior itself can be seen from its delivery frequency, delivery range time and the goods value in rupiah, or that usually called RFM Analysis. This research use K-Means clustering to compute RFM analysis result. K-Means will produce segmentation where the supplier belong to one and only segment.

The outcome of this research supplier segmentation that will be presented in visualization with internet based and hopefully can help PTPN X PG Meritjan to segment the supplier for helping

company to make the right strategy for communicating for each segment. The company has different importance degree for each RFM variables, recency is 0.67, frequency is 0.23 and monetary is 0.1. Supplier segmentation that generated from the data mining process consist 3 segments, whereas the best cluster is cluster number 3 with recency value is 156 days, the amount of sugarcane delivery is 4 times and its monetary value is Rp53,351,836

Keywords: supplier segmentation, clustering, k-means, RFM analysis

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa sehingga penulis dapat menyelesaikan buku tugas akhir dengan judul :

SEGMENTASI SUPPLIER MENGGUNAKAN METODE K-MEANS *CLUSTERING* (STUDI KASUS : PTPN X PG MERITJAN)

yang merupakan salah satu syarat kelulusan pada Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Dalam pengerjaan tugas akhir yang berlangsung selama satu semester, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada yang senantiasa terlibat secara langsung mrahemberikan bantuan dan dukungan dalam pengerjaan tugas akhir ini :

- Allah SWT yang telah melimpahkan rahmat dan anugrahnya sehingga Tugas Akhir ini dapat selesai
- Kepada keluarga yang selalu mendoakan dan memberi dukungan yang besar untuk menyelesaikan Tugas Akhir ini.
- Bapak Dr. Ir. Aris Tjahyanto, M.Kom, selaku Kepala Departemen Sistem Informasi ITS, yang telah menyediakan fasilitas terbaik untuk kebutuhan penelitian mahasiswa.
- Bapak Edwin Riksakomara, S.Kom., M.T. selaku dosen pembimbing yang telah memberikan waktunya untuk memberi masukan dan membimbing penulisan tugas akhir ini
- Kepada pihak PTPN X PG Meritjan yang telah memberi atas ketersediaan waktu dan tenaga dalam memberikan data penelitian.
- CCG, Oriehanna, Garin, Catherine, Nevy, Nadya, Aya dan Dara yang selalu memberi semangat dan dukungan dalam mengerjakan tugas akhir
- Teman-teman laboratorium RDIB dan BELTRANIS, yang selalu mendukung satu sama lain dalam penyelesaian tugas akhir

Penulis menyadari bahwa Tugas Akhir ini masih belum sempurna dan memiliki banyak kekurangan di dalamnya. Oleh karena itu, penulis merima saran dan kritik yang membangun guna kesempurnaan Tugas Akhir ini. Semoga Tugas Akhir ini bermanfaat bagi kita semua. Amin.

Surabaya, 2017

Penulis

DAFTAR ISI

ABSTRAK.....	vii
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL.....	xviii
DAFTAR SCRIPT.....	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
1.6 Relevansi.....	6
1.7 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	9
2.1 Penelitian Sebelumnya.....	9
2.2 Dasar Teori.....	11
2.2.1 PT Perkebunan Nusantara X Meritjan.....	11
2.2.2 Pengelolaan Hubungan Supplier.....	14
2.2.3 Penggalan Data.....	16
2.2.4 Analisis RFM.....	19
2.2.5 Transformasi Data.....	20
2.2.6 Analytic Hierarchial Process (AHP).....	21
2.2.7 Clustering.....	24
2.2.8 K-Means.....	26
2.2.9 Euclidian Distance.....	28
2.2.10 Metode Elbow.....	28
2.2.11 Uji Performa.....	29
2.2.12 Normalisasi Min-Max.....	31
2.2.13 Customer Lifetime Value.....	32
2.2.14 Uji Korelasi.....	32
2.2.15 Linear Regresi.....	33
2.2.16 Tukey Method.....	34
BAB III METODOLOGI TUGAS AKHIR.....	35
3.1 Tahapan Pelaksanaan Tugas Akhir.....	35

3.2	Uraian Metodologi.....	35
3.2.1	Identifikasi Permasalahan.....	35
3.2.2	Studi Literatur.....	35
3.2.3	Pengumpulan Data.....	36
3.2.4	Data Selection.....	37
3.2.5	Ekstraksi RFM.....	37
3.2.6	Data Preparation	37
3.2.7	Data Transformation.....	37
3.2.8	Clustering Data.....	38
3.2.9	Pembobotan RFM Menggunakan AHP	39
3.2.10	Perhitungan Nilai CLV	39
3.2.11	Uji Korelasi dan Perhitungan Lahan Optimal	40
3.2.12	Denormalisasi Data	40
3.2.13	Analisis Hasil Cluster	40
3.2.14	Visualisasi Cluster	40
3.2.15	Penarikan Kesimpulan dan Saran	41
3.2.16	Pembuatan Laporan Tugas Akhir	41
BAB IV PERANCANGAN		43
4.1	Pengumpulan Data.....	43
4.2	Data Selection.....	45
4.3	Ekstraksi RFM.....	45
4.3.1	Ekstraksi Variabel Recency.....	48
4.3.2	Ekstraksi Variabel Frequency.....	48
4.3.3	Ekstraksi Variabel Monetary	48
4.4	Data Preparation	48
4.4.1	Data Transformation.....	49
4.4.2	Data Cleansing.....	49
4.4.3	Normalisasi Menggunakan Min-Max.....	49
4.5	Penentuan Jumlah Cluster	50
4.6	Use Case Visualisasi Supplier	50
4.7	Rancangan Visualisasi.....	55
BAB V IMPLEMENTASI		59
5.1	Ekstraksi RFM.....	59
5.1.1	Ekstraksi Variabel Recency.....	59
5.1.2	Ekstraksi Variabel Frequency.....	59
5.1.3	Ekstraksi Variabel Monetary	60
5.2	Data Preparation	60

5.2.1	Data Transformation.....	60
5.2.2	Data Cleansing	60
5.2.3	Normalisasi Menggunakan Min-Max.....	62
5.3	Penentuan Jumlah Cluster	62
5.4	Proses Clustering.....	63
5.4.1	Clustering Menggunakan K-Means.....	63
5.4.2	Uji Performa Menggunakan SSE	64
5.4.3	Uji Performa Menggunakan CValid	64
5.4.4	Verifikasi Cluster	64
5.5	Uji Korelasi dan Perhitungan Lahan Optimal	65
5.6	Denormalisasi.....	66
5.7	Visualisasi Cluster.....	67
5.7.1	3D Scatter plot.....	67
5.7.2	Pie chart.....	68
5.7.3	Box and whisker.....	69
5.7.4	Shiny Dashboard	70
BAB VI HASIL DAN PEMBAHASAN.....		73
6.1	Ekstraksi RFM.....	73
6.2	Data Preparation.....	74
6.2.1	Data Transformation.....	74
6.2.2	Data Cleansing	75
6.2.3	Normalisasi Menggunakan Min-Max.....	77
6.3	Penentuan Jumlah Cluster	79
6.4	Clustering	80
6.4.1	Clustering Menggunakan K-Means.....	80
6.4.2	Uji Performa Menggunakan SSE	81
6.4.3	Uji Performa Menggunakan CValid	81
6.4.4	Verifikasi Hasil Clustering.....	82
6.5	Pembobotan RFM Menggunakan AHP.....	83
6.5.1	Perhitungan AHP.....	84
6.5.2	Uji Konsistensi AHP	84
6.6	Perhitungan Nilai CLV.....	85
6.7	Analisa Uji Korelasi dan Perhitungan Lahan Optimal	86
6.8	Analisa Clustering.....	89

6.8.1	Cluster 1	90
6.8.2	Cluster 2	91
6.8.3	Cluster 3	91
6.9	Analisa Hasil Visualisasi.....	92
6.9.1	Scattered Plot.....	92
6.9.2	Pie chart.....	94
6.9.3	Box and whisker.....	95
6.10	Kesimpulan Penelitian.....	99
6.11	Rekomendasi	100
BAB VII KESIMPULAN DAN SARAN		101
7.1	Kesimpulan.....	101
7.2	Saran.....	101
BIODATA PENULIS.....		103
DAFTAR PUSTAKA.....		105
LAMPIRAN A		A-1
LAMPIRAN B		B-3
LAMPIRAN C		C-3
LAMPIRAN D		D-1
LAMPIRAN E.....		E-3
LAMPIRAN F.....		F-1
LAMPIRAN G		G-1
LAMPIRAN H		H-1
LAMPIRAN I.....		I-1
LAMPIRAN J.....		J-1

DAFTAR GAMBAR

Gambar 2.1 Elbow method.....	28
Gambar 3.1 Metodologi Penelitian	36
Gambar 4.1 Scatterplot Sebelum Cleansing.....	49
Gambar 4.2 Use Case Diagram Visualisasi.....	51
Gambar 4.3 Rancangan halaman dashboard	56
Gambar 4.4 Rancangan halaman cluster	56
Gambar 4.5 Rancangan halaman data	57
Gambar 5.1 Scatterplot Sebelum Cleansing.....	61
Gambar 5.2 Grafik 3D Scatter plot	68
Gambar 5.3 Grafik Pie chart	69
Gambar 5.4 Grafik Box and whisker.....	70
Gambar 5.5 Halaman Dashboard A	71
Gambar 5.6 Halaman Cluster	71
Gambar 5.7 Halaman Data	72
Gambar 6.1 Perbandingan Histogram Sebelum dan Sesudah Transformasi pada Frequeny	75
Gambar 6.2 Perbandingan Histogram Sebelum dan Sesudah Transformasi pada Monetary	75
Gambar 6.3 Perbandingan Ada Tidaknya Outlier pada Frequency.....	76
Gambar 6.4 Perbandingan Ada Tidaknya Outlier pada Monetary	76
Gambar 6.5 Grafik Penghapusan Data bernilai NA	77
Gambar 6.6 Elbow method.....	80
Gambar 6.7 Plot Area- Monetary	87
Gambar 6.8 Grafik 3D Scatter plot Recency.....	93
Gambar 6.9 Grafik 3D Scatter plot Frequency.....	94
Gambar 6.10 Grafik 3D Scatter plot Monetary	94
Gambar 6.11 Grafik Prosentasi Anggota Setiap Clutser	95
Gambar 6.12 Box and whisker Recency	96
Gambar 6.13 Box and whisker Frequency	98
Gambar 6.14 Box and whisker Monetary	99

DAFTAR TABEL

Tabel 2.1 Penelitian Sebelumnya Paper 1	9
Tabel 2.2 Penelitian Sebelumnya Paper 2	9
Tabel 2.3 Penelitian Sebelumnya Paper 3	10
Tabel 2.4 Penelitian Sebelumnya Paper 4	11
Tabel 2.5 Distribusi Data dan Transformasi Data yang Tepat	21
Tabel 2.6 Pairwise comparison.....	22
Tabel 2.7 Random Consistency Index	24
Tabel 2.8 Algoritma K-Means.....	27
Tabel 2.9 Algoritma Elbow method	29
Tabel 2.10 Tabel Interpretasi Uji Korelasi	33
Tabel 4.1 Data Mentah Tabel Pengadaan Tebu.....	43
Tabel 4.2 Perhitungan Monetary	44
Tabel 4.3 Tabel Profil Supplier	44
Tabel 4.4 Variabel Terpilih dari Tabel Penerimaan Tebu	45
Tabel 4.5 Potongan Data Mentah Transaksi Pengadaan Tebu	46
Tabel 4.6 Potongan Data Mentah Tabel Profil Supplier	47
Tabel 4.7 Use case description melihat grafik 3d scatter plot supplier	51
Tabel 4.8 Use case description melihat data cluster supplier	52
Tabel 4.9 Use case description melihat grafik pie chart persebaran supplier	52
Tabel 4.10 Use case description melihat grafik box and whisker variabel RFM.....	53
Tabel 4.11 Use case description melihat supplier terburuk dan terbaik RFM	54
Tabel 4.12 Use case description melihat rata-rata variabel RFM per cluster	55
Tabel 6.1 Ekstraksi Recency	73
Tabel 6.2 Ekstraksi Frequency	73
Tabel 6.3 Ekstraksi Monetary.....	74
Tabel 6.4 Hasil Transformasi Data.....	74
Tabel 6.5 Tabel Hasil Penghapusan Outlier	76
Tabel 6.6 Hasil Penghapusan Data yang Memiliki Nilai NA.....	77
Tabel 6.7 Normalisasi Recency	78
Tabel 6.8 Normalisasi Recency Setelah Disesuaikan.....	78

Tabel 6.9 Normalisasi Frequency.....	78
Tabel 6.10 Normalisasi Monetary	79
Tabel 6.11 Integrasi Tabel.....	79
Tabel 6.12 Hasil Clustering.....	80
Tabel 6.13 Hasil Uji SSE	81
Tabel 6.14 Hasil Uji performa Menggunakan CValid	81
Tabel 6.15 Hasil perhitungan distance pada setiap centroid ..	82
Tabel 6.16 Hasil Perbandingan Jumlah Anggota Cluster.....	83
Tabel 6.17 Hasil Wawancara Perbandingan Kriteria	83
Tabel 6.18 Bobot Variabel RFM.....	84
Tabel 6.19 Hasil Penjumlahan Tiap Variabel	85
Tabel 6.20 Hasil Perkalian Data dengan Bobot	86
Tabel 6.21 Hasil Uji Korelasi.....	87
Tabel 6.22 Hasil Prediksi Linear Regresi.....	88
Tabel 6.23 Tabel Lahan yang Dapat Dioptimalkan	88
Tabel 6.24 Tabel Lahan Terburuk.....	89
Tabel 6.25 Deskripsi Data.....	89
Tabel 6.26 Range Nilai Variabel Per Cluster	90
Tabel 6.27 Analisis Cluster 1	90
Tabel 6.28 Analisis Cluster2	91
Tabel 6.29 Analisis Cluster 3	91
Tabel 6.30 Tabel Range Recency.....	95
Tabel 6.31 Tabel Range Frequency.....	97
Tabel 6.32 Tabel Range Monetary	98

DAFTAR SCRIPT

Script 5.1 Ekstraksi Recency	59
Script 5.2 Ekstraksi Frequency	60
Script 5.3 Ekstraksi Monetary	60
Script 5.4 Transformasi data menggunakan metode algoritmik	60
Script 5.5 Script Penghapusan Outlier.....	62
Script 5.6 Script Normalisasi Min-Max	62
Script 5.7 Script Tahapan Cluster.....	63
Script 5.8 Script Grafik Elbow Method.....	63
Script 5.9 Script K-Means Clustering	63
Script 5.10 Script memperlihatkan wss	64
Script 5.11 Script CValid	64
Script 5.12 Script menunjukan distance terdekat	65
Script 5.13 Script Uji Korelasi	65
Script 5.14 Script linear regresi	65
Script 5.15 Script plot linear regresi.....	65
Script 5.16 Script perhitungan dan pemilihan lahan optimal .	66
Script 5.17 Script pemilihan lahan terburuk.....	66
Script 5.18 Script denormalisasi tahapan normalisasi	66
Script 5.19 Script denormalisasi tahapan data transformasi ...	67
Script 5.20 Script Plot 3D Scatterplot	67
Script 5.21 Script Pie chart.....	68
Script 5.22 Script Box and whisker	69
Script 6.1 Hasil uji kesamaan hasil cluster	82

BAB I

PENDAHULUAN

Pada bab ini, akan dijelaskan tentang Latar Belakang Masalah, Perumusan Masalah, Batasan Masalah, Tujuan Tugas Akhir, dan Relevansi atau Manfaat Kegiatan Tugas Akhir

1.1 Latar Belakang Masalah

PTPN X Meritjan adalah sebuah perusahaan milik negara di bidang agribisnis. Fokus perusahaan adalah pengolahan tebu menjadi gula yang nantinya akan dipasarkan di seluruh pelosok Indonesia. Walaupun target pasar perusahaan ini terhitung luas, perusahaan ini tidak memiliki lahan sendiri. Perusahaan bekerjasama dengan supplier untuk selalu memasok bahan baku mereka, yaitu tebu. Jumlah pemasok bahan baku utama perusahaan berjumlah lebih dari 500 yang terdiri dari para petani tebu. Oleh karena itu, proses pengadaan bahan baku ini menjadi proses yang sangat penting bagi perusahaan.

Selain itu, peran fungsi pengadaan bahan baku mengalami perubahan yang drastis. Sebelumnya, bagian pengadaan dahulu memiliki tugas untuk meyakinkan bahwa produk tersedia pada waktu yang ditentukan. Namun kini, perusahaan telah mengetahui bahwa terdapat potensi kontribusi bagian pengadaan untuk mencapai target biaya yang dikeluarkan dalam proses pengadaan. Pada bagian pengadaan, perusahaan dapat melakukan optimalisasi jumlah bahan baku yang dipesan sehingga dapat mengurangi harga produk. Selain itu setelah ditelusuri lebih lanjut, bagian pengadaan juga memiliki peran yang penting dalam mengoptimasi biaya total seluruh proses.

Menurut Remko, 50 – 80% dari total biaya adalah biaya yang dikeluarkan oleh bagian pengadaan. [1] Selain itu, perusahaan sangat bergantung dengan supplier mereka sehingga menyebabkan beberapa risiko termasuk terbatasnya kontrol perusahaan terhadap ketersediaan, jadwal pengiriman dan kualitas produk yang dihasilkan. Terlebih lagi, supplier bahan utama produk, dapat menentukan untuk memberhentikan

pasokan bahan baku atau menaikkan harga yang berada di luar kendali perusahaan. Kekurangan bahan baku dapat berdampak buruk pada produksi dan menyebabkan peningkatan harga bahan baku tersebut. Peningkatan harga baku tentunya akan mempengaruhi seluruh proses produksi. [2] Oleh karena itu, kolaborasi dan integrasi dengan supplier atau pihak eksternal sangat penting untuk dilakukan untuk mengurangi resiko ketersediaan bahan baku.

Kolaborasi dapat dikatakan sukses apabila supplier mau untuk memberikan nilai yang lebih pada perusahaan dengan meningkatkan servis, fleksibilitas dan berbagi sumber daya dengan perusahaan. Untuk mendapatkan kepercayaan supplier, perusahaan harus membangun hubungan dengan pihak supplier sama seperti hubungan yang dibangun dengan pelanggannya. [3] Pengelolaan hubungan supplier (SRM) bertujuan untuk membantu perusahaan untuk mengelola hubungan perusahaan dengan setiap suppliernya dan juga membantu perusahaan untuk memanfaatkan kapabilitas supplier serta potensi supplier untuk mencapai tujuan bisnis perusahaan melalui kolaborasi.

Pengelolaan hubungan supplier adalah pendekatan yang digunakan untuk berhubungan dengan supplier pada tingkat yang mencerminkan prioritas perusahaan dan bagaimana kebutuhan perusahaan dapat dicapai. Melalui pengelolaan hubungan supplier perusahaan meningkatkan keunggulan kompetitif perusahaan, dimana perusahaan dapat melakukan *outsource* atau mendapatkan saran dari supplier mengenai pembuatan produk baru, salah satu cara terbaik dalam melakukan produksi. [4] 80% responden dari buku *Negotiation and Managing Key Supplier Relationships* mengatakan bahwa terjadi peningkatan nilai yang diberikan oleh supplier sebanyak 25% setelah mengimplementasikan pengelolaan hubungan supplier. [5]

Pengimplementasian SRM juga tidak mudah, karena proses ini membutuhkan sumber daya yang berupa uang dan waktu. [6] Sehingga, tidak semua hubungan dengan supplier memiliki

dampak positif dengan supplier atau dapat dibidang tidak semua supplier menjadi partner bisnis. Segmentasi supplier dibutuhkan untuk membedakan strategi dalam membangun hubungan dengan supplier. Apabila supplier bernilai tinggi bagi perusahaan, maka perusahaan dapat memutuskan untuk membangun perkongsian dengan supplier tersebut dan bila tidak menguntungkan hubungan dengan supplier dapat dihentikan. [1]

Segmentasi supplier adalah proses membedakan supplier untuk menentukan tingkat hubungan dengan supplier tertentu. Segmentasi supplier juga membantu perusahaan untuk menentukan sumber daya (waktu dan biaya) yang digunakan untuk mengelola supplier pada tingkat yang berbeda. [7]. Segmentasi supplier merupakan jembatan antara pengelolaan hubungan supplier dan pemilihan supplier.

PTPN X Meritjan telah mengetahui pentingnya supplier untuk kemajuan bisnis mereka dengan menerapkan SRM, akan tetapi PTPN X Meritjan merasa bahwa SRM yang mereka lakukan belum mendapatkan dampak yang signifikan. Hal ini disebabkan PTPN X Meritjan memberikan perlakuan yang sama pada setiap supplier. Selain itu, perusahaan juga mengelompokkan supplier berdasarkan lokasi dan hubungan sanak saudara. Padahal, berdasarkan wawancara yang dilakukan sebelumnya, perusahaan berharap untuk mengetahui supplier mana saja yang berkontribusi besar pada perusahaan. Pengelompokan seperti itu tidak dapat menggambarkan kondisi atau kontribusi supplier pada perusahaan.

Solusi untuk menangani masalah tersebut adalah mengelompokkan dengan karakteristik bisnis supplier yang dinamis. Supplier yang memiliki perilaku sama akan dikelompokkan menjadi satu. [8] Perilaku bisnis supplier dapat dilihat melalui frekuensi pengiriman, rentan waktu pembelian dan nilai barang dalam bentuk uang yang dapat disebut dengan RFM Analysis. RFM Analysis merupakan sebuah analisis yang dapat membantu membedakan pihak yang berpengaruh dari

data yang besar berdasarkan tiga variabel yaitu interval satu transaksi dengan saat ini, frekuensi dan nilai produk dalam bentuk uang [9]. Ketiga variabel ini didapatkan berdasarkan transaksi yang dilakukan dari waktu ke waktu. Pola perilaku bisnis supplier juga dapat berubah sehingga dengan mengelompokkan berdasarkan dengan karakteristiknya, hasil pengelompokan menjadi lebih dinamis dan lebih sesuai dengan keadaan bisnis sesungguhnya. [8]

Segmentasi supplier dapat dilakukan dengan melakukan *clustering*. *Clustering* adalah sebuah metode yang digunakan untuk mengidentifikasi *cluster* yang terdapat dalam data, dimana *cluster* adalah sekumpulan objek data yang sama antara satu sama lain. [10]. *Clustering* akan digunakan untuk mencari supplier-supplier mana saja yang memiliki perilaku yang sama. Setelah itu akan dilakukan pengelompokan berdasarkan karakteristiknya.

Metode *clustering* yang digunakan untuk melakukan segmentasi supplier pada PTPN X Meritjan adalah K-Means *clustering*. K-Means *clustering* adalah teknik *clustering* berdasarkan prototip yang membuat pemisahan tingkat satu pada data objek. [11] Metode ini digunakan karena hasil pengelompokan yang diinginkan anggota kelompoknya hanya benar-benar berada hanya ada di kelompok tersebut. Selain itu, metode ini mudah diinterpretasikan, mudah diimplementasikan dan dapat beradaptasi pada data yang tersebar [12]. Namun, metode ini memiliki kekurangan dimana nilai k harus ditentukan terlebih dahulu untuk menghasilkan hasil *clustering* yang terbaik. Metode elbow dapat membantu pencarian nilai k yang optimal dengan membandingkan nilai k dengan nilai SSE atau presentase biaya pada nilai *cluster* yang telah ditentukan. Hasil presentase akan ditampilkan melalui grafik yang mempermudah untuk dipahami. Apabila nilai SSE atau presentase biaya pada grafik tersebut turun drastis dan membentuk siku, maka nilai tersebut yang digunakan sebagai nilai k [13].

Tugas akhir ini akan menghasilkan segmentasi supplier PTPN X PG Meritjan untuk menampilkan perilaku supplier berdasarkan analisis RFM . Hasil segmentasi akan ditampilkan melalui visualisasi yang berbasis web dengan metode *k-means clustering* x. Visualisasi dilakukan untuk memudahkan perusahaan untuk memahami hasil segmentasi supplier.

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, perumusan masalah yang diangkat dalam tugas akhir ini dibagi menjadi yaitu :

1. Bagaimana implementasi *clustering* yang menampilkan perilaku bisnis supplier berdasarkan model RFM?
2. Bagaimana visualiasi *clustering* ditampilkan?

1.3 Batasan Masalah

Dalam pengerjaan tugas akhir ini, penulis memberikan batasan-batasan masalah dalam usulan tugas akhir ini agar pembahasan tugas akhir terorganisir dengan baik. Berikut adalah batasan tugas akhir :

- a. Data yang diolah dalam tugas akhir ini adalah data transaksi pengadaan tebu yang dimiliki PTPN X Meritjan Kediri pada tanggal 9 Juni 2016 hingga 6 November 2016
- b. Penelitian ini terbatas pada segmentasi supplier berdasarkan lahan tanam
- c. Penentuan variabel *clustering* menggunakan metode RFM
- d. Penentuan jumlah *cluster* menggunakan metode elbow
- e. Penelitian ini menggunakan software R sebagai *tools clustering*
- f. Visualisasi *cluster* berbasis web

1.4 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan yang ingin dicapai yaitu sebagai berikut :

1. Untuk membuat *clustering* yang menampilkan perilaku bisnis supplier berdasarkan model RFM
2. Untuk membuat visualisasi *clustering* berbasis web-based

1.5 Manfaat Penelitian

Manfaat penelitian yang akan diperoleh berdasarkan tujuan adalah untuk memudahkan PTPN X PG Meritjan dalam membangun strategi berhubungan dan penggunaan supplier yang tepat bagi supplier perusahaan.

1.6 Relevansi

Tugas akhir ini berkaitan dengan mata kuliah Sistem Cerdas dan Manajemen Rantai Pasok dan Hubungan Pelanggan yang tercakup pada Laboratorium Rekayasa Data Dan Intelegensi Bisnis.

1.7 Sistematika Penulisan

Pada penulisan tugas akhir, terdapat sistematika penulisan yang digunakan. Sistematika ini terbagi menjadi tujuh bab yang akan dijabarkan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat hal-hal yang melatarbelakangi pengerjaan tugas akhir ini yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, perumusan masalah, tujuan, manfaat penelitian, relevansi dan yang terakhir adalah sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang studi sebelumnya yang mendukung tugas akhir serta teori-teori yang berhubungan dengan segmentasi supplier, *clustering*, RFM model dan k-

means yang akan digunakan untuk menyelesaikan permasalahan pada tugas akhir.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai alur pengerjaan tugas akhir yang berawal dari identifikasi masalah, studi literatur, pengumpulan data, praproses data, penentuan variabel RFM, *clustering* data, visualisasi data, penarikan kesimpulan dan saran, dan diakhiri dengan pembuatan laporan tugas akhir.

BAB IV PERANCANGAN

Pada bab ini akan dibahas mengenai perancangan pengoalahan data dengan melakukan praproses data, analisis RFM dan juga penentuan jumlah *clustering* (nilai k) yang akan digunakan pada K-means *clustering*.

BAB V IMPLEMENTASI

Pada bab ini tentang pengoalah data yang didapat dari tahap sebelumnya dan implemntasi *clustering* dengan menggunakan R.

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan ditampilkan hasil *clustering* dan uji performa hasil *clustering* berdasarkan metode K-Means yang diharapkan dapat menjadi penyelesaian permasalahan.

BAB VII PENUTUP

Pada bab ini akan dijelaskan kesimpulan dan saran untuk penelitian sebelumnya dari pengerjaan tugas akhir yang menjawab rumusan masalah yang telah dijabarkan pada bab Pendahuluan.

(halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan mengenai konsep segmentasi supplier serta kriteria apa saja yang memengaruhinya. Selain itu, akan disertakan penjelasan singkat mengenai *clustering* yang akan digunakan dalam membantu proses segmentasi supplier.

2.1 Penelitian Sebelumnya

Bagian ini akan menjelaskan penelitian-penelitian sebelumnya yang digunakan sebagai dasar penelitian tugas akhir. Hal-hal yang didapat dari penelitian sebelumnya dijabarkan pada tabel dibawah ini:

Tabel 2.1 Penelitian Sebelumnya Paper 1

Judul Penelitian	<i>Feature-based Unsupervised Clustering for Supplier Categorization</i> [1]
Penulis (Tahun)	Danish Irfan, Xu Xiaofei, Deng Shengchun dan Ye Yunming (2008)
Deskripsi Umum Penelitian	Kategorisasi supplier dapat dipermudah dengan menggunakan <i>feature-based clustering</i> . <i>Feature-based clustering</i> merupakan metode pengkategorisasian supplier berdasarkan karakteristik dinamis supplier. Hal ini juga membantu perusahaan mengurangi biaya yang dikeluarkan oleh perusahaan
Keterkaitan Penelitian	Referensi latar belakang alasan penggunaan <i>clustering</i> berdasarkan performa supplier dibanding dengan variabel lokasi atau produk.

