



TUGAS AKHIR - KS141501

RANCANG BANGUN APLIKASI PENILAIAN PENERAPAN OPEN DATA BERDASARKAN FRAMEWORK OPEN DATA BAROMETER (STUDI KASUS: KATEGORI DATA KESEHATAN)

APPLICATION DEVELOPMENT FOR OPEN DATA IMPLEMENTATION ASSESSMENT BASED ON THE OPEN DATA BAROMETER FRAMEWORK (STUDY CASE: HEALTH CATEGORY DATA)

TUBAGUS HENDRO PRAMONO
NRP 0521 14 4000 0013

Dosen Pembimbing

- 1. Nur Aini Rakhmawati, S.Kom, M.Sc.Eng, Ph.D**
- 2. Irmasari Hafidz, S.Kom, M.Sc.**

Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018

TUGAS AKHIR - KS141501

**RANCANG BANGUN APLIKASI PENILAIAN
PENERAPAN OPEN DATA BERDASARKAN
FRAMEWORK OPEN DATA BAROMETER (STUDI
KASUS: KATEGORI DATA KESEHATAN)**

TUBAGUS HENDRO PRAMONO
NRP 05211440000013

Dosen Pembimbing

- 1. Nur Aini Rakhmawati, S.Kom, M.Sc.Eng, Ph.D**
- 2. Irmasari Hafidz, S.Kom, M.Sc.**

Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018

TUGAS AKHIR - KS141501

**APPLICATION DEVELOPMENT FOR OPEN DATA
IMPLEMENTATION ASSESSMENT BASED ON THE
OPEN DATA BAROMETER FRAMEWORK (STUDY
CASE: HEALTH CATEGORY DATA)**

**TUBAGUS HENDRO PRAMONO
NRP 05211440000013**

Supervisor

- 1. Nur Aini Rakhmawati, S.Kom, M.Sc.Eng, Ph.D**
- 2. Irmasari Hafidz, S.Kom, M.Sc.**

**Departement of Information Systems
Faculty of Information Technology and Communication
Institut Teknologi Sepuluh Nopember
Surabaya 2018**

LEMBAR PENGESAHAN

RANCANG BANGUN APLIKASI PENILAIAN PENERAPAN OPEN DATA BERDASARKAN FRAMEWORK OPEN DATA BAROMETER (STUDI KASUS: KATEGORI DATA KESEHATAN)

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember

Oleh:

TUBAGUS HENDRO PRAMONO

NRP. 0521 14 4000 0013

Surabaya, 16 Juli 2018

**KEPALA
DEPARTEMEN SISTEM INFORMASI**

Dr. Ir. Aris Tjahyanto, M.Kom.
NIP 19650310 199102 1 001

LEMBAR PERSETUJUAN

RANCANG BANGUN APLIKASI PENILAIAN PENERAPAN OPEN DATA BERDASARKAN FRAMEWORK OPEN DATA BAROMETER (STUDI KASUS: KATEGORI DATA KESEHATAN)

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

Oleh:

TUBAGUS HENDRO PRAMONO

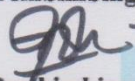
NRP. 0521 14 4000 0013

Disetujui Tim Penguji: Tanggal Ujian: 6 Juli 2018
Periode Wisuda: September 2018


**Nur Aini Rakhmawati, S.Kom, M.Sc.Eng,
Ph.D**


(Pembimbing I)

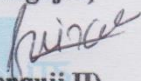
Irmasari Hafidz, S.Kom, M.Sc


(Pembimbing II)

**Renny Pradina Kusumawardani, S.T,
M.T, SCJP**


(Penguji I)

Faizal Johan Atletiko, S.Kom, M.T


(Penguji II)

RANCANG BANGUN APLIKASI PENILAIAN PENERAPAN OPEN DATA BERDASARKAN FRAMEWORK OPEN DATA BAROMETER (STUDI KASUS: KATEGORI DATA KESEHATAN)

Nama Mahasiswa : Tubagus Hendro Pramono

NRP : 0521144000013

Departemen : Sistem Informasi FTIK-ITS

Dosen Pembimbing :

1. Nur Aini Rakhmawati, S.Kom, M.Sc.Eng, Ph.D

2. Irmasari Hafidz, S.Kom, M.Sc

ABSTRAK

Penerapan konsep open data yang sesuai dengan prinsip open data merupakan hal penting bagi setiap pemerintah daerah untuk mewujudkan pemerintahan yang lebih terbuka, partisipatif, dan inovatif sesuai dengan tujuan Open Government Indonesia. Hal ini juga mendukung kebijakan pemerintah yaitu undang-undang nomor 14 tahun 2008 terkait keterbukaan informasi publik. Namun penerapan dari open data ini belum sepenuhnya dapat dinilai dan dibandingkan untuk mengetahui kesesuaian penerapan open data terhadap standar yang ada.

Penilaian dari penerapan open data pada tingkat negara setiap tahunnya telah dilakukan oleh Open Data Barometer. Namun, metodologi yang digunakan dalam melakukan penilaian tersebut melibatkan survei pakar yang ditujukan kepada peneliti dan pemerintah sehingga membutuhkan biaya yang besar dan waktu yang cukup lama untuk menghasilkan penilaian. Padahal jumlah pemerintah daerah di Indonesia lebih banyak daripada negara yang terlibat dalam penilaian

Open Data Barometer sehingga lebih banyak biaya dan waktu yang harus dialokasikan untuk mendapatkan hasil penilaian.

Penggunaan CKAN yang merupakan sistem manajemen open data sebagai alat dalam mengambil metadata dari dataset pada portal data milik pemerintah daerah merupakan salah satu cara untuk mendapatkan informasi yang dibutuhkan dalam melakukan penilaian penerapan open data. Selain itu, teknologi web-scraping juga dapat digunakan untuk mendapatkan informasi terkait open data milik pemerintah daerah Indonesia yang berasal dari sumber yang lain. Kesehatan yang merupakan kepentingan setiap orang merupakan kategori yang cukup relevan untuk dilakukan penilaian berdasarkan penerapan Open Data oleh pemerintah daerah.

Hasil dari penelitian ini merupakan aplikasi berbasis web yang dapat menampilkan hasil penilaian yang dilakukan terhadap open data milik setiap pemerintah daerah di Indonesia. Aplikasi berhasil melakukan penilaian berdasarkan pertanyaan dataset sesuai Open Data Barometer dengan beberapa asumsi yang dibuat. Aplikasi berhasil melakukan kategorisasi menggunakan 17 kata parser yang terbentuk dari proses penghitungan IDF. Berdasarkan nilai yang didapatkan oleh setiap pemerintah daerah di Indonesia, diketahui bahwa 18,6% pemerintah daerah memiliki dataset kategori kesehatan yang diperbarui, 53,6% pemerintah daerah memiliki dataset kategori kesehatan yang tidak diperbarui, dan 27,7% pemerintah daerah tidak memiliki dataset kategori kesehatan.

Kata Kunci: Open Government, Open Data, CKAN, Assessment

APPLICATION DEVELOPMENT FOR OPEN DATA IMPLEMENTATION ASSESSMENT BASED ON THE OPEN DATA BAROMETER FRAMEWORK (STUDY CASE: HEALTH CATEGORY DATA)

Student Name : Tubagus Hendro Pramono
NRP : 0521144000013
Department : Information Systems FTIK-ITS
Supervisors :
1. Nur Aini Rakhmawati, S.Kom, M.Sc.Eng, Ph.D
2. Irmasari Hafidz, S.Kom, M.Sc

ABSTRACT

Implementation of open data concept in accordance with the open data principles is an important thing for every local government to create a more open, participative, and innovative government in accordance with the goals of Open Government Indonesia. It also supports government policy, namely Act 14 of 2008 about Public Information Disclosure. However, the implementation of this open data has not been fully assessed and benchmarked to know the suitability of open data implementation against existing standards.

Assessment of open data implementation at the country level has been done by the Open Data Barometer every year. However, the methodology used in conducting such assessments involves expert surveys directed at researchers and governments. The methodology requires substantial cost and time to produce results, whereas the number of local governments in Indonesia is more than the countries involved in the assessment of Open Data Barometer, more cost and time

should be allocated to acquire the assessment results. The use of CKAN which is an open data management system as a tool in taking metadata from dataset at local government data portal is one way to retrieve the information needed in conducting an assessment of open data implementation. In addition, web-scraping technology can also be used to obtain information related to open data belonging to the local government of Indonesia originating from other sources. Health is a category that is relevant enough to be assessed based on the Open Data implementation by the local government.

The results of this research are a web-based application that can display the results of open data assessment conducted on every local government in Indonesia. The application successfully performs an assessment based on dataset question according to Open Data Barometer with some assumptions made. The application successfully categorizes the dataset using 17 parser words acquired from the IDF calculation process. Based on the results gained by every local government in Indonesia, it is known that 18.6% of local governments have updated health category datasets, 53.6% of local governments have outdated health category datasets, and 27.7% of local governments have no health category dataset.

Keywords: *Open Government, Open Data, CKAN, Assessment*

KATA PENGANTAR

Puji syukur penulis haturkan kepada Allah SWT yang telah melimpahkan rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “Rancang Bangun Aplikasi Penilaian Penerapan *Open Data* berdasarkan *Framework* Open Data Barometer (Studi Kasus: Kategori Data Kesehatan)” sebagai salah satu syarat kelulusan pada Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi Sepuluh Nopember dengan tepat waktu. Semoga apa yang tertulis dalam buku Tugas Akhir ini dapat bermanfaat kepada para pembacanya, dan dapat memberikan kontribusi dalam perkembangan ilmu pengetahuan khususnya bagi bangsa Indonesia.

Dalam penyusunan Tugas Akhir ini tentunya sangat banyak bantuan yang penulis terima dari berbagai pihak baik dalam bentuk doa, semangat, arahan, kritik, saran, dan berbagai bantuan lainnya. tanpa mengurangi rasa hormat penulis secara khusus ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orangtua penulis yang selalu memberikan dukungan materiil dan non-materiil demi terselesainya pengerjaan tugas akhir ini.
2. Ibu Nur Aini Rakhmawati, S.Kom., M.Sc., Eng. Ph.D dan Ibu Irmasari Hafidz, S.Kom, M.Sc selaku dosen pembimbing yang telah membimbing dan membantu penulis selama pengerjaan tugas akhir ini.
3. Ibu Renny Pradina Kusumawardani S.T, M.T, SCJP, Bapak Faizal Johan Atletiko, S.Kom, M.T, dan Bapak Radityo Prasetyanto Wibowo, S.Kom, M.Kom selaku dosen penguji yang telah memberikan masukan-masukan guna menyempurnakan Tugas Akhir ini.
4. Seluruh Dosen Departemen Sistem Informasi ITS yang telah memberikan ilmu pengetahuan yang bermanfaat dan pengalaman yang berharga bagi penulis.