Tabel 2.2 Penelitian Sebelumnya Paper 2

Judul Penelitian	<i>Classifying the segmentation of customer value via RFM model and RS theory</i> [2]
Penulis (Tahun)	Ching-Hsue Cheng, You Shyang Chen (2009)
Deskripsi Umum Penelitian	Penelitian ini bertujuan untuk mengelompokan customer pada perusahaan mereka. Pengelompokan ini didasari oleh keinginan

	<p>peningkatan CRM perusahaan. Penelitian ini menggabungkan variabel RFM dan algoritma K-Means pada teori <i>rough set</i> untuk menghasilkan peraturan CRM. RFM variabel digunakan sebagai variabel input kualitatif. Jumlah kelas yang digunakan pada penelitian ada 3 macam yaitu 3, 5 dan 7 yang nantinya akan dibandingkan. Hasil penelitian ini membuktikan bahwa gabungan variabel RFM, algoritma K-Means dan RS theory lebih efisien dalam melakukan pengelompokan berdasarkan nilai pelanggan.</p>
Keterkaitan Penelitian	<p>Referensi penggunaan analisis RFM pada <i>data sample</i>, penggunaan K-Means sebagai metode <i>clustering</i> dan juga alur pengerjaan tugas akhir.</p>

Tabel 2.3 Penelitian Sebelumnya Paper 3

Judul Penelitian	<p><i>Supplier Categorization with K-Means Type Subspace Clustering</i> [3]</p>
Penulis (Tahun)	<p>Xingjun Zhang, Joshua Zhexue Huang, Depei Qian, Jun Xu dan Liping Jing (2006)</p>
Deskripsi Umum Penelitian	<p>Penelitian ini membenahi pengelompokan supplier yang awalnya berasal dari data statistik (lokasi dan product item) menjadi gabungan antara data statistik dan data perilaku (frekuensi pembelian, kualitas dan waktu pengiriman). Data perilaku diambil dari data transaksi. <i>Clustering</i> yang dihasilkan bersifat dinamis karena perilaku dari supplier sering berganti. FW- <i>Clustering</i> digunakan karena data yang besar dan juga banyak data yang bernilai 0.</p>
Keterkaitan Penelitian	<p>Sebagai referensi studi pustaka utama dalam menggunakan metode <i>clustering</i> dari data perilaku supplier yang diambil dari data transaksi serta penggunaan metode k-means untuk <i>clustering</i>.</p>

Tabel 2.4 Penelitian Sebelumnya Paper 4

Judul Penelitian	<i>Customer segmentation issue and strategies for an automobil dealership with two clusteringg technique [4]</i>
Penulis (Tahun)	Chih-Fong Tsai, Ya-Han Hu dan Yu-Hsin Lu (2013)
Deskripsi Umum Penelitian	Segmentasi pelanggan digunakan untuk mengelompokkan pelanggan dengan karakteristik yang sama dan mengidentifikasi perbedaan antara satu kelompok dengan kelompok lainnya untuk menghasilkan strategi marketing. Penelitian ini menggunakan dua teknik pengelompokan yaitu k-means dan expectation maximization. Pelanggan akan dibagi menjadi 4 kelompok yaitu loyal, potensial, VIP dan <i>churn</i> . Variabel yang digunakan dalam pengelompokan ini adalah tingkat kepuasan, pendapatan servis, selang servis, frekuensi transaksi dan kompensasi.
Keterkaitan Penelitian	Penggunaan algoritma k-means sebagai metode <i>clustering</i> .

2.2 Dasar Teori

Pada bagian ini akan dijelaskan mengenai teori-teori yang digunakan oleh penulis dalam tugas akhir ini.

2.2.1 PT Perkebunan Nusantara X Meritjan

PT Perkebunan Nusantara X atau yang disingkat sebagai PTPN X adalah salah satu perusahaan yang berada dibawah naungan pemerintah yang bergerak dalam bidang agribisnis, pengolahan dan pemasaran hasil perkebunan tersebut . PTPN X secara resmi didirikan pada tahun 1996. PTPN X memiliki fokus pada pengolahan gula dan tembakau. Hasil produksi gula akan dipasarkan didalam negeri melalui persaingan bebas dan terkoordinir. Ada juga pembeli produk tetes yang terbagi menjadi dua yaitu, pabrikan dan pembelian melalui tender. Untuk unit industri gula PTPN X memiliki 11 unit pabrik gula yang berada di wilayah Jawa Timur. Kesebelas unit pabrik gula

tersebut adalah PG Kremboong, PG Watoetoelis, PG Toelangan, PG Gempolkrep, PG Djombang Baru, PG Tjoekir, PG Lestari, PG Meritjan, PG Pesantren Baru, PG Ngadirejo dan PG Modjopanggoong.

PG Meritjan adalah salah satu unit industry gula yang terletak di daerah Kediri. PG Meritjan memiliki kapasitas sebesar 2800 TCD yang artinya PG Meritjan dapat menggiling tebu sebanyak 2800 ton tebu perharinya. Wilayah kerja dari PG Meritjan meliputi daerah Kabupaten Nganjuk, Kota Kediri dan Kabupaten Kediri. PG Meritjan dipimpin oleh Ir. Alan Purwandiarto .Msi. sebagai general manager.

Pada tahun 1903, Pabrik Gula Meritjan Kediri didirikan oleh Nderland Indiche Landbouw Maatshaaplj (NLK) dan perusahaan ini bekerja hingga tahun 1935. Perusahaan kemudian ditutup pada tahun 1935 karena perang dunia kedua terjadi. Pada tahun 1942-1945 Pabrik Gula Meritjan diduki Jepang dan dijadikan pabrik senjata. Sehingga, pada tahun 1945 -1948 Pabrik Gula Meritjan dibagi menjadi dua (2) bagian yaitu [5]:

- a. Bagian I sebagai pabrik senjata
- b. Bagian II sebagai perkebunan tebu yang dikelola oleh Badan Penyelenggara Perusahaan Tebu Negara (BPPGN) dan tebunya kemudian digiling oleh Pabrik Gula Pesantren.

Pada tahun 1948 hingga 1957, Pabrik Gula Meritjan diambil alih oleh Belanda. Pada tanggal 9 Desember 1957 sesuai dengan Surat Keputusan (SK) Penguasa Militer/Menteri Pertanian No. 1063/PTNT/1957, semua perusahaan diambil alih oleh Pemerintah RI yang kemudian dikelola dibawah satu badan yaitu Perusahaan Perkebunan Negara (PPN) baru yang berkedudukan di setiap daerah Swatantra Tingkat I.

Pada tahun 1958 diadakan pembagian Prae Unit dimana Pabrik Gula Meritjan masuk ke dalam Prae Unit Gula A. Tertanggal 26 April 1961, berdasarkan PP No. 166/1961, bentuk Prae Unit dihapuskan dan diubah menjadi bentuk Kesatuan dimana

Pabrik Gula Meritjan masuk ke dalam Kesatuan Jawa-Timur II dan berlangsung hingga tahun 1963. Kemudian, berdasarkan PP No. I tahun 1963, Pabrik Gula Meritjan dikuasai oleh BPU-PPN yang berkedudukan di Surabaya

Pada PP No. 14/1968 BPU-PPN, nama Perusahaan Perkebunan Negara (PPN) berganti menjadi Perusahaan Negara Perkebunan (PNP) yang berkedudukan di Surabaya dan badan hukumnya beralih pada Direksi PNP XXI. Berdasarkan PP No. 23 tanggal 11 Mei 1973, bentuk PNP diubah menjadi Perusahaan Perseroan yang dikenal dengan nama PT Perkebunan XXI-XXII (Persero) yang mengelola 12 unit pabrik gula, 2 rumah sakit dan satu kantor pusat. Pada tahun 1984, sesuai dengan akta yang dihadapkan notaris Sdr. Imas Fatimah, SH di Jakarta No 109 pada tanggal 8 Maret 1985 disebut PT Perkebunan XXI-XXII (Persero) dengan badan hukumnya berada di kantor Direksi PT Perkebunan XXI-XXII (Persero) yang berada di Jalan Jembatan Merah 3-9 Surabaya.

Sesuai dengan Peraturan Pemerintah No. 15/1996 pada tanggal 14 Februari 1996, diadakan peleburan PTP termasuk PTP XXI-XXII, PTP XIX dan XXVII menjadi satu PT. Perkebunan Nusantara X (Persero). Pendirian PTPN X (Persero) sesuai akte notaris Harun Kamil, SH. No. 43 pada tanggal 1 Maret 1996 dan disahkan oleh Menteri Kehakiman RI dengan surat keputusan No. CZ-8338 IH 01.01. pada tahun 1996 dan Pabrik Gula Meritjan merupakan sebuah Strategi Bisnis Unit (SBU) Gula. [5]

Pada tanggal 2 Oktober 2014, menteri BUMN Dahlan Iskan meresmikan Holding BUMN Perkebunan yang berangotakan PTPN I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIV dengan PTPN III sebagai induk holding BUMN Perkebunan. Dasar hukum perubahan PTPN X (Persero) menjadi PTPN X adalah Keputusan Para Pemegang Saham Perusahaan Perseroan PT Perkebunan Nusantara X Nomor: PTPN X/RUPS/01/X/2014 dan Nomor SK-57/D1.MBU/10/2014 tentang Perubahan Anggaran Dasar. [5].

2.2.2 Pengelolaan Hubungan Supplier

Pengelolaan hubungan supplier (SRM) adalah pendekatan yang digunakan untuk berhubungan dengan supplier pada tingkat yang mencerminkan prioritas perusahaan dan bagaimana kebutuhan perusahaan dapat dicapai. SRM merupakan proses diferensiasi yang mengetahui bahwa tidak semua supplier sama sehingga strategi untuk setiap supplier akan berbeda. SRM berfokus dalam membangun nilai bersama dengan supplier berdasarkan kepercayaan, komunikasi dan kolaborasi. [6] Produk dan servis yang diberikan oleh pihak supplier akan sangat mempengaruhi bagaimana perusahaan melihat dan berinteraksi dengan supplier tersebut. [7]

Memanfaatkan kemampuan dari supplier adalah tujuan utama dari proses SRM. Hubungan yang baik dengan supplier dapat meningkatkan keunggulan kompetitif perusahaan, dimana perusahaan dapat melakukan *outsource* atau mendapatkan saran dari supplier mengenai pembuatan produk baru, cara terbaik dalam melakukan produksi dan lain-lain. Mendapatkan pengetahuan, sumber, kemampuan dan ide adalah kunci dari proses SRM. Tujuan kedua dari SRM adalah mengurangi biaya. SRM membantu perusahaan dan supplier menyadari dengan melakukan kerjasama kedua belah pihak akan sama-sama mendapatkan keuntungan. Ketersediaan bahan baku adalah tujuan ketiga dari SRM. Semakin maraknya kompetisi dan permintaan pelanggan dekade ini membuat perusahaan sulit meyakinkan pelanggan mengenai ketersediaan. Hubungan dengan supplier yang baik dapat mengurangi ketidakpastian akan bahan baku dan menghindari adanya permintaan khusus saat terjadi fluktuasi jumlah bahan baku melalui peramalan dan perencanaan yang terintegrasi. [6]

SRM memiliki beberapa aktivitas, pada penelitian ini yang dibahas adalah segmentasi supplier. Segmentasi supplier adalah proses membedakan supplier untuk menentukan tingkat hubungan dengan supplier tertentu. Segmentasi supplier juga membantu perusahaan untuk menentukan sumber daya yang digunakan untuk mengelola supplier pada tingkat yang berbeda.

Penentuan sumber daya sangat penting sehingga supplier bekerja sesuai dengan tingkat segmentasi yang ditentukan. [8]

Segmentasi supplier didorong oleh tujuan bisnis, sehingga evaluasi terhadap supplier seharusnya berfokus pada kemampuan supplier mencapai tujuan tersebut sehingga menghasilkan nilai. Untuk melakukan segmentasi supplier terdapat 3 proses yang harus dilakukan yaitu [9] :

1. Merancang segmentasi supplier berdasarkan tujuan bisnis
2. Melakukan segmentasi supplier berdasarkan kemampuannya yang didasarkan tujuan bisnis
3. Membuat model interaksi dan rencana kegiatan

Dalam melakukan segmentasi supplier, perusahaan harus melihat kedua arah, yaitu *outward* (keluar) dan *inward* (kedalam). Bagian *outward* (keluar) berisikan evaluasi analisis mengenai kapabilitas supplier. Metode statistik dapat digunakan untuk mengidentifikasi kelompok kapabilitas yang berbeda pada basis pesedian perusahaan. Akan tetapi untuk penilaian yang lebih baik supplier juga harus melihat bagian *inward*, yaitu prioritas kemampuan yang dibutuhkan untuk menggapai tujuan bisnis. [9]

Pada segmentasi supplier, evaluasi secara analisis merupakan pendekatan bottom-up, sementara prioritas strategis atau bagian inward merupakan pendekatan top-down. Sangat disarankan untuk menggabungkan kedua pendekatan tersebut untuk mendapatkan manfaat dari keduanya. Penggabungan ini disebut dengan pendekatan *hybrid*. Pendekatan *bottom-up* akan melakukan *clustering* untuk mengetahui supplier yang memiliki karakteristik yang sama. Sementara pendekatan *top-down* melakukan *clustering* untuk menetapkan segmen mana yang paling penting. [9]

Hasil dari segmentasi merupakan sekumpulan rencana kegiatan. Rencana kegiatan bagi supplier yang memiliki nilai yang tinggi akan memasukan rencana untuk meningkatkan

kolaborasi kerjasama. Untuk supplier yang memiliki potensi, rencana kegiatan akan memasukan rencana untuk mendorong supplier untuk mencapai potensinya. Bagi supplier lainnya, rencana kegiatannya dapat memasukan pemutusan hubungan dengan supplier. [9]

Segmentasi supplier memiliki beberapa manfaat bagi perusahaan, yang pertama adalah meyakinkan alokasi waktu dan sumber daya yang terbatas secara optimal. Kemudian segmentasi juga dapat membantu menentukan hubungan seperti apa yang harus dibentuk untuk berbagai macam supplier. Ketiga adalah membantu mengelola dan mengklarifikasi ekspektasi supplier. Selain itu segmentasi juga dapat memandu membuat strategi negosiasi serta struktur pengelolaan hubungan. [10].

2.2.3 Penggalian Data

Penggalian data adalah proses menemukan pengetahuan yang menarik, seperti pola, hubungan, perubahan, keanehan dan struktur tertentu dari data besar yang disimpan di basis data, data warehouse atau penyimpanan informasi lainnya. Karena banyaknya data yang berbentuk elektronik dan kebutuhan pengubahan data tersebut menjadi informasi dan pengetahuan, penggalian data menjadi salah satu solusi yang dilirik di dunia informasi akhir-akhir ini. [11]

Penggalian data terkenal sebagai persamaan dari *knowledge discovery*, walaupun penggalian data, menurut peneliti adalah langkah untuk *knowledge discovery*. Secara umum, proses *knowledge discovery* adalah sebagai berikut [11]:

1. *Data cleaning*, dimana mengatasi data pencilan, keliru, hilang dan yang tidak berhubungan
2. *Data integration*, dimana data dari berbagai sumber dikumpulkan dan diintegrasikan menjadi satu.
3. *Data selection*, proses dimana data yang berhubungan dengan tugas analisis diambil dari basis data
4. *Data transformation*, proses dimana data diubah atau digabung menjadi bentuk yang sesuai untuk proses

penggalian data dengan melakukan operasi agregasi dan perangkuman

5. *Penggalian data*, proses ini merupakan proses utama dari *knowledge discovery*, dimana metode diterapkan pada data untuk menghasilkan pola data
6. *Pattern Evaluation*, proses untuk mengidentifikasi pola yang menarik yang merepresentasikan pengetahuan berdasarkan *interestingness measure*.
7. *Knowledge presentation*, proses dimana teknik visualisasi dan representasi pengetahuan digunakan untuk menyampaikan pengetahuan pada user.

Dengan adanya *relation database system* dan *data warehouse*, keempat proses pertama: *data cleaning*, *data integration*, *data selection* dan *data transformation* dapat dilakukan dengan membangun *data warehouse* dan melakukan operasi OLAP pada *data warehouse*. Proses lainnya, yaitu *penggalian data*, *pattern evaluation* dan *knowledge presentation* terkadang digabung dalam satu proses iterative yang disebut *penggalian data*. [11]

Tugas penggalian data dapat dikategorikan menjadi dua bagian yaitu *descriptive* dan *predictive*. Penggalian data *descriptive* menjelaskan data set dengan ringkas dan menyampaikan sifat umum dari data. Penggalian data *predictive* membuat satu atau sekumpulan model yang kemudian diambil dugaan mengenai data set yang ada. Lalu penggalian data *predictive* akan memprediksi karakteristik data set terbaru. [11]

Penggalian data dapat membantu tugas penggalian data sebagai berikut :

1. *Class description*

Class description menyediakan ringkasan yang jelas mengenai kumpulan data dan membedakannya dengan yang lain. Ringkasan tersebut disebut dengan *class characterization*. *Class description* menjelaskan tentang total, rata-rata, kuartil dan lain-lain. [11]

2. *Association*

Asosiasi adalah pencarian hubungan asosiasi atau korelasi diantara item. Asosiasi dinyatakan dengan peraturan untuk menunjukkan kondisi variabel yang muncul bersamaan di kumpulan data. [11]

3. *Classification*

Klasifikasi menganalisis sejumlah *training data* (objek yang kelasnya telah diketahui) dan membuat model untuk setiap kelasnya berdasarkan ciri-ciri dari data tersebut. *Decision tree* dibuat berdasarkan proses klasifikasi yang digunakan untuk memahami kelas pada basis data dan untuk klasifikasi data di masa depan. [11]

4. *Prediction*

Fungsi ini digunakan untuk memprediksi nilai dari data yang hilang atau nilai distribusi dari variabel tertentu pada sekumpulan objek. Prediksi mencari variabel yang berhubungan dengan variabel yang dicari dan memperkirakan nilai distribusi berdasarkan sekumpulan data yang sama dengan objek yang dipilih.

5. *Clustering*

Analisis *clustering* digunakan untuk mengidentifikasi *cluster* yang terdapat dalam data, dimana *cluster* adalah sekumpulan objek data yang sama antara satu sama lain. Kesamaan dapat dinyatakan dengan *distance function*, yang dispesifikasikan oleh user atau ahli. Metode *clustering* yang baik menghasilkan *cluster* yang berkualitas tinggi dimana kesamaannya tinggi.

6. *Time-series analysis*

Analisis *time series* digunakan untuk menganalisa data *time-series* yang besar untuk mencari keberaturan dan karakteristik yang menarik, termasuk mencari urutan yang sama dan mencari pola urutan, deviasi dan tren, dan perioditas.

Penggalian data digunakan pada penelitian ini untuk membantu pengolahan data supplier dari PTPN X Meritjan. Tugas penggalian data pada penelitian ini termasuk pada *descriptive* penggalian data karena penggalian data digunakan untuk menjelaskan data yang sudah ada, tidak digunakan untuk memprediksi data di masa depan. *Clustering* adalah kategori tugas penggalian data yang dipilih pada penelitian ini.

2.2.4 Analisis RFM

Analisis RFM merupakan sebuah analisis yang dapat membantu membedakan pihak yang berpengaruh dari data yang besar berdasarkan tiga variabel yaitu interval satu transaksi dengan saat ini, frekuensi dan jumlah uang. Definisi dari RFM adalah sebagai berikut [2]:

(1.) *Recency (R)*

Rentan dari satu transaksi dengan saat ini dilambangkan dengan R atau yang disebut dengan *recency*. Semakin pendek intervalnya, maka semakin besar nilai R.

(2.) *Frequency (F)*

F merepresentasikan frekuensi, yaitu jumlah transaksi dalam periode tertentu pada periode tertentu, contohnya dua kali dalam satu tahun atau dua kali dalam satu bulan. Semakin tinggi frekuensinya semakin besar nilai F.

(3.) *Monetary Value (M)*

M merepresentasikan *monetary*, yaitu nilai produk dalam bentuk uang pada periode tertentu. Semakin banyak jumlah uang pada periode tersebut maka semakin tinggi nilai M.

Berdasarkan penelitian yang dilakukan oleh Wu dan Lin [2], semakin tinggi nilai R dan nilai F, semakin tinggi juga pihak tersebut melakukan transaksi kembali dengan perusahaan. Terlebih lagi semakin tinggi M, semakin mungkin pihak

tersebut memberikan atau membeli produk atau servis dengan perusahaan lagi.

Dalam pemberian bobot pada variabel RFM terdapat dua opini yang berbeda. Menurut Hughes [12] , ketiga variabel RFM memiliki kepentingan yang sama, sehingga bobot ketiganya sama. Sementara menurut Stone [13], ketiga variabel memiliki kepentingan yang berbeda tergantung dengan karakteristik perusahaan. Sehingga, bobot ketiga variabelnya sehingga tidak setara. Baik setara atau tidak, pemberian bobot disesuaikan dengan kebutuhan perusahaan. [2]

RFM model memiliki beberapa kelebihan, yang pertama, RFM tidak mengeluarkan biaya yang banyak untuk mendapatkan analisis perilaku dan mudah untuk mengukur perilaku supplier. Kedua, RFM dapat memprediksi keuntungan perusahaan dalam jangka pendek. Kelebihan yang ketiga adalah sangat mudah untuk membuat model dengan variabel RFM dengan menggunakan jumlah variabel yang sedikit. Terakhir, RFM adalah metode yang digunakan untuk mengukur kekuatan dari hubungan dengan supplier dan juga RFM dapat mengidentifikasi supplier yang berharga.

Selain kelebihan, RFM juga memiliki kekurangan. Kekurangan yang pertama adalah RFM analisis hanya berfokus pada supplier terbaik. Selain itu, kekurangan lainnya adalah RFM hanya dapat menganalisis supplier saat ini saja tidak dapat diaplikasikan pada supplier baru dikarenakan data transaksinya belum ada. [14]

2.2.5 Transformasi Data

Transformasi data biasanya dilakukan untuk membuat distribusi data menjadi normal. Alasan lainnya adalah untuk membuat grafik data lebih informatif, identifikasi *outlier* yang lebih baik dan meningkatkan sensitivitas dari statiscal test. [15]

Transformasi terkait dengan mengaplikasikan fungsi matematika pada setiap data. Transformasi dibutuhkan apabila data benar-benar condong. Pada Tabel 2.5 dijelaskan tipe

transformasi apa yang dibutuhkan sesuai dengan kecondongannya. [16]

Tabel 2.5 Distribusi Data dan Transformasi Data yang Tepat

Data Distribution	Transformation Method	Formula
Moderately positive skewness	Square Root	$NEWX = \sqrt{X}$
Substantially positive skewness	Logarithmic (Log 10)	$NEWX = LG10(X)$
Substantially positive skewness (with zero values)	Logarithmic (Log 10)	$NEWX = LG10(X + C)$
Moderately negative skewness	Square-Root	$NEWX = \sqrt{K - X}$
Substantially negative skewness	Logarithmic (Log 10)	$NEWX = LG10(K - X)$
C = a constant added to each score so that the smallest score is 1.		
K = a constant from which		

2.2.6 Analytic Hierarchial Process (AHP)

Analytic Hierarchy Process merupakan pendekatan pembuatan pilihan dengan berbagai macam kriteria yang dikenalkan oleh Saaty pada tahun 1977. AHP merupakan *decision support tool* yang dapat digunakan untuk menyelesaikan permasalahan keputusan yang kompleks. Metode ini menggunakan tujuan, kriteria, sub-kriteria dan alternatif yang memiliki struktur *multi-level hierarchical*. Data yang bersangkutan didapatkan menggunakan *pairwise comparison*. Perbandingan ini akan mendapatkan bobot kepentingan dari kriteria pemilihan dan pengukuran performa relatif pada kriteria pemilihan individual. Bila perbandingan tidak benar-benar konsisten, AHP akan memberikan mekanisme untuk meningkatkan konsistensi [17]. Berikut adalah tahapan penggunaan AHP [18] :

1. Melakukan *pairwise comparison*

Pembuat keputusan membuat *pairwise comparison* dari kepentingan variabel RFM menggunakan skala pada Tabel 2.6:

Tabel 2.6 *Pairwise comparison*

Skala	Definisi	Keterangan
1	Kedua kriteria sama penting	Kedua kriteria memiliki pengaruh yang sama
3	Kriteria yang satu sedikit lebih penting	Penilaian sedikit memihak pada satu kriteria dibandingkan dengan kriteria pasangannya
5	Kriteria yang satu lebih penting	Penilaian sedikit sangat memihak pada satu kriteria dibandingkan dengan kriteria pasangannya
7	Kriteria yang satu jelas lebih penting	Salah satu kriteria sangat berpengaruh dan dominasinya tampak secara nyata
9	Kriteria yang satu mutlak jauh lebih penting	Salah satu kriteria sangat jelas lebih penting dibandingkan dengan kriteria pasangannya

2. Penilaian konsistensi dari penilaian *pairwise*

Kemudian lakukan derajat ketidak konsistenan menggunakan index ketidak konsistenan. Bila pembuat keputusan tersebut konsisten, maka hasil yang didapatkan adalah nilai 0 pada index ketidak konsistenan. Namun, hal tersebut jarang tercapai karena manusia tidak konsisten dan berat sebelah. Sehingga bila index mencapai angka 0.1 telah dianggap baik.

3. Perhitungan bobot relatif

Pada tahap ini akan dilakukan penentuan bobot pada setiap elemen keputusan. Data didapatkan berdasarkan hasil wawancara dengan penguji menggunakan kuisison dan hasilnya akan ditampilkan dalam bentuk matriks *pairwise comparison*.

Kuisisioner yang telah diisi oleh penguji akan menggunakan metodologi AHP dihitung dengan tahap dibawah ini [19]:

- Pertama-tama hitunglah jumlah nilai pada setiap kolom dalam matriks *pairwise comparison*
- Kemudian, bagi setiap nilai dalam kolom pada setiap jumlah nilai perkolom, yang menghasilkan matriks normalisasi
- Selanjutnya hitung rata-rata nilai pada setiap baris. Pada tahap in dilakukan perubahan nilai dari pecahan ke nilai decimal
- Baris rata-rata menampilkan preferensi untuk setiap pilihan. Preferensi ini dituliskan dalam matriks satu kolom yang dinamakan dengan *vector*. Hitung *vector* untuk pada setiap pilihan
- Setelah itu gabungkan *vector* menjadi satu *single preference matrix*
- Hitung nilai bobot pada setiap variabel/pilihan dengan rumus:

$$W = \frac{X_1 + \dots + X_n}{n} \quad (2.1) \quad [19]$$

n = jumlah pilihan/variabel

- Urutkan setiap pilihan sesuai dengan nilai yang dihasilkan
- Hitunglah nilai *Consistency Index* (C_i) untuk melihat apakah hasil tersebut konsisten, dengan rumus sebagai berikut :

$$C_i = \frac{x - n}{n - 1} \quad (2.2) \quad [20]$$

n = jumlah variabel/pilihan x = rata-rata bobot sebelumnya

- Lalu carilah nilai R_1 (*Random Consistency Index*) dengan dengan melihat tabel R_1 sesuai dengan jumlah variabel yang dibandingkan seperti yang terlihat pada Tabel 2.7

Tabel 2.7 Random Consistency Index

n	2	3	4	5	6	7	8	9	10
RI	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.51

- Hitunglah *consistency ratio*, apabila *consistency ratio* bernilai baik bila nilainya < 0.10 dan apabila > 0.10 maka hasilnya tidak konsisten. Berikut adalah rumus C_r :

$$C_r = \frac{C_i}{R_i} \quad (2.3) \quad [20]$$

$C_i =$ Consistency Index

$R_i =$ Random Consistency Index

2.2.7 Clustering

Clustering adalah analisis untuk mengidentifikasi *cluster* yang ada pada data, dimana *cluster* adalah koleksi objek data yang mirip satu dengan lainnya [11]. *Clustering* digunakan untuk menata data pada struktur yang bermanfaat, atau dengan kata lain disebut dengan taksonomi. *Clustering* adalah salah satu alat analisis yang bertujuan untuk melakukan penyortiran objek yang berbeda dalam suatu kelompok dimana derajat asosiasinya antar dua objek akan maksimal bila kedua objek tersebut dalam kelompok yang sama, bila tidak maka nilainya akan minimal. *Cluster* juga dapat digunakan untuk mencari struktur data tanpa memberikan penjelasan [21].

Clustering mengelompokkan data pada subset yang memiliki karakteristik yang sama dan apabila berbeda akan dimasukkan pada kelompok lain. Data tersebut disusun pada representasi karakteristik dari sampel tersebut. Struktur dari *cluster* direpresentasikan sebagai kelompok dari subset $C = C_1, \dots, C_k$ dari S , sehingga $S = \bigcup_{i=1}^k C_i$ dan $C_i \cap C_j = \emptyset$ untuk $i \neq j$, sehingga data pada S hanya masuk benar-benar masuk dalam satu subset. [22]

Menurut Farley dan Raftery [22], *clustering* dibagi menjadi 2 kelompok besar, yaitu metode *hierarchical* dan *partitioning*. Metode *hierarchical* adalah sebuah *cluster* bertingkat yang

disusun sebagai sebuah tree (kecuali *leaf node*). Sementara *clustering partitional* adalah pembagian dari sekumpulan data objek pada sub set (*cluster*) yang tidak tumpang tindih sehingga setiap data objek berada pada satu subset. [23]

Pada *Clustering Hierarhical*, setiap node (*cluster*) dalam *tree* adalah gabungan dari anakannya atau *subcluster* dan akar dari *tree* adalah *cluster* yang memuat semua objek. Seringkali, *leaves* dari *tree* adalah *cluster* tunggal dari sebuah data objek. *Clustering hierarhical* dapat dikatakan sebagai urutan *clustering partitional* dan *clustering partitional* dapat diraih dengan memotong *hierarchical tree* pada tingkat tertentu. [23]

Clustering memiliki empat tahapan, yaitu membuat desain penelitian, menurunkan *cluster*, interpretasi *cluster* dan yang terakhir adalah uji performa *cluster*. Berikut adalah penjelasan tiap tahapannya [24] :

1. Membuat Desain Penelitian

Pembuatan desain penelitian bertujuan untuk mempermudah proses *clustering* dengan mendeteksi pencilan, standarisasi data dan pengukuran kesamaan objek.

Pencilan adalah nilai dari sekumpulan data yang berbeda dibanding biasanya dan juga tidak merepresentasikan karakteristik data tersebut [25]. Adanya pencilan dapat mengubah struktur sebenarnya dari data dan hasil yang didapatkan tidak sesuai dengan struktur sebenarnya.