5. Keluarga besar E-Home yang telah menemani kehidupan kampus penulis selama tiga tahun. Memberikan warna pada masa kuliah dengan berbagai tipe mahasiswa yang dapat ditemui.
6. Member UKM Cinta Lab yang menjadi teman mengerjakan, bertukar pikiran, dan berbagi keluh kesah selama pengerjaan tugas akhir.
7. Komplotan BegundalHood yang telah membantu mengurangi beban pikiran dengan memberikan informasi-informasi bermanfaat dan kebahagiaan.
8. Serta seluruh pihak-pihak lain yang tidak dapat disebutkan satu per satu yang telah banyak membantu penulis selama perkuliahan hingga dapat menyelesaikan tugas akhir ini.

Penulis sadar bahwa Tugas Akhir ini masih jauh dari kata sempurna, sehingga saran dan kritik yang membangun dari pembaca merupakan *feedback* yang berarti untuk perbaikan ke depan. Semoga Tugas Akhir ini dapat bermanfaat bagi perkembangan ilmu pengetahuan dan semua pihak.

AFTAR ISI

LEMBAR PENGESAHAN.....	vii
LEMBAR PERSETUJUAN.....	ix
ABSTRAK	xi
ABSTRACT	xiii
KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL.....	xxiii
DAFTAR KODE.....	xxv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Maslaah.....	3
1.4 Tujuan	4
1.5 Manfaat	4
1.6 Relevansi	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Sebelumnya	7
2.1.1 Rancang Bangun Perangkat Lunak untuk <i>Monitoring Website</i> Pemerintahan dengan Pembobotan Kriteria Berbasis <i>Analytical Hierarchy</i> <i>Process (AHP)</i>	7
2.1.2 Towards Surabaya as An Open Data City.....	8
2.1.3 Benchmarking <i>open government: An open data</i> <i>perspective</i>	8

2.1.4	<i>The contextual benchmark method: Benchmarking e-government services</i>	9
2.2	Dasar Teori	9
2.2.1	Open Data.....	9
2.2.2	Open Data Barometer	10
2.2.3	Open Government Data	11
2.2.4	Lisensi Data.....	13
2.2.5	Badan Pusat Statistik	14
2.2.6	Metadata	15
2.2.7	Web-Scraping	17
2.2.8	Term Frequency Inverse Document Frequency ...	18
2.2.9	Comprehensive Knowledge Archive Network	19
BAB III METODOLOGI		23
3.1	Rancangan Arsitektur Sistem.....	23
3.2	Tahapan Pengerjaan Tugas Akhir	27
3.2.1	Studi Literatur.....	28
3.2.2	Akuisisi Data	28
3.2.3	Kategorisasi dan Penilaian Open Data	30
3.2.4	Perancangan dan Pembuatan Visualisasi Aplikasi...	35
3.2.5	Penyusunan Buku Tugas Akhir	37
BAB IV PERANCANGAN		39
4.1	Akuisisi Data	39
4.1.1	Pengumpulan Daftar Situs	39
4.1.2	Pemetaan Data	40
4.1.3	Penyimpanan Data	42
4.1.4	Pengambilan Data.....	48

4.2	Kategorisasi dan Penilaian <i>Open Data</i>	52
4.2.1	Pelabelan Subkategori terhadap Dataset	52
4.2.2	Pra-Proses Data	53
4.2.3	Penentuan Kata <i>Parser</i>	55
4.2.4	Penilaian terhadap setiap Dataset	56
4.2.5	Penilaian Open Data setiap Pemerintah Daerah... ..	56
4.3	Visualisasi Aplikasi	57
4.3.1	Pembuatan Aplikasi Visualisasi	57
4.3.2	Desain Antarmuka	58
BAB V IMPLEMENTASI		61
5.1	Lingkungan Implementasi	61
5.2	Implementasi Akuisisi Data.....	62
5.2.1	Pengumpulan Daftar Situs	62
5.2.2	Penyimpanan Data.....	63
5.2.3	Implementasi Pembuatan <i>Crawler</i>	63
5.3	Implementasi Kategorisasi dan Penilaian Open Data	67
5.3.1	Pelabelan Dataset	67
5.3.2	Pra-Proses Data	69
5.3.3	Teknik Penentuan Kata <i>Parser</i>	70
5.3.4	Pembuatan Kode Penilaian Dataset	71
5.3.5	Pembuatan Kode Penilaian Penerapan <i>Open Data</i> Pemerintah Daerah	72
5.4	Pembuatan <i>Bash Script Scheduler</i>	73
5.5	Implementasi Pembuatan Aplikasi Visualisasi	74
5.5.1	Statistik Penilaian Bulanan	74
5.5.2	Nilai Open Data Terbaik Pemerintah Daerah	75
5.5.3	Informasi Terkait Open Data Pemerintah Daerah	76

5.5.4 Jumlah Dataset berdasarkan Subkategori.....	78
5.5.5 Tren Nilai <i>Open Data</i> suatu Pemerintah Daerah..	79
5.5.6 Pengelolaan Kata <i>Parser</i>	80
BAB VI HASIL DAN PEMBAHASAN	81
6.1 Analisis <i>Parser</i>	81
6.1.1 Memuat data	81
6.1.2 Pra-proses Data.....	82
6.1.3 Pembuatan kata <i>parser</i> berdasarkan nilai IDF	84
6.1.4 Hasil Kategorisasi Dataset Menggunakan Kata <i>Parser</i>	85
6.2 Analisis Penilaian Open Data	91
6.2.1 Jumlah Dataset Ternilai	91
6.2.2 Karakteristik Penilaian Dataset.....	93
6.2.3 Karakteristik Penilaian Penerapan <i>Open Data</i> Pemerintah Daerah	94
6.3 Pengujian Non Fungsional.....	97
6.3.1 Pengujian <i>Load Time</i> Situs	97
6.3.2 Pengujian Akses Perangkat.....	98
6.3.3 Pengujian Akses Peramban.....	101
BAB VII KESIMPULAN DAN SARAN	103
7.1 Kesimpulan.....	103
7.2 Saran	104
DAFTAR PUSTAKA.....	105
LAMPIRAN A – Asumsi terhadap Pertanyaan Dataset	A-1
LAMPIRAN B – Daftar Stopwords Tambahan	B-1
LAMPIRAN C – Daftar 25 kata DF/IDF Terendah setiap Subkategori	C-1
BIODATA PENULIS.....

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Kode CKAN	20
Gambar 3.1 Arsitektur Sistem Opendata Egovbench	24
Gambar 3.2 Tahap Pengerjaan Tugas Akhir.....	27
Gambar 3.3 Tahapan Akuisisi Data.....	28
Gambar 3.4 Tahapan Kategorisasi dan Penilaian Open Data	30
Gambar 4.1 Alur Pengumpulan Situs BPS	39
Gambar 4.2 Alur Pengumpulan Situs Open Data	40
Gambar 4.3 Alur Pemetaan Data.....	40
Gambar 4.4 Alur Pengambilan Data.....	49
Gambar 4.5 Alur Pelabelan Data.....	52
Gambar 4.6 Alur Pra-Proses Data	53
Gambar 4.7 Alur Penentuan Kata Parser	55
Gambar 4.8 Alur Penilaian Dataset	56
Gambar 4.9 Alur Penilaian Penerapan <i>Open Data</i> Pemerintah Daerah.....	57
Gambar 4.10 Rancangan Visualisasi Statistik Bulanan	58
Gambar 4.11 Rancangan Visualisasi Nilai <i>Open Data</i> Terbaik	58
Gambar 4.12 Rancangan Visualisasi Nilai <i>Open Data</i> menurut Jenis Pemerintah Daerah	59
Gambar 4.13 Rancangan Visualisasi Informasi terkait Pemerintah Daerah	59
Gambar 4.14 Rancangan Visualisasi berdasarkan Subkategori Dataset	59
Gambar 4.15 Rancangan Visualisasi Perkembangan Nilai <i>Open Data</i>	60
Gambar 4.16 Rancangan Visualisasi Detail Nilai <i>Open Data</i> Pemerintah Daerah	60
Gambar 5.1 Visualisasi Statistik Bulanan Open Data.....	75
Gambar 5.2 Visualisasi Pemerintah Daerah dengan Nilai <i>Open Data</i> Terbaik	76
Gambar 5.3 Visualisasi Informasi <i>Open Data</i> Pemerintah Daerah.....	77

Gambar 5.4 Visualisasi Jumlah Dataset berdasarkan Subkategori	79
Gambar 5.5 Visualisasi Tren Nilai <i>Open Data</i> Pemerintah Daerah	80
Gambar 5.6 Halaman Pengelolaan Kata <i>Parser</i>	80
Gambar 6.1 Rata-rata dan Standar Deviasi Nilai Peneraan <i>Open Data</i> Pemerintah Daerah pada Periode Juni 2018.....	96
Gambar 6.2 Tampilan pada Komputer dengan Resolusi 1280x800.....	98
Gambar 6.3 Tampilan pada Mobile dengan Resolusi 720x1280 (Potrait)	99
Gambar 6.4 Tampilan pada Mobile dengan Resolusi 720x1280 (Landscape).....	99
Gambar 6.5 Tampilan pada Tablet dengan Resolusi 1024x768 (Potrait)	100
Gambar 6.6 Tampilan pada Tablet dengan Resolusi 1024x768 (Landscape).....	100
Gambar 6.7 Tampilan pada Peramban Google Chrome	101
Gambar 6.8 Tampilan pada Peramban Mozilla Firefox	101
Gambar 6.9 Tampilan pada Peramban Microsoft Edge	102

DAFTAR TABEL

Tabel 3.1 Kategori Dataset berdasarkan Klaster menurut Open Data Barometer	31
Tabel 3.2 Penilaian Open Data	34
Tabel 4.1 Ketersediaan Data yang diambil	41
Tabel 4.2 Alur Akses Basis Data	43
Tabel 4.3 Document Key pada MongoDB	44
Tabel 4.4 Tabel dan Atributnya pada MySQL	45
Tabel 4.5 Data yang disimpan oleh <i>Web-Scraper</i> BPS	50
Tabel 4.6 Data Daftar Dataset yang disimpan CKAN <i>Crawler</i>	51
Tabel 4.7 Data Metadata Dataset yang disimpan CKAN <i>Crawler</i>	52
Tabel 5.1 Spesifikasi Perangkat Lunak	61
Tabel 5.2 Perangkat Lunak Pengembangan Aplikasi	61
Tabel 5.3 Jumlah Daftar Situs yang Terkumpul	62
Tabel 5.4 Contoh Informasi Hasil <i>Scraping</i> Situs BPS	64
Tabel 5.5 Contoh Informasi Hasil <i>Crawling</i> Daftar Dataset CKAN	65
Tabel 5.6 Contoh Informasi Hasil <i>Crawling</i> Metadata Dataset CKAN	67
Tabel 5.7 Hasil Pelabelan Dataset	68
Tabel 5.8 Contoh Hasil Pra-Proses Data	69
Tabel 5.9 Daftar Kata <i>Parser</i> yang Terpilih	71
Tabel 6.1 Distribusi Label berdasarkan Subkategori	81
Tabel 6.2 Contoh Dataset Tidak Terkategorisasi	82
Tabel 6.3 Jumlah Kata setiap Tahapan Pra-Proses	83
Tabel 6.4 Ringkasan Hasil Pra-Proses Data	84
Tabel 6.5 Jumlah Kata <i>Parser</i> setiap Subkategori	84
Tabel 6.6 Confusion Matrix Percobaan Kategorisasi 1	85
Tabel 6.7 Evaluasi Performa Percobaan Kategorisasi 1	86
Tabel 6.8 Confusion Matrix Percobaan Kategorisasi 2	87
Tabel 6.9 Evaluasi Performa Percobaan Kategorisasi 2	87
Tabel 6.10 Confusion Matrix Percobaan Kategorisasi 3	88
Tabel 6.11 Evaluasi Performa Percobaan Kategorisasi 3	89