Standarisasi dilakukan agar perhtungan *cluster* lebih mudah, sehingga menghilangkan kemungkinan perbedaan variabel. Perbedaan variabel sangat berpengaruh pada penyebaran nilai dan berdampak pada hasil akhirnya. Standarisasi terdapat dua jenis yaitu standarisasi variabel dan standarisasi berdasarkan observasi.

2. Menurunkan *Cluster*

Dalam tahap ini, peneliti diharuskan menentukan metode apa yang akan digunakan dan memutuskan jumlah *cluster* yang akan dibentuk. Metode dan banyaknya *cluster* akan berdampak pada hasil dan interpretasi hasil *clustering*.

Algoritma *clustering* terbagi menjadi dua yang terdiri dari *hierarchical* dan *partitioning*. *Cluster partitioning* memasukan objek dalam *cluster* tertentu sehingga menghasilkan beberapa jumlah *cluster*. *Cluster hierarhichal* menganggap bahwa tiap objek adalah *cluster* sendiri. Kemudian obyek memiliki kesamaan digabungkan menjadi *cluster* baru dan proses ini berlanjut hingga tidak ada kesamaan lagi antar *cluster* (tunggal).

3. Proses *Cluster*

Pada tahap ini akan dilakukan proses *clustering* menggunakan metode yang telah ditentukan. Setelah itu akan dianalisis hasil *cluster* untuk melakukan pemberian nama.

4. Uji Performa *Cluster*

Interpretasi *cluster* bersifat subyektif sehingga dibutuhkan uji performa dari hasil *clustering*. Uji performa dilakukan untuk mencari tahu apakah hasil telah merepresentasikan populasi secara umum. Uji performa dapat dilakukan dengan pendekatan langsung dengan cara membandingkan hasil *cluster* dengan asumsi *cluster* yang telah ada. Hal ini dapat dilakukan membagi sampel menjadi dua bagian kemudian dibandingkan. [24]

2.2.8 K-Means

K-Means adalah teknik *clustering* berdasarkan prototip yang membuat pemisahan tingkat satu pada data objek. K-means mengartikan prototip sebagai centroid, yaitu nilai rata-rata dari kelompok nilai dan dapat diaplikasikan pada objek di ruang dimensi-n yang berkelanjutan.

Teknik K-Means sangatlah simpel, pertama-tama tentukan K centroid awal, dimana K adalah parameter yang ditentukan oleh peneliti. Setiap poin ke ditetapkan pada centroid terdekat dan setiap kumpulan poin ditentukan pada centroid adalah *cluster*. Kemudian hal ini dilakukan kembali dan memperbarui langkah hingga *cluster* tidak berubah atau hingga centroidnya tetap. Tabel 2.8 adalah tahapan dari algoritma K-Means. [23]

Tabel 2.8 Algoritma K-Means

Algoritma : Algoritma K-Means Dasar	
1	Tentukan poin K sebagai centroid awal
2	Ulangi
3	Buatlah <i>cluster</i> K dengan menempatkan setiap point pada centroid terdekat
4	Hitung ulang centroid setiap <i>cluster</i>
5	Hingga centroid tidak berganti

K-Means dapat digunakan diberbagai macam tipe data. K-Means juga sangat efisien walaupun terkadang dijalankan beberapa kali. Selain itu, metode ini mudah diinterpretasikan, mudah diimplementasikan dan dapat beradaptasi pada data yang tersebar. [22] Namun, K-Means tidak dapat mengatasi *cluster* yang memiliki ukuran dan kepadatan yang berbeda. K-Means juga memiliki kesulitan untuk mengelompokan data yang mengandung pencilan. Pendeteksian dan penghapusan pencilan akan membantu proses K-Means. [23]

Metode *clustering* K-Means digunakan dalam penelitian ini karena hasil yang diharapkan dari *clustering* adalah kelompok-kelompok supplier yang objek didalamnya benar-benar hanya milik kelompok tersebut. Untuk menghindari hasil yang buruk dari *clustering* K-Means, data sebelumnya telah melakukan praproses data untuk menghilangkan pencilan pada data. K-Means digunakan dalam penelitian ini juga karena metode ini mudah diimplementasikan dan proses perhitungan termasuk cepat.

2.2.9 Euclidian Distance

Euclidian distance merupakan jarak garis lurus antara dua data. Euclidean distance menghitung akar perbedaan antara objek yang berpasangan. Jarak ini digunakan dalam proses k-mean clustering untuk menentukan anggota sebuah kelompok cluster. [26]

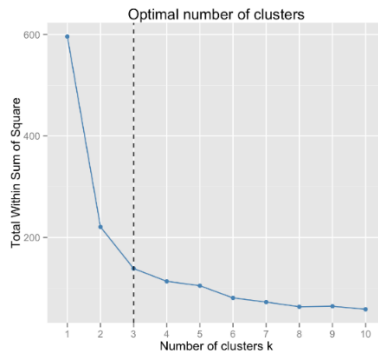
$$\text{dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (2.4) \quad [26]$$

$x, y = \text{titik satu data}$ $a, b = \text{titik data lainnya}$

Keuntungan penggunaan metode ini adalah jarak antara dua objek tidak akan terganggu dengan adanya objek baru yang bisa jadi merupakan *outlier*. Akan tetapi jaraknya bisa sangat terpengaruhi dengan perbedaan skala antar dimensi dimana jarak dihitung. [27]

2.2.10 Metode Elbow

Metode elbow adalah metode interpretasi dan uji performa konsistensi dalam analisis *cluster* untuk menemukan jumlah yang tepat dari *cluster*. Metode ini adalah metode yang melihat SSE sebagai fungsi dari jumlah *cluster*. Nilai SSE dijelaskan oleh *cluster* yang diplotkan berdasarkan jumlah *cluster*. *Cluster* pertama akan memberikan informasi namun pada titik tertentu kenaikan marjinal akan turun secara dramatis dan memberikan sebuah lekukan pada grafik yang disebut dengan “kriteria siku”.



Gambar 2.1 Elbow method

Nilai tersebutlah yang menjadi nilai k atau jumlah *cluster* yang dipilih. Metode ini berawal dengan menentukan nilai k sebesar 2, kemudian akan ditingkatkan satu pada tiap langkahnya. Kemudian akan dihitung *clustering* dan SSE yang dihasilkan. Pada nilai k tertentu, SSE akan turun secara drastis dan setelah itu akan mencapai titik konstan dimana nilai tidak akan berubah walaupun nilai k dinaikkan. Tabel 2.9 adalah algoritma metode *elbow* untuk menentukan nilai k pada k -means. [28]

Tabel 2.9 Algoritma Elbow method

Algoritma : Metode elbow untuk menentukan nilai K dari K-means	
1.	Menginisiasi $k=1$
2.	Mulai
3.	Tambahkan nilai dari k
4.	Hitung SSE
5.	Bila pada poin tertentu SSE dari solusi turun secara drastis
6.	Berikut adalah nilai k
7.	Selesai

2.2.11 Uji Performa

Uji performa *clustering* adalah untuk mencari kelompok *cluster* yang cocok dengan jumlah *cluster* tanpa informasi kelas. Pada uji performa *clustering* terdapat dua tipe uji performa *clustering*. Tipe pertama adalah uji performa eksternal, tipe ini menggunakan informasi terdahulu mengenai data [29]. Tipe ini digunakan untuk mengetahui algoritma terbaik untuk proses *clustering* sesuai dengan *data set* tertentu. Sementara, tipe data yang kedua adalah uji performa internal yang hanya menggunakan data yang ada. Uji performa ini digunakan untuk mengevaluasi seberapa bagus struktur *cluster*. Pada kenyataannya, data eksternal untuk uji performa eksternal tidak ada, sehingga seringkali uji performa internal digunakan.

Uji performa internal diukur berdasarkan dua kriteria. Pertama adalah kepadatan yang mengukur seberapa dekat objek yang berhubungan dalam satu *cluster*. Kedua adalah pemisahan yang mengukur seberapa bedanya antara satu *cluster* dengan *cluster*

lainnya [30] . Uji performa internal dapat menggunakan SSE dan uji performa lainnya.

2.2.11.1 Sum Squared Error

Sum Squared Error adalah kriteria termudah untuk mengukur *clustering*. SSE adalah salah satu teknik *statistic* yang digunakan untuk mencari apakah obyek cocok pada satu *cluster*. Perhitungan SSE adalah sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (2.5) \quad [22]$$

$C_k = \text{kelompok data dalam cluster } k$ μ_k
 $= \text{rata - rata cluster } k$

Bila obyek sangat cocok dengan *cluster* tersebut maka nilai SSE adalah nol atau yang berarti tidak ada error atau sangat cocok. Namun hal itu jarang terjadi, Oleh karena itu, *clustering* yang baik adalah yang memiliki nilai SSE yang serendah mungkin. Semakin rendah nilai SSE maka semakin sama. SSE yang tinggi maka memiliki derajat perbedaan antara obyek dan *cluster* yang dituju [9]. SSE digunakan untuk uji performa K-means, karena perhitungan K-means menggunakan *Euclidean distance* untuk mengalokasikan objek pada *cluster*. Pada setiap kelompok, asumsinya setiap data objek yang berada di satu kelompok memiliki SSE yang sama. [31]

2.2.11.2 Dunn Index

Dunn index merupakan rasion jarak terkecil antara data observasi yang ada di dalam satu *cluster* dengan jarak *cluster* dengan *cluster* lainnya. Nilai dunn index semakin tinggi semakin baik. [32]

$$Dunn = \min_{1 \leq i \leq c} \left\{ \min \left\{ \frac{d(c_i, c_j)}{\max_{1 \leq k \leq c} (d(X_k))} \right\} \right\} \quad (2.6) \quad [29]$$

$d(c_i, c_j) = \text{intercluster distance}$ $(d(X_k)) = \text{intracluster distance}$

2.2.11.3 Silhouette Width

Silhouette width adalah rata-rata dari nilai Silhouette pada setiap observasi. Nilai Silhouette mengukur derajat kepercayaan pada penempatan *clustering*, apabila nilainya mendekati 1 maka penempatan *cluster* sudah tepat dan apabila mendekati -1 maka penempatannya buruk. [32]

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (2.7) \quad [32]$$

A_i = rata-rata jarak i dengan observasi lain dalam satu *cluster*
 B_i = rata-rata jarak antara i dengan observasi pada *cluster* terdekat

2.2.11.4 Connectivity

Connectivity menunjukkan derajat ketersamungan dari *cluster*. Bila nilainya sama maka akan menghasilkan nilai 0 dan bila berbeda nilainya 1. Hasil dari perhitungan *connectivity* memiliki nilai antara 0 dan ∞ dan harus diminimalkan [32]

$$Conn = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (2.8) \quad [32]$$

$X_{i,nn_{i(j)}}$ = nilai 0 atau 1 apakah satu *cluster* atau tidak
 j = observasi *cluster* lain terdekat

2.2.12 Normalisasi Min-Max

Normalisasi merupakan proses yang digunakan untuk standarisasi semua variabel dataset dan memberikan bobot yang sama [1]. Normalisasi dapat mengeliminasi data yang redundan dan pencilan, sehingga data yang dihasilkan valid, dapat diandalkan dan meningkatkan akurasi hasil. Metode normalisasi yang digunakan tergantung dengan data yang akan dinormalisasi, pada penelitian ini metode yang digunakan adalah min-max.

Normalisasi min-max adalah proses perubahan data yang diukur menjadi nilai anatara 0.0 dan 1.0. Dimana nilai terendah adalah 0.0 dan tertinggi adalah 1.0. Normalisasi min-max akan

memperudahkan perbandingan nilai yang memiliki satuan ukuran yang berbeda. Berikut adalah rumus normalisasi min-max :

$$MM(X_{ij}) = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}} \quad (2.9) \quad [33]$$

$X_{min} = \text{nilai } x \text{ terkecil} \quad X_{max} = \text{nilai } x \text{ terbesar}$

2.2.13 Customer Lifetime Value

Customer lifetime value adalah idea dimana nilai ekonomi yang diambil dari hubungan perusahaan dengan pelanggannya. CLV digunakan untuk mengidentifikasi pelanggan yang menguntungkan dan untuk membangun strategi untuk menargetkan pelanggan.

Untuk menghitung nilai CLV, terlebih dahulu harus diukur RFM. [34] RFM digunakan untuk menghindari pelanggan yang kurang menguntungkan dan menggunakan sumber daya pada pelanggan yang lebih menguntungkan [35]. Berdasarkan Stone, setiap variabel RFM diberikan bobot yang berbeda sesuai dengan karakteristik perusahaan atau yang biasa disebut dengan *weighted RFM*. [13]

$$CLV = NR \times WR + NF \times WF + NM \times WM \quad (2.10) \quad [36]$$

N : normalisasi R: *recency* F: *frequency* M: *Monetary* W: *Weight*
 Pada penelitian ini objek yang akan diteliti adalah supplier. Perhitungan CLV akan diadopsi untuk mengetahui peringkat supplier sesuai dengan *clusternya*.

2.2.14 Uji Korelasi

Uji korelasi adalah teknik yang digunakan untuk menganalisis hubungan antara dua variabel atau lebih. Variabel saling berkorelasi apabila perubahan satu variabel menyebabkan perubahan variabel lainnya. Korelasi terdiri dua tipe yaitu korelasi negatif dan positif. Korelasi positif adalah hubungan antara dua variabel searah. Sementara, korelasi negatif dimana kedua variabel saling berkontradiksi [37]. Tabel 2.10 adalah tabel interpretasi nilai uji korelasi, apakah variabel tersebut saling berkorelasi apa tidak [38].

Tabel 2.10 Tabel Interpretasi Uji Korelasi

Size of Correlation	Interpretation
0.9 to 1.0 (-0.9 to -1.0)	<i>Very high positive (negative) correlation</i>
0.7 to 0.9 (-0.9 to -1.0)	<i>High positive (negative) correlation</i>
0.5 to 0.7 (-0.5 to -0.7)	<i>Moderate positive (negative) correlation</i>
0.3 to 0.5 (-0.3 to -0.5)	<i>Low positive (negative) correlation</i>
0.0 to 0.3 (-0.0 to -0.3)	<i>Negligible correlation</i>

Ada tiga jenis uji korelasi, yaitu pearson, spearman dan kendall. Pearson digunakan apabila data adalah linear dan memiliki distribusi yang normal [39]. Sementara, kendall dan spearman digunakan apabila datanya bersifat *monotonic*. [37].

Akan tetapi menurut Kendall dan Gibson, hasil dari spearman kurang handal dibandingkan kendall. Selain itu spearman tidak bisa menghitung apabila ranking antara keduanya sama. [40] Rumus pearson adalah [41]:

$$r = \frac{SS_{XY}}{\sqrt{(SS_{XX})(SS_{YY})}} \quad (2.11) \quad [42]$$

SS_{XY} : sum crossed product SS_{XX} : sum of squares for variabel X
 SS_{YY} : sum of square for variabel Y

2.2.15 Linear Regresi

Linear regresi digunakan untuk mengetahui analisis mengenai hubungan antara dua kejadian, kegiatan atau masalah. Selain itu juga, digunakan untuk memprediksi nilai masa depan. Apabila nilai X dan Y memiliki hubungan maka, nilai X dapat digunakan untuk memprediksi nilai Y.

Nilai X yang digunakan dalam perhitungan ini adalah variabel bebas atau variabel peramal. Nilai X digunakan untuk meramalkan nilai Y. Nilai Y adalah nilai yang akan diramalkan dan nama lainnya adalah variabel dependen. Terdapat dua ukuran dalam regresi yaitu koefisien regresi yang mengukur besarnya pengaruh X terhadap Y. Kedua, adalah koefisien korelasi yaitu koefisien yang mengukur kuatnya hubungan X dan Y. [43]

$$Y = a + b(x) \quad (2.12) \quad [43]$$

y : variabel terikat x : variabel bebas a: *intercept*/konstanta
b : koefisien regresi /*slope*

2.2.16 Tukey Method

Tukey Method adalah sebuah metode yang digunakan untuk mendeteksi *outlier*. Metode ini akan menghasilkan sebuah grafik yaitu boxplot yang menampilkan media, Q1, Q3, nilai minimum dan maximum. Tukey method cocok untuk data yang besar dan tidak bergantung dengan distribusi datanya. Berikut adalah aturan untuk metode ini [44] :

1. IQR (*Inter Quartile Range*) adalah jarak antara kuartil bawah (Q1) dan atas (Q3)
2. *Inner fences* diletakan 1.5 kali IQR bawah Q1 dan diatas Q3
3. *Outer fences* diletakan 3 kali IQR bawah Q1 dan diatas Q3
4. Semua yang tercakup pada *inner* dan *outer fences* merupakan *outlier* dan *outlier* juga mungkin berada di atas batas *outer fences*

BAB III

METODOLOGI TUGAS AKHIR

Bab ini akan membahas tahapan yang dilakukan penulis dalam pengerjaan tugas akhir dengan penjelasannya dan juga jadwal pengerjaan tugas akhir. Hal ini digunakan agar pengerjaan tugas akhir terakhir dan sistematis.

3.1 Tahapan Pelaksanaan Tugas Akhir

Pada bagian ini akan dijelaskan mengenai proses pengerjaan maupun data dan metode yang digunakan dalam pengerjaan dari tugas akhir ini. Tahapan pelaksanaan tugas akhir digambarkan seperti Gambar 3.1

3.2 Uraian Metodologi

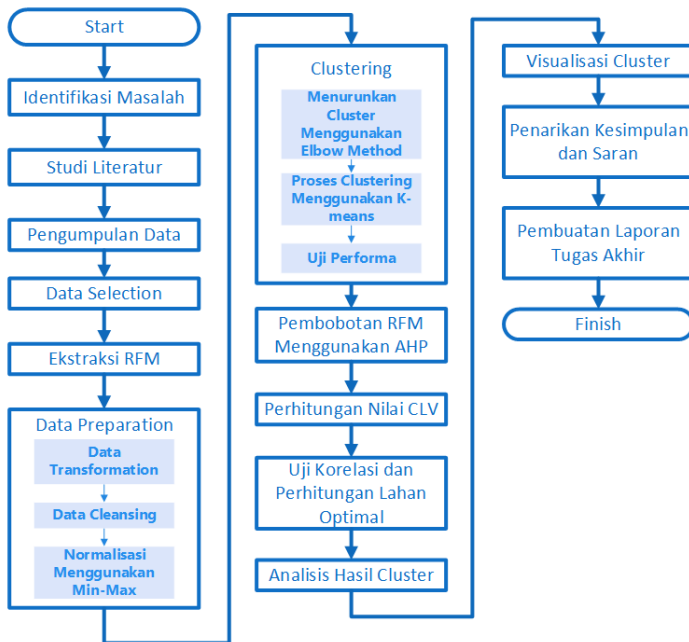
Berdasarkan pada diagram alur metodologi pada sub bab sebelumnya, di bawah ini merupakan penjelasan dari setiap prosesnya.

3.2.1 Identifikasi Permasalahan

Tahap ini adalah tahapan awal dari pengerjaan tugas akhir. Pada tahap ini dilakukan analisis terhadap objek yang ditentukan yaitu PTPN X Meritjan Kediri. Berdasarkan analisa yang dilakukan pada kondisi objek penelitian kemudian masalah yang akan diangkat ditetapkan. Permasalahan yang diangkat adalah segmentasi supplier tebu. Setelah penentuan masalah akan ditetapkan perumusan masalah, pendefinisian atasan, tujuan dan manfaat sehingga pengerjaan tugas akhir lebih terarah.

3.2.2 Studi Literatur

Tahap ini membahas mengenai studi-studi terdahulu yang membahas segmentasi supplier, *clustering*, *k-means*, normalisasi, RFM analysis, PTPN X PG Mertijan dan penggalan data yang dibahas di tugas akhir. Studi literatur dilakukan berdasarkan narasumber dengan melakukan wawancara, buku, penelitian sebelumnya dan dokumen terkait lainnya.



Gambar 3.1 Metodologi Penelitian

3.2.3 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data-data yang berhubungan dengan segmentasi supplier PTPN X PG Meritjan yang terbagi menjadi empat bagian yaitu :

1. Data profil supplier tebu dari PTPN X PG Meritjan
2. Data transaksi penerimaan tebu dari supplier pada tahun 2016
3. Data preferensi RFM
4. Data pendukung lainnya yang berkaitan dengan tugas akhir

Data dikumpulkan menggunakan metode primer, meminta data mentah pada perusahaan dan juga metode sekunder melalui wawancara dan kuisioner. Keluaran pada tahap ini adalah data yang dibutuhkan untuk proses *clustering*.

3.2.4 Data Selection

Data selection dilakukan untuk mempermudah proses *clustering* sehingga data yang diolah tidak terlalu banyak. Data yang dipilih yang digunakan proses *clustering* saja. Pada tahapan ini data yang telah dikumpulkan akan dipilih variabel yang memiliki keterkaitan dengan proses *clustering* dan analisis RFM. Dari variabel yang ada pada data pengadaan tebu, variabel yang dipilih adalah id supplier, nama, tanggal penerimaan tebu dan hasil gula.

3.2.5 Ekstraksi RFM

Pada tahapan ini data yang dipilih akan dianalisis menggunakan analisis RFM. Masukan dari proses ini adalah data yang telah dilakukan praproses data dan keluarannya berupa variabel RFM. Data yang sudah dikumpulkan dalam *data selection* kemudian akan diubah kedalam variabel RFM. Transformasi ini menggunakan bantuan *query SQL* yang dijalankan pada *tools navicat*. Data tanggal penerimaan terbu akan diubah menjadi atribut *recency* dan *frequency*. Sementara nilai gula dalam bentuk uang akan diubah menjadi variabel *monetary*. Keluaran dari tahapan ini adalah variabel *recency*, *frequency* dan *monetary*. Proses ekstraksi ini dilakukan pada tanggal 20 Maret 2017.

3.2.6 Data Preparation

Ketiga variabel RFM yang telah diekstraksi dari tahapan sebelumnya disiapkan dan diubah ke bentuk yang mudah untuk dilakukan proses *clustering*. *Data preparation* terdiri dari transformasi data, penghapusan *outlier* dan normalisasi.

3.2.7 Data Transformation

Pada tahapan ini dilakukan perubahan data menggunakan metode Logarithmic. Tahapan ini dilakukan agar persebaran data tidak condong sebelah. Metode logarithmic dipilih karena data mentah memiliki *positive skewness*.

3.2.7.1 *Data Cleansing*

Pada tahapan ini hasil ekstraksi variabel RFM dibersihkan sehingga tidak akan mengganggu proses *clustering*. Proses *clustering* ini terdiri dari penghilangan *outlier* atau data yang memiliki nilai yang jauh berbeda. Metode yang digunakan adalah Tukey Method. Tukey Method digunakan karena data yang digunakan dalam penelitian ini besar dan distribusi datanya tidak normal

3.2.7.2 **Normalisasi Variabel RFM Menggunakan Min-Max**

Ketiga variabel RFM memiliki *range* nilai yang berbeda anantara satu sama lainnya, dimana *monetary* dapat memiliki nilai hingga jutaan sementara *recency* dan *frequency* memiliki nilai hanya berkisar pada puluhan saja. Pada tahap ini akan dilakukan normalisasi sehingga ketiga variabel RFM memiliki nilai yang sama. Metode normalisasi yang digunakan adalah Metode Min-Max, dimana semua variabel akan diubah menjadi nilai nol hingga satu. Setelah dilakukan normalisasi data diintegrasikan menjadi satu tabel. Masukan proses ini adalah variabel RFM dan keluarannya adalah variabel RFM yang memiliki *range* nilai yang sama dan terintegrasi dalam satu tabel.

3.2.8 *Clustering Data*

Setelah melakukan pengumpulan data, data supplier kemudian di kelompokkan menggunakan metode *clustering*. Metode *clustering* yang digunakan metode K-Means. *Clustering* terdiri dari 3 tahapan yaitu menurunkan *cluster*, proses *cluster* dan yang terakhir adalah uji performa. Masukan dari proses ini adalah data yang memiliki variabel RFM. Keluarannya berupa *cluster* yang telah optimum.

3.2.8.1 **Menurunkan Cluster Menggunakan Elbow Method**

Metode *clustering* yang akan digunakan adalah non hiraki *cluster* atau *partition clustering* karena tujuan utamanya untuk mendapatkan hasil *cluster* yang benar-benar hanya jadi anggota di suatu kelompok. Pada tahap ini akan dilakukan penentuan

nilai k atau jumlah kelompok yang digunakan menggunakan *elbow method*. Setelah itu dilakukan proses *clustering* menggunakan metode k -means.

3.2.8.2 Proses Cluster Menggunakan K-Means

Pada tahap ini akan dilakukan proses *clustering* menggunakan metode K-Means. Hasil *cluster* yang didapatkan akan diberikan label nama sesuai dengan perilaku suppliernya.

3.2.8.3 Uji Performa Cluster

Pada tahap ini hasil *clustering* akan dibandingkan dengan keadaanya, apakah telah benar-benar merepresentasikan keadaan sebenarnya. Uji performa dilakukan dengan menggunakan perhitungan SSE, Dunn Index, *Silhouette Width* dan *Connectivity*. Apabila nilai dunn index tinggi maka hasilnya baik sementara *connectivity* bila rendah maka baik.

3.2.9 Pembobotan RFM Menggunakan AHP

Pada tahap ini hasil *clustering* akan dibobotkan sesuai dengan kepentingan yang sama. Hasil pembobotan akan digunakan untuk mencari nilai *customer lifetime value* yang digunakan untuk memberi ranking pada supplier.

Metode yang digunakan dalam pembobotan ini adalah *Analytic Hierarchical Process (AHP)*. Masukan pada tahapan ini adalah hasil kuisioner atau wawancara pada pihak yang terkait mengenai ketiga variabel RFM. Pada tahap ini ketiga variabel akan dibandingkan satu sama lainnya dan dicari mana yang memiliki kepentingan yang lebih tinggi. Setelah melakukan proses AHP, hasilnya akan dicek apakah konsisten dengan melihat nilai *consistency rate*. Bila nilai *consistency rate* kurang dari 0.1 maka hasil AHP termasuk konsisten. Setelah itu bobot yang dihasilkan AHP dikalikan pada nilai rata-rata RFM pada setiap *cluster*.

3.2.10 Perhitungan Nilai CLV

Pada tahapan ini rata-rata variabel RFM yang telah dikalikan bobot dijumlahkan menjadi satu. Penjumlahan nilai tersebut yang di sebut oleh CLV. Nilai CLV ini didapatkan berdasarkan

RFM yang disesuaikan dengan objek penelitiannya yaitu supplier. Setelah itu nilai CLV diurutkan dari besar ke kecil. Nilai CLV yang besar berarti supplier tersebut adalah supplier terbaik diantara yang lainnya. Keluaran dari tahapan ini adalah ranking supplier.

3.2.11 Uji Korelasi dan Perhitungan Lahan Optimal

Pada tahapan ini dilakukan uji korelasi pada variabel monetary dan area, untuk mengetahui apakah luas area mempengaruhi nilai monetary. Selain itu juga akan dicari perhitungan lahan optimal dengan menggunakan metode *Linear Regression*, agar perusahaan mengetahui lahan mana saja yang bisa dioptimalkan untuk menghasilkan gula.

3.2.12 Denormalisasi Data

Pada tahapan ini data yang telah dinormalisasi dan di transformasi dikembalikan ke nilainya yang semula. Tahapan ini dilakukan untuk mempermudah analisis cluster.

3.2.13 Analisis Hasil Cluster

Pada tahapan ini, hasil *cluster* dianalisis dengan membandingkan nilai atribut RFM dengan *cluster* lainnya. Selain itu ranking dari CLV dan jumlah anggota supplier yang termasuk dalam *cluster* tersebut.

3.2.14 Visualisasi Cluster

Pada tahap visualisasi *cluster*, hasil *cluster* akan ditampilkan melalui *dashboard* yang berbasis web. Alat yang digunakan adalah *package* dari R yaitu R-shiny untuk menampilkan visualisasi grafik *clustering*. Grafik yang ditampilkan berupa scattered plot tiga dimensi, *pie chart* dan *Box and whisker*. Selain grafik, akan ditampilkan juga anggota per *cluster*-nya beserta informasi hasil analisis yang dilakukan pada tahap sebelumnya. Masukan proses ini *cluster* yang dihasilkan oleh *tools* R dan keluarannya berupa halaman website yang menampilkan *cluster* dalam bentuk grafik.

3.2.15 Penarikan Kesimpulan dan Saran

Pada tahap ini ditarik kesimpulan berdasarkan *clustering* atau segmentasi yang dilakukan pada data supplier. Hasil *clustering* akan menghasilkan kelompok-kelompok supplier yang memiliki ciri yang sejenis. Kesimpulan yang dihasilkan akan menjawab tujuan penelitian. Serta memberikan saran untuk penelitian selanjutnya.

3.2.16 Pembuatan Laporan Tugas Akhir

Pembuatan laporan tugas akhir adalah tahapan akhir pada pengerjaan tugas akhir ini. Tahapan ini dilakukan sebagai dokumentasi seluruh proses pengerjaan tugas akhir. Laporan yang dibuat telah mengikuti format dokumen yang telah ditentukan.

(halaman sengaja dikosongkan)

BAB IV PERANCANGAN

Pada bab ini akan dijelaskan mengenai rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian dilakukan.

4.1 Pengumpulan Data

Pengumpulan data dilakukan secara langsung pada perusahaan PTPN X PG Meritjan, Kediri. Data yang digunakan adalah data transaksi pengadaan tebu pada tanggal 9 Juni 2016 hingga 6 November 2016 berupa *soft copy*. Selain itu wawancara dilakukan untuk mempermudah pemahaman alur bisnis perusahaan dan kendala yang pernah dialami. Data yang didapatkan telah berbentuk *excel workbook* dengan variabelnya seperti Tabel 4.1.