Tabel 6.12 Confusion Matrix Percobaan Kategorisasi 4.....	90
Tabel 6.13 Jumlah Dataset dan Pemerintah Daerah Pemilik Dataset yang Terambil dan Ternilai	92
Tabel 6.14 Kombinasi Nilai Dataset yang Terjadi Berdasarkan Sumber	93
Tabel 6.15 Nilai Penerapan <i>Open Data</i> yang didapatkan Pemerintah Daerah	94
Tabel 6.16 Nilai Akhir Penerapan <i>Open Data</i> berdasarkan Kepemilikan Dataset Ternilai.....	95
Tabel 6.17 Pengujian Load Time Situs.....	98

DAFTAR KODE

Kode 5.1 <i>Pseudocode Web-Scraper</i> Situs BPS	64
Kode 5.2 <i>Pseudocode Package List Crawler</i> CKAN	65
Kode 5.3 <i>Pseudocode Informasi Dataset Crawler</i> CKAN.....	66
Kode 5.4 <i>Pseudocode Pra-Proses Data</i>	69
Kode 5.5 <i>Pseudocode</i> penghitungan DF/IDF kata pada judul dataset	70
Kode 5.6 <i>Pseudocode</i> Penilaian Dataset	72
Kode 5.7 <i>Pseudocode</i> Penilaian <i>Penerapan</i> Open Data	73
Kode 5.8 <i>Bash Script</i> opendata.sh	73
Kode 5.9 Crontab untuk menjalankan opendata.sh.....	73
Kode 5.10 <i>Query</i> Statistik Bulanan Open Data	74
Kode 5.11 <i>Query</i> Nilai <i>Open Data</i> Terbaik.....	76
Kode 5.12 <i>Query</i> Informasi Open Data Pemerintah Daerah..	77
Kode 5.13 <i>Query</i> Jumlah Dataset tiap Subkategori	78
Kode 5.14 <i>Query</i> Mengambil Nilai <i>Open Data</i> Pemerintah Daerah.....	79
Kode 5.15 <i>Query</i> Pengambilan Kata Parser	80

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

Bab ini akan menguraikan gambaran umum penelitian meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan relevansi terhadap pengerjaan tugas akhir. Tujuan dan manfaat dilakukannya penelitian diharapkan dapat memberikan gambaran awal mengenai penelitian.

1.1 Latar Belakang

Ekspektasi masyarakat terhadap pemerintah semakin meningkat. Sehingga pemerintah tidak bisa hanya mengukur perkembangan dari negara hanya berdasarkan produk domestik bruto, namun juga melakukan penilaian terhadap kualitas hidup dari masyarakat[1]. Sesuai dengan poin kedua dalam agenda prioritas yang disebut sebagai program Nawa Cita, Presiden Jokowi berjanji bahwa "Kami (Jokowi-JK) akan membuat pemerintah selalu hadir dengan membangun tata kelola pemerintahan yang bersih, efektif, demokratis, dan terpercaya"[2]. Menurut Undang-undang no. 14 tahun 2008 tentang Keterbukaan Informasi Publik, warga negara berhak untuk mengetahui rencana pembuatan kebijakan publik, program kebijakan publik guna mendorong partisipasi masyarakat dalam proses pengambilan kebijakan publik[3]. Dengan begitu, masyarakat dapat berperan aktif dalam mewujudkan penyelenggaraan negara yang baik. Bagi pemerintah, undang-undang tersebut mewajibkan badan publik untuk meningkatkan pengelolaan dan pelayanan informasi dalam penyediaan akses terhadap informasi publik.

Sebagai upaya realisasi salah satu program Nawa Cita tersebut, dibentuklah Open Government Indonesia untuk mewujudkan pembangunan pemerintahan yang lebih terbuka, partisipatif, dan inovatif[4]. *Open government* memungkinkan masyarakat untuk berpartisipasi dalam proses pengambilan keputusan. Indonesia sebagai salah satu negara pelopor pada platform

Open Government Partnership telah mengembangkan tujuan dari *open government* berdasarkan kebutuhannya sesuai dengan Reformasi[1]. Salah satu produk yang merupakan hasil dari pembuatan Open Government Indonesia adalah Satu Data Indonesia, sebuah portal resmi *open data* Indonesia yang berisi data kementerian, lembaga pemerintahan, pemerintah daerah, dan semua instansi lain yang menghasilkan data tentang Indonesia dan dapat diakses pada <http://data.go.id> [5]. Portal data Indonesia ini menjadi penilaian kesesuaian konsep *open data* yang diterapkan terhadap *Open Data Charter*, sebuah prinsip dari *open data*. Penilaian tersebut dilakukan dalam proyek Open Data Barometer yang digagas oleh World Wide Web Foundation. Open data barometer bertujuan untuk menemukan penerapan dan dampak dari inisiasi open data di seluruh penjuru dunia. Sampai akhir tahun 2016 kemarin, terdapat total 115 yuridiksi yang dinilai pada proyek tersebut, termasuk Indonesia. Peringkat Indonesia terhadap 115 yuridiksi yang menjadi obyek penilaian terdapat pada urutan ke 38 tahun 2016 kemarin[6]. Penilaian tersebut dilakukan dengan metode survei pakar yang ditujukan kepada peneliti dan pemerintah. Penilaian dengan metode tersebut membutuhkan waktu yang lama yaitu sekitar 5 bulan untuk memperoleh hasilnya[7].

Pemerintah daerah Indonesia juga sudah mulai melakukan penerapan konsep *open government*. Penerapan tersebut dilakukan dengan cara membuat situs yang khusus digunakan untuk mempublikasikan *open data* milik mereka. Namun belum ada alat yang digunakan untuk melakukan penilaian terhadap penerapan yang telah dilakukan. Dari permasalahan tersebut, maka dibutuhkan sebuah *platform* yang mampu melakukan penilaian dan memvisualisasikan hasil penerapan *open government* dari setiap pemerintah daerah. Dengan kondisi jumlah pemerintah daerah Indonesia yang lebih banyak dibandingkan dengan jumlah negara yang terlibat pada proyek Open Data Barometer, maka tentunya waktu yang dibutuhkan juga lebih lama. Oleh karena itu, penggunaan teknologi *CKAN Action API* dan *web-scrapers* sebagai alat untuk melakukan otomatisasi pengambilan informasi mengenai data yang

dibutuhkan untuk menjawab survei mempersingkat waktu yang dibutuhkan. Hasil dari informasi yang diambil tersebut diproses berdasarkan metodologi yang digunakan oleh Open Data Barometer, kemudian divisualisasikan untuk mengetahui hasil penerapan yang telah dilakukan oleh pemerintah daerah. Melalui penelitian ini diharapkan pemerintah daerah dapat mengetahui kondisi penerapan yang telah dilakukannya dan dapat melakukan peningkatan implementasi agar mendapatkan nilai yang lebih baik.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang, maka perumusan permasalahan yang menjadi fokus dan akan diselesaikan dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana menentukan kategorisasi dan penilaian terhadap penerapan *open data* milik setiap pemerintah daerah Indonesia?
2. Bagaimana cara melakukan akuisisi terhadap *open data* milik setiap pemerintah daerah Indonesia?
3. Bagaimana memvisualisasikan hasil penilaian *open data* milik pemerintah daerah Indonesia?

1.3 Batasan Maslaah

Berdasarkan permasalahan yang dihadapi, terdapat batasan masalah yang ditetapkan sebagai fokus pengerjaan tugas akhir. Batasan yang ditetapkan dalam tugas akhir ini adalah sebagai berikut:

1. Hasil akhir dari penelitian ini adalah pengembangan aplikasi web yang memiliki fitur menampilkan nilai penerapan *open data* pemerintah daerah Indonesia.
2. *Open data* yang dinilai merupakan milik pemerintah daerah Provinsi, Kabupaten, dan Kota yang ada di Indonesia.
3. *Open data* yang dinilai merupakan *open data* yang termasuk dalam kategori kesehatan menurut Open Data Barometer.

4. Penilaian *open data* dilakukan berdasarkan ketersediaan data yang dapat diambil menggunakan teknik *scraping*.
5. *Open data* milik setiap pemerintah daerah Indonesia didapatkan dari situs BPS dan portal data berbasis CKAN milik setiap pemerintah daerah Indonesia.

1.4 Tujuan

Berdasarkan perumusan masalah dan batasan masalah yang telah disebutkan sebelumnya, tujuan yang ingin dicapai dari pengerjaan tugas akhir ini adalah pembuatan *platform* yang dapat melakukan pengkategorisasian dan penilaian *open data* yang dimiliki oleh pemerintah daerah Indonesia serta memvisualisasikan hasil dari penilaian tersebut. Dengan dibuatnya aplikasi ini, diharapkan pemerintah daerah di Indonesia akan lebih memperhatikan pentingnya keterbukaan informasi untuk publik dan masyarakat/penduduk di daerah tersebut dapat ikut berperan dalam mendorong pemerintah daerah agar dapat memberikan hak akses atas data yang berhak untuk diketahuinya.

1.5 Manfaat

Manfaat yang dapat diperoleh dari pengerjaan tugas akhir ini adalah sebagai berikut:

1. Memfasilitasi pemerintah daerah di Indonesia untuk mengetahui penerapan *open data* yang telah dilakukan dibandingkan pemerintah daerah lainnya, sehingga dapat menjadi acuan pemerintah daerah dalam meningkatkan nilai keterbukaan informasi publik berdasarkan hasil penilaian.
2. Membangun aplikasi penilaian *open data* milik pemerintah daerah Indonesia, yang mana dapat dikembangkan untuk menjadi aplikasi yang lebih

baik oleh para peneliti selanjutnya, khususnya mahasiswa Departemen Sistem Informasi.

3. Menjadi dasar aplikasi pengkategorisasian dan penilaian *open data* yang dapat diterapkan untuk melakukan penilaian *open data* pada kategori maupun objek lain.