Tabel 4.1 Data Mentah Tabel Pengadaan Tebu

Variabel	Type Data	Keterangan
Purch.Doc	Number	ID dokumen yang dibuat
Tebu (Kw)	Number	Jumlah tebu yang dikirimkan
Rendemen Sementara	Number	Kandungan gula dalam tebu
Hasil Gula	Number	Gula yang dihasilkan dari tebu tersebut
Gula Tetes	Number	Gula tetes yang dihasilkan
Nomor Polisi Truk	Text	Nomor kendaraan truk yang mengangkat
Kecamatan	Text	Kecamatan data tersebut ditanamkan
Block Name	Text	ID lahan tebu ditanam
Tanggal Timbang	Date	Tanggal tebu masuk

Berdasarkan variabel di atas data transaksi yang didapatkan pada tahun 2016 adalah 53.375 terhitung dari bulan Juni hingga November. Potongan data mentah pada *excel workbook* akan terlihat pada Tabel 4.5

Akan tetapi hasil gula ini masih dalam bentuk kuintal, sehingga perlu dijadikan kilogram terlebih dahulu dan dikalikan dengan harga gula yang beredar. Harga gula pada saat ekstraksi RFM berlangsung adalah Rp 12.5000,- [45] Tabel 4.2 adalah perhitungan nilai gula dalam kg menjadi rupiah

Tabel 4.2 Perhitungan Monetary

Hasil Gula (kw)	Hasil Gula (kg)	Hasil Gula (Rp)
4.1408	414.08	5,176,000
5.3214	532.14	6,651,750
4.4188	441.88	5,523,500
4.8519	485.19	6,064,875
4.3592	435.92	5,449,000

Selain data transaksi pengadaan tebu, data yang digunakan adalah data profil supplier yang akan digabungkan dengan data transaksi pengadaan tebu. Tabel 4.3 adalah variabel dan tipenya.

Tabel 4.3 Tabel Profil Supplier

Variabel	Type Data	Keterangan
<i>Block Name</i>	Text	ID lahan tebu ditanam
Nama Petani	Text	Nama petani tebu
<i>Complex Name</i>	Text	Daerah tebu ditanam
Area Tanam	Number	Luas area tanam
Satuan Luas	Text	Satuan luas
Kabupaten	Text	Nama kabupaten
Kecamatan	Text	Nama kecamatan
Desa	Text	Nama desa

Jumlah supplier yang terdaftar pada tabel ini sebanyak 4871 supplier. Potongan data mentah pada *excel workbook* akan

terlihat pada Tabel 4.5. Tabel profil supplier yaitu Tabel 4.6 dan tabel akan diintegrasikan menjadi satu tabel .

4.2 Data Selection

Pada tahapan ini akan dilakukan pemilihan variabel yang digunakan dalam proses *clustering*. Dari variabel data yang ada pada Tabel 4.1, variabel yang digunakan adalah tanggal penerimaan, *block name* dan hasil gula. Sementara variabel yang digunakan pada tabel supplier adalah nama dan *blockname*. Tabel 4.4 adalah tabel setelah melakukan setelah dilakukan pemilihan variabel dan integrasi kedua tabel tersebut.

Tabel 4.4 Variabel Terpilih dari Tabel Penerimaan Tebu

Block Name	Nama	Hasil Gula (Rp)	Tanggal Timbang
10000023	HARTI	5,176,000	6/8/2016
10000084	SOLEKAN	6,651,750	6/8/2016
10000084	SOLEKAN	5,523,500	6/8/2016
10000126	RAJIYO	6,064,875	6/8/2016
10000126	RAJIYO	5,449,000	6/9/2016
10000158	MULYANI	4,890,000	6/8/2016
10000158	MULYANI	5,021,625	6/8/2016
10001944	ANDIRA SARI	5,447,500	6/8/2016

4.3 Ekstraksi RFM

Pada tahap ini akan dilakukan analisis sesuai dengan *RFM analysis* dimana akan dilakukan ekstraksi variabel, normalisasi dan pembobotan variabel RFM. Normalisasi akan menggunakan metode Min-Max sedangkan pembobotan variabel akan menggunakan metode AHP yang melalui tahap wawancara sebelumnya.

Tabel 4.5 Potongan Data Mentah Transaksi Pengadaan Tebu

Purch.Doc.	GR Qty/Ku Tebu	Rendemen Sementara	Hasil Gula	Gula Tetes	Nomor Polisi Truk	Kecamatan	Block Name	Tanggal Timbang
8100042097	57	7.2500	5,176,000	171.0000	AG8467VB	BERBEK	10000023	6/8/2016
8100042102	71	7.4800	6,651,750	213.0000	AG8963VE	BERBEK	10000084	6/8/2016
8100042103	60	7.3500	5,523,500	180.0000	AG9867VD	BERBEK	10000084	6/8/2016
8100042106	62	7.8100	6,064,875	186.0000	AG8153UW	LOCERET	10000126	6/8/2016
8100042107	55	7.9100	5,449,000	165.0000	AG8153UW	LOCERET	10000126	6/9/2016
8100042108	54	7.2300	4,890,000	162.0000	AE9600NB	BERBEK	10000158	6/8/2016

Tabel 4.6 Potongan Data Mentah Tabel Profil Supplier

Block Name	Nama Petani	Complex Name	Area Tanam	Satuan Luas	Kabupaten	Kecamatan No	Desa
10000017	BAMBANG MUJIATMOKO	KWAGEAN	0.6900	Hectare	NGANJUK	LOCERET	KWAGEAN
10000018	BAMBANG MUJIATMOKO	KWAGEAN	0.2270	Hectare	NGANJUK	LOCERET	KWAGEAN
10000024	HENDRO SUKIMIN	PATRANREJO	0.4920	Hectare	NGANJUK	BERBEK	PATRANREJO
10000025	HENDRO SUKIMIN	PATRANREJO	0.5140	Hectare	NGANJUK	BERBEK	PATRANREJO
10002532	ACHMAD SULKAN	SEKARPUTIH	1.1300	Hectare	NGANJUK	BAGOR	SEKARPUTIH
10000027	YOYOK.H	PATIHAN	1.1680	Hectare	NGANJUK	LOCERET	PATIHAN
10000029	HARTOYO	MUNGKUNING	0.1870	Hectare	NGANJUK	LOCERET	MUNGKUNING

Pada tahapan ini akan dilakukan ekstraksi variabel RFM berdasarkan proses *data selection*. Variabel *recency* menggunakan tanggal timbang, variabel *frequency* menggunakan variabel tanggal timbang dan variabel *monetary* menggunakan variabel hasil gula. Proses ekstraksi menggunakan *tools* navicat berdasarkan *SQL query*. Proses ekstraksi variabel RFM dilakukan pada 20 Maret 2017.

4.3.1 Ekstraksi Variabel *Recency*

Recency adalah selisih antara tanggal hari ini dengan transaksi terakhir. Variabel yang digunakan untuk menghasilkan variabel *recency* adalah tanggal timbang. Nama tanggal timbang sebelumnya diubah menjadi tanggal untuk mempermudah proses ekstraksi. Sedangkan untuk mengidentifikasi supplier, variabel yang digunakan adalah *block name*.

4.3.2 Ekstraksi Variabel *Frequency*

Frequency adalah jumlah banyaknya transaksi yang dilakukan pada satu periode tertentu. Variabel yang digunakan untuk menghasilkan variabel *frequency* adalah tanggal timbang. Nama tanggal timbang sebelumnya diubah menjadi tanggal untuk mempermudah proses ekstraksi. Sedangkan untuk mengidentifikasi supplier, variabel yang digunakan adalah *block name*.

4.3.3 Ekstraksi Variabel *Monetary*

Monetary adalah nilai produk atau barang dalam nilai uang atau dalam kasus ini adalah rupiah. Variabel yang digunakan untuk menghasilkan variabel *monetary* adalah hasil gula. Hasil gula sebelumnya diubah menjadi gula untuk mempermudah proses ekstraksi. Sedangkan untuk mengidentifikasi supplier, variabel yang digunakan adalah *blockname*.

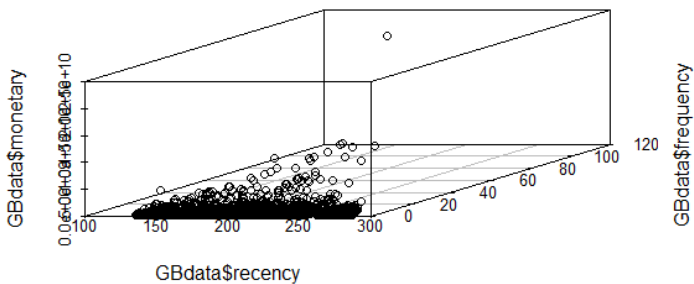
4.4 Data Preparation

Pada tahapan ini dilakukan pemetaan data RFM untuk mengetahui keadaan data saat ini. Kemudian, dilakukan

transformasi data menggunakan metode algoritmik. Selanjutnya, kemudian dilakukan penghapusan *outlier* dan terakhir lakukan normalisasi menggunakan metode min-max.

4.4.1 Data Transformation

Data yang dimiliki tidak tersebar dan bias pada satu variabel. Data dipetakan menggunakan 3D *scatter plot* yang dapat menggambarkan keadaan 3 variabel yang digunakan yaitu *recency*, *frequency* dan *monetary* seperti Gambar 4.1 yang menunjukkan bahwa data tidak tersebar melainkan berkumpul menjadi satu. Oleh karena itu perlu dilakukan transformasi data. Metode transformasi data yang digunakan adalah *Logarithmic*, dikarenakan data yang dimiliki kecondongan positif yang sangat besar pada variabel *frequency* dan *monetary*.



Gambar 4.1 Scatterplot Sebelum Cleansing

4.4.2 Data Cleansing

Clustering sangat sensitif dengan data *outlier* oleh karena itu perlu dilakukan *data cleansing* terlebih dahulu. Untuk melakukan *data cleansing*, data terlebih dahulu dipetakan untuk melihat apakah ada data *outlier*. Proses *data cleansing* menggunakan *tukey method* yang ada pada *tools R*. Pertama-tama bandingkan petakan dulu keadaan data sebelum dilkaukan penghapusan *outlier*.

4.4.3 Normalisasi Menggunakan Min-Max

Ketiga variabel RFM memiliki *range* nilai yang berbeda yang dapat membuat proses *clustering* sulit nantinya, sehingga ketiga

variabel tersebut perlu dilakukan proses normalisasi. Proses normalisasi ini menggunakan metode min-max, yaitu proses perubahan data yang mengubah nilai menjadi 0 hingga 1. Rumus metode min-max adalah sebagai berikut :

$$MM(X_{ij}) = \frac{X_{ij} - X_{min}}{X_{max} - X_{min}} \quad [33]$$

X_{min} = nilai x terkecil X_{max} = nilai x terbesar

Nilai *recency* semakin baik bila memiliki nilai yang kecil, namun pada penentuan nilai CLV, semua nilai akan di tambahkan menjadi satu. Nilai *recency* bertentangan dengan nilai yang lain sehingga sebelumnya nilai *recency* akan dibalik dengan melakukan pengurangan 1 dengan hasil normalisasi *recency*.

4.5 Penentuan Jumlah Cluster

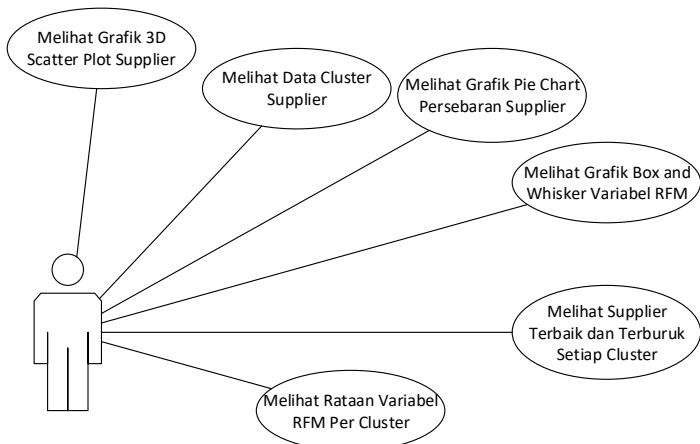
Metode yang digunakan untuk menentukan nilai k adalah dengan menggunakan *elbow method*. Dimana nilai k yang dipilih dilihat dari grafik perbandingan nilai k dengan nilai SSE yang dihasilkan. Apabila, nilai k turun secara drastis dan diikuti grafik yang stabil maka pada titik tersebut yang dipilih sebagai nilai k. *Tools* yang digunakan adalah R studio. Bila ada keraguan mengenai titik yang dipilih, titik dengan nilai SSE terkecil adalah titik yang dipilih sebagai nilai k. Berdasarkan nilai SSE k yang digunakan adalah 3.

4.6 Use Case Visualisasi Supplier

Gambar 4.2 adalah gambaran *use case* dari visualisasi segmentasi supplier pada PTPN X PG Meritjan, terdapat 6 use case utama yaitu :

1. Melihat grafik 3D *Scatter plot* Supplier
2. Melihat data *cluster* supplier
3. Melihat grafik *pie chart* persebaran supplier
4. Melihat grafik *box and whisker* variabel RFM
5. Melihat Supplier Terbaik dan Terburuk Setiap *Cluster*
6. Melihat Rataan Variabel RFM Per *Cluster*

Tabel 4.7 merupakan penjelasan mengenai use case melihat grafik 3D *Scatter plot* yang berisikan tujuan, *overview*, *actor*, kondisi sebelum dan sesudah serta *flow* yang akan terjadi apabila berhasil, *flow* alternatif dan *flow* lainnya.



Gambar 4.2 Use Case Diagram Visualisasi

Tabel 4.7 Use case description melihat grafik 3d scatter plot supplier

UC – 1 Melihat Grafik 3D Scatter plot Supplier		
Purpose	Melihat Grafik 3D <i>Scatter plot</i> Supplier	
Overview	Dimulai ketika user berada di halaman dashboard ataupun <i>cluster</i>	
Actors	Pengguna visualisasi	
Pre-Condition	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
Post Condition	Grafik 3D <i>Scatter plot</i>	
Typical Course Event	Actor	System
	1. Mengklik icon maximize pada kotak 3D <i>Scatterplot</i>	2. Memanggil data supplier yang telah tersimpan 3. Menghasilkan grafik 3D <i>scatter plot</i>

UC – 1 Melihat Grafik 3D Scatter plot Supplier	
Alternate Flow of Events	-
Exceptional Flow of Events	Visualisasi tertutup karena error

Tabel 4.8 merupakan penjelasan mengenai use case melihat data *cluster* supplier yang berisikan tujuan, *actor*, kondisi sebelum dan sesudah serta *flow* yang akan terjadi.

Tabel 4.8 Use case description melihat data cluster supplier

UC – 2 Melihat data cluster supplier		
Purpose	Melihat data <i>cluster</i> supplier	
Overview	Dimulai ketika user berada di halaman <i>cluster</i>	
Actors	Pegguna visualisasi	
Pre Condition	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
Post Condition	Tabel data <i>cluster</i> supplier	
Typical Course Event	Actor	System
	1. Mengklik jenis <i>cluster</i> yang diinginkan	2. Memanggil data supplier yang telah tersimpan 3. Menampilkan tabel data <i>cluster</i> supplier
Alternate Flow of Events	-	
Exceptional Flow of Events	Visualisasi tertutup karena error	

Tabel 4.9 merupakan penjelasan mengenai use case melihat grafik *pie chart* persebaran supplier

Tabel 4.9 Use case description melihat grafik pie chart persebaran supplier

UC – 3 Melihat grafik pie chart persebaran supplier	
Purpose	Melihat grafik <i>pie chart</i> persebaran supplier
Overview	Dimulai ketika user berada di halaman dashboard

UC – 3 Melihat grafik <i>pie chart</i> persebaran supplier		
Actors	Pengguna visualisasi	
Pre Condition	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
Post Condition	Grafik <i>pie chart</i> persebaran supplier dihasilkan	
Typical Course Event	Actor	System
	1. Mengklik icon maximize pada kotak <i>pie chart</i>	2. Memanggil data supplier yang telah tersimpan 3. Menghitung jumlah supplier per <i>cluster</i> 4. Menampilkan <i>pie chart</i>
Alternate Flow of Events	-	
Exceptional Flow of Events	Visualisasi tertutup karena error	

Tabel 4.10 merupakan penjelasan mengenai use case melihat grafik *box and whisker* variabel RFM yang berisikan tujuan, *actor*, kondisi sebelum dan sesudah serta *flow* yang akan terjadi.

Tabel 4.10 Use case description melihat grafik *box and whisker* variabel RFM

UC – 4 Melihat grafik <i>box and whisker</i> variabel RFM		
Purpose	Melihat grafik <i>box and whisker</i> variabel RFM	
Overview	Dimulai ketika user berada di halaman dashboard	
Actors	Pengguna visualisasi	
Pre Condition	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
Post Condition	Grafik <i>box and whisker</i> variabel RFM dihasilkan	
Typical Course Event	Actor	System
	1. Memilih variabel	2. Memanggil data supplier yang telah tersimpan

UC – 4 Melihat grafik <i>box and whisker</i> variabel RFM		
	yang diinginkan	3. Memilih variabel yang digunakan 4. Menampilkan grafik <i>box and whisker</i>
<i>Alternate Flow of Events</i>	-	
<i>Exceptional Flow of Events</i>	Visualisasi tertutup karena error	

Tabel 4.11 merupakan penjelasan mengenai use case melihat supplier terburuk dan terbaik setiap *cluster* yang berisikan tujuan, *actor*, kondisi sebelum dan sesudah serta *flow* yang akan terjadi.

Tabel 4.11 Use case description melihat supplier terburuk dan terbaik RFM

UC – 5 Melihat supplier terburuk dan terbaik setiap <i>cluster</i>		
<i>Purpose</i>	Melihat supplier terburuk dan terbaik setiap <i>cluster</i>	
<i>Overview</i>	Dimulai ketika user berada di halaman <i>cluster</i>	
<i>Actors</i>	Pegguna visualisasi	
<i>Pre Condition</i>	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
<i>Post Condition</i>	Valuebox berisikan nama dan nilai RFMnya	
<i>Typical Course Event</i>	<i>Actor</i>	<i>System</i>
	1. Memilih jenis <i>cluster</i>	2. Memanggil data supplier yang telah tersimpan 3. Melakukan sorting data sesuai nilai CLV 4. Menampilkan nama dan clv dalam valuebox
<i>Alternate Flow of Events</i>	-	
<i>Exceptional Flow of Events</i>	Visualisasi tertutup karena error	

Tabel 4.12 merupakan penjelasan mengenai use case melihat rata-rata variabel RFM per *cluster* yang berisikan tujuan, *actor*, kondisi sebelum dan sesudah serta *flow* yang akan terjadi.

Tabel 4.12 Use case description melihat rata-rata variabel RFM per *cluster*

UC – 6 Melihat rata-rata variabel RFM per <i>cluster</i>		
Purpose	Melihat rata-rata variabel RFM per <i>cluster</i>	
Overview	Dimulai ketika user berada di halaman <i>cluster</i>	
Actors	Pengguna visualisasi	
Pre Condition	<ul style="list-style-type: none"> • Pengguna membuka visualisasi • Data supplier telah ada pada sistem 	
Post Condition	Valuebox berisikan rata-rata variabel RFM per <i>cluster</i>	
Typical Course Event	Actor	System
	1. Memilih jenis <i>cluster</i> yang diinginkan	2. Memanggil data supplier yang telah tersimpan 3. Melakukan proses perhitungan rata-rata tiap variabel 4. Menampilkan nilai rata-rata tiap variabel
Alternate Flow of Events	-	
Exceptional Flow of Events	Visualisasi tertutup karena error	

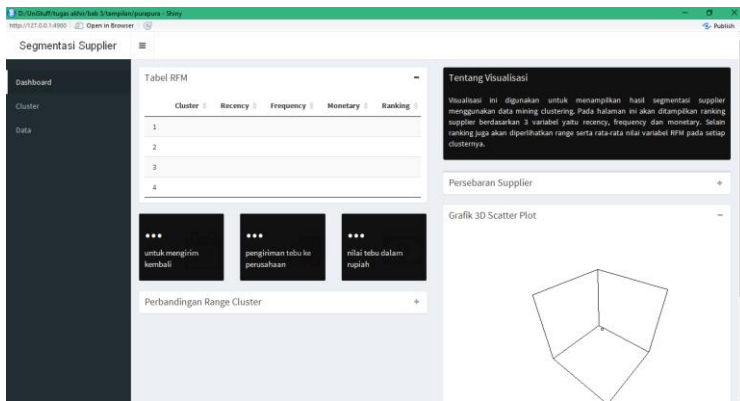
4.7 Rancangan Visualisasi

Pada bagian ini akan dijelaskan mengenai rancangan visualisasi segmentasi supplier yang terdiri dari 3 halaman yaitu halaman dashboard, halaman *cluster* dan halaman data. *Tools* yang digunakan untuk visualisasi adalah shiny dashboard yang ada pada *tools* R.

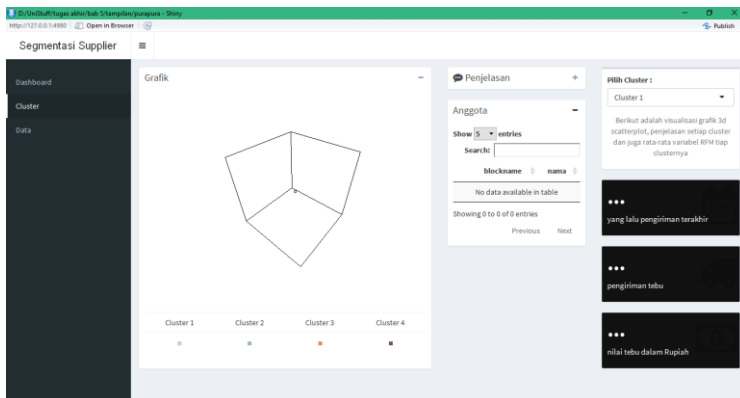
Gambar 4.3 menunjukkan rancangan awal halaman dashboard. Pada halaman ini akan dimuat grafik *scatterplot* yang dapat

menunjukkan *cluster* yang ada, kemudian disertakan dengan informasi mengenai data secara keseluruhan.

Sementara untuk rancangan halaman *cluster* seperti yang ditunjukkan Gambar 4.4 halaman ini akan memuat *box* untuk menampilkan grafik *scatter plot* serta pilihan *select box* untuk memilih informasi *cluster* yang ditampilkan

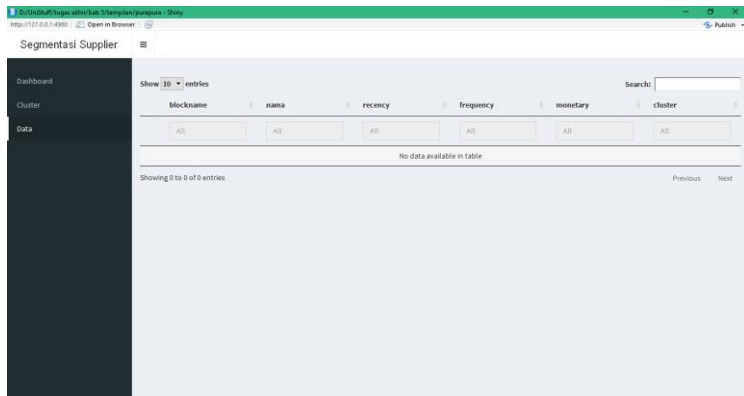


Gambar 4.3 Rancangan halaman dashboard



Gambar 4.4 Rancangan halaman cluster

Gambar 4.5 menunjukkan rancangan halaman data yang memuat tabel berisikan informasi supplier berupa blockname, nama, nilai RFM dan *clusternya*.



Gambar 4.5 Rancangan halaman data

(halaman ini sengaja dikosongkan)

BAB V IMPLEMENTASI

Pada tahapan ini akan dilakukan proses pengolahan data dari data yang dihasilkan dari tahapan perancangan. Teknik yang digunakan dalam pengolahan data adalah *clustering*.

5.1 Ekstraksi RFM

Pada tahapan ini akan dilakukan ekstraksi variabel RFM berdasarkan proses *data selection*. Variabel *recency* menggunakan tanggal timbang, variabel *frequency* menggunakan variabel tanggal timbang dan variabel *monetary* menggunakan variabel hasil gula. Proses ekstraksi menggunakan *tools* navicat berdasarkan *SQL query*.

5.1.1 Ekstraksi Variabel *Recency*

Recency adalah selisih antara tanggal hari ini dengan transaksi terakhir. Variabel yang digunakan untuk menghasilkan variabel *recency* adalah tanggal timbang. Selisish tanggal transaksi terakhir dengan tanggal saat ini dapat menggunakan *SQL query* DATEDIFF. *Script* yang digunakan terlihat seperti *Script 5.1*

```
SELECT blockname, DATEDIFF(NOW(),max(tanggal))as 'recency'  
from gabungannama  
group by blockname  
order by blockname;
```

Script 5.1 Ekstraksi Recency

5.1.2 Ekstraksi Variabel *Frequency*

Frequency adalah jumlah banyaknya transaksi yang dilakukan pada satu periode tertentu. Variabel yang digunakan untuk menghasilkan variabel *frequency* adalah tanggal timbang. Jumlah transaksi pada periode dapat dihasilkan menggunakan *SQL query* COUNT. *Script* yang digunakan terlihat seperti *Script 5.2*

```
select blockname, count(distinct(tanggal)) as 'frequency'
from `gabunganama
where EXTRACT(month FROM tanggal) BETWEEN 1 and 12
group by blockname
order by blockname;
```

Script 5.2 Ekstraksi Frequency

5.1.3 Ekstraksi Variabel *Monetary*

Monetary adalah nilai produk atau barang dalam nilai uang atau dalam kasus ini adalah rupiah. Variabel yang digunakan untuk menghasilkan variabel *monetary* adalah hasil gula dalam bentuk rupiah. Jumlah gula dalam bentuk rupiah pada periode ini dapat dihasilkan menggunakan *SQL query* SUM. *Script* yang digunakan terlihat seperti *Script 5.3*

```
select DISTINCT blockname, sum(gula) as 'monetary'
from gabunganama
group by blockname
order by blockname;
```

Script 5.3 Ekstraksi Monetary

5.2 Data Preparation

Data preparation terdiri dari transformasi data, penghapusan *outlier* dan normalisasi. Tools yang digunakan dalam tahapan ini adalah R.

5.2.1 Data Transformation

Perubahan data menggunakan metode logarithmic. Variabel *frequency* dan *monetary*, memiliki kecondongan positif yang sangat besar. Oleh karena itu, ketiga variabel ini diubah nilainya menggunakan *function* **log10()**. *Script* yang digunakan terlihat seperti *Script 5.4*

```
BStrans$recency <-log10(BStrans$recency)
BStrans$frequency <-log10(BStrans$frequency)
BStrans$monetary <-log10(BStrans$monetary)
```

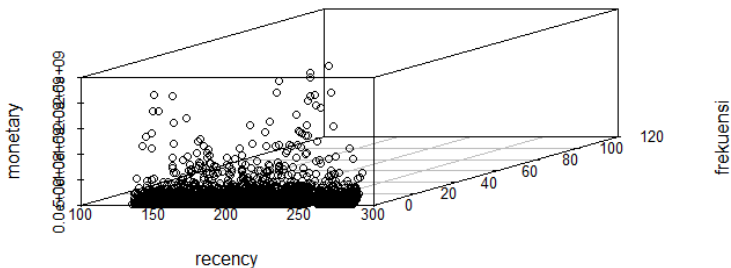
Script 5.4 Transformasi data menggunakan metode algoritmik

5.2.2 Data Cleansing

Untuk melihat perbedaan antara data yang telah dihilangkan *outliernya* dan keadaan aslinya, data dipetakan menggunakan

3D *scatter plot* yang dapat menggambarkan keadaan 3 variabel yang digunakan yaitu *recency*, *frequency* dan *monetary*.

Seperti yang dilihat pada Gambar 5.1 data yang dimiliki masih tidak tersebar karena adanya *outlier*. Hal ini dapat menyebabkan hasil *clustering* yang dihasilkan buruk. Oleh karena itu akan dilakukan penghapusan *outlier* terlebih dahulu.