1.6 Relevansi

Tugas akhir ini berkaitan dengan pembangunan aplikasi web untuk melakukan penilaian penerapan *open data* Pemerintah Daerah Indonesia sesuai dengan undang-undang mengenai keterbukaan informasi publik dengan cara melakukan akuisisi data dari portal *open data* kemudian divisualisasikan kedalam bentuk grafis sehingga sesuai dengan fokus Lab Akuisisi Data dan Diseminasi Informasi. Tugas Akhir ini layak dijadikan sebagai tugas akhir pada tingkat S1, karena tugas ini merupakan pemecahan masalah yaitu penyediaan *platform* yang dapat digunakan oleh setiap pengguna untuk mengetahui nilai pemerintah daerah terhadap data terbuka untuk publik serta sebagai motivasi bagi pemerintah daerah dalam meningkatkan kualitas *e-government* mereka.

Selain itu tugas akhir ini berkaitan dengan mata kuliah Pemrograman Berbasis Web, Konstruksi dan Pengujian Perangkat Lunak, Desain Basis Data, dan Pemrograman Integratif sehingga layak untuk dijadikan sebagai tugas akhir departemen Sistem Informasi.

Halaman ini sengaja dikosongkan

BAB II

TINJAUAN PUSTAKA

Bab ini akan menjelaskan mengenai penelitian sebelumnya yang terkait dengan tugas akhir dan membahas dasar teori yang perlu dipahami untuk dijadikan acuan atau landasan dalam pengerjaan tugas akhir ini.

2.1 Penelitian Sebelumnya

2.1.1 Rancang Bangun Perangkat Lunak untuk Monitoring Website Pemerintahan dengan Pembobotan Kriteria Berbasis *Analytical Hierarchy Process* (AHP)

Penelitian ini berjudul "*Rancang Bangun Perangkat Lunak untuk Monitoring Website Pemerintahan dengan Pembobotan Kriteria Berbasis Analytical Hierarchy Process (AHP)*" oleh Biondi Hasbi Handoko[8]. Peneliti mengembangkan aplikasi berbasis web *Egovbench* yang melakukan pemeringkatan situs web milik pemerintah daerah berdasarkan kelengkapan dari situs terhadap kriteria yang ditetapkan oleh undang-undang dengan metode pembobotan *Analytical Hierarchy Process*. Output dari penelitian ini adalah aplikasi visualisasi *ranking* situs web milik pemerintah daerah berdasarkan kelengkapan isi dari situs web yang dimilikinya.

Tugas akhir ini akan melanjutkan penelitian mengenai aplikasi *Egovbench* dengan menambahkan penilaian terhadap Pemerintah Daerah berdasarkan penerapan *open data* yang berasal dari situs web dan portal data berbasis CKAN milik pemerintah daerah, serta situs Badan Pusat Statistik(BPS) dari setiap Pemerintah Daerah.

2.1.2 Towards Surabaya as An Open Data City

Penelitian ini menjelaskan mengenai kondisi ketersediaan data secara daring oleh pemerintah daerah Kota Surabaya. Pada penelitian ini, dijelaskan bahwa data yang dimiliki oleh Kota Surabaya masih tersebar dan tersedia dalam format yang berbeda-beda(heterogen) sehingga belum terintegrasi[9]. Peneliti mengusulkan sebuah *Open Data City Pipeline* untuk Kota Surabaya dalam melakukan penyediaan data secara daring agar data dapat lebih dimanfaatkan untuk keperluan analisa dengan pembuatan arsitektur sistem *Open Data Pipeline* Surabaya. Data yang tersedia dikumpulkan dan diubah menggunakan *Parser* dan *Annotator* sehingga dapat disimpan kedalam bentuk basis data. Data tersebut selanjutnya dianalisa dengan menggunakan algoritma klasterisasi dan agregasi. Hasil analisa tersebut divisualisasikan untuk memberikan pemahaman dan dipublikasikan kedalam format *Linked Open Data*.

2.1.3 Benchmarking open government: An open data perspective

Penelitian ini mengusulkan sebuah model bernama *Open Government Benchmark*(OpenGovB) untuk melakukan penilaian terhadap konsep *Open Government*[10]. Model ini menggunakan dua ukuran sebagai dasar penilaian dari konsep *Open Government*. Ukuran pertama adalah *e-Government Openness Index*(eGovOI) yang menunjukkan usaha dan kesiapan dari pemerintah dalam mempublikasikan *open dataset* dengan memperhatikan waktu, kebutuhan pengguna, dan peran pengguna pada pemerintahan. Ukuran kedua adalah tingkat kematangan yang menunjukkan perkembangan dan keterbukaan dari pemerintah terhadap konsep yang baru.

Pada penelitian ini, penulis mengadaptasi pengukuran *e-Government Openness Index* (eGovOI) pada bagian *data openness indicator* terhadap informasi yang bisa didapatkan oleh penulis terkait *open data* milik pemerintah dari situs web yang menjadi sumber data.

2.1.4 *The contextual benchmark method: Benchmarking e-government services*

Penelitian ini mengusulkan sebuah metode bernama *Contextual Benchmark Method* atau disebut sebagai CBM sebagai metode pendekatan dalam melakukan *e-government benchmark*[11]. Metode ini mengkombinasikan kebutuhan *benchmarking* terhadap konteks yang diinginkan. Konteks yang dimaksud berupa penyesuaian dengan kebutuhan lokal, dan *benchmarking* dilakukan sesuai dengan data terhadap objek yang dibandingkan. Suatu *framework* dibuat dengan membagi konteks dalam beberapa level atau bagian, dan menetapkan teknik atau aspek yang dapat menilai setiap konteks yang ditetapkan. CBM dapat digunakan untuk menganalisa permasalahan yang sesuai terhadap *framework* dengan mengetahui pada level atau aspek manakah suatu organisasi atau objek *benchmark* memiliki nilai yang lebih rendah dan harus ditingkatkan.

Pada penelitian ini, penulis mengadaptasi metode CBM dengan konteks *open data* pada level pemerintah daerah yang dibagi menjadi beberapa kategori dataset dengan aspek karakteristik yang ada pada dataset yang dinilai. Metode ini disesuaikan dengan modifikasi terhadap standar penilaian yang menjadi acuan penilaian *open data* yang digunakan pada Open Data Barometer.

2.2 Dasar Teori

2.2.1 Open Data

Open data merupakan suatu data yang dapat diakses, digunakan, diubah, dan dibagikan oleh siapapun untuk keperluan apapun[12]. Data yang termasuk kedalam kategori *open data* memiliki sifat keterbukaan sebagai berikut[13]:

- **Ketersediaan dan Akses:** Data harus tersedia utuh dan bebas biaya. Data sebaiknya dapat diunduh melalui internet. Data juga harus tersedia dalam

bentuk yang mudah digunakan dan dapat diolah kembali.

- **Penggunaan dan Pendistribusian:** Data yang digunakan harus tersedia dalam lisensi yang memperbolehkan penggunaan dan pendistribusian kembali termasuk pengkombinasian dengan dataset lain. Data harus dapat dibaca oleh mesin.
- **Terbuka untuk umum:** Setiap orang bebas menggunakan dan mendistribusikan kembali dataset. Tidak diperkenankan adanya diskriminasi atas bidang usaha, orang, atau kelompok.

Data yang dimaksud dalam *open data* meliputi data yang memiliki potensial penggunaan dan penerapan seperti data Budaya, Sains, Keuangan, Statistik, Cuaca, dan Lingkungan[14].

2.2.2 Open Data Barometer

Open Data Barometer merupakan sebuah proyek yang bertujuan untuk menemukan penerapan dan dampak dari inisiasi open data di seluruh penjuru dunia. Proyek ini merupakan kolaborasi penelitian antara World Wide Web Foundation dan Open Data for Development. Analisa dilakukan terhadap tren global dengan menyediakan data komparatif antar negara dan region menggunakan metodologi yang mengkombinasikan data kontekstual, penilaian teknis, dan indikator sekunder[15]. Open Data Barometer menilai setiap pemerintahan berdasarkan perkembangannya terhadap *Open Data Charter*, sebuah *framework* untuk menerapkan prinsip *open data* sebagai budaya praktik dari pemerintah yang terbuka[6]. *Open Data Charter* berguna dalam memberikan arahan untuk mempublikasikan data, enam prinsip dalam mempublikasikan data tersebut meliputi[16]:

- Terbuka secara default
- Berbasis waktu dan komprehensif
- Dapat diakses dan digunakan
- Dapat dibandingkan dan dioperasikan (dapat dibaca mesin)

- Digunakan untuk meningkatkan tata kelola pemerintahan dan keterlibatan masyarakat
- Digunakan untuk keperluan inovasi dan perkembangan

Open Data Barometer membagi struktur penilaian berdasarkan 3 sub-indeks dan beberapa komponen setiap indeksnya yaitu[7]:

- a. Readiness
 - Government Policies
 - Government Action
 - Entrepreneurs & Business
 - Citizen & Civil Society
- b. Implementation
 - Accountability dataset cluster
 - Innovation dataset cluster
 - Social Policy dataset cluster
- c. Impacts
 - Political
 - Economic
 - Social

Masing-masing struktur penilaian tersebut didapatkan dari 4 jenis sumber data yang dinilai menggunakan survei kuesioner yaitu:

- Government self assessment
- Peer-reviewed expert survey responses
- Detailed dataset assessment
- Secondary data

2.2.3 Open Government Data

Open Government merupakan sebuah konsep yang menjelaskan bahwa masyarakat memiliki hak akses atas dokumen dan laporan kerja pemerintah untuk dapat memberikan pengawasan publik[17]. Secara umum, Open Government adalah sebuah inisiatif untuk menyediakan informasi pemerintah pada internet. Open Government terbentuk atas hasil memorandum Presiden Barack Obama yang menyatakan bahwa pemerintahan Amerika Serikat harus

memiliki sifat transparan, partisipatif, dan kolaboratif[18]. Presiden Joko Widodo juga ingin membangun pemerintahan Indonesia yang lebih terbuka, partisipatif, dan inovatif dengan dibentuknya Open Government Indonesia. Open Government Indonesia dibentuk dengan upaya membuka pintu sebesar-besarnya bagi seluruh lapisan masyarakat untuk dapat berkolaborasi dan berperan aktif di dalam menentukan prioritas penyelenggaraan pemerintahan di Indonesia[4].

Istilah *open government data* terbentuk dari 2 elemen yaitu *open data* dan *government data*. Memiliki arti sebagai setiap data atau informasi yang dibuat atau dikeluarkan oleh pemerintah, yang dapat digunakan atau dimanfaatkan secara bebas oleh siapapun. *Open government data* dianggap sebagai alat dalam mewujudkan *open government*, kunci dalam meningkatkan layanan, transparansi, dan keterlibatan publik[19]. Dalam mendukung dan mengarahkan pengembangan *open government data*, digagas sebuah prinsip *open government data* dari hasil pertemuan Open Government Working Group Meeting pada bulan Desember 2007. Prinsip *open government data* tersebut meliputi[20]:

1. **Complete:** Seluruh data publik tersedia. Data publik merupakan data yang tidak terbatas oleh privasi, keamanan, dan hak khusus.
2. **Primary:** Data yang diberikan sama seperti data asli yang dikumpulkan, dengan level granularity tertinggi, bukan merupakan hasil agregat atau modifikasi.
3. **Timely:** Data tersedia sesegera mungkin sesuai kebutuhan untuk menjaga nilai data.
4. **Accessible:** Data dapat diakses oleh seluruh pengguna untuk tujuan yang beragam.
5. **Machine processable:** Data tersedia secara terstruktur agar dapat diproses secara otomatis.
6. **Non-discriminatory:** Data tersedia untuk siapapun, tanpa syarat registrasi.