Gambar 5.1 Scatterplot Sebelum Cleansing

Script 5.5 Script Penghapusan Outlier yang digunakan untuk penghapusan *outlier* dibuat oleh Klodian Dhana [46]

```

outlierKD <- function(dt, var) {
  var_name <- eval(substitute(var),eval(dt))
  tot <- sum(!is.na(var_name))
  na1 <- sum(is.na(var_name))
  m1 <- mean(var_name, na.rm = T)
  par(mfrow=c(2, 2), oma=c(0,0,3,0))
  boxplot(var_name, main="With outliers")
  hist(var_name, main="With outliers", xlab=NA, ylab=NA)
  outlier <- boxplot.stats(var_name)$out
  mo <- mean(outlier)
  var_name <- ifelse(var_name %in% outlier, NA, var_name)
  boxplot(var_name, main="Without outliers")
  hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
  title("Outlier Check", outer=TRUE)
  na2 <- sum(is.na(var_name))
  m2 <- mean(var_name, na.rm = T)
  response <- readline(prompt="Do you want to remove outliers and to
replace with NA? [yes/no]: ")
  if(response == "y" | response == "yes"){
    dt[as.character(substitute(var))] <- invisible(var_name)
  }
}

```

```

assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)
message("Outliers successfully removed", "\n")
return(invisible(dt))
} else{
  message("Nothing changed", "\n")
  return(invisible(var_name)) }}
outlierKD(dat,var)

```

Script 5.5 Script Penghapusan Outlier

5.2.3 Normalisasi Menggunakan Min-Max

Proses normalisasi ini menggunakan metode min-max, yaitu proses perubahan data yang mengubah nilai menjadi 0 hingga 1. Proses normalisasi dibantu dengan *tools* R menggunakan Script 5.6

Pada *script* ini dilakukan inisiasi data terlebih dahulu, kemudian hilangkan variabel yang tidak berhubungan. Setelah itu, lakukan pergantian nilai menggunakan rumus min-max yang kemudain akan dimasukan pada kolom variabelnya.

```

BSnorm$recency<-((BSnorm$recency-min(BSnorm$recency))/(max(BSnorm$recency)-min(BSnorm$recency)))
BSnorm$recencyB <-(1-BScnorm$recency)
BSnorm$frequency<-((BSnorm$frequency-min(BScnorm$frequency))/(max(BScnorm$frequency)-min(BScnorm$frequency)))
BSnorm$monetary<-((BSnorm$monetary-min(BScnorm$monetary))/(max(BScnorm$monetary)-min(BScnorm$monetary)))

```

Script 5.6 Script Normalisasi Min-Max

5.3 Penentuan Jumlah Cluster

Penentuan nilai k menggunakan *elbow method*. Pertama-tama hasilkan random number menggunakan *set.seed*. Lalu tentukan nilai k maksimal yang dihasilkan pada contoh ini adalah 15, untuk penentuan jumlah k maksimal tidak ada ketentuan karena metode *elbow method* adalah metode *try and error*. Lalu inisiasikan data yang digunakan yaitu data bobot dengan nama data. *Script* untuk tahap tersebut terlihat pada Script 5.7

```
> set.seed(123) #membuat random data
> k.max<-15 #menentukan nilai k yang maksimal
> data<-BSelbow #inisiasi data
```

Script 5.7 Script Tahapan Cluster

Setelah itu lakukan *clustering* dan pembuatan grafik yang menampilkan perbandingan jumlah *cluster* dan SSE yang dihasilkan. *Script* yang digunakan ditampilkan pada Script 5.8

```
VidSSE <- sapply(1:k.max,
                 function(k){kmeans(BSelbow, k, nstart=10)$tot.withinss})
plot(1:k.max,VidSSE,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")
View(VidSSE)
```

Script 5.8 Script Grafik Elbow Method

5.4 Proses *Clustering*

Pada tahap ini akan dilakukan proses *clustering* menggunakan metode K-Means. Hasil *cluster* yang didapatkan akan diberikan label nama sesuai dengan perilaku supliernya.

5.4.1 *Clustering* Menggunakan K-Means

Proses *clustering* menggunakan bantuan *tools* R. Metode yang digunakan adalah K-Means dengan nilai k yang telah ditentukan yaitu 3. Script 5.9 adalah *script* yang digunakan untuk melakukan proses *clustering* :

```
BSccluster <- kmeans(BSelbow,3)
aggregate(BSelbow,by=list(BSccluster$cluster),FUN=mean)
BShasilclus<-data.frame(BSelbow,BSccluster$cluster)
write.csv(BShasilclus,"D:/UniStuff/tugas akhir/bismillah/BShasilclus.csv")
View(BShasilclus)
```

Script 5.9 Script K-Means Clustering

5.4.2 Uji Performa Menggunakan SSE

Lakukan perbandingan nilai SSE yang dapat dihasilkan dengan meneruskan *script* dari *elbow method* sebelumnya dengan Script 5.10

```
>wss <- sapply(1:k.max,
  function(k){kmeans(data, k, nstart=10)$tot.withinss})
>View(wss)
```

Script 5.10 Script memperlihatkan wss

5.4.3 Uji Performa Menggunakan clValid

R memiliki sebuah package yang bernama clValid(), package ini dapat membantu untuk melakukan uji performa nilai k yang dipilih, perhitungan uji performanya berdasarkan nilai dunn, sillhoutte dan connectivity. *Script* yang digunakan untuk melakukan uji performa nampak pada Script 5.11

```
library(clValid)
BSDataIntern <-data.frame(BSelbow)
BSInternal<- clValid(BSDataIntern, nClust= 2:15, clMethods="kmeans",
  validation="internal",maxitems = 5000)
summary(BSInternal)
```

Script 5.11 Script clValid

5.4.4 Verifikasi Cluster

Untuk membuktikan apakah satu anggota benar-benar termasuk pada cluster tersebut, dicarilah Euclidian distance antara data tersebut dengan masing-masing centroid. Setelah itu dicari jarak cluster mana yang memiliki jarak terdekat. *Script* yang digunakan nampak pada Script 5.12

```
BSeucli1$freq <-((BSeucli1$frequency-BSc1$frequency)^2)
BSeucli1$mon <-((BSeucli1$monetary-BSc1$monetary)^2)
BSeucli1$rec <-((BSeucli1$recency-BSc1$recency)^2)
BSeucli1$dist1 <-sqrt(BSeucli1$rec+BSeucli1$freq+BSeucli1$mon)
View(BSeucli1)
#membuat data frame berisi distance
BSdist <-data.frame(BSeucli1$dist1,BSeucli2$dist2,BSeucli3$dist3)

#menunjukkan nama column dengan nilai minimal
BSdist$min<-apply(BSdist,1,which.min)
```

```
#menunjukkan nilai minimalnya
BSmindist <- BSdist
BSmindist$min=NULL
BSmindist$minVal<-apply(BSdist,1,min)

#cek apakah identikal
identical(BSpersamaan$min,BSpersamaan$cluster)
```

Script 5.12 Script menunjukkan distance terdekat

5.5 Uji Korelasi dan Perhitungan Lahan Optimal

Pada tahap ini akan dilakukan uji korelasi untuk analisa tambahan mengenai hasil gula yang dihasilkan dengan luas area lahan tanam. Metode uji korelasi yang digunakan adalah metode pearson. Uji korelasi dilakukan menggunakan *tools* R. Script 5.13 adalah *script* R yang digunakan untuk melakukan uji korelasi.

```
> library(ggpubr)
> View(BStanahreg)
> cor.test(BStanahreg$monetary, BStanahreg$area)
```

Script 5.13 Script Uji Korelasi

Setelah dilakukan uji korelasi, akan dilakukan perhitungan prediksi nilai gula dengan menggunakan metode linear regresi. Perhitungan prediksi nilai gula menggunakan *tools* R dengan function **lm()** dan **predict()** seperti Script 5.14.

```
BStanahreg<- data.frame(BSujitanah$monetary,BSujitanah$area)
names(BStanahreg)<-c("y","x")
relasi <-lm(y~x)
new <- data.frame(x = BStanahreg$x)
BStanahreg$forecast <- predict(relasi,new)
```

Script 5.14 Script linear regresi

Untuk melihat plot dari linear regresi dapat menggunakan

```
plot(BStanahreg$x,BStanahreg$y)
abline(relasi)
```

Script 5.15 Script plot linear regresi

Kemudian dilakukan perhitungan dan seleksi lahan yang memiliki lahan yang dapat dioptimalkan untuk mengirimkan

tebu pada perusahaan dengan bantuan sql query melalui tools r menggunakan fungsi **sqldf()** seperti yang terlihat pada Script 5.16. Lahan optimal adalah lahan yang memiliki distance yang kurang dari nilai rata-rata distance dengan titik seharusnya.

```
BStanahreg$selisih <- abs(y-forecast)
mean(BStanahreg$selisih)
0.138805
BStanahpotensi <- sqldf("select * from BStanahreg where selisih < 0.138805")
View(BStanahpotensi)
```

Script 5.16 Script perhitungan dan pemilihan lahan optimal

Kemudian untuk mengetahui lahan yang memiliki performa yang buruk terlihat seperti Script 5.17

```
BSunderperf <-data.frame(sqldf::sqldf("select* from BStanahreg where
monetary < forecast"))
BSworst <-data.frame(sqldf::sqldf("select* from BSunderperf where selisih
ih > 0.138805"))
View(BSworst)
836/3809 # persenan yang keseluruhan
0.2194802
write.csv(BSworst,"D:/UniStuff/tugas akhir/bismillah/BSworst.csv")
```

Script 5.17 Script pemilihan lahan terburuk

5.6 Denormalisasi

Pada tahapan ini dilakukan pembalikan nilai dari normalisasi dan transformasi data untuk mempermudah proses analisis clustering. Proses denormalisasi terbagi menjadi dua bagian yaitu denormalisasi tahapan normalisasi dan denormalisasi data transform. Denormalisasi normalisasi terlihat pada Script 5.18 dan denormalisasi data transform terlihat seperti Script 5.19

```
#denormalisasi normalisasi min-max
BSdenorm$recencyB <- (1-BSdenorm$recency)
BSdenorm$frequencyD <- (BSdenorm$frequency *(max(BSnona$frequency)-min(BSnona$frequency)))+min(BSnona$frequency)
BSdenorm$recencyD <- (BSdenorm$recencyB *(max(BSnona$recency)-min(BSnona$recency)))+min(BSnona$recency)
BSdenorm$monetaryD <- (BSdenorm$monetary *(max(BSnona$monetary)-min(BSnona$monetary)))+min(BSnona$monetary)
```

Script 5.18 Script denormalisasi tahapan normalisasi

```
#denormalisasi log transform
BSdenormExp$recencyDE <-10^BSdenormExp$recencyD
BSdenormExp$monetaryDE <-10^BSdenormExp$monetaryD
BSdenormExp$frequencyDE <-10^BSdenormExp$frequencyD
```

Script 5.19 Script denormalisasi tahapan data transformasi

5.7 Visualisasi Cluster

Pada tahapan ini, hasil *cluster* akan dipresentasikan dalam bentuk grafik untuk mempermudah pemahaman dan analisa pada hasil *cluster* yang telah dilakukan sebelumnya. Grafik yang digunakan dalam penelitian ini adalah 3D *scatterplot*, *pie chart* dan *Box and whisker*. Visualisasi dihasilkan dengan bantuan *tools* Shiny dan Shiny Dashboard.

5.7.1 3D Scatter plot

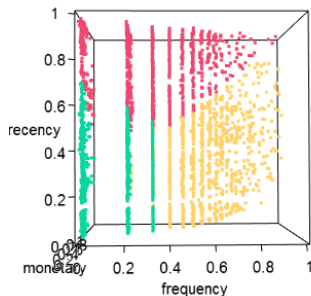
Grafik ini digunakan untuk menampilkan persebaran data sesuai dengan *clusternya*. Grafik ini juga dapat menampilkan alasan data-data tersebut termasuk dalam satu *cluster*. Selain itu juga, representasi dalam bentuk 3D dikarenakan terdapat 3 variabel yang mendasari pembentukan *cluster* yaitu *recency*, *frequency* dan *monetary*. *Package* yang digunakan untuk menggambarkan plot ini adalah *plot3D*. Script 5.20 adalah *script* yang digunakan untuk menampilkan plot tersebut.

```
plot3d(pat$normrec,
       pat$normf,
       pat$normm,
       col =c("#bbcdb", "#9ebd9e", "#dd855c", "#745151")[as.numeric(
pat$skls)],
       xlab = "recency",
       ylab = "frequency",
       zlab = "monetary" )
aspect3d(1,1,1)
```

Script 5.20 Script Plot 3D Scatterplot

Grafik 3D *scatter plot* akan nampak seperti Gambar 5.2 Grafik 3D Scatter plot. Sumbu x pada grafik ini adalah variabel *recency*. Sumbu y adalah variabel *frequency* dan sumbu z nya adalah variabel *monetary*. Titik yang tersebar pada grafik ini merepresentasikan nilai RFM yang dimiliki supplier tersebut.

Sementara, warna yang berbeda ditunjukkan untuk merepresentasikan masing-masing *cluster*.



Gambar 5.2 Grafik 3D Scatter plot

5.7.2 Pie chart

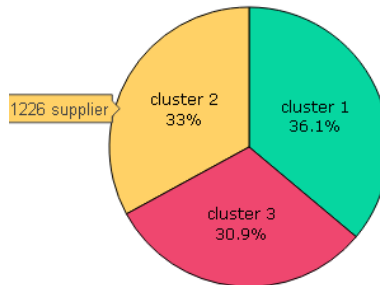
Grafik ini digunakan untuk menunjukkan presentase banyak anggota yang termasuk dalam satu *cluster*. Sehingga, perusahaan dapat menjadikan presentasi ini sebagai salah satu faktor dalam menentukan strategi untuk mengelola supplier kedepannya. *Package* yang digunakan untuk membuat grafik ini adalah *plotly*. Script 5.21 Script Pie chart adalah *script* yang digunakan untuk menghasilkan *pie chart* :

```
library(plotly)
dirColors <-c("1"="#bbcdb", "2"="#9ebd9e", "3"="#dd855c")
plot_ly(pie, labels = ~cluster, values = ~supplier, type = 'pie',
        textposition = 'inside',
        textinfo = 'label+percent',
        insidetextfont = list(color = '#FFFFFF'),
        hoverinfo = 'text',
        text = ~paste(supplier, 'supplier'),
        marker = list(colors = dirColors,
                     line = list(color = '#FFFFFF', width = 1)))
```

Script 5.21 Script Pie chart

Warna yang berbeda merepresentasikan *cluster* yang berbeda, data yang digunakan berdasarkan data banyaknya supplier yang

menjadi anggota dalam *cluster* tersebut. Pada grafik ini juga ditampilkan prosentase banyaknya anggota *cluster* tersebut dikeseluruhan supplier. *Pie chart* yang dihasilkan akan terlihat seperti Gambar 5.3 :



Gambar 5.3 Grafik Pie chart

5.7.3 Box and whisker

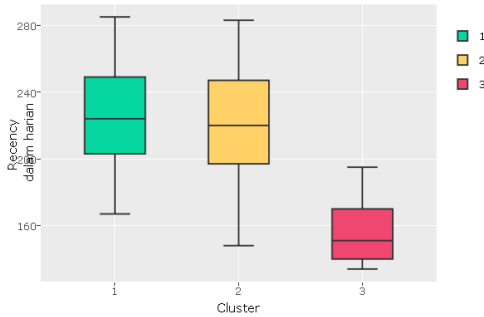
Grafik ini digunakan untuk membandingkan *range* setiap variabel *recency*, *frequency* dan *monetary* pada setiap *cluster*. Selain itu juga dengan menggunakan grafik ini, dapat diketahui nilai variabel RFM yang dimiliki oleh *cluster* tersebut seperti apa, apakah supplier kebanyakan memiliki nilai RFM yang sama atau tidak. Untuk membuat grafik ini menggunakan *package* *ggplot2* pada R. Script 5.22 adalah *script box and whisker*

```
library(ggplot2)
ggplot(tnprp,
  aes(x = tnprp$klaster, y = tnprp$recency, fill=tnprp$klaster))+
  geom_boxplot()+
  scale_fill_manual(values=c('#bbcdb', '#9ebd9e', '#dd855c'))+
  scale_y_continuous(name = "Recency")+
  scale_x_discrete(name = "Cluster")+
  labs(fill = "Cluster")
```

Script 5.22 Script Box and whisker

Data yang digunakan adalah variabel *recency*, *frequency* dan *monetary* pada setiap *cluster*. Sumbu x pada grafik ini adalah jenis *clusternya* dan sumbu y adalah variabel RFM. Pada grafik

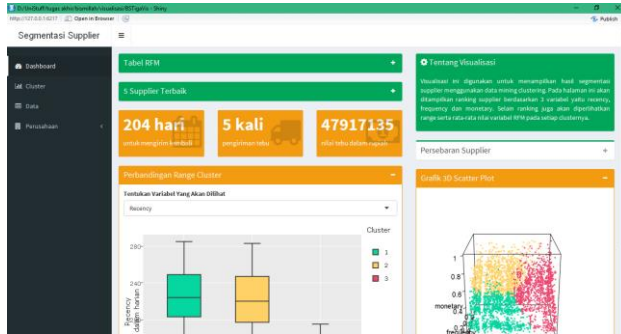
ini dilakukan pembedaan warna sesuai dengan *clusternya*. Untuk mempermudah pembacaan diberikan juga legend yang menampilkan warna yang sesuai dengan *clusternya*. Grafik *Box and whisker* yang dihasilkan akan tampak seperti Gambar 5.4



Gambar 5.4 Grafik *Box and whisker*

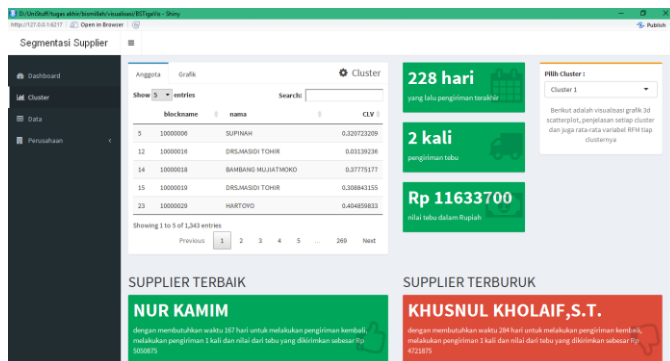
5.7.4 Shiny Dashboard

Shiny dashboard merupakan salah satu *package* yang ada pada *shiny* yang bertujuan untuk membantu visualisasi apa yang telah dikerjakan di R. *Shiny dashboard* pada penelitian ini digunakan untuk menampilkan keadaan data dan hasil dari *clustering*. Visualisasi ini terdiri dari tiga halaman. Untuk *script* yang digunakan akan disertakan pada halaman lampiran. Halaman pertama adalah visualisasi data secara umum. Pada halaman ini memuat tiga grafik yaitu grafik 3D *scatterplot*, *pie chart* dan *Box and whisker*. Selain grafik pada halaman ini juga disediakan halaman keterangan mengenai visualisasi dan juga tabel yang memuat informasi *cluster* secara keseluruhan. Tampilan *dashboard* terlihat seperti Gambar 5.5



Gambar 5.5 Halaman Dashboard A

Pada halaman selanjutnya yaitu halaman *cluster*, terdapat informasi yang berfokus pada satu *cluster* saja. Terdapat 3D *scatter plot* yang dapat menunjukkan alasan tersebut menjadi satu *cluster*, penjelasan mengenai *cluster*, anggota *cluster* tersebut dan juga keterangan mengenai variabel RFMnya secara khusus tentang *cluster* tersebut. Gambar halaman *cluster* nampak seperti Gambar 5.6



Gambar 5.6 Halaman Cluster

Pada halaman data, terdapat tabel yang berisikan nilai *recency*, *frequency* dan *monetary*. Tabel tersebut disertakan kolom *search* dan fungsi *sort* untuk melihat data yang digunakan. Tampilan halaman data tampak seperti Gambar 5.7

The screenshot shows a web browser window with the URL 'http://192.168.182.17'. The page title is 'Segmentasi Supplier'. On the left, there is a sidebar menu with options: 'Dashboard', 'Cluster', 'Data', and 'Perincikan'. The main content area displays a table with the following data:

	id	blockname	nama	recency	frequency	monetary	CLV	cluster
1	10000002	SUPPAH		137	6	28887150	0.815031025	3
2	10000003	SUPPAH		148	2	7202000	0.86424257	3
3	10000004	SUPPAH		136	2	9330000	0.728172989	3
4	10000005	SUPPAH		138	3	10682750	0.747825989	3
5	10000006	SUPPAH		201	1	3188375	0.320723209	1
6	10000007	RALHO		229	7	8053875	0.584210504	2
7	10000011	TARKADJI		238	4	19733375	0.288440588	2
8	10000012	TARKADJI		221	9	12480000	0.388922684	2
9	10000013	TARKADJI		232	16	12478000	0.438593306	2
10	10000014	TARKADJI		234	9	50851125	0.277880901	2

At the bottom of the table, it says 'Showing 1 to 10 of 3,720 entries'. There are navigation links for 'Previous', '1', '2', '3', '4', '5', '372', and 'Next'.

Gambar 5.7 Halaman Data

(halaman ini sengaja dikosongkan)

BAB VI HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan hasil dari perancangan dan implementasi yang telah disusun dalam tugas akhir ini. Bagian ini menjelaskan lingkungan uji coba, hasil *pra-processing* dari data, hasil dari setiap tahapan pembentukan model, hingga hasil peramalan.

6.1 Ekstraksi RFM

Pada tahapan ini akan dilakukan ekstraksi variabel RFM berdasarkan proses *data selection*. Variabel *recency* menggunakan tanggal timbang, variabel *frequency* menggunakan variabel tanggal timbang dan variabel *monetary* menggunakan variabel hasil gula. Tabel 6.1 adalah potongan hasil ekstraksi data *recency*.

Tabel 6.1 Ekstraksi Recency

blockname	recency (hari)
10000002	137
10000003	146
10000004	136
10000005	138
10000006	201

Pada Tabel 6.2 ditampilkan potongan hasil ekstraksi data variabel *frequency* yang dihasilkan

Tabel 6.2 Ekstraksi Frequency

blockname	frekuensi
10000002	6
10000003	2
10000004	2
10000005	3
10000006	1

Pada Tabel 6.3 ditampilkan potongan hasil ekstraksi data variabel *monetary*.

Tabel 6.3 Ekstraksi Monetary

blockname	<i>monetary</i>
10000002	26,808,750
10000003	7,202,000
10000004	7,630,500
10000005	10,682,750
10000006	3,180,375

6.2 Data Preparation

Pada bagaian ini akan dijelaskan hasil dari ketiga proses yang dilakukan pada data preparation yaitu data transformation, data cleansing dan normalisasi.

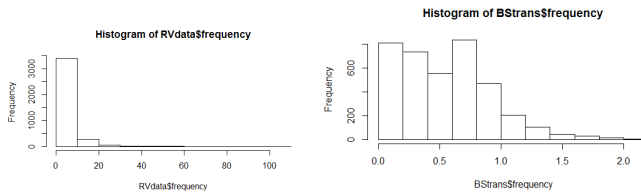
6.2.1 Data Transformation

Pada tahapan ini data mentah ditransformasi dengan metode logatimik karena adanya kecondongan pada variabel frequency dan monetary. Transformasi dilakukan pada ketiga variabel RFM. Tabel 6.4 merupakan hasil transformasi data.

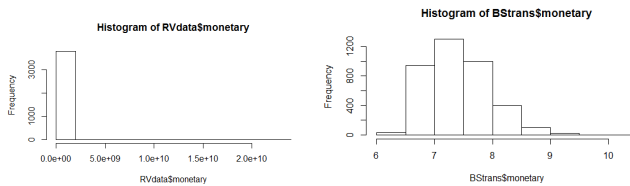
Tabel 6.4 Hasil Transformasi Data

blockname	frequency	monetary	recency
10000001	1.612783857	8.710080643	2.181843588
10000002	0.77815125	7.428276565	2.136720567
10000003	0.301029996	6.857453117	2.164352856
10000004	0.301029996	6.882552997	2.133538908
10000005	0.477121255	7.028683065	2.139879086

Setelah dilakukan proses transformasi data, dilakukan perbandingan sebelum dan sesudah data dengan grafik histogram seperti terlihat pada Gambar 6.1 dan Gambar 6.2 untuk melihat apakah distribusi data lebih baik dari sebelumnya.



Gambar 6.1 Perbandingan Histogram Sebelum dan Sesudah Transformasi pada Frequency



Gambar 6.2 Perbandingan Histogram Sebelum dan Sesudah Transformasi pada Monetary

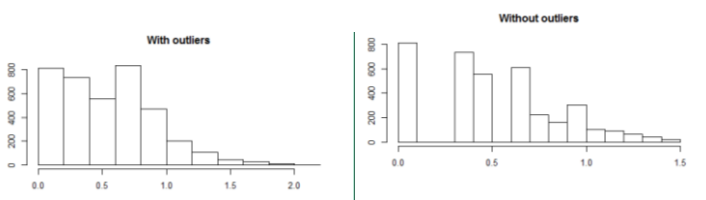
Dapat dilihat setelah dilakukan transformasi distribusi data tidak terlalu condong kekiri dan nilainya lebih tersebar.

6.2.2 Data Cleansing

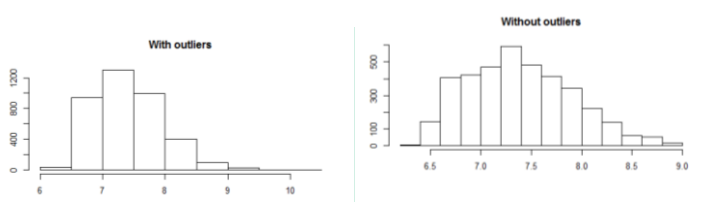
Data cleansing dilakukan untuk mempermudah proses *cluster*. *Script* yang digunakan pada *tools R* tidak hanya menghapus *outlier*, namun juga memberikan grafik perbandingan histogram antara jumlah banyak anggota dan nilai variabelnya. Pada variabel *recency* tidak ada *outlier* yang perlu dihapus. Sementara pada variabel *frequency*, proporsi *outlier* yang ada pada variabel tersebut adalah 1.7%. Sebanyak 66 data yang termasuk *outlier* dari variabel *frequency*. Seperti yang terlihat pada Gambar 6.3 data dapat terlihat persebarannya, bila sebelumnya terlihat bahwa anggota *supplier* memiliki range yang luas hingga 2, kini hanya menjadi nilainya 1.5.

Sementara pada variabel *monetary*, sebanyak 1% dari datanya adalah *outlier* yaitu 39 data. Pada Gambar 6.4 terlihat perbandingan yang sangat jauh dimana grafik dengan *outlier*

berkumpul rangenya sangat luas kini range nilainya telah dipersempit.



Gambar 6.3 Perbandingan Ada Tidaknya Outlier pada Frequency



Gambar 6.4 Perbandingan Ada Tidaknya Outlier pada Monetary

Pada Tabel 6.5 ditampilkan potongan data dimana nilai RFM telah digantikan dengan NA

Tabel 6.5 Tabel Hasil Penghapusan Outlier

no	Block name	recency	frequency	monetary
1	10000001	2.181843588	NA	8.710080643
2	10000002	2.136720567	0.77815125	7.428276565
3	10000003	2.164352856	0.301029996	6.857453117
4	10000004	2.133538908	0.301029996	6.882552997
5	10000005	2.139879086	0.477121255	7.028683065

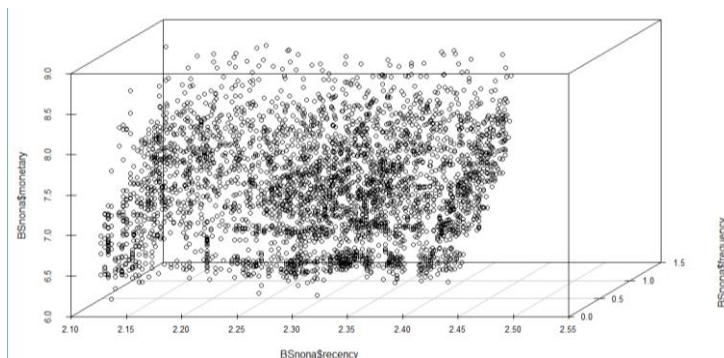
Setelah melakukan pembersihan *outlier* data terlihat lebih tersebar dimana akan memudahkan proses *clustering*. Namun, karena adanya NA membuat nilai tiap supplier akan terlihat ambigu sehingga perlu dilakukan penghapusan data yang memiliki nilai NA. Data yang dihapus berjumlah 89 data atau

sebesar 2% dari keseluruhan dan jumlah data menjadi 3720 dan tabel RFM akan terlihat pada Tabel 6.6

Tabel 6.6 Hasil Penghapusan Data yang Memiliki Nilai NA

no	blockname	recency	frequency	monetary
1	10000002	2.14	0.78	7.43
2	10000003	2.16	0.30	6.86
3	10000004	2.13	0.30	6.88
4	10000005	2.14	0.48	7.03
5	10000006	2.30	0.00	6.50

Setelah itu akan dilakukan pemetaan data pada grafik 3D *scatter plot* untuk melihat persebaran data menggunakan *tools R*. Gambar 6.5 adalah grafik dari hasil pembersihan *outlier*



Gambar 6.5 Grafik Penghapusan Data bernilai NA

6.2.3 Normalisasi Menggunakan Min-Max

Proses normalisasi ini menggunakan metode min-max, yaitu proses perubahan data yang mengubah nilai menjadi 0 hingga 1. Tabel 6.7 adalah hasil normalisasi menggunakan metode-normalisasi pada variabel *recency*.

Tabel 6.7 Normalisasi Recency

blockname	recency (normalisasi)
10000002	0.02933962
10000003	0.11365122
10000004	0.019631747
10000005	0.038976889
10000006	0.537289393

Variabel *recency* mengalami penyesuaian nilai karena berkebalikan dengan variabel *monetary* dan *frequency*, sehingga nilainya dibalik dengan pengurangan 1 – nilai normalisasi, hasilnya terlihat pada Tabel 6.8

Tabel 6.8 Normalisasi Recency Setelah Disesuaikan

blockname	recency (normalisasi)
10000002	0.97066038
10000003	0.88634878
10000004	0.980368253
10000005	0.961023111
10000006	0.462710607

Sementara untuk hasil normalisasi menggunakan metode-normalisasi pada variabel *frequency* adalah seperti terlihat pada Tabel 6.9

Tabel 6.9 Normalisasi Frequency

blockname	frekuensi (normalisasi)
10000002	0.521772319
10000003	0.201849087
10000004	0.201849087
10000005	0.319923233
10000006	0

Tabel 6.10 adalah hasil normalisasi menggunakan metode-normalisasi pada variabel *monetary*

Tabel 6.10 Normalisasi Monetary

Blockname	<i>Monetary</i> (normalisasi)
10000002	0.452809375
10000003	0.239635975
10000004	0.249009498
10000005	0.303581614
10000006	0.107071025

Setelah itu ketiga variabel dijadikan menjadi satu tabel sehingga tabelnya terlihat seperti yang terlihat pada Tabel 6.11

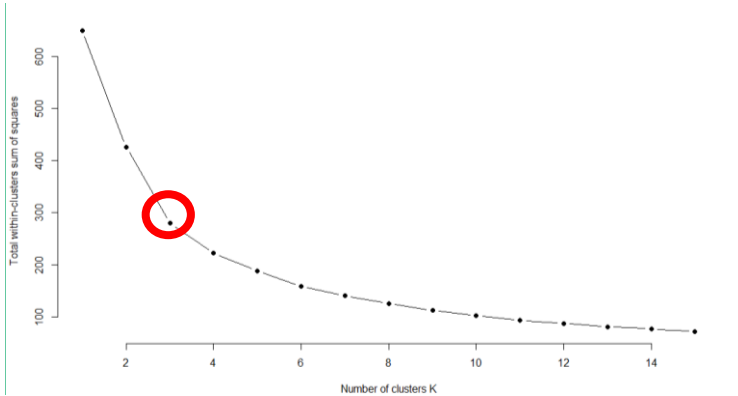
Tabel 6.11 Integrasi Tabel

blockname	<i>recency</i> (normalisasi)	frekuensi (normalisasi)	<i>monetary</i> (normalisasi)
10000002	0.97066038	0.521772319	0.452809375
10000003	0.88634878	0.201849087	0.239635975
10000004	0.980368253	0.201849087	0.249009498
10000005	0.961023111	0.319923233	0.303581614
10000002	0.462710607	0	0.107071025

6.3 Penentuan Jumlah Cluster

Penentuan nilai k menggunakan *elbow method*. Gambar 6.6 adalah hasil *elbow method* menggunakan *tools R*

Berdasarkan Gambar 6.6 penurunan terjadi secara drastis dan diikuti grafik yang stabil pada titik 3.



Gambar 6.6 Elbow method

6.4 Clustering

Pada tahap ini akan dilakukan proses *clustering* menggunakan metode K-Means. Hasil *cluster* yang didapatkan akan diberikan label nama sesuai dengan perilaku suppliernya.

6.4.1 Clustering Menggunakan K-Means

Proses *clustering* menghasilkan sebuah tabel yang disimpan dalam bentuk csv agar mudah dikelola. Hasil *cluster* dalam bentuk tabel akan terlihat pada Tabel 6.12

Tabel 6.12 Hasil Clustering

blockname	recency	frequency	monetary	k
10000002	0.521772319	0.452809375	0.97066038	3
10000003	0.201849087	0.239635975	0.88634878	3
10000004	0.201849087	0.249009498	0.980368253	3
10000005	0.319923233	0.303581614	0.961023111	3
10000006	0	0.107071025	0.462710607	1

Setelah itu data dipisahkan pada masing-masing *clusternya*. *Cluster* pertama memiliki 1343 anggota, kedua memiliki 1226 anggota dan ketiga memiliki 1151 anggota.

6.4.2 Uji Performa Menggunakan SSE

Apabila *script* SSE dijalankan maka hasil dari nilai SSE pada setiap titik adalah terlihat pada Tabel 6.13

Tabel 6.13 Hasil Uji SSE

Jumlah K	SSE	Selisih
1	649.3595547	
2	426.2241096	223.135445
3	279.5535454	146.6705643
4	222.4915729	57.06197247
5	187.954802	34.53677092

Tabel 6.12 menunjukkan adanya penurunan nilai yang besar dari 1-2 dan 2-3, akan tetapi dari titik 3 dan selanjutnya terjadi penurunan yang stabil. Sesuai dengan konsep elbow method, nilai K yang digunakan adalah 3. Selain itu juga nilai SSE yang memiliki nilai k sama dengan 3 lebih kecil dibandingkan dengan nilai SSE yang dimiliki oleh k =2

6.4.3 Uji Performa Menggunakan CValid

Uji uji performa dengan menggunakan *package* **cValid** (), digunakan untuk memuji performa nilai k yang digunakan. Connectivity mencari nilai yang terkecil karena mengartikan bahwa data tersebut telah dikelompokkan sesuai dengan clusternya. Dunn mencari nilai yang tertinggi karena mengartikan bahwa cluster tersebut semakin berbeda dengan cluster lainnya. Silhoutte memilih nilai yang tertinggi karena menunjukkan bahwa derajat kepercayaan mengenai kebenaran penempatan cluster semakin tinggi. Hasil yang dikeluarkan seperti Tabel 6.14

Tabel 6.14 Hasil Uji performa Menggunakan CValid

Validation Measure	Cluster	Score
Connectivity	2	175.8302
Dunn	10	0.0094
Silhoutte	3	0.3507

Bila dilihat dari nilainya, silhouette memiliki nilai yang sama dengan k yang dipilih menggunakan elbow method.

6.4.4 Verifikasi Hasil Clustering

Pada tahapan ini dilakukan perhitungan jarak setiap data dengan centroid masing-masing cluster dan cluster yang dipilih adalah yang memiliki cluster yang terdekat. Pengukuran jarak menggunakan menggunakan Euclidean distance.

Tabel 6.15 Hasil perhitungan distance pada setiap centroid

blockname	1	2	3	minVal	min
10000002	0.79	0.64	0.26	0.26	3
10000003	0.58	0.74	0.26	0.26	3
10000004	0.68	0.81	0.29	0.29	3
10000005	0.68	0.72	0.22	0.22	3
10000006	0.27	0.75	0.59	0.27	1

Hasil pemilihan cluster dibandingkan dengan hasil cluster dari `kmeans()` dan dicek apakah nilainya sama atau tidak menggunakan fungsi `identical()`. Script 6.1 adalah hasil dari menggunakan fungsi tersebut.

```
> identical(BSpersamaan$min, BSpersamaan$cluster)
[1] TRUE
```

Script 6.1 Hasil uji kesamaan hasil cluster

Hasil uji tersebut menunjukkan TRUE, benar adanya **bahwa nilai cluster dengan perhitungan jarak manual dan cluster hasil kmeans memiliki nilai yang sama**. Sehingga, bisa disimpulkan bahwa penempatan cluster telah benar.

Selain itu juga dilakukan perhitungan hasil cluster secara 5 kali untuk mengetahui apakah anggotanya sama apabila dilakukan proses clustering kembali. Tabel 6.16 adalah perbandingan jumlah anggota cluster dalam beberapa kali percobaan.

Tabel 6.16 Hasil Perbandingan Jumlah Anggota Cluster

Cluster	coba 1	coba 2	coba 3	coba 4	coba 5
1	1343	1226	1226	1343	1226
2	1226	1151	1151	1151	1151
3	1151	1343	1343	1226	1343

Dapat dilihat bahwa jumlah anggota setiap cluster itu sama pada setiap kali percobaan. Jumlahnya mutlak 1343, 1226 dan 1151, hanya saja hasil tersebut ditempatkan di cluster yang berbeda pada percobaan lainnya. **Hal ini menandakan bahwa anggota cluster tersebut benar-benar milik cluster tersebut, karena jumlah anggota cluster tidak berganti namun tetap.**

Sehingga, bila disimpulkan baik dengan percobaan melalui pengukuran Euclidian distance dan juga percobaan clustering secara berulang, hasilnya menunjukkan hasil anggota data tersebut telah benar benar di cluster tersebut.

6.5 Pembobotan RFM Menggunakan AHP

Pada tahap ini akan dilakukan pembobotan pada tiap variabel yang mencerminkan kepeningan perusahaan. Pembobotan ini dilakukan berdasarkan wawancara/ kuisisioner yang dilakukan pada pihak yang terkait dalam supplier perusahaan. Wawancara dilakukan pada 4 orang yang berhubungan dengan supplier tebu, yaitu bagian tanaman, *quality assurance* dan pengkreditan & bagi hasil. Tabel 6.17 adalah hasil wawancara yang didapatkan oleh masing-masing bagian.

Tabel 6.17 Hasil Wawancara Perbandingan Kriteria

Perbandingan Kriteria		Keterangan
<i>Recency</i>	<i>Frequency</i>	<i>Recency</i> 4 kali lebih penting daripada <i>frequency</i>
<i>Frequency</i>	<i>Monetary</i>	<i>Frequency</i> 3 kali lebih penting daripada <i>monetary</i>
<i>Monetary</i>	<i>Recency</i>	<i>Monetary</i> 1/5 kali lebih penting daripada <i>recency</i>

6.5.1 Perhitungan AHP

Hasil kuisioner tersebut dituangkan dalam bentuk tabel *pairwise comparison* untuk memudahkan perbandingan antara *recency*, *frequency* dan *monetary*. Setelah itu akan dilakukan perhitungan *preference vector* untuk mengetahui prioritas pada variabel RFM. Pertama-tama nilai pada setiap kolom ditambahkan dan mendapatkan nilai total kesuluruhan perkolomnya. Kemudian nilai pada *pairwise comparison* dibagi dengan nilai total dari tiap kolomnya, hasil dari tahap ini disebut tabel yang telah dinormalisasi.

Kemudian, rata-ratakan semua nilai pada satu baris pada tabel normalisasi. *Preference vector* atau bobot dari variabel tersebut adalah hasilnya. Tabel 6.18 adalah hasil perhitungan bobot

Tabel 6.18 Bobot Variabel RFM

<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
0.67	0.23	0.1

Sehingga bobot untuk *recency* adalah 0.67 kemudian *frequency* adalah 0.23 dan *monetary* adalah 0.1. Setelah menemukan bobot dari variabel RFM hasil AHP akan diuji kekonsistennya. Bila dilihat dari nilainya maka derajat kepentingan tertinggi dimiliki oleh *recency*, *frequency* dan *monetary*.

6.5.2 Uji Konsistensi AHP

Pengujian konsisten atau tidaknya suatu variabel menggunakan *Consistency Ratio* (CR) pada hasil survey. Hasil CR dari hasil kuisioner AHP tidak boleh melebihi dari 0.1. Bila, melebihi nilai 0.1 maka nilai tersebut disebut tidak konsisten. Nilai CR didapatkan dengan membagi nilai *consistency index* (CI) dengan *random consistency index* (RI). Kriteria yang digunakan dalam AHP ini terdapat 3 kriteria dan sesuai dengan tabel RI maka nilai RI nya adalah 0.58. Sementara untuk perhitungan CI maka pertama-tama perlu didapatkan dulu total pengalihan nilai variabel matriks *pairwise comparison* dengan bobot tiap variabelnya. Penjumlahan tersebut dilakukan pada tiap barisnya sehingga menghasilkan Tabel 6.19

Tabel 6.19 Hasil Penjumlahan Tiap Variabel

	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	Total
<i>Recency</i>	0.67	0.92	0.52	2.11
<i>Frequency</i>	0.17	0.23	0.31	0.71
<i>Monetary</i>	0.13	0.08	0.10	0.31

Kemudian, total tersebut dikalikan dengan bobot pada setiap variabelnya. Hasil ketiganya kemudian dirata-ratakan. Hasil rata-rata ini akan digunakan untuk mencari nilai CI yang direpresentasikan sebagai nilai x . Berikut adalah perhitungan untuk mencari nilai x ;

$$\begin{aligned} \text{nilai} &= \frac{((2.11 \times 0.67) + (0.71 \times 0.23) + (0.31 \times 0.1))}{3} \\ &= \mathbf{3.086949408} \end{aligned}$$

Setelah mendapatkan nilai x maka perhitungan untuk mencari *consistency index* (CI). Sehingga nilai CI yang didapatkan adalah :

$$C_i = \frac{3.086949408 - 3}{3 - 1} = \mathbf{0.0435}$$

Nilai CR didapatkan dengan membagi nilai CI dengan nilai RI, sehingga nilai CRT adalah :

$$CR = \frac{CI}{RI} = \frac{0.435}{0.58} = \mathbf{0.075}$$

Consistency rate yang dihasilkan adalah 0.075. Nilai tersebut kurang dari 0.1 sehingga dapat dikatakan bahwa hasil dari AHP telah konsisten.

6.6 Perhitungan Nilai CLV

Pada tahap ini ketiga rata-rata variabel RFM yaitu *recency*, *frequency* dan *monetary* dikalikan dengan bobot yang didapatkan dari perhitungan AHP. Bobot untuk *recency* adalah 0.67, *frequency* adalah 0.23 dan *monetary* adalah 0.1.

Kemudian ketiga variabel RFM ditambahkan. Setelah dilakukan perkalian data akan berubah menjadi Tabel 6.20

Tabel 6.20 Hasil Perkalian Data dengan Bobot

<i>Cluster</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	<i>CLV</i>	<i>Rank</i>
1	0.21	0.03	0.03	0.27	3
2	0.23	0.13	0.06	0.42	2
3	0.54	0.07	0.05	0.66	1

Cluster supplier terbaik adalah *cluster* 3 dimana nilai *recency* dari *cluster* ini tertinggi dibandingkan dengan *cluster* lainnya sehingga menyebabkan nilai CLVnya tinggi. Kemudian diikuti dengan *cluster* 2 dengan posisi kedua, dimana *cluster* ini memiliki nilai *frequency* dan *monetary* yang tertinggi dibanding *cluster* lainnya. Kemudian diikuti *cluster* terburuk adalah *cluster* 1 dimana *cluster* ini memiliki nilai RFM yang terkecil dibanding *cluster* lainnya CLV yang terkecil.

Pembobotan yang dilakukan pada variabel *recency*, frekuensi dan *monetary* sangat berpengaruh dengan ranking *cluster* saat ini. Berdasarkan hasil AHP *recency* memiliki bobot tertinggi diantara dua variabel lainnya. Variabel *frequency* memiliki bobot tertinggi kedua dan variabel *monetary* menempati peringkat terakhir dalam peringkat kepentingan dengan bobot. Sehingga kurang lebih ranking *cluster* berdasarkan CLV sama dengan ranking *cluster* berdasarkan *recency*.

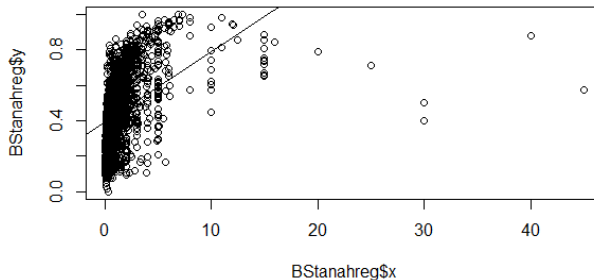
6.7 Analisa Uji Korelasi dan Perhitungan Lahan Optimal

Selain melakukan analisa terhadap hasil *clustering* dengan menggunakan RFM analisis, dilakukan juga analisa lainnya, yaitu analisa hubungan antara luas area tanam dengan variabel *monetary*. Analisa ini dilakukan setelah melakukan uji korelasi. Tabel 6.21 adalah nilai yang didapatkan dari uji korelasi dan interpretasi hubungan variabel tersebut :

Tabel 6.21 Hasil Uji Korelasi

Variabel		Cor	Korelasi	Interpretasi
area	Monetary	0.4134983	positif	Korelasi rendah

Berdasarkan Tabel 6.21 dapat dilihat bahwa **hubungan antara area terhadap nilai monetary sangat rendah**. Untuk membuktikannya, dapat digunakan linear regression untuk mengetahui jumlah gula yang harusnya dihasilkan dengan luas area tertentu. Gambar 6.7 adalah plot yang dihasilkan penggambaran data actual dan garis adalah yang seharusnya terjadi.



Gambar 6.7 Plot Area- Monetary

Berdasarkan perhitungan *linear regression* fungsi yang dihasilkan untuk memperoleh prediksi gula adalah :

$$\text{prediksiMon} = 0.38897 + 0.04003(\text{area})$$

Rumus tersebut menunjukkan variabel monetary (hasil gula) yang tidak dipengaruhi oleh gula atay saat luas area sebesar 0, maka monetarynya sebesar 0.389 yang ditunjukkan oleh *intercept*nya. Nilai *slope* sebesar 0.040 mengartikan monetary memiliki hubungan positif dengan luas area, karena koefisiennya nilainya positif. Setiap peningkatan luas area akan berpengaruh dengan peningkatan nilai monetary sebesar 0.04.

Tabel 6.22 menunjukkan hasil dari prediksi monetary berdasarkan fungsi yang dibentuk dari *intercept* dan *slope*. Pada Tabel 6.22 juga terdapat selisih nilai actual dan prediksi yang telah dimutlakan untuk mencari tahu jarak antara titik aktual dengan titik prediksinya.

Tabel 6.22 Hasil Prediksi Linear Regresi

Blockname	Monetary	area	forecast	selisih
10000002	0.45	0.60	0.41	0.04
10000003	0.24	0.17	0.40	0.16
10000004	0.25	0.26	0.40	0.15
10000005	0.30	0.22	0.40	0.09
10000006	0.11	0.27	0.40	0.29

Dari hasil selisih itu dicari lahan yang memiliki selisih yang kurang dari rata-rata selisih secara keseluruhan yaitu 0.138805. Lahan tersebut yang bisa dikatakan sebagai lahan yang dapat dioptimalkan hasilnya. Tabel 6.23 adalah potongan data lahan yang dapat dioptimalkan.

Tabel 6.23 Tabel Lahan yang Dapat Dioptimalkan

Blockname	monetary	area	forecast	selisih
10000002	0.45	0.60	0.41	0.04
10000005	0.30	0.22	0.40	0.09
10000011	0.40	0.27	0.40	0.00
10000015	0.45	0.48	0.41	0.05
10000017	0.44	0.69	0.42	0.02

Setelah dihitung terdapat 2027 lahan yang dapat dioptimalkan untuk menghasilkan hasil gulanya. Lahan yang dioptimalkan sebanyak 54% dari keseluruhan.

Setelah itu ingin diketahui lahan mana yang kinerjanya buruk dengan melihat lahan yang jauh dari nilai rata-rata. Sesuai dengan Tabel 6.24 terdapat 836 lahan atau 22% yang memiliki performa yang buruk. Lahan yang memiliki performa yang

buruk adalah lahan yang memiliki nilai monetary kurang dari peramalan dan juga nilainya kurang dari batasan toleransi.

Tabel 6.24 Tabel Lahan Terburuk

blockname	monetary	area	forecast	selisih
10000003	0.239636	0.17	0.395774983	0.156139008
10000004	0.24901	0.261	0.399418074	0.150408576
10000006	0.10707	0.266	0.399618244	0.292547218
10000016	0.18843	0.203	0.397096104	0.208662861
10000018	0.257164	0.227	0.398056919	0.140892072

6.8 Analisa Clustering

Setelah dilakukan pengelompokan, setiap nilai *recency*, *frekuensi* dan *monetary* dikembalikan dengan nilai awalnya sebelum normalisasi. Hal ini dilakukan untuk mempermudah dalam melakukan analisis hasil *cluster* tersebut.

Kemudian data pada setiap *cluster* diolah untuk melihat *range* nilai setiap variabel pada setiap *cluster*. *Range* nilai dicari menggunakan fungsi **min ()** dan **max ()** pada *tools excel*. Tabel 6.25 menampilkan *range* pada setiap variabelnya berdasarkan data keseluruhan.

Tabel 6.25 Deskripsi Data

	<i>recency</i>	frekuensi	<i>monetary</i>
average	204	5	47,917,135
min	134	1	1,643,500
max	285	31	782,550,000

Sementara bila dipetakan berdasarkan *cluster*, tabel hasil *cluster* akan terlihat pada Tabel 6.26

Tabel 6.26 Range Nilai Variabel Per Cluster

<i>cluster</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>	<i>rank</i>
1	167-285	1-3	1,845,125-293,635,250	3
2	148-283	2-31	12,644,625-771,163,000	2
3	134-195	1-28	1,643,500-782,550,000	1

Supplier terbaik adalah yang memiliki nilai *recency* terendah, karena itu menandakan transaksi terakhir yang mereka lakukan lebih dekat dengan tanggal saat analisis *cluster* dilakukan. Selain itu juga, perusahaan memprioritaskan variabel *recency* dalam penilaian supplier dengan bobot yang diberikan pada variabel ini adalah 67%. Sementara itu variabel lainnya frekuensi dan *monetary* akan semakin baik apabila nilainya semakin tinggi. Frekuensi yang tinggi menandakan pengiriman tebu yang sering dan *monetary* yang tinggi bahwa tebu yang dibawa oleh supplier tersebut memiliki nilai yang tinggi.

6.8.1 Cluster 1

Tabel 6.27 merupakan tabel kesimpulan hasil analisis *cluster* 1 yang berisikan karakteristik dan ranking supplier berdasarkan nilai CLV.

Tabel 6.27 Analisis Cluster 1

Ranking CLV	3	
Jumlah Anggota	1343	
Karakteristik	Recency	167-285 (228 hari)
	Frequency	1-3 (2 kali)
	Monetary	Rp1,845,125-Rp293,635,250 (Rp11,633,700)

Cluster satu merupakan *cluster* terburuk dibandingkan *cluster* lainnya berdasarkan nilai CLVnya. Transaksi terakhir dilakukan supplier *cluster* ini berkisar 228 hari dari ekstraksi variabel RFM atau 8 bulan yang lalu yang jatuh pada bulan Agustus. Frekuensi pengiriman tebu juga sangat rendah yaitu hanya 2 kali dalam periode tersebut. Nilai tebu dalam rupiahnya bila dibandingkan dengan supplier lainnya juga sangat rendah

yaitu hanya berjumlah Rp11,633,700. Bila dibandingkan dengan rata-rata keseluruhan supplier, *cluster* ini ketiga variabelnya masing-masing berada dibawah rata-rata.

6.8.2 Cluster 2

Tabel 6.28 merupakan tabel kesimpulan hasil analisis *cluster* 1 yang berisikan karakteristik dan ranking supplier berdasarkan nilai CLV.

Tabel 6.28 Analisis Cluster2

Ranking CLV	2	
Jumlah Anggota	1226	
Karakteristik	Recency	148-283 (222 hari)
	Frequency	2-31(8 kali)
	Monetary	Rp12,644,625-Rp771,163,000 (Rp82,560,944)

Cluster kedua merupakan *cluster* yang memiliki jumlah pengiriman tebu terbanyak dan nilai tebu dalam rupiah yang tertinggi dibandingkan supplier lainnya. Jumlah pengiriman yang dilakukan *cluster* ini berkisar sebanyak 8 kali dan nilai dari tebu yang dikirimkan adalah sebesar Rp82,560,944. Namun transaksi terakhir yang dilakukan *cluster* ini dibandingkan dengan *cluster* 3 lebih lama jaraknya, yaitu 7 bulan yang lalu dari waktu ekstraksi yaitu bulan September. Menjadikan nilai CLV nya lebih rendah dibandingkan dengan *cluster* 1 karena nilai *recency*nya lebih rendah walaupun nilai *frequency* dan *monetary*nya tinggi.

6.8.3 Cluster 3

Tabel 6.29 Analisis Cluster 3

Ranking CLV	1	
Jumlah Anggota	1151	
Karakteristik	Recency	134-195 (156 hari)
	Frequency	1-28 (4 kali)
	Monetary	Rp1,643,500-Rp782,550,000 (Rp53,351,836)

Cluster tiga merupakan *cluster* terbaik dibanding dua *cluster* lainnya. Hal ini dikarenakan transaksi pengiriman terakhir dilakukan 156 hari atau kurang lebih 5 bulan lalu dari waktu ekstraksi RFM yaitu bulan Oktober. Dibandingkan *cluster* lainnya *cluster* ini adalah *cluster* yang terakhir melakukan transaksi pengiriman tebu. Sementara itu, jumlah pengiriman tebu pada *cluster* ini pada periode Juni - November 2016 sebanyak 4 kali. Frekuensi pengiriman pada *cluster* ini lebih baik dari *cluster* 1 tapi tidak sebaik *cluster* 2. Untuk nilai tebu yang dikirimkan oleh *cluster* ini rata-rata sebanyak Rp53,351,836. Perbandingan dengan *cluster* lainnya sama dengan variabel frekuensi, dimana *cluster* 2 lebih baik daripada *cluster* ini. *Cluster* ini menjadi *cluster* terbaik bagi perusahaan dikarenakan nilai *recency*-nya yang tinggi dimana perusahaan lebih mementingkan supplier yang melakukan transaksi terkini.

6.9 Analisa Hasil Visualisasi

Hasil *clustering* kemudian ditampilkan dalam bentuk grafik untuk mempermudah pembacaan data. Berikut adalah analisa data dalam bentuk visualisasi.

6.9.1 Scattered Plot

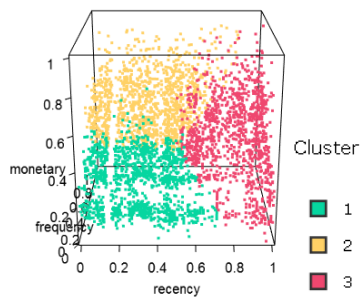
Grafik 3D *scatter plot* digunakan untuk melihat persebaran supplier dan bagaimana supplier tersebut berkelompok. 3D *scatter plot* terdiri dari tiga sumbu yaitu nilai *recency*, nilai *frequency* dan nilai *monetary*.

Gambar 6.8 merupakan 3d *scatter plot* dengan sudut pandang dari variabel *recency*. *Cluster* 1 yang direpresentasikan dengan warna hijau nampak berkumpul pada range *recency* yang rendah yaitu berkisar dari 0 – 0.5 mengartikan bahwa *cluster* ini transaksi terakhir yang dilakukan pada waktu yang lebih lampau dibandingkan *cluster* lainnya. *Cluster* 2 yang digambarkan dengan warna kuning, berkumpul pada range *recency* yang rendah yaitu berkisar dari 0 – 0.5, namun beberapa datanya tersebar dalam range yang tinggi, menunjukkan *cluster* ini memiliki nilai *recency* yang lebih tinggi dari *cluster* 1. Sementara *cluster* 3 yang digambarkan dengan

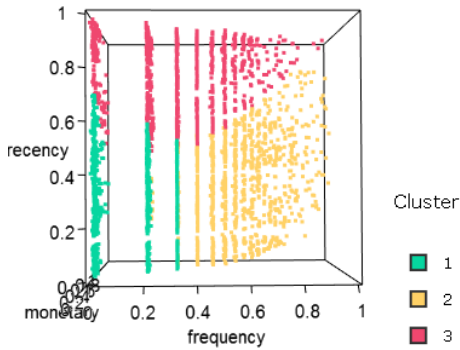
warna merah, terlihat range nilai *recency* yang tinggi yaitu berkisar dari 0.5 hingga 1 mengartikan bahwa *cluster* ini adalah *cluster* yang melakukan transaksi dalam waktu dekat

Gambar 6.9 merupakan 3d *scatterplot* dengan sudut pandang *frequency*. *Cluster* 1 yang digambarkan dengan warna hijau terletak pada range nilai yang rendah, berkisar dari 0-0.3 mengartikan frekuensi pengirimannya jarang. *Cluster* 2 yang direpresentasikan dengan warna kuning terletak pada range nilai yang tinggi dari 0.4-1, mengartikan cluster ini memiliki nilai *frequency* yang tinggi. *Cluster* 3 yang digambarkan dengan warna merah nampak tersebar luas, akan tetapi kebanyakan nilainya berada pada range yang rendah, mengartikan bahwa frekuensi pengiriman tebunya lebih rendah daripada cluster 2.

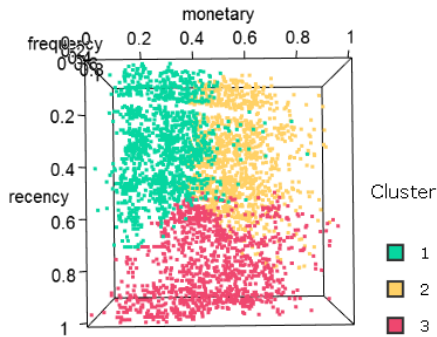
Gambar 6.10 merupakan 3d *scatterplot* dari sudut pandang *monetary*. Warna hijau yang menggambarkan *cluster* 1, berkumpul pada range nilai *monetary* yang rendah, mengartikan nilai rupiah dari tebu yang dikirimkan oleh supplier tersebut paling kecil dibandingkan cluster lainnya. *Cluster* 2 yang direpresentasikan dengan warna kuning, supplier pada *cluster* ini berkumpul pada range *monetary* yang tinggi, mengartikan bahwa nilai tebu yang dikirimkan oleh supplier ini paling tinggi dibandingkan dengan cluster lainnya. Sementara *cluster* 3 yang direpresentasikan dengan warna merah tersebar mengartikan bahwa supplier pada *cluster* ini nilainya beragam.



Gambar 6.8 Grafik 3D Scatter plot Recency



Gambar 6.9 Grafik 3D Scatter plot Frequency



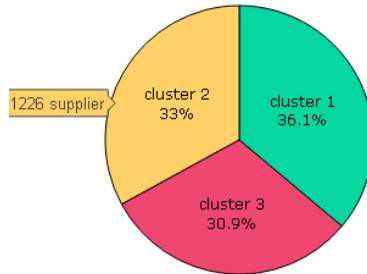
Gambar 6.10 Grafik 3D Scatter plot Monetary

6.9.2 Pie chart

Grafik ini digunakan untuk menjelaskan persebaran supplier pada *clusternya*. Grafik dibuat untuk menggunakan bantuan package **ggplot()**.

Prosentasi supplier pada masing-masing *clusternya* ditampilkan dengan grafik *pie chart* seperti yang nampak pada Gambar 6.11. *Cluster* 1 yang digambarkan dengan warna hijau memiliki anggota sebesar 36.1% dari jumlah supplier keseluruhan, bila dilihat prosesntasena, *cluster* ini adalah *cluster* yang memiliki anggota terbanyak dengan anggota sebesar 1343 supplier. *Cluster* 2 yang digambarkan dengan warna kuning memiliki

1226 anggota atau 33% dari keseluruhan. *Cluster 3* merupakan *cluster* dengan anggota paling sedikit dibanding dua *cluster* lainnya. Warna merah merepresentasikan *cluster 3*, anggota sebanyak 1151 supplier atau 30.9% dari keseluruhan.