7. **Non-proprietary:** Data tersedia dalam format yang tidak memerlukan kebutuhan khusus.
8. **License-free:** Data tidak terikat hak cipta, paten, merk dagang. Batasan untuk alasan privasi, keamanan, dan hak khusus diperbolehkan.

2.2.4 Lisensi Data

Lisensi data merupakan sebuah instrumen hukum yang menjelaskan apa yang dapat dan tidak dapat dilakukan terhadap data riset yang memiliki lisensi tersebut. Sebaliknya, *wavier* juga merupakan sebuah instrumen hukum namun digunakan sebagai cara untuk menyerahkan hak atas hak cipta yang dimilikinya[21]. Setiap data yang dipublikasikan ke repository biasanya harus memiliki lisensi terhadap data sesuai dengan ketentuan yang ada. *Open license* merupakan sebuah lisensi yang menyatakan bahwa seseorang dapat mengakses, menggunakan kembali, mendistribusikan ulang data dengan beberapa atau tanpa suatu batasan[22]. Sebuah *open license* harus memenuhi syarat sebagai berikut[23]:

- *Reuse*. Lisensi yang diberikan memperbolehkan terjadinya penyalinan, modifikasi, dan penggunaan kembali sesuai dengan persyaratan aslinya.
- *Redistribution*. Lisensi yang diberikan memperbolehkan terjadinya distribusi ulang terhadap data yang dipakai tanpa dikenai biaya apapun.
- *No Discrimination against Persons, Groups, or Fields of Endeavor*. Lisensi yang diberikan tidak boleh mendiskriminasi seseorang atau suatu kelompok. Tidak boleh ada batasan penggunaan data pada suatu aspek.

Standar lisensi yang umum digunakan dalam sebuah open license adalah 2 standar berikut[24]:

- a. Creative Commons License
Creative Commons License merupakan standar lisensi yang cocok digunakan untuk memberikan lisensi terhadap konten kreatif seperti teks, fotografi, *slides*, dan semacamnya. Saat ini versi terakhir *Crative Commons License* yang ada adalah

versi 4.0 yang sudah berstandar internasional. Berdasarkan level lisensi, berikut adalah tipe dari lisensi *Creative Commons*:

- **Public Domain** : Creative Commons Zero (CC0)
- **Attribution** : Creative Commons Attribution 4.0 International (CC-by)
- **Attribution & Share-alike** : Creative Commons Attribution-ShareAlike 4.0 International (CC-by-sa)

b. Open Data Commons License

Open Data Commons License merupakan standar lisensi yang cocok digunakan untuk memberikan lisensi terhadap dataset ataupun database. Berdasarkan level lisensi, berikut adalah tipe dari lisensi *Open Data Commons*:

- **Public Domain** : Open Data Commons Public Domain Dedication and License (PDDL)
- **Attribution** : Open Data Commons Attribution License (ODC-by)
- **Attribution & Share-alike** : Open Data Commons Open Database License (ODbL)

2.2.5 Badan Pusat Statistik

Badan Pusat Statistik adalah Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden. Sebelumnya, BPS merupakan Biro Pusat Statistik, yang dibentuk berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomor 7 Tahun 1960 tentang Statistik. Sebagai pengganti kedua UU tersebut ditetapkan UU Nomor 16 Tahun 1997 tentang Statistik[25]. BPS merupakan Lembaga Pemerintah Non Departemen di Indonesia yang mempunyai fungsi pokok sebagai penyedia data statistik dasar, baik untuk pemerintah maupun untuk masyarakat umum, secara nasional maupun regional. BPS terdapat di setiap provinsi, kabupaten dan kota di seluruh Indonesia. Dinamakan perwakilan BPS di daerah, karena BPS merupakan instansi vertikal, yakni instansi pemerintah pusat yang berada di daerah, sehingga bukan merupakan bagian dari instansi milik daerah, Tugas lain BPS di

daerah adalah melakukan koordinasi dengan pemerintah daerah dalam rangka penyelenggaraan statistik regional[26].

BPS yang berada pada tiap daerah memiliki berbagai data statistik terkait daerahnya. Data yang dimiliki tersebut tersedia pada sub domain situs web milik BPS yaitu <http://bps.go.id>. Data tersebut dibagi kedalam 3 kategori yaitu Sosial dan Kependudukan, Ekonomi dan Perdagangan, Pertanian dan Pertambangan. Setiap BPS daerah mempublikasikan datanya berdasarkan sub-kategori yang ada, data tersebut tersedia dalam berbagai bentuk. Selain data terkait daerah, situs web tersebut juga menyediakan berita resmi statistik dan juga informasi publik.

2.2.6 Metadata

Metadata merupakan suatu data yang menjelaskan data lain. Metadata memberikan informasi terkait sebuah konten dari data[27]. Misalkan data mengenai gambar dapat berisi besar/resolusi dari gambar, kapan gambar dibuat, ukuran file,, dan data lainnya. Misalkan pada data berupa dokumen dapat berisi penulis, tanggal pembuatan, deskripsi singkat dokumen, dan informasi lain terkait dokumen. The World Wide Web Consortium (W3C) sendiri mendefinisikan metadata sebagai informasi mengenai data pada web yang dapat dibaca oleh mesin[28]. Metadata sebuah dokumen dapat berada didalam dokumen tersebut, dokumen terpisah ataupun ditransfer bersama dokumen (sebagai header)[29]. Metadata memberikan informasi dasar mengenai data sehingga pencarian dan interaksi dengan data menjadi lebih mudah[30]. Pada sebuah laman web biasanya terdapat metadata yang biasa disebut sebagai meta tag. Meta tag berisi informasi terkait laman web yang tidak ditampilkan oleh web browser. Informasi tersebut meliputi program yang digunakan dalam pembuatan laman, deskripsi laman, kata kunci yang relevan dengan laman, yang digunakan oleh mesin pencari sebagai alat untuk melakukan indeks pada laman web[31]. Arsitektur dari metadata merepresentasikan sejumlah pernyataan yang independen[29]. Menurut W3C, metadata memiliki peran sebagai berikut[32]:

1. Menghubungkan versi alternatif dari laman web yang sesuai dengan yang tidak sesuai
2. Menemukan dan mendeskripsikan laman alternatif dari versi laman yang dikembangkan, biasanya dirancang untuk penderita disabilitas.
3. Mendeskripsikan sebuah sub-komponen dari suatu laman web

Menurut David Haynes, metadata memiliki fungsi sebagai berikut[33]:

- *Resource Description*: Metadata berfungsi mendeskripsikan suatu data.
- *Information Retrieval*: Metadata berfungsi sebagai konteks semantik pada data, yang mempermudah mesin pencari dalam menemukan data yang relevan.
- *Management of Informations*: Metadata berfungsi dalam melakukan pengelolaan informasi, sehingga lebih mudah untuk diperlakukan sesuai dengan cara mengelolanya.
- *Right Management, Ownership, and Authenticity*: Metadata berfungsi dalam melakukan pengelolaan terkait hak cipta dari sumber informasi, sehingga tidak terjadi plagiarisme.
- *Interoperability*: Metadata berfungsi sebagai media dalam pertukaran dan penggunaan informasi dari dua atau lebih sumber yang berbeda.

Ada pandangan yang mengungkapkan bahwa metadata merupakan salah satu cara untuk meningkatkan efektivitas penggunaan Linked Open Data. Sampai saat ini, metadata yang digunakan dalam Open Data biasanya merupakan *discovery* metadata yang berfungsi sebagai alat untuk membantu menemukan open data yang relevan berdasarkan pencarian atau query[34].

2.2.7 Web-Scraping

Web scraping adalah suatu proses pengambilan dokumen semi terstruktur yang berasal dari internet berupa halaman web dalam bahasa mark up seperti HTML atau XHTML, dokumen tersebut dianalisis dan data tertentu yang sesuai diambil untuk digunakan pada kepentingan lainnya[35]. *Web scraping* melakukan pengindeksan dengan cara menelusuri dokumen HTML atau XHTML dalam sebuah situs pada bagian tag HTML yang dimiliki oleh dokumen, informasi tersebut diambil untuk ditirukan pada aplikasi web-scraping yang dibuat[36]. Aplikasi *web scraping* sering juga disebut sebagai *intelligent, automated*, atau *autonomous agents* berfokus hanya pada pengambilan / ekstraksi data dengan ukuran bervariasi[37]. *web scraping* mirip dengan teknik *universal* yang digunakan oleh mesin pencari dalam melakukan pengindeksan, perbedaannya adalah web scraping berfokus pada transformasi web tidak terstruktur yang biasanya berformat HTML menjadi suatu format data terstruktur yang dapat disimpan pada basis data untuk dianalisa[38]. Web scraping memiliki langkah-langkah sebagai berikut[37]:

1. *Create Scraping Template*: Mempelajari dokumen HTML dari situs web yang akan diambil informasinya melalui tag HTML yang berisi informasi yang akan diambil.
2. *Explore Site Navigation*: Mempelajari teknik navigasi pada website yang akan diambil informasinya untuk ditirukan oleh aplikasi *web scraper* yang dibuat.
3. *Automate Navigation and Extraction*: Pembuatan aplikasi *web scraper* yang digunakan untuk melakukan otomatis pengambilan informasi dari web yang ditentukan berdasarkan hasil pembelajaran pada langkah 1 dan 2.
4. *Extracted Data and Package History*: Penyimpanan informasi yang telah didapatkan oleh *web scraper* kedalam basis data yang selanjutnya digunakan untuk kepentingan lain.

2.2.8 Term Frequency Inverse Document Frequency

Term Frequency Inverse Document Frequency atau biasa disebut sebagai TF-IDF merupakan sebuah metode untuk menentukan kata manakah yang lebih cocok untuk digunakan dalam suatu kueri. TF-IDF menghitung bobot untuk setiap kata pada dokumen berdasarkan proposi terbalik dari frekuensi suatu kata muncul pada suatu dokumen terhadap persentase dari jumlah dokumen yang mengandung kata tersebut[39].