Gambar 6.11 Grafik Prosentasi Anggota Setiap Clutser

6.9.3 *Box and whisker*

Grafik ini digunakan untuk menunjukkan range nilai variabel tiap *clusternya*. Grafik *Box and whisker* untuk variabel *recency* ditampilkan pada Gambar 6.12

Tabel 6.30 menampilkan nilai minimal, quartil satu, median atau rata-ratanya, quartil 3 dan nilai maximum dari variabel *frequency* pada setiap *clusternya*.

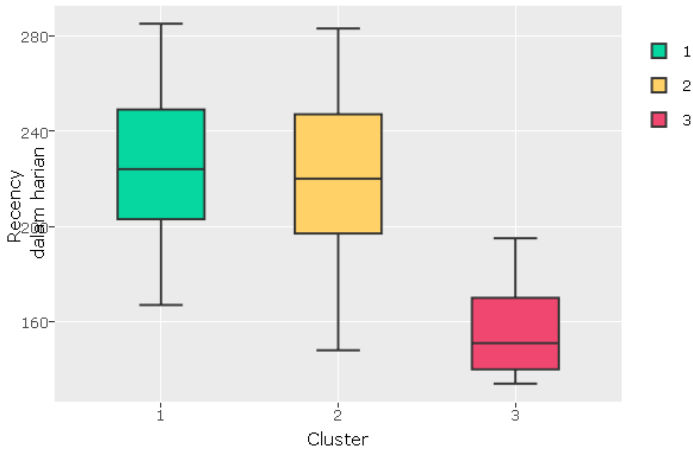
Tabel 6.30 Tabel Range Recency

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Max	285	283	195
Q3	249	247	170
Median	224	220	151
Q1	203	197	140
Min	167	148	134

Berdasarkan Gambar 6.12 berikut adalah analisis untuk variabel *recency* :

- Box *cluster 3* berada dibawah boxclusterlainnya menandakan bahwa nilai *recency* cluster 3 memiliki nilai *recency* yang pendek

- Bagian atas box cluster 3 lebih panjang daripada bagian bawahnya, hal ini menandakan bahwa data ini termasuk *skewed right* yang mengartikan nilai dibawah mean lebih seragam sementara 25% diatas mean nilainya lebih beragam
- Cluster 2 memiliki box and whisker yang lebih panjang dari cluster lainnya mengartikan nilai dari cluster lebih beragam
- Cluster 1 terletak diatas cluster atas cluster lainnya yang mengartikan bahwa cluster ini selisih waktu transaksi terkahir dan waktu ekstraksi RFM sangat besar atau bisa dibilang transaksi terkahir terjadi saat lampau



Gambar 6.12 Box and whisker Recency

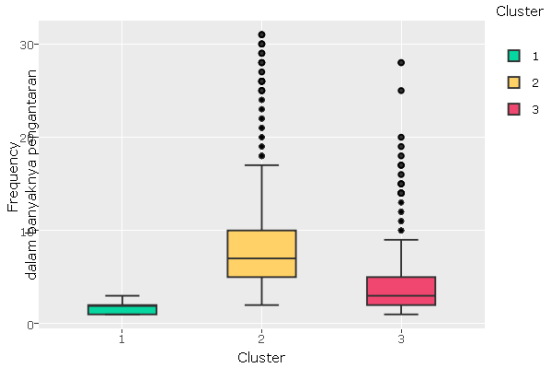
Tabel 6.31 menampilkan nilai minimal, quartil satu, median atau rata-ratanya, quartil 3 dan nilai maximum dari variabel *frequency* pada setiap *clusternya*.

Tabel 6.31 Tabel Range Frequency

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Max	3	17	9
Q3	2	10	5
Median	2	7	3
Q1	1	5	2
Min	1	2	1

Sementara untuk variabel *frequency*, berikut adalah analisis yang didapat dari Gambar 6.13:

- Dapat dilihat bahwa terdapat titik pada cluster 2 dan 3, hal ini menunjukkan bahwa terdapat *outlier* dimana nilai dari *outlier* tersebut lebih dari nilai maksimum atau lebih dari $3/2$ IQR
- Cluster 1 berada dibawah cluster lainnya mengartikan nilai frequencynya lebih rendah dari cluster lainnya
- Box cluster 2 berada diatas cluster lainnya, mengartikan bahwa cluster ini memiliki nilai frequency lebih besar dari yang lain
- Cluster 2 memiliki panhang box and whisker yang lebih panjang dari yang lain, mengartikan bahwa nilai frequency ini paling beragam
- Cluster 1 tidak memiliki box dan whisker bagian bawah mengartikan nilai minimal, quartil 1 sama dengan nilai median
- Box bagian atas cluster 3 lebih panjang dibandingkan box bagian bawah mengartikan bahwa 25% data diatas mean nilainya beragam
- Cluster 1 merupakan box and whisker terpendek mengartikan nilai frequency pada cluster ini paling seragam



Gambar 6.13 Box and whisker Frequency

Tabel 6.32 menampilkan nilai minimal, quartil satu, median atau rata-ratanya, quartil 3 dan nilai maximum dari variabel *frequency* pada setiap *clusternya*.

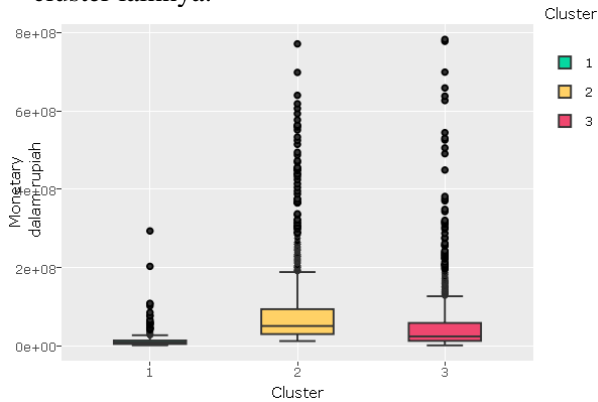
Tabel 6.32 Tabel Range Monetary

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Max	27,525,625	188,764,625	127,003,000
Q3	14,122,781	94,009,000	58,827,531
Median	8,987,875	51,096,000	24,554,500
Q1	5,159,968	30,468,875	13,136,093
Min	1,845,125	12,644,625	1,643,500

Terakhir untuk variabel *monetary*, berikut adalah analisis yang ditarik dari Gambar 6.14:

- Semua cluster terdapat titik pada bagian atas yang berarti semua cluster memiliki nilai yang lebih dari nilai maksimumnya atau lebih dari $3/2$ IQ
- Cluster 1 memiliki box and whisker yang terpendek mengartikan nilai yang ada di cluster ini paling seragam
- Baik cluster 2 dan cluster 3 memiliki bagian box atas yang lebih besar daripada bagian bawah mengartikan bahwa 25% data dibawah median memiliki nilai yang beragam

- Cluster 1 berada dibawah cluser lainnya mengartikan nilai monetary cluster ini terendah dibanding yang lain
- Cluster 2 terletak diatas cluster lainnya mengartikan nilai monetary cluster ini lebih tinggi dibandingkan cluster lainnya.



Gambar 6.14 Box and whisker Monetary

6.10 Kesimpulan Penelitian

Setelah melakukan proses *clustering* dan visualisasi dapat disimpulkan bahwa :

1. Dengan menggunakan *elbow method*, jumlah *cluster* yang digunakan pada penelitian ini berjumlah 3 *cluster*
2. *Cluster* 1 memiliki anggota supplier sebesar 1343 yang merupakan *cluster* dengan prosentase supplier terbanyak yaitu 36.1% *cluster* 2 memiliki anggota sebesar 1226 dan *cluster* 3 memiliki anggota paling sedikit dibanding lainnya dengan prosentase 30.9% dengan anggota sebanyak 1151
3. Pengaruh *recency* paling besar dibanding kedua variabel lainnya yaitu sebesar 67%. Sementara pengaruh variabel *frequency* sebesar 23% dan *monetary* sebesar 10 %. Nilai ini didapatkan setelah melakukan perhitungan AHP

4. *Cluster 3 cluster* terbaik bagi perusahaan berdasarkan nilai CLV, diikuti *cluster 2* dan terakhir adalah *cluster 1*
5. *Cluster 1* adalah *cluster* terburuk dibandingkan *cluster* lainnya, baik nilai *recency*, *frequency* dan *monetary*nya sangat kecil dan bila dibandingkan dengan nilai keseluruhan nilainya dibawah rata-rata
6. *Cluster 2* adalah *cluster* dengan frekuensi pengiriman tersering dan memiliki nilai tebu yang tinggi
7. *Cluster 3* adalah *cluster* yang melakukan transaksi paling terkini dan menjadikan *cluster* ini *cluster* terbaik
8. Berdasarkan uji korelasi, hubungan nilai *monetary* dengan luas area rendah namun positif
9. Lahan yang dioptimalkan sebanyak 54% dari keseluruhan.
10. Terdapat 22% dari lahan keseluruhan yang memiliki performa yang buruk

6.11 Rekomendasi

Berdasarkan analisis yang dilakukan pada supplier perusahaan, rekomendasi yang dapat diberikan perusahaan adalah perusahaan perlu menaruh perhatian pada *cluster 1*, karena *cluster* ini memiliki porsi terbesar yaitu 36.1% dalam supplier secara keseluruhan, namun nilai CLVnya masih sangat rendah, sehingga supplier perlu memberikan strategi untuk meningkatkan nilai CLV anggota supplier *cluster 1*.

BAB VII

KESIMPULAN DAN SARAN

Pada bagian ini berisi kesimpulan dari semua proses yang telah dilakukan dalam tugas akhir serta saran yang dapat diberikan untuk pengembangan kedepannya yang lebih baik.

7.1 Kesimpulan

Berdasarkan uji coba yang telah dilakukan dalam tugas akhir ini, maka kesimpulan yang dapat diambil adalah :

1. *Clustering* dapat membantu perusahaan menampilkan perilaku bisnis supplier dilihat dari waktu melakukan pengiriman tebu kembali, jumlah pengiriman tebu yang dilakukan dalam periode tertentu dan juga nilai tebu dalam nilai rupiah (RFM).
2. Visualisasi 3D *scatterplot* dapat membantu perusahaan melihat persebaran supplier berdasarkan *clusternya*. *Pie chart* dapat membantu supplier mengetahui prosesntasi supplier pada tiap *cluster* dan *Box and whisker* dapat membantu menganalisis range variabel RFM pada setiap *clusternya*.

7.2 Saran

Dalam pengerjaan tugas akhir ini terdapat beberapa kekurangan yang dapat diperbaiki dan ditingkatkan lagi. Oleh karena itu, berikut adalah beberapa saran yang dapat dipertimbangkan :

1. Penelitian dapat dilanjutkan dengan klasifikasi berdasarkan hasil clustering pada penelitian ini untuk mempermudah pemberian label cluster pada data transaksi baru
2. Perlu ditambahkan atribut jarak atau location yang digunakan dalam clustering untuk analisis cluster yang lebih optimal

(halaman ini sengaja dikosongkan)

BIODATA PENULIS



Penulis lahir di Bandung pada tanggal 07 bulan Maret tahun 1995. Penulis merupakan anak pertama dari dua bersaudara. Penulis telah menempuh Pendidikan formal pada SD Al-Azhar 9 Kemang Pratama, SMP 8 Al-Azhar Kemang Paratama, SMA Labschool Jakarta dan kini menepuh pendidikan perguruan tinggi di Institut Teknologi Sepuluh Nopember pada Departemen Sistem Informasi tahun 2013 dengan jalur undangan dengan nomor induk (NRP) 5213100051. Selama perkuliahan, penulis aktif di organisasi dan kepanitiaan, antara lain menjadi staff *Information Media* di BEM FTIF, sebagai staff design acara ITS Expo dan juga menjadi editor SESINDO yaitu Seminar Nasional Sistem Informasi Indonesia selama dua periode berturut-turut.

(halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- [1] S. T. G. M. Deepali Virmani, "Normalization based k means clustering algothm," 2015.
- [2] Y.-S. C. Ching-Hsue Cheng, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert system with application*, vol. 36, pp. 4176-4184, 2009.
- [3] J. Z. H. D. Q. J. X. d. L. J. Xingjun Zhang, "Supplier Categorization with K-Means Type Subspace Clustering," 2006.
- [4] Y.-H. H. d. Y. -H. L. Chih-Fong Tsai, "Customer segmentation issues and strategies for an automobil dealership with two clustering technique," *Expert Systems*, 2013.
- [5] PG Meritjan, "Wilayah Kerja," PG Meritjan, [Online]. Available:
<http://pgmeritjan.com/index.php/profilemr/wilayah-kerja>. [Accessed 25 November 2016].
- [6] Pricewaterhouse Coopers Accountants, Supplier Relationship Management, Pricewaterhouse Coopers Accountants, 2013.
- [7] Procurement Leaders/ Sigaria Ltd, STRATEGY GUIDE: SUPPLIER RELATIONSHIP MANAGEMENT, London: Procurement Leaders/ Sigaria Ltd, 2013.
- [8] V. Scacchitti, "Keys to effective supplier segmentation," *Inside Supply Management*, 2011.

- [9] P. W. Schroder, "A better way to engage with supplier," in *CSCMP's Supply Chain*, Agile Business Media, 2012.
- [10] J. C. Black, "Effective Supplier Relationship Management," in *Northeast Supply Chain Conference*, Southborough.
- [11] "What is Data Mining," 1999.
- [12] H. A. M, *Strategic database marketing*, Chicago: Probus Publishing, 1994.
- [13] S. B, "Successful direct marketing methods," Lincolnwood, NTC Business Book, 1995, pp. 35-57.
- [14] S.-Y. L. H.-H. W. Jo-Ting Wei, "A review of the application of RFM model," *African journal of business management*, vol. 4, no. 19, 2010.
- [15] P. Seth Roberts, "Transform your data," *Nutrition*, vol. 24, pp. 492-494, 2008.
- [16] D. C. Howell, *Statistical methods for psychology* (6th ed.), CA: Thomson, 2007.
- [17] S. H. M. Evangelos Triantaphyllou, "USING THE ANALYTIC HIERARCHY PROCESS FOR DECISION MAKING IN ENGINEERING APPLICATIONS: SOME CHALLENGES," *Inter'l Journal of Industrial Engineering: Applications and Practice*, vol. 2, no. 1, pp. 35-44, 1995.
- [18] Y.-Y. Duen-RenLiua, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information&Management*, vol. 42, p. 387-400, 2005.

- [19] A. K. d. A. A. Seyed Mahdi Rezaeinia, "An integrated AHP-RFM method to banking customer segmentation," *Int. J. Electronic Customer Relationship Management*, vol. 6, no. 2, pp. 153-169, 2012.
- [20] B. W. T. III, *Introduction to Management Science*, Pearson Education Inc, 2013.
- [21] Quest Software Inc, "Cluster Analysis," 5 August 2015. [Online]. Available: <https://documents.software.dell.com/statistics/textbook/cluster-analysis>. [Accessed 3 January 2017].
- [22] L. R. a. O. Maimon, "Clustering Methods," in *Data Mining and Knowledge Discovery Handbook*, Tel-Aviv, Tel-Aviv University, pp. 321-351.
- [23] "Cluster Analysis : Basic Concepts and Algorithms," 2005, pp. 487-568.
- [24] A. Cluster, "Analisis Cluster," 30 December 2008. [Online]. Available: <https://prayudho.wordpress.com/2008/12/30/analisis-cluster/>. [Accessed 30 December 2016].
- [25] Zulia, "Data Mining (Outlier)," 25 Maret 2012. [Online]. Available: <https://zuliattunnartin.wordpress.com/2012/03/25/data-mining-outlier/>. [Accessed 4 January 2017].
- [26] A. Y. ., A. R. Archana Singh, "K-means with Three different Distance Metrics," *International Journal of Computer Applications*, vol. 67, no. 10, 2013.
- [27] D. A. K. G. Dibya Jyoti Bora, "Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab," *International journal of computer science and*

information technologies, vol. 5, no. 2, pp. 2501-2506, 2014.

- [28] P. B. a. A. Kumar, "EBK- Means : A clustering technique based on elbow method and k-means in WSN," *International Journal of Computer Applications*, vol. 105, no. 9, 2014.
- [29] I. A. A. A. a. E. M. Q. Eréndira Rendón, "Internal versus External cluster validation indexes," *INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS*, vol. 1, no. 5, pp. 27-34, 2011.
- [30] Z. L. H. X. X. G. J. W. Yanchi Liu, "Understanding of Internal Clustering Validation Measures," *IEEE International Conference on Data Mining*, pp. 911-916, 2010 .
- [31] J. Templin, 10 September 2006. [Online]. Available: http://jonathantemplin.com/files/clustering/psyc993_09.pdf. [Accessed 2 Maret 2017].
- [32] V. P. D. S. D. Guy Brock, "clValid, an R package for cluster validation," *Journa lof Statistical Software*, 2011.
- [33] E. Martina, 11 September 2013. [Online]. Available: <https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwjI2bOHRyUAhUMQo8KHf76CHsQFgg4MAM&url=https%3A%2F%2Fpdfs.semanticscholar.org%2F855a%2Fec7e4697dabc2f8e7c77e307256d651886ce.pdf&usg=AFQjCNHBSyr1mkd-y610GrDgOvILbtit>. [Accessed 20 Februari 2017].
- [34] Y.-Y. Duen-RenLiu, "IntegratingAHPanddataminingforproductrecommendati

- on based on customer lifetime value," *Information & Management*, vol. 42, p. 387–400, 2005.
- [35] J. Goodman, "Leveraging the customer database to your com-," *Direct Marketing*, vol. 55, no. 8, pp. 26-27, 1992.
- [36] K. Z. ., S. A. ., S. Mahboubeh Khajvand, "Estimating Customer Lifetime Value Based o RFM Analysis of Customer Purchase Behavior : Case Study," *Procedia Computer Science* , vol. 3, pp. 57-63, 2011.
- [37] S. M. Mugi Wahidin, "Uji Korelasi Spearman Rank dan Uji Korelasi Kendal Tau," 7 Juni 2014. [Online]. Available: <http://slideplayer.info/slide/3040351/>. [Accessed 26 Mei 2017].
- [38] P. I. O. Dawud Adebayo Agunbiade, "Effect of Correlation Level on the Use of Auxiliary Variable in Double Sampling for Regression Estimation," *Open Journal of Statistics*, vol. 3, pp. 312-318, 2013.
- [39] Leaerd Statistic, "Pearson Product-Moment Correlation," Leaerd Statistic, [Online]. Available: <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php>. [Accessed 26 Mei 2017].
- [40] R. Newson, "Parameters behind “non-parametric” statistics: Kendall ,Somers and median differences," *The Stata Journal*, pp. 1 - 20, 2001.
- [41] statistical-research.com, "Kendall's tau," 05 September 2012. [Online]. Available: https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=16&cad=rja&uact=8&ved=0ahUKEwiv-rKbi7_UAhXFro8KHfQpDP8QFgh1MA8&url=https%3

A%2F%2Fstatistical-research.com%2Fwp-content%2Fuploads%2F2012%2F09%2Fkendall-tau1.pdf&usg=AFQjCNEt8M-MFnUE_LLjeC9wzGw. [Accessed Juni 2017].

- [42] M. A. .. Wicks, "Association Between Variables," 1998, pp. 795 - 828.
- [43] Andi, "REGRESI LINEAR SEDERHANA," 22 Mei 2016. [Online]. Available: https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&cad=rja&uact=8&ved=0ahUKEwiXhpiF6_vUAhWIN48KHTodCMUQFghgMAG&url=http%3A%2F%2Fisbat.lecturer.pens.ac.id%2Fmaterikuliah%2FModul%2520Teori%2FMetode%2520Numerik%2FMetode%2520Regresi.pdf&usg=AFQ.
- [44] S. Seo, "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets," 2006.
- [45] DISPERINDAG JAWA TIMUR, "SISKAPERBAPO," [Online]. Available: <http://siskaperbapo.com/harga/tabel>. [Accessed 20 Maret 2017].
- [46] K. Dhana, "Identify, describe, plot, and remove the outliers from the dataset," 30 April 2016. [Online]. Available: https://googleweblight.com/?lite_url=https://datascienceplus.com/identify-describe-plot-and-removing-the-outliers-from-the-dataset/&ei=zqc-kaFn&lc=id-ID&s=1&m=96&host=www.google.co.id&ts=1490661681&sig=AJsQQ1C4Hq8iDmTKVOrCmtEl4JWAqp6X3Q. [Accessed 4 April 2017].

LAMPIRAN A

Potongan Tabel Pemetaan Variabel RFM dan Cluster

(halaman ini sengaja dikosongkan)

Potongan Tabel Pemetaan Variabel RFM dan *Cluster*

blockname	recency	frequency	monetary	cluster
10000002	137	6	26,808,750	3
10000003	146	2	7,202,000	3
10000004	136	2	7,630,500	3
10000005	138	3	10,682,750	3
10000006	201	1	3,180,375	1
10000007	229	7	65,019,875	2
10000011	239	4	19,733,375	2
10000012	231	9	52,400,000	2
10000013	232	16	124,758,500	2
10000014	234	9	50,851,125	2
10000015	273	4	26,972,875	2
10000016	281	1	5,252,250	1
10000017	203	5	24,369,500	2
10000018	202	2	8,024,000	1
10000019	219	2	9,554,625	1
10000020	180	20	104,379,250	2
10000022	276	3	58,983,375	2
10000023	283	3	67,041,375	2
10000024	233	7	35,348,500	2
10000025	182	8	36,650,500	2
10000026	236	15	202,098,000	2
10000027	137	3	66,097,875	3
10000029	196	2	8,192,625	1
10000030	190	6	27,101,625	2
10000031	186	7	40,110,500	2
10000032	266	2	8,794,000	1
10000033	283	2	14,582,750	1
10000036	234	4	18,718,375	2

A-2

blockname	recovery	frequency	monetary	cluster
10000037	201	1	7,496,000	1
10000039	277	3	14,624,000	1
10000041	198	1	3,809,500	1
10000042	200	1	3,730,750	1
10000051	265	5	30,998,500	2
10000052	279	7	46,911,750	2
10000053	185	3	12,918,000	3
10000054	261	3	19,666,250	1
10000055	252	4	21,563,500	2
10000056	247	12	70,398,125	2
10000058	172	6	33,575,500	3
10000059	244	3	17,217,250	1
10000060	226	4	20,594,250	2
10000061	138	4	49,560,625	3
10000062	138	5	54,388,875	3
10000063	148	4	21,050,625	3
10000065	263	2	10,143,875	1
10000067	224	3	19,912,750	1
10000069	220	17	200,569,500	2
10000070	222	4	41,449,375	2
10000071	225	4	34,480,250	2
10000072	219	2	13,869,250	1
10000073	220	2	10,531,125	1
10000074	223	2	9,345,250	1
10000075	219	3	13,386,125	1
10000076	226	2	24,876,375	1
10000078	177	1	3,031,000	1



LAMPIRAN B

Potongan Hasil Normalisasi



(halaman ini sengaja dikosongkan)

Potongan Hasil Normalisasi

blockname	recency	frequency	monetary	cluster
10000002	0.97	0.52	0.45	3
10000003	0.89	0.20	0.24	3
10000004	0.98	0.20	0.25	3
10000005	0.96	0.32	0.30	3
10000006	0.46	0.00	0.11	1
10000007	0.29	0.57	0.60	2
10000011	0.23	0.40	0.40	2
10000012	0.28	0.64	0.56	2
10000013	0.27	0.81	0.70	2
10000014	0.26	0.64	0.56	2
10000015	0.06	0.40	0.45	2
10000016	0.02	0.00	0.19	1
10000017	0.45	0.47	0.44	2
10000018	0.46	0.20	0.26	1
10000019	0.35	0.20	0.29	1
10000020	0.61	0.87	0.67	2
10000022	0.04	0.32	0.58	2
10000023	0.01	0.32	0.60	2
10000024	0.27	0.57	0.50	2
10000025	0.59	0.61	0.50	2
10000026	0.25	0.79	0.78	2
10000027	0.97	0.32	0.60	3
10000029	0.50	0.20	0.26	1
10000030	0.54	0.52	0.45	2
10000031	0.57	0.57	0.52	2
10000032	0.09	0.20	0.27	1
10000033	0.01	0.20	0.35	1
10000036	0.26	0.40	0.39	2

B-2

blockname	recency	frequency	monetary	cluster
10000037	0.46	0.00	0.25	1
10000039	0.04	0.32	0.35	1
10000041	0.48	0.00	0.14	1
10000042	0.47	0.00	0.13	1
10000051	0.10	0.47	0.48	2
10000052	0.03	0.57	0.54	2
10000053	0.57	0.32	0.33	3
10000054	0.12	0.32	0.40	1
10000055	0.16	0.40	0.42	2
10000056	0.19	0.72	0.61	2
10000058	0.67	0.52	0.49	3
10000059	0.21	0.32	0.38	1
10000060	0.31	0.40	0.41	2
10000061	0.96	0.40	0.55	3
10000062	0.96	0.47	0.57	3
10000063	0.87	0.40	0.41	3
10000065	0.11	0.20	0.30	1
10000067	0.32	0.32	0.40	1
10000069	0.34	0.83	0.78	2
10000070	0.33	0.40	0.52	2
10000071	0.31	0.40	0.49	2



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Inst
Tekn
Sep

LAMPIRAN C

KUISIONER AHP



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Inst
Tekn
Sep



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Inst
Tekn
Sep



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Inst
Tekn
Sep



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Inst
Tekn
Sep

(halaman ini sengaja dikosongkan)



**KUISIONER PENELITIAN
SURVEY PERINGKATAN KEPENTINGAN
KRITERIA PADA PERILAKU SUPPLIER PTPN X PG
MERITJAN DALAM PENGADAAN TEBU**



Kuisisioner ini bertujuan untuk mendapatkan masukan pendapat/pemikiran dari Narasumber terkait dengan peringkatan kriteria yang akan digunakan dalam pemberian bobot pada variabel segmentasi supplier. Kami menjamin kerahasiaan data/informasi yang diberikan serta tidak akan membawa implikasi apapun bagi responden.

IDENTITAS RESPONDEN

Nama Responden : _____

Jabatan : _____

Tanda Tangan : _____

PETUNJUK PENGISIAN KUISIONER

1. Berilah tanda checklist (√) untuk menunjukkan persepsi atau penilaian bapak/ibu terhadap kriteria *supplier* bagaimanakah yang dipilih lebih penting dengan melihat perilaku *supplier* dalam melakukan transaksi dengan perusahaan. (Bandingkan kriteria yang ada pada kolom sebelah kiri dengan kriteria yang ada pada kolom sebelah kanan)
2. Skala numerik akan menunjukkan suatu perbandingan dari tingkat kepentingan dua kriteria dengan penjelasan setiap skalanya yaitu:

Skala Nilai	Tingkat Preferensi	Penjelasan
1	Sama pentingnya (<i>equal importance</i>)	Kedua kriteria memiliki pengaruh yang sama
3	Sedikit lebih penting (<i>Relatively more important</i>)	Penilaian dan pengalaman sedikit memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.
5	Lebih penting (<i>More important</i>)	Penilaian dan pengalaman memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.

C-2

7	Jelas lebih penting (<i>Very important relative</i>)	Salah satu kriteria lebih diprioritaskan dan relatif lebih penting dibandingkan kriteria pasangannya.
9	Mutlak sangat penting (<i>Extremely more important</i>)	Salah satu kriteria sangat jelas lebih penting dibandingkan kriteria pasangannya.
2, 4, 6, 8		Diberikan bila terdapat keraguan penilaian di antara dua tingkat kepentingan yang berdekatan.

3. Jika kriteria pada sebelah kiri lebih penting dari pada kriteria sebelah kanan maka, pilih skala yang dekat dengan sebelah kiri dan jika sebaliknya pilih skala yang dekat dengan sebelah kanan.
4. Kriteria pembobotan terhadap *supplier* terdapat tiga macam yaitu:
 - *Recency* : jarak antara waktu terakhir transaksi dengan waktu saat ini.
 - *Frequency* : seberapa sering jumlah transaksi yang dilakukan oleh *supplier* pada periode tertentu.
 - *Monetary* : jumlah uang yang dihabiskan *supplier* saat transaksi pada periode tertentu

CONTOH PENGISIAN KUISIONER

Kriteria																	Kriteria	
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Frequency
															√			

Artinya:

Pada pernyataan tersebut, tanda checklist (√) di kolom jawaban sebelah kiri menunjukkan bahwa: “*Recency*” ¹/₇ kali lebih penting dibandingkan dengan “*Frequency*” dalam hal peringkat kriteria untuk *supplier*. Dengan kata lain “*Frequency*” 7 kali lebih penting dibandingkan dengan “*Recency*” dalam hal peringkat kriteria untuk *supplier*.

DAFTAR PERTANYAAN

Berikan tanda checklist (√) untuk menunjukkan persepsi atau penilaian bapak/ibu terhadap perbandingan kriteria *supplier* yang dipilih lebih penting dengan melihat perilaku *supplier* dalam melakukan transaksi dengan perusahaan. “Manakah kriteria yang lebih penting antara kriteria pada kolom sebelah kiri dengan kriteria pada kolom sebelah kanan?”