Term Frequency digunakan dalam mengukur banyak jumlah kata yang muncul dalam satu dokumen. Rumus penghitungan *term frequency* adalah:

$$tf(t, d) = 1 + \log(f_{t,d})$$

Keterangan :

$tf(t, d)$ = *term frequency*

$f_{t,d}$ = jumlah suatu kata muncul dalam satu dokumen

Inverse Document Frequency digunakan dalam menghitung pentingnya suatu kata. Rumus penghitungan *inverse document frequency* adalah:

$$idf_t = \left| \log\left(\frac{N}{df}\right) \right|$$

Keterangan :

idf_t = *inverse document frequency*

df = total kata (t) yang muncul pada dokumen (d)

N = total dokumen

Sehingga untuk menghitung nilai TF-IDF digunakan rumus:

$$tf - idf = (tf) \times (idf)$$

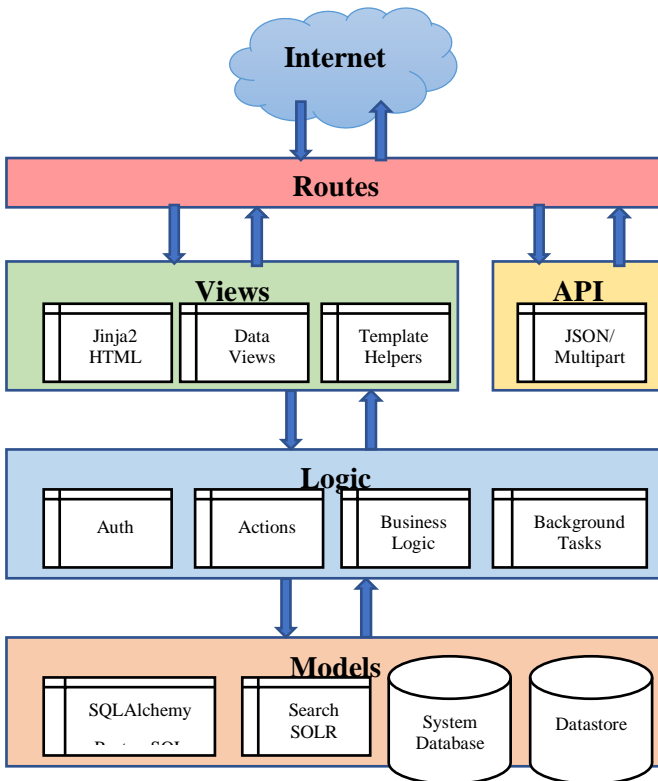
2.2.9 Comprehensive Knowledge Archive Network

Comprehensive Knowledge Archive Network atau biasa disebut sebagai CKAN merupakan sebuah sistem manajemen data untuk membuat situs web *open data* yang dapat diakses. CKAN digunakan dalam mengelola dan mempublikasikan data agar dapat dicari dan digunakan kembali oleh orang lain. CKAN merupakan perangkat lunak *open source* yang dikembangkan oleh komunitas aktif yang tergabung pada CKAN Association[40]. CKAN memberikan setiap dataset suatu halaman yang berisi banyak metadata sehingga menjadi data yang lebih bernilai dan mudah dicari[41].

CKAN dibangun dengan menggunakan *Python* pada *backend* dan *Javascript* pada *frontend*, menggunakan *framework Pylons* dan *SQLAlchemy* sebagai ORM. Database yang digunakan adalah *PostgreSQL* dengan mesin pencari menggunakan *SOLR*[40].

CKAN memiliki banyak fitur, fitur yang akan digunakan dalam penelitian ini adalah *CKAN Action API* yang merupakan *RPC-Style API* yang digunakan untuk memanfaatkan fitur utama CKAN sebagai pengguna API. Fitur lain yang dimiliki CKAN diantaranya: *Datastore*; *Extend*; *Federate*; *FileStore*; *Geospatial*; *Metadata*; *Publish and Manage Data*; *Search and Discovery*; *Themable*; dan *Visualization*; yang dapat dibaca lebih lanjut pada laman <https://ckan.org/features/> [42].

Berikut merupakan arsitektur kode dari CKAN:



Gambar 2.1 Arsitektur Kode CKAN

1. Routes
Bagian ini mendefinisikan koneksi yang terjadi antara URL CKAN dengan *Views* yang memproses *request* dan mengembalikan *response*.
2. Views
Bagian ini membaca dan mengubah data sesuai dengan *request* yang terjadi menggunakan *action functions* dan mengirimkan *responses* dengan template Jinja2 HTML.

Template helpers digunakan untuk kode yang sering digunakan kembali atau kode yang rumit.

3. API

Merupakan bagian yang dapat melakukan *Action functions* pada Logic yang dapat dimanfaatkan untuk mengakses perintah yang dimiliki oleh CKAN menggunakan aplikasi lain/eksternal.

4. Logic

Pada bagian ini, seluruh fungsi-fungsi milik aplikasi disimpan dan dijalankan. Fungsi-fungsi tersebut mencakup fungsi otentikasi, validasi, *action*, perubahan data (*create, delete, get, update*), dan juga *exception*.

5. Models

Pada bagian ini, fungsi-fungsi yang bertugas untuk mengakses basis data pada sistem CKAN yaitu SQLAlchemy dijalankan. Setiap fungsi yang ada pada bagian ini dipanggil atau digunakan melalui bagian *logic* pada arsitektur.

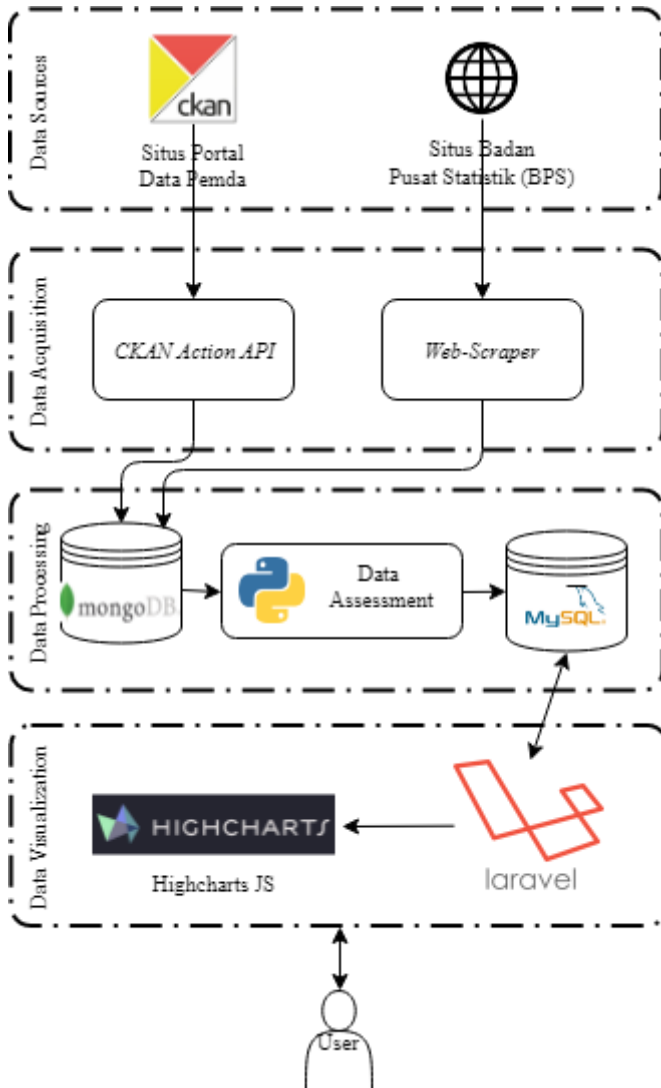
Halaman ini sengaja dikosongkan

BAB III METODOLOGI

Pada bab metodologi akan menjelaskan bagaimana langkah pengerjaan tugas akhir dengan disertakan deskripsi dari setiap penjelasan untuk masing-masing tahapan beserta jadwal kegiatan pengerjaan tugas akhir.

3.1 Rancangan Arsitektur Sistem

Pada sub bab ini akan dijelaskan tentang rancangan Arsitektur Sistem. Arsitektur Sistem berisi desain dari sistem aplikasi yang akan dibuat, terdiri dari komponen-komponen yang membentuk suatu aplikasi. Arsitektur sistem untuk aplikasi *Egovbench* dijelaskan seperti pada gambar berikut:



Gambar 3.1 Arsitektur Sistem Opendata EgoVench

Penjelasan arsitektur sistem:

Pada aplikasi *Egovbench* terdapat 3 proses utama yaitu :

1. *Data Acquisition*

Data Acquisition bertujuan untuk mengambil data atau informasi milik pemerintah daerah yang diambil dari berbagai sumber. Data atau informasi yang didapatkan diolah dan disimpan sesuai dengan kriteria yang telah ditentukan.

2. *Data Processing*

Data Processing merupakan tahapan untuk mengolah data yang telah diambil sesuai dengan metode yang dipakai. Pada penilaian terhadap *open data*, standar yang digunakan disesuaikan dengan cara penilaian yang dilakukan oleh Open Data Barometer.

3. *Data Visualization*

Data Visualization merupakan proses menampilkan hasil penilaian yang telah dilakukan sehingga dapat memperlihatkan perkembangan nilai dari setiap pemerintah daerah dan perbandingannya terhadap pemerintah daerah yang lain. Visualisasi yang dilakukan merupakan detail nilai yang didapatkan oleh setiap pemerintah daerah.

Egovbench memiliki komponen-komponen sebagai berikut :

1. *Web-Crawler*

Merupakan alat yang digunakan untuk mengambil URL dari daftar situs web milik pemerintah yang dimiliki sesuai dengan *regular expression(regex)* yang ditetapkan.

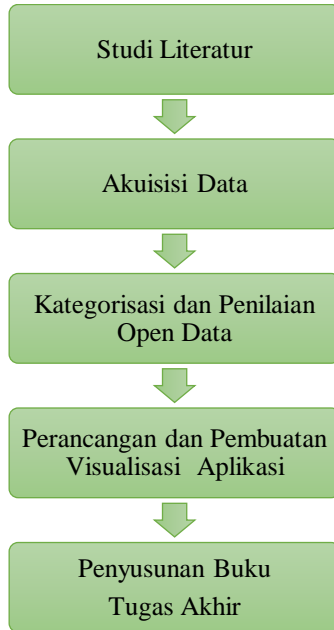
2. *Web-Scraper*

Merupakan alat yang digunakan untuk mengambil suatu informasi yang dikehendaki dari suatu laman web yang terdaftar. *Web-Scraper* ini digunakan untuk mengambil metadata yang dibutuhkan untuk melakukan penilaian terhadap dataset yang ada.

3. *CKAN Action API*
Merupakan sebuah API yang dimiliki oleh *platform open data CKAN* yang memungkinkan pengambilan daftar dataset dan informasi metadata terkait dataset pada portal *open data* milik pemerintah daerah. Informasi didapatkan dalam format JSON dan diambil sesuai dengan keperluan penilaian terhadap *open data*.
4. *MySQL*
Merupakan teknologi basis data relasional yang digunakan untuk menyimpan informasi situs web, sosial media, portal *open data* yang dimiliki oleh pemerintah daerah. Hasil penilaian juga disimpan pada basis data ini.
5. *MongoDB*
Merupakan teknologi basis data non relasional yang digunakan menyimpan informasi hasil *crawling* yang dilakukan.
6. *Laravel*
Merupakan *framework* bahasa pemrograman PHP dengan pola arsitektur *Model-View-Controller* yang digunakan sebagai pembentuk situs *Egovbench*.
7. *Data Assessment*
Merupakan sebuah penilaian terhadap data yang telah dikumpulkan berdasarkan metode yang digunakan. Scoring dilakukan untuk mendapatkan nilai yang akan ditampilkan kepada pengguna.
8. *Highchart JS*
Merupakan *library javascript* yang digunakan dalam memvisualisasikan hasil penilaian yang dilakukan pada setiap *e-government* milik pemerintah daerah.