Kriteria																					Kriteria
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Frequency			
Frequency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Monetary			
Monetary	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Recency			

-Terima kasih atas waktu yang telah diberikan Bapak/Ibu untuk mengisi kuisisioner ini

C-4

(halaman ini sengaja dikosongkan)



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember

LAMPIRAN D

Perhitungan AHP



(halaman ini sengaja dikosongkan)

Hasil Wawancara

No	Jabatan	RF	FM	MR
1	Asisten Manager Tanaman	5F	3F	4M
2	Asisten Muda 1	4R	3F	5R
3	Asisten Manager Quality Assurance	4R	5M	4R
4	Asisten Manager Perkreditan dan Admin Hasil	4R	3F	5R
Kesimpulan		4R	3F	5R

Tambahkan semua isi kolom

	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
<i>Recency</i>	1	4	5
<i>Frequency</i>	1/4	1	3
<i>Monetary</i>	1/5	1/3	1
<i>Column Sum</i>	1.45	5.33	9.00

Pembagian nilai dengan *column sum*

$$W_R = (0.69 + 0.75 + 0.56) / 3 = \mathbf{0.67}$$

$$W_F = (0.17 + 0.19 + 0.33) / 3 = \mathbf{0.23}$$

$$W_M = (0.14 + 0.06 + 0.11) / 3 = \mathbf{0.10}$$

	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	Preference Vector
<i>Recency</i>	0.69	0.75	0.56	0.67
<i>Frequency</i>	0.17	0.19	0.33	0.23
<i>Monetary</i>	0.14	0.06	0.11	0.10

Pengalihan nilai dengan PV lalu dijumlahkan per baris

	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>	total
<i>Recency</i>	0.67	0.92	0.52	2.11
<i>Frequency</i>	0.17	0.23	0.31	0.71
<i>Monetary</i>	0.13	0.08	0.10	0.31

Pembagian total dengan PV dan nilainya di rata-ratakan

Criteria	Total	Preference Vector	Result
<i>Recency</i>	2.11	0.67	3.170547288
<i>Frequency</i>	0.71	0.23	3.067702936
<i>Monetary</i>	0.31	0.10	3.022598002
Average			3.086949408

LAMPIRAN E
SCRIPT UI DAN SERVER VISUALISASI

(halaman ini sengaja dikosongkan)

Kode Program UI

```
library(rgl)
library(DT)

header <- dashboardHeader(
  title = "Segmentasi Supplier"
)

sidebar <- dashboardSidebar(
  br(),
  sidebarMenu(
    id = "tabs",
    menuItem("Dashboard", tabName = "dashboard",
    icon = icon("dashboard")),
    menuItem("Cluster", icon = icon("bar-chart-
o"), tabName = "widgets"),
    menuItem("Data", icon = icon("table"), tabName
= "user"),
    menuItem("Perusahaan", icon=icon("building"),
tabName = "Perusahaan",
      a(
        href =
"http://www.ptpn10.com/", target
=
"_blank", "Perusahaan"
      )
    )
  )

body <- dashboardBody(
  tabItems(
    tabItem("dashboard",
      fluidRow(
        column(width=7,
          box(
            width=NULL,
            status = "success",
            solidHeader = TRUE,
            title = "Tabel RFM",
            collapsible = T, collapsed
= T,
            dataTableOutput("kesi")
          ),
          box(
            width=NULL,
            status = "success",
            solidHeader = TRUE,
```

```

Terbaik",
= T,
title = "5 Supplier",
collapsible = T, collapsed
dataTableOutput("baek")
),
fixedRow(
valueBoxOutput("reci"),
valueBoxOutput("fri"),
valueBoxOutput("moni")
),
box(width=NULL,
solidHeader = TRUE,
title="Perbandingan
Range Cluster",
status="warning",
collapsible =
T, collapsed = F,
selectInput("variabel",
label =
"Tentukan variabel Yang Akan Dilihat",
choices =
c("Recency", "Frequency",
"Monetary"),
selected =
"Recency"),
plotlyOutput("boki")
),
column(width=5,
box(title
=tagList(shiny::icon("gear"), " Tentang
Visualisasi"),
background = "green",
width=NULL,
p(aligned = "justify"
,"Visualisasi ini digunakan untuk menampilkan
hasil segmentasi
supplier menggunakan data mining
clustering. Pada
halaman ini akan ditampilkan ranking
supplier berdasarkan 3
variabel yaitu recency, frequency
dan monetary. Selain
ranking juga akan diperlihatkan range
serta rata-rata nilai
variabel RFM pada
setiap clusternya. ")

```



```

    ),
    box(title="Persebaran
Supplier", collapsible = T,
    collapsed =
T, status="success", width=NULL,
    plotlyOutput("piec")),
    box (width=NULL,
    solidHeader = TRUE,
status="warning",
    collapsible =
T, collapsed = F,
    title="Grafik 3D
Scatter plot",

rglwidgetOutput('thewidget2', height = 400,
width=400),

    uiOutput("lgdi")
)
)
),
tabItem("widgets",
fluidRow(
column(width =6,
tabBox(width=NULL,
title =
tagList(shiny:::icon("gear"), "Cluster"),
tabPanel(
width=NULL,
title = "Anggota",

dataTableOutput("lol")
),
tabPanel(
width = NULL,
title="Grafik ",

rglwidgetOutput('thewidget1'),
uiOutput("lgd")
)
),
column(width =3,
valueBoxOutput("recency",
width=NULL),
valueBoxOutput("frq",
width=NULL),

```

```

valueBoxOutput("mnt",
width=NULL)
),
column(width =3,
box(width = NULL, status =
"warning",
selectInput("var",
label =
"Pilih cluster :",
choices = c(
"cluster 1",
"cluster 2", "cluster 3"),
selected =
"cluster 1"),
helpText(align="center","Berikut adalah
visualisasi grafik 3d scatterplot,
setiap cluster dan juga rata-rata
clusternya")
penjelasan
variabel
RFM
tiap
)
),
fluidRow(
column(width =6,
valueBoxOutput("max",
width=NULL)
),
column(width =6,
valueBoxOutput("min",
width=NULL)
)
),
fluidRow(
box(width = 10, status ="success",
title=tagList(shiny::icon("commenting"),
"Penjelasan"),
textOutput("text1"),
collapsible = T,collapsed = T,
solidHeader = TRUE
)
),
tabItem("user",
br(),
dataTableOutput("mytable3"))

```

```

))

dashboardPage(
  skin = "black",
  header,
  sidebar,
  body
)

```

Kode Program Sever

```

library(shinydashboard)
library(shiny)
library(rgl)
library(RColorBrewer)
library(ggplot2)
library(plotly)
library(DT)
library(sqldf)

#DATAFIX
kes<-read.csv("data/GBKesimpulan.csv")
dataRaw <-read.csv("data/GBDataHasil.csv")
dataNorm <-read.csv("data/GBNormHasil.csv")

#DEKLARASI WARNA
dirColors <-c("1"="#06D6A0", "2"="#FFD166",
"3"="#EF476F")

options(rgl.useNULL = TRUE)
shinyServer(function(input, output, session) {

  #MENGHASILKAN GRAFIK 3D SCATTER PLOT
  open3d()
  par3d(windowRect = c(0, 20, 50, 50))
  plotids<-plot3d(dataNorm$recency,
                 dataNorm$frequency,
                 dataNorm$monetary,
                 col =c('#06D6A0', '#FFD166',
'#EF476F')[as.numeric(dataNorm$cluster)],
                 xlab = "recency",
                 ylab = "frequency",
                 zlab = "monetary"
  )
  aspect3d(1,1,1)
  scene1 <- scene3d()
  rgl.close()
  save <- options(rgl.inShiny = TRUE)
  on.exit(options(save))

```

```

#RENDER GRAFIK SCATTER PLOT
output$thewidget1 <- renderRglwidget(
  rglwidget(scene1))

output$thewidget2 <- renderRglwidget(
  rglwidget(scene1))

#RENDER TABEL SEMUA
output$mytable3 = renderDataTable(dataRaw,
  options =
list(aLengthMenu = c(10, 30, 50),
  iDisplayLength = 10,
  width = '100%'),
  filter =
"top")

  output$kesi = renderDataTable(kes,
  options = list
  (dom='t',
  columnDefs =
list(list
(classname = 'dt-center',
targets = 0:5 )))

#RENDER DATA TABLE LIMA SUPPLIER TERBAIK
output$baek = renderDataTable(
  lima <- sqldf("select
blockname,nama,recency,frequency,monetary,cluster
from(select * from dataRaw
group by blockname ORDER BY CLV
desc) LIMIT 5"),
  options = list
  (dom='t',
  columnDefs = list(list
(classname = 'dt-center',
targets = 0:5 )))

#RENDER DATA TABLE PER CLUSTER
output$lo1 = renderDataTable(
  yeay <-switch(input$var,
  "cluster 1" = tabone<-
subset(dataRaw, cluster == 1, select =
c(blockname,nama,CLV)),
  "cluster 2" =tatwo<-
subset(dataRaw, cluster == 2, select =
c(blockname,nama,CLV)),

```

```

subset(dataRaw, cluster == 3, select =
c(blockname, nama, CLV))
),
options = list(aLengthMenu = c(3, 5, 10),
iDisplayLength = 5, width =
'100%'))

#RENDER GRAFIK BOX AND WHISKER
output$boki <- renderPlotly({
  yeay <-switch(input$variabel,
    "Recency" = dataRaw$recency,
    "Frequency" =
dataRaw$frequency,
    "Monetary" = dataRaw$monetary)
  judul <- switch(input$variabel,
    "Recency" = "Recency \n dalam
harian",
    "Frequency" = "Frequency \n
dalam banyaknya pengantaran",
    "Monetary" = "Monetary\n
dalam rupiah")
  dataRaw$cluster<-factor(dataRaw$cluster,
labels = c("1", "2", "3"))

  p<-ggplot(dataRaw,
aes(x = dataRaw$cluster, y = yeay,
fill=dataRaw$cluster))+
  geom_boxplot()+
  scale_fill_manual(values=c('#06D6A0',
'#FFD166', '#EF476F'))+
  scale_y_continuous(name = judul)+
  scale_x_discrete(name = "cluster")+
  labs(fill = "cluster")

  p <- ggplotly(p)}})

#SUBSET DATA PER CLUSTER
one<-subset(dataRaw, cluster == 1)
two<-subset(dataRaw, cluster == 2)
three<-subset(dataRaw, cluster == 3)

#RENDER VALUE BOX RECENCY PER CLUSTER
output$recency <- renderValueBox({
  valueBox(
    value = paste(round(mean(
      hi<- switch(input$var,
        "cluster 1" = one$recency,
        "cluster 2" = two$recency,

```

```

                                "cluster"      3"      =
three$recency))), "hari"),
  subtitle = "yang lalu pengiriman terakhir",
  icon = icon("calendar"),
  color= "green", href = NULL
)})

#RENDER VALUE BOX FREQUENCY PER CLUSTER
output$frq <- renderValueBox({
  valueBox(
    value = paste(round(mean(
      hi<- switch(input$var,
                  "cluster 1" = one$frequency,
                  "cluster 2" = two$frequency,
                  "cluster
                                3"      =
three$frequency))), "kali"),
    subtitle = "pengiriman tebu",
    icon = icon("truck"),
    color= "green", href = NULL
  ))

#RENDER VALUE BOX MONETARY PER CLUSTER
output$mnt <- renderValueBox({
  valueBox(
    value = paste("Rp",
                  round(mean(hi<-
switch(input$var,
"cluster 1" = one$monetary,
"cluster 2" = two$monetary,
"cluster 3" = three$monetary,
"cluster 4" = four$monetary))),
    subtitle = "nilai tebu dalam Rupiah",
    icon = icon("money"),
    color= "green", href = NULL
  ))

#RENDER VALUE BOX SUPPLIER TERBAIK PER CLUSTER
output$max <- renderValueBox({
  valueBox(
    value = yuuu <-switch(input$var,
                          "cluster"      1"      =
sqldf("select nama from (select blockname,
nama,max(CLV) as max, cluster from one)"),
                          "cluster"      2"      =
sqldf("select nama from (select blockname,
nama,max(CLV)as max, cluster from two)"),

```

```

                                "cluster 3" =
sqldf("select nama from (select blockname,
nama,max(CLV)as max, cluster from three)")),
    subtitle = paste("dengan membutuhkan waktu
",
                                yuuu <-switch(input$var,
                                "cluster 1"
= sqldf("select recency from (select *,max(CLV)as
max, cluster from one)")),
                                "cluster 2"
= sqldf("select recency from (select *,max(CLV)as
max, cluster from two)")),
                                "cluster 3"
= sqldf("select recency from (select *,max(CLV)as
max, cluster from three)")),
    "hari untuk melakukan
pengiriman kembali, melakukan pengiriman",
    yuuuh <-switch(input$var,
    "cluster 1"
= sqldf("select frequency from (select
*,max(CLV)as max, cluster from one)")),
    "cluster 2"
= sqldf("select frequency from (select
*,max(CLV)as max, cluster from two)")),
    "cluster 3"
= sqldf("select frequency from (select
*,max(CLV)as max, cluster from three)")),
    "kali dan nilai dari tebu
yang dikirimkan sebesar Rp",
    yis <-switch(input$var,
    "cluster 1" =
sqldf("select monetary from (select *,max(CLV)as
max, cluster from one)")),
    "cluster 2" =
sqldf("select monetary from (select *,max(CLV)as
max, cluster from two)")),
    "cluster 3" =
sqldf("select monetary from (select *,max(CLV)as
max, cluster from three)"))
    )),
    icon = icon("thumbs-o-up"),
    color= "green", href = NULL
)
})

#RENDER VALUE BOX SUPPLIER TERBURUK PER CLUSTER
output$min <- rendervalueBox({
  valueBox(
    value = yuuu <-switch(input$var,

```



```

#RENDER VALUEBOX MONETARY PER CLUSTER
output$mnt <- renderValueBox({
  valueBox(
    value = paste("Rp",round(mean(hi<-
switch(input$var,
"Cluster 1" = one$monetary,
"Cluster 2" = two$monetary,
"Cluster 3" = three$monetary))))),
    subtitle = "nilai tebu dalam Rupiah",
    icon = icon("money"),
    color= "green", href = NULL
  )
})

#PERHITUNGAN RATA-RATA KESELURUHAN
nyuba <-subset(dataRaw,
select=c(recency,frequency,monetary))
rata<-round(colMeans(nyuba))
ca<-t(data.frame(rata))

#RENDER VALUEBOX RECENCY SEMUA
output$reci <- renderValueBox({
  valueBox(
    value= paste(ha <- ca[1:1],"hari"),
    subtitle= "untuk mengirim kembali", icon =
icon("calendar"),
    color = "yellow", width=2, href = NULL
  )
})

#RENDER VALUEBOX FREQUENCY SEMUA
output$fri <- renderValueBox({
  valueBox(
    value=paste(ha <- ca[2:2],"kali"),
    subtitle= "pengiriman tebu", icon =
icon("truck"),
    color = "yellow",width=2, href = NULL
  )
})

#RENDER VALUEBOX MONETARY SEMUA
output$moni <- renderValueBox({
  valueBox(
    value=paste(ha <- ca[3:3]),subtitle= "nilai
tebu dalam rupiah", icon = icon("money"),
    color = "yellow", width=2, href = NULL
  )
})

```

```

#PENJUMLAHAN ANGGOTA SUPPLIER
piech<-data.frame(cluster=c("cluster
1","cluster 2","cluster 3"),
supplier=c(nrow(one),nrow(two),nrow(three)))

#RENDER PIE CHART
output$piec <- renderPlotly(
  plot_ly(piech, labels = ~cluster, values =
~supplier, type = 'pie',
          textposition = 'inside',
          textinfo = 'label+percent',
          insidetextfont = list(color =
'#000000'),
          hoverinfo = 'text',
          text = ~paste(supplier, 'supplier'),
          marker = list(colors = dirColors,
                        line = list(color =
'#000000', width = 1))))

#PEMBUATAN LEGEND
output$lgd <- renderUI({
  tags$table(class = "table",
             tags

```

```

    ))
  })

#PEMBUATAN LEGEND 2
output$lgdi <- renderUI({
  # Create a Bootstrap-styled table
  tags$table(class = "table",
    tags$tbody(
      tags$tr(align="center",
        tags$td("Cluster 1"),
        tags$td("Cluster 2"),
        tags$td("Cluster 3")
      ),
      tags$tr(align="center",
        tags$td(span(style
sprintf(
  height:0.5em;
display:inline-block;",
  dirColors[1]
))),
tags$td(span(style
sprintf(
  height:0.5em;
display:inline-block;",
  dirColors[2]
))),
tags$td(span(style
sprintf(
  height:0.5em;
display:inline-block;",
  dirColors[3]
)))
    ))
  })

#KATA-KATA
output$text1 <- renderText({
  kata <- switch(input$var,
    "Cluster 1" =
      "Cluster satu merupakan
cluster terbaik dibanding dua cluster lainnya.
Hal ini dikarenakan transaksi
pengiriman terakhir dilakukan 169 hari atau kurang
lebih 6 bulan lalu dari
waktu ekstraksi RFM yaitu bulan
Oktober.

```

Dibandingkan *cluster* lainnya *cluster* ini adalah *cluster* yang terakhir melakukan transaksi pengiriman tebu.

Sementara itu, jumlah pengiriman tebu pada *cluster* ini pada periode Juni - November 2016 sebanyak 3 kali.

Frekuensi pengiriman pada *cluster* ini lebih baik dari *cluster* 2 tapi tidak sebaik *cluster* 3.

Untuk nilai tebu yang dikirimkan oleh *cluster* ini rata-rata sebanyak Rp19,412,902.

Perbandingan dengan *cluster* lainnya sama dengan variabel frekuensi, dimana *cluster* 3 lebih baik daripada *cluster* ini.

Cluster ini menjadi *cluster* terbaik bagi perusahaan dikarenakan nilai recencynya yang tinggi

dimana perusahaan lebih mementingkan supplier yang melakukan transaksi terkini.",

"*Cluster* 2" =

"*Cluster* dua merupakan *cluster* terburuk dibandingkan *cluster* lainnya berdasarkan nilai CLVnya.

Transaksi terakhir dilakukan ini berkisar 242 hari dari ekstraksi variabel RFM atau

8 bulan yang lalu yang jatuh pada bulan Juli. Frekuensi pengiriman tebu juga sangat rendah yaitu hanya

2 kali dalam periode tersebut. Nilai tebu dalam rupiahnya bila dibandingkan dengan supplier

lainnya juga sangat rendah yaitu hanya berjumlah Rp16,404,445.

Bila dibandingkan dengan rata-rata keseluruhan supplier, *cluster* ini ketiga variabelnya masing-masing berada dibawah rata-rata.",

"*Cluster* 3" =

"*Cluster* ketiga merupakan *cluster* yang memiliki jumlah pengiriman tebu terbanyak dan nilai tebu dalam rupiah yang tertinggi

dibandingkan supplier lainnya. Jumlah pengiriman yang dilakukan *cluster* ini berkisar sebanyak 8 kali

dan nilai dari tebu yang dikirimkan adalah sebesar Rp61,411,380.

Namun transaksi terakhir yang dilakukan *cluster* ini dibandingkan dengan *cluster* 1 lebih lama jaraknya, yaitu 7 bulan yang lalu dari waktu ekstraksi yaitu bulan September. Menjadikan nilai CLV nya lebih rendah dibandingkan dengan *cluster* 1 karena nilai recencynya lebih rendah walaupun nilai frequency dan monetarynya tinggi."

)
})

})

(halaman ini sengaja dikosongkan)



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember

LAMPIRAN F

HASIL UJI PERFORMA



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS
Institut
Teknologi
Sepuluh Nopember



ITS



ITS



ITS



ITS

ITS

(halaman ini sengaja dikosongkan)

Uji Performa

K	connectivity	dunn	silhoutte	K
2	175.8302	0.0043	0.321	2
3	197.6036	0.0055	0.3507	3
4	340.3282	0.0076	0.3154	4
5	338.8135	0.0053	0.3151	5
6	377.9794	0.0061	0.3274	6
7	420.2067	0.0073	0.3378	7
8	436.7345	0.0033	0.3368	8
9	488.6448	0.005	0.3248	9
10	491.0599	0.0094	0.3362	10
11	577.9183	0.0083	0.3181	11
12	504.6623	0.004	0.3388	12
13	529.2567	0.0086	0.3328	13
14	534.9349	0.0089	0.3301	14
15	543.1504	0.0086	0.3265	15

(halaman ini sengaja dikosongkan)

LAMPIRAN G

HASIL DATA TRANSFORMASI



(halaman ini sengaja dikosongkan)

Hasil Transformasi Data dengan Metode Logartima

blockname	frequency	monetary	recency
10000001	1.61	8.71	2.18
10000002	0.78	7.43	2.14
10000003	0.30	6.86	2.16
10000004	0.30	6.88	2.13
10000005	0.48	7.03	2.14
10000006	0.00	6.50	2.30
10000007	0.85	7.81	2.36
10000011	0.60	7.30	2.38
10000012	0.95	7.72	2.36
10000013	1.20	8.10	2.37
10000014	0.95	7.71	2.37
10000015	0.60	7.43	2.44
10000016	0.00	6.72	2.45
10000017	0.70	7.39	2.31
10000018	0.30	6.90	2.31
10000019	0.30	6.98	2.34
10000020	1.30	8.02	2.26
10000021	1.62	8.32	2.20
10000022	0.48	7.77	2.44
10000023	0.48	7.83	2.45
10000024	0.85	7.55	2.37
10000025	0.90	7.56	2.26
10000026	1.18	8.31	2.37
10000027	0.48	7.82	2.14
10000029	0.30	6.91	2.29
10000030	0.78	7.43	2.28
10000031	0.85	7.60	2.27
10000032	0.30	6.94	2.42

G-2

blockname	frequency	monetary	recency
10000033	0.30	7.16	2.45
10000036	0.60	7.27	2.37
10000037	0.00	6.87	2.30
10000039	0.48	7.17	2.44
10000041	0.00	6.58	2.30
10000042	0.00	6.57	2.30
10000051	0.70	7.49	2.42
10000052	0.85	7.67	2.45
10000053	0.48	7.11	2.27
10000054	0.48	7.29	2.42
10000055	0.60	7.33	2.40
10000056	1.08	7.85	2.39
10000058	0.78	7.53	2.24
10000059	0.48	7.24	2.39
10000060	0.60	7.31	2.35
10000061	0.60	7.70	2.14
10000062	0.70	7.74	2.14
10000063	0.60	7.32	2.17
10000065	0.30	7.01	2.42
10000067	0.48	7.30	2.35
10000069	1.23	8.30	2.34
10000070	0.60	7.62	2.35
10000071	0.60	7.54	2.35
10000072	0.30	7.14	2.34
10000073	0.30	7.02	2.34

LAMPIRAN H

PERHITUNGAN LINEAR REGRESI



(halaman ini sengaja dikosongkan)

Hasil Perhitungan Linear Regresi

blockname	monetary	area	forecast	selisih
10000002	0.45	0.60	0.41	0.04
10000003	0.24	0.17	0.40	0.16
10000004	0.25	0.26	0.40	0.15
10000005	0.30	0.22	0.40	0.09
10000006	0.11	0.27	0.40	0.29
10000007	0.60	0.85	0.42	0.17
10000011	0.40	0.27	0.40	0.00
10000012	0.56	0.59	0.41	0.15
10000013	0.70	1.02	0.43	0.27
10000014	0.56	0.53	0.41	0.15
10000015	0.45	0.48	0.41	0.05
10000016	0.19	0.20	0.40	0.21
10000017	0.44	0.69	0.42	0.02
10000018	0.26	0.23	0.40	0.14
10000019	0.29	0.24	0.40	0.11
10000020	0.67	1.61	0.45	0.22
10000022	0.58	0.88	0.42	0.16
10000023	0.60	0.81	0.42	0.18
10000024	0.50	0.49	0.41	0.09
10000025	0.50	0.51	0.41	0.09
10000026	0.78	2.41	0.49	0.30
10000027	0.60	1.17	0.44	0.16
10000029	0.26	0.19	0.40	0.14
10000030	0.45	0.85	0.42	0.03
10000031	0.52	0.89	0.42	0.09
10000032	0.27	0.39	0.40	0.13
10000033	0.35	0.13	0.39	0.04
10000036	0.39	0.59	0.41	0.02

H-2

blockname	monetary	area	forecast	selisih
10000037	0.25	0.39	0.40	0.16
10000039	0.35	0.20	0.40	0.04
10000041	0.14	1.50	0.45	0.31
10000042	0.13	0.38	0.40	0.27
10000051	0.48	0.47	0.41	0.07
10000052	0.54	0.68	0.42	0.13
10000053	0.33	0.41	0.41	0.07
10000054	0.40	0.29	0.40	0.00
10000055	0.42	0.39	0.40	0.01
10000056	0.61	0.92	0.43	0.18
10000058	0.49	0.54	0.41	0.08
10000059	0.38	0.22	0.40	0.02
10000060	0.41	0.31	0.40	0.01
10000061	0.55	0.49	0.41	0.14
10000062	0.57	0.52	0.41	0.16
10000063	0.41	1.00	0.43	0.02
10000065	0.30	0.12	0.39	0.10
10000067	0.40	0.28	0.40	0.00
10000069	0.78	2.50	0.49	0.29
10000070	0.52	0.60	0.41	0.11
10000071	0.49	0.48	0.41	0.09
10000072	0.35	0.21	0.40	0.05

LAMPIRAN I

PERHITUNGAN EUCLIDEAN DISTANCE



(halaman ini sengaja dikosongkan)

Perhitungan Euclidean Distance

blockname	1	2	3	minVal	min	cluster
10000002	0.79	0.64	0.26	0.26	3	3
10000003	0.58	0.74	0.26	0.26	3	3
10000004	0.68	0.81	0.29	0.29	3	3
10000005	0.68	0.72	0.22	0.22	3	3
10000006	0.27	0.75	0.59	0.27	1	1
10000007	0.54	0.06	0.59	0.06	2	2
10000011	0.31	0.26	0.59	0.26	2	2
10000012	0.58	0.1	0.63	0.1	2	2
10000013	0.8	0.28	0.77	0.28	2	2
10000014	0.58	0.11	0.64	0.11	2	2
10000015	0.41	0.35	0.76	0.35	2	2
10000016	0.33	0.76	0.9	0.33	1	1
10000017	0.4	0.2	0.39	0.2	2	2
10000018	0.16	0.5	0.42	0.16	1	1
10000019	0.08	0.47	0.51	0.08	1	1
10000020	0.89	0.41	0.63	0.41	2	2
10000022	0.44	0.39	0.78	0.39	2	2
10000023	0.48	0.42	0.81	0.42	2	2
10000024	0.49	0.11	0.6	0.11	2	2
10000025	0.6	0.26	0.36	0.26	2	2
10000026	0.83	0.31	0.8	0.31	2	2
10000027	0.76	0.68	0.21	0.21	3	3
10000029	0.2	0.51	0.39	0.2	1	1

(halaman ini sengaja dikosongkan)

LAMPIRAN J
TABEL LAHAN PERFORMA RENDAH

(halaman ini sengaja dikosongkan)

Tabel Lahan yang Memiliki Performa yang Rendah

blockname	monetary	area	forecast
10000003	0.239635975	0.17	0.395774983
10000004	0.249009498	0.261	0.399418074
10000006	0.107071025	0.266	0.399618244
10000016	0.188433243	0.203	0.397096104
10000018	0.257164847	0.227	0.398056919
10000037	0.246125196	0.388	0.404502387
10000041	0.136345554	1.496	0.448860016
10000042	0.132957689	0.382	0.404262183
10000078	0.099268783	0.177	0.396055221
10000080	0.168322716	0.121	0.393813319
10000082	0.25229966	0.148	0.394894236
10000088	0.184747908	0.132	0.394253693
10000113	0.156549231	0.133	0.394293727
10000116	0.137067707	0.109	0.393332912
10000117	0.140475449	0.09	0.392572267
10000186	0.173212017	0.074	0.391931723
10000189	0.109079262	0.537	0.410467447
10000191	0.20534522	1.12	0.433807247
10000195	0.244341232	0.317	0.401659976
10000209	0.197285099	0.484	0.408345647
10000256	0.197296063	0.188	0.396495595
10000257	0.147098313	0.219	0.397736648
10000271	0.156445831	0.397	0.404862693
10000280	0.131505733	0.673	0.415912066
10000288	0.192534246	0.314	0.401539874
10000290	0.187383812	0.632	0.414270674
10000317	0.226088956	0.998	0.428923103
10000335	0.171168362	0.137	0.394453863

J-2

10000336	0.171168362	0.106	0.39321281
10000351	0.197467741	0.312	0.401459806
10000360	0.163866341	0.43	0.406183813
10000361	0.252271019	0.405	0.405182964
10000385	0.248847347	0.303	0.4010995
10000388	0.177496414	0.276	0.400018583
10000389	0.171597144	0.286	0.400418923
10000390	0.156568023	0.081	0.392211961
10000415	0.252200696	0.143	0.394694067
10000416	0.145189592	0.116	0.39361315
10000419	0.201698127	0.588	0.412509179
10000422	0.245941182	0.154	0.39513444
10000423	0.119361506	0.147	0.394854202
10000428	0.158190069	0.11	0.393372946
10000429	0.138350073	0.101	0.39301264
10000432	0.264787511	0.407	0.405263032
10000436	0.153480465	0.358	0.403301368
10000441	0.171917988	0.11	0.393372946
10000443	0.166229387	0.108	0.393292878
10000446	0.254590136	0.362	0.403461504
10000448	0.171168362	0.226	0.398016885
10000451	0.194565091	0.141	0.394613999
10000455	0.27449051	0.721	0.417833696
10000460	0.153480465	0.145	0.394774134
10000468	0.16913776	0.799	0.420956345
10000471	0.13756497	0.944	0.426761269
10000484	0.200974483	0.245	0.398777531
10000491	0.179881979	0.198	0.396895934
10000506	0.158106326	0.393	0.404702557