3.2 Tahapan Pengerjaan Tugas Akhir

Pada sub bab ini akan dijelaskan mengenai metodologi dalam pelaksanaan tugas akhir. Metodologi tersebut dapat dilihat pada gambar berikut:



Gambar 3.2 Tahap Pengerjaan Tugas Akhir

Berdasarkan gambar diatas berikut adalah penjelasan dari tiap-tiap proses yang dilakukan dalam menyelesaikan tugas akhir ini.

3.2.1 Studi Literatur

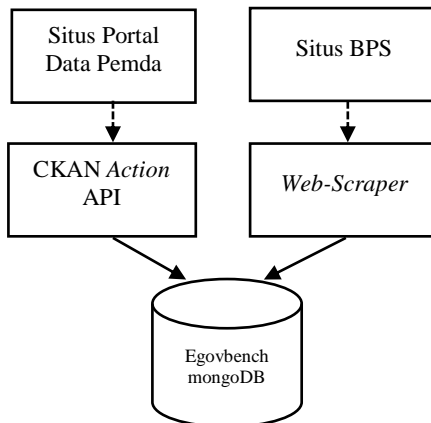
Pada proses studi literatur dilakukan pencarian terhadap materi-materi yang terkait dengan penelitian tugas akhir. Materi-materi yang dicari meliputi :

1. Konsep *open data*
2. Konsep lisensi data
3. Konsep *metadata*
4. Konsep TF-IDF
5. Kategorisasi dan penilaian penerapan *open data* berdasarkan ketersediaan dataset
6. Teknik akuisisi data menggunakan *CKAN Action API*
7. Pemrograman PHP menggunakan *framework* Laravel
8. Teknik pembuatan grafik visualisasi menggunakan *Highcharts JS*.

Penilaian yang dilakukan terhadap *open data* mengacu pada proyek internasional Open Data Barometer[13].

3.2.2 Akuisisi Data

Pada proses akuisisi data, data yang akan digunakan dalam penilaian *open data* diambil dan disimpan kedalam basis data dalam beberapa tahapan sebagai berikut:



Gambar 3.3 Tahapan Akuisisi Data

a. Mendaftar situs portal data dan BPS

Pengumpulan situs portal data dilakukan dengan cara melakukan pencarian secara manual pada mesin pencari Google dengan memasukkan kata kunci “site:.go.id open data <nama pemda>” ataupun “site:.go.id” portal data <nama pemda>”. Sedangkan untuk situs BPS, pencarian juga dilakukan secara manual pada mesin pencari Google dengan memasukkan kata kunci “site:bps.go.id <nama pemda>”. Seluruh data situs portal data dan BPS milik pemerintah daerah selanjutnya disimpan sebagai sumber data yang akan diambil.

b. Mengambil data dari sumber data

Berdasarkan sumber data yang telah didaftar pada aktivitas sebelumnya, *metadata* dari dataset yang tersedia pada sumber data diambil menggunakan 2 alat yaitu:

1. *CKAN Action API*

Pengambilan informasi dataset menggunakan API dilakukan dengan cara melakukan *request* ke situs portal data milik pemerintah daerah terhadap dataset yang dimiliki.

2. *Web-Scraper*

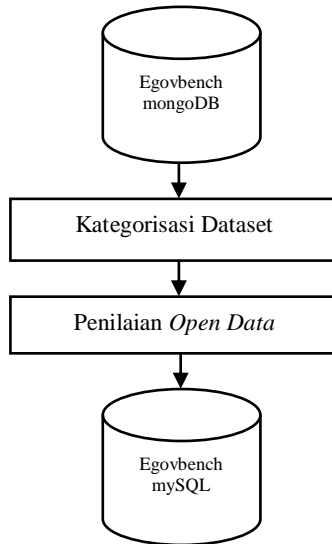
Pengambilan informasi dataset menggunakan *web-scrapers* dilakukan dengan cara membaca tag HTML dari dataset pada situs BPS milik setiap pemerintah daerah yang tersedia. Tag HTML yang sesuai dengan kebutuhan metadata diambil oleh *web-scrapers*.

c. Penyimpanan Hasil Akuisisi Data

Hasil data yang dikumpulkan oleh *crawler* disimpan kedalam basis data MongoDB. Hal itu dilakukan untuk mengetahui hasil dari akuisisi data yang telah dilakukan, apakah sesuai dengan yang diinginkan atau tidak. Selain itu, penggunaan MongoDB sebagai basis data untuk menyimpan informasi dataset dikarenakan atribut yang tersedia untuk setiap dataset dapat berbeda-beda.

3.2.3 Kategorisasi dan Penilaian Open Data

Pada proses kategorisasi dan penilaian open data, hasil dari akuisisi data diolah untuk menghasilkan nilai berdasarkan standar Open Data Barometer. Berikut tahapan yang dilakukan:



Gambar 3.4 Tahapan Kategorisasi dan Penilaian Open Data

Setelah dataset berhasil diakuisisi sesuai dengan informasi yang dibutuhkan, dilakukan proses kategorisasi dan penilaian terhadap dataset. Kategorisasi dan penilaian ini dilakukan berdasarkan *framework* Open Data Barometer. Hasil dari penilaian ini merupakan nilai penerapan *open data* yang telah dilakukan oleh pemerintah daerah tersebut.

a. Kategorisasi Dataset

Berdasarkan metadata yang telah dikumpulkan pada aktivitas sebelumnya, dilakukan kategorisasi yang sesuai dengan panduan penelitian edisi keempat oleh Open Data Barometer yang dibagi kedalam klaster berikut[7, p. 12]:

Tabel 3.1 Kategori Dataset berdasarkan Kluster menurut Open Data Barometer

Kluster	Penjelasan	Kode	Kategori Data
Inovasi	Data yang biasanya dimanfaatkan oleh para pengusaha dan bernilai bagi perusahaan.	D1	Map Data
		D9	Public Transport Timetables
		D10	International Trade Data
		D13	Crime Statistics
		D16	Public Contracts
Kebijakan Sosial	Data yang berguna pada perencanaan, penyampaian, dan kritik terhadap kebijakan sosial dan meningkatkan keterlibatan dan pemberdayaan masyarakat	D4	Detailed Census Data
		D11	Health Sector Performance
		D12	Primary and Secondary education performance data
		D14	Environmental Data

Akuntabilitas	Data untuk mengawasi atau memegang akuntabilitas dari pemerintah dan perusahaan.	D2	Land Ownership Data
		D5	Detailed Government Budget
		D6	Detailed Government Spend
		D7	Company Register
		D8	Legislation
		D15	National Election Results
		D17	Information on Public Officials (NEW)

Pada tugas akhir ini, penelitian difokuskan untuk melakukan kategorisasi dataset pada kategori kesehatan. Dalam panduan, dijelaskan bahwa kategori kesehatan meliputi statistik yang diperoleh dari data administrasi yang dapat digunakan untuk mengindikasikan performa dari suatu layanan atau keseluruhan sistem kesehatan. Dataset yang termasuk dalam kategori kesehatan meliputi[13, p. 49]:

1. *Mortality and Survival Rates*
2. *Levels of Vaccination*
3. *Levels of Access to Health Care*
4. *Health Care Outcomes for Particular Groups*
5. *Patient Satisfaction with Health Services*
6. *Waiting Times for Medical Treatment*
7. *Spend per Admission*

Adapun dalam melakukan kategorisasi terhadap dataset yang telah diambil, dilakukan beberapa tahapan pra-proses data dengan tujuan mendapatkan kata yang tepat untuk menjadi kata kunci pengkategorisasian data. Tahapan pra-proses tersebut adalah:

1. *Cleansing*
2. *Casefolding*
3. *Tokenizing*
4. *Stopwords Removal*

Setelah data dilakukan pra-proses tersebut, maka dicarilah kata kunci yang paling sesuai untuk setiap dataset yang telah disebutkan menggunakan metode TF-IDF. Hasil TF-IDF tersebut digunakan sebagai *parsing* untuk mengkategorikan setiap dataset yang telah diakuisisi.

b. Penilaian Open Data

Setelah dataset yang diambil dikategorisasikan pada aktivitas sebelumnya, maka dilakukan penilaian karakteristik terhadap kategori dataset kesehatan tersebut. Penilaian karakteristik dataset dilakukan dengan memberikan nilai 1 untuk YES dan 0 untuk NO sesuai dengan panduan penelitian edisi keempat Open Data Barometer[13, pp. 54–65]. Setelah nilai tersebut diberikan, maka dilakukan pembobotan terhadap setiap nilai pertanyaan yang telah dihasilkan. Detail pembobotan sesuai dengan metodologi yang digunakan dalam Open Data Barometer edisi keempat seperti pada tabel 3.2 berikut[7, pp. 6–7]:

Tabel 3.2 Penilaian Open Data

Kode	Pertanyaan	Bobot	Logika Kondisi
a	Does the data exist?	5	IF a = No THEN 0 ELSE 5
b	Is it available online from government in any form?	10	IF a = No THEN 0 ELSE (IF b = Yes THEN 10 ELSE 0)
c	Is the dataset provided in machine readable format?	15	IF b = No THEN 0 ELSE (IF c = Yes THEN 15 ELSE 0)
d	Is the machine readable data available in bulk?	15	IF c = No THEN 0 ELSE (IF d = Yes THEN 15 ELSE 0)
e	Is the dataset available free of charge?	15	IF c = No THEN 0 ELSE (IF e = Yes THEN 15 ELSE 0)
f	Is the data openly licensed?	15	IF c = No THEN 0 ELSE (IF f = Yes THEN 15 ELSE 0)
g	Is the dataset up to date?	10	IF c= No THEN 0 ELSE (IF g = No THEN -5)

			ELSE (IF c = Yes AND g = Yes THEN 10)
h	Is the publication of the dataset sustainable?	5	IF c = No THEN 0 ELSE (IF h = Yes THEN 5 ELSE 0)
i	Was it easy to find information on this dataset?	5	IF c = No THEN 0 ELSE (IF i = Yes THEN 5 ELSE 0)
j	Are (linked) data URIs provided for key elements in the dataset?	5	IF c = No THEN 0 ELSE (IF j = Yes THEN 5 ELSE 0)

3.2.4 Perancangan dan Pembuatan Visualisasi Aplikasi

Pada tahap ini dilakukan pengembangan aplikasi *Egovbench* yang telah dibuat sebelumnya. Pengembangan aplikasi dilaksanakan dengan melakukan penambahan modul yang dirancang berdasarkan kebutuhan fungsional aplikasi sebagai alat untuk melakukan penilaian dan *benchmarking open data* milik pemerintah daerah Indonesia. Berikut merupakan tahapan dalam melakukan pengembangan aplikasi *Egovbench*:

a. Analisis Kebutuhan Aplikasi

Pembuatan aplikasi harus sesuai dengan tujuan penelitian yang telah direncanakan. Analisis kebutuhan dilakukan sehingga menghasilkan daftar kebutuhan fungsional sebagai berikut:

1. Aplikasi dapat mengambil metadata *open dataset* yang berasal dari portal data milik pemerintah daerah
2. Aplikasi dapat mengambil metadata *open dataset* yang berasal dari situs BPS.
3. Aplikasi dapat menyimpan hasil akuisisi data.
4. Aplikasi dapat melakukan penilaian terhadap *open dataset* yang telah dikumpulkan.
5. Aplikasi dapat menampilkan hasil penilaian *open dataset* yang telah dikumpulkan.

b. Desain Aplikasi

Pada tahap ini, dilakukan pembuatan desain aplikasi berdasarkan arsitektur sistem dan kebutuhan fungsional aplikasi yang telah dibuat. Desain aplikasi akan menjelaskan gambaran umum dari aplikasi sebelum melakukan rancang bangun. Desain aplikasi yang dibuat mencakup:

1. Desain alur sistem
2. Desain basis data
3. Desain antar muka

c. Rancang Bangun Aplikasi (*Coding*)

Pada tahap ini, modul aplikasi *Egovbench* mulai dibuat sesuai dengan desain aplikasi yang telah dibuat. Beberapa bagian rancang bangun aplikasi yang dibuat meliputi:

1. Rancang Bangun Akuisisi Data
2. Rancang Bangun Penilaian *Open Data*
3. Rancang Bangun Visualisasi

d. Pengujian Aplikasi

Pada tahap ini, modul aplikasi yang telah dibuat diuji untuk mengetahui kesesuaian fitur aplikasi dengan tujuan pembuatan aplikasi. Pengujian aplikasi yang dilakukan adalah pengujian kata *parser* dan penilaian terhadap *open data* untuk memastikan bahwa aplikasi dapat berfungsi sesuai dengan tujuan pembuatannya. Selain itu, aplikasi diuji pada beberapa perangkat untuk mengetahui kesesuaian hasil visualisasi yang didapatkan.

3.2.5 Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan dokumentasi pada setiap tahapan pelaksanaan Tugas Akhir mulai dari studi literatur hingga perancangan dan pembuatan visualisasi dengan website. Penyusunan laporan dikerjakan selama penelitian berlangsung dan juga sebagai aktivitas akhir dalam kegiatan pengerjaan Tugas Akhir ini.

Halaman ini sengaja dikosongkan

BAB IV PERANCANGAN

Pada bab ini membahas terkait alur perancangan terkait beberapa hal yang diperlukan dalam proses pembuatan aplikasi sesuai dengan alur yang dijelaskan pada bab metodologi. Adapun perancangan ini diperlukan sebagai panduan dalam melakukan penelitian tugas akhir, yang dijelaskan sebagai berikut.

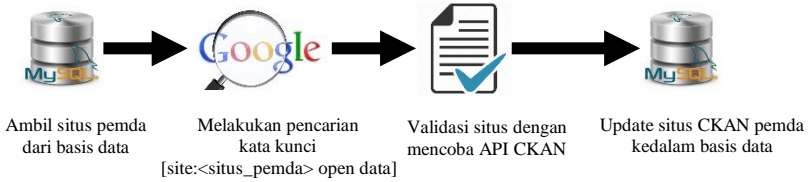
4.1 Akuisisi Data

4.1.1 Pengumpulan Daftar Situs

Dalam melakukan pengumpulan daftar situs, hal yang pertama dilakukan adalah mendapatkan daftar pemerintah daerah yang ada di Indonesia. Hal itu dilakukan dengan cara mengakses data daftar pemerintah daerah yang telah dimiliki dari basis data Egovbench sebelumnya. Selain itu, untuk mendapatkan situs open data milik pemerintah daerah, dibutuhkan situs web utama dari setiap pemerintah daerah yang dapat diperoleh juga dari basis data Egovbench sebelumnya. Selanjutnya, dilakukan pencarian situs menggunakan mesin pencari google dengan memanfaatkan kata kunci berdasarkan nama pemerintah daerah dan situs pemerintah daerah. Validasi terhadap situs dilakukan secara manual dengan membuka situs hasil pencarian yang sesuai. Setelah dipastikan situs tersebut valid, maka dimasukkan kedalam daftar situs pada basis data MySQL



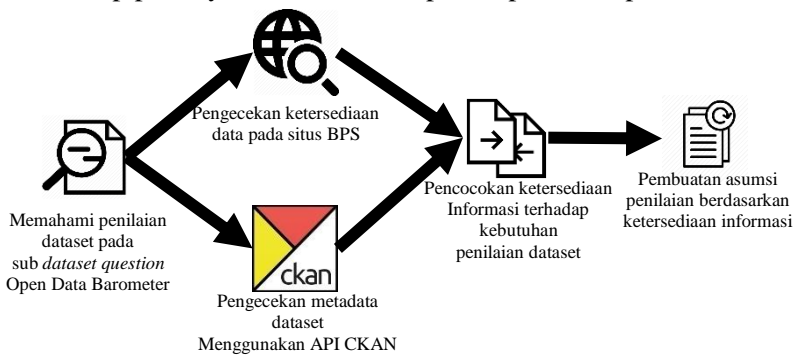
Gambar 4.1 Alur Pengumpulan Situs BPS



Gambar 4.2 Alur Pengumpulan Situs Open Data

4.1.2 Pemetaan Data

Dalam melakukan pemetaan data, dilakukan studi literatur terhadap penilaian yang dilakukan oleh Open Data Barometer edisi keempat pada dataset yang didapatkan pada subbab *Dataset Question* pada *Open Data Barometer 4th Edition – Research Handbook*[13, pp. 54–65]. Penilaian terkait dataset dilakukan dengan menjawab sekumpulan pertanyaan terkait dataset sesuai dengan kondisi dataset yang ada, pada penelitian ini pertanyaan dijawab dengan cara mendapatkan informasi yang tersedia dari situs asal dataset seperti yang dijelaskan pada Tabel 4.1. Asumsi dibuat terhadap setiap pertanyaan dataset sesuai apa yang bisa didapatkan, melihat kondisi nyata dari situs yang menjadi sumber dataset, apabila situs telah memiliki standar yang sama, maka penilaian dapat disamaratakan untuk seluruh pemerintah daerah, penjelasan terkait asumsi penulis terhadap pertanyaan dataset dilampirkan pada Lampiran A.



Gambar 4.3 Alur Pemetaan Data

Tabel 4.1 Ketersediaan Data yang diambil

NO	OPEN DATA BAROMETER	PORTAL DATA BERBASIS CKAN	BADAN PUSAT STATISTIK
1	<i>Does the data exist?</i>	Dataset	Dataset
2	<i>Is it available online from government in any form?</i>	Dataset	Dataset
3	<i>Is the dataset provided in machine readable format?</i>	Dinilai berdasarkan atribut "format" pada JSON	Apabila dataset tersedia, maka dapat memenuhi kriteria ini karena format yang tersedia berupa xls.
4	<i>Is the machine readable data available in bulk?</i>	Portal open data tidak menyediakan fitur download banyak dataset secara langsung, maka tidak dapat memenuhi kriteria.	Situs BPS tidak menyediakan fitur download banyak dataset secara langsung, maka tidak dapat memenuhi kriteria.
5	<i>Is the dataset available free of charge?</i>	Dataset tersedia secara gratis tanpa pembayaran apapun, maka apabila dataset tersedia dapat memenuhi kriteria.	Dataset tersedia gratis tanpa pembayaran apapun, maka apabila dataset tersedia dapat memenuhi kriteria
6	<i>Is the data openly licensed?</i>	Dinilai berdasarkan atribut "licence/license title" pada JSON	Tidak ada keterangan yang menjelaskan lisensi dataset pada situs BPS, maka tidak dapat memenuhi kriteria
7	<i>Is the dataset up to date?</i>	Dinilai berdasarkan atribut "created" pada JSON, ditentukan berdasarkan sensus berkala yang	Dinilai berdasarkan tahun pada judul dataset BPS, ditentukan berdasarkan sensus berkala yang dilakukan pemerintah

		dilakukan pemerintah	
8	<i>Is the publication of the dataset sustainable?</i>	Dinilai berdasarkan atribut “created” pada JSON, sehingga memenuhi kriteria apabila dataset juga memenuhi kriteria sebelumnya	Sensus dilakukan secara berkala, maka dapat memenuhi kriteria apabila dataset juga memenuhi kriteria sebelumnya.
9	<i>Was it easy to find information on this dataset?</i>	Informasi terkait dataset ini dapat ditemukan dengan mudah melalui portal open data, maka apabila dataset tersedia dapat memenuhi kriteria.	Informasi terkait dataset ini dapat ditemukan dengan mudah melalui situs BPS, maka apabila dataset tersedia dapat memenuhi kriteria.
10	<i>Are (linked) data URIs provided for key elements in the dataset?</i>	Portal open data di Indonesia masih belum ada yang menggunakan linked data, maka tidak dapat memenuhi kriteria.	Situs BPS tidak menggunakan linked data, maka tidak dapat memenuhi kriteria ini

4.1.3 Penyimpanan Data

Dalam melakukan penilaian terhadap open data milik pemerintah daerah Indonesia, dataset kategori kesehatan yang berasal dari situs BPS dan *Open Data* berbasis CKAN merupakan objek dari penilaian. Data yang dibutuhkan merupakan metadata dari dataset yang dapat menjawab pertanyaan-pertanyaan terkait penilaian penerapan Open Data yang dimiliki oleh Open Data Barometer. Dalam pengumpulan metadata dari dataset terkait, dibutuhkan daftar situs BPS dan situs *open data* berbasis CKAN milik pemerintah daerah sebagai *input* dari proses *scraping* yang dilakukan. Data mentah yang dikumpulkan disimpan kedalam basis data MongoDB, sedangkan daftar situs BPS dan CKAN, daftar kata hasil

penghitungan TF-IDF, dan hasil penilaian penerapan *open data* akan dimasukkan kedalam basis data MySQL. Berikut merupakan penjelasan terkait data yang dikumpulkan dan disimpan:

Tabel 4.2 Alur Akses Basis Data

No.	Input	Proses	Output
1	Nama pemerintah daerah dan situs pemerintah daerah daerah (MySQL)	Mendaftar situs BPS dan <i>open data</i> berbasis CKAN milik pemerintah daerah	Daftar situs BPS dan <i>open data</i> berbasis CKAN milik pemerintah daerah (MySQL)
2	Situs BPS dan <i>open data</i> berbasis CKAN (MySQL)	<i>Scraping</i> informasi dataset	Informasi dataset dari situs milik pemerintah daerah (MongoDB)
3	Informasi dataset BPS (MongoDB)	<i>Preprocessing</i> dan penghitungan IDF	Kata <i>Parser</i> untuk kategori terkait (MySQL)
4	Informasi dataset BPS (MongoDB)	Kategorisasi dan penilaian dataset BPS	Nilai dataset BPS (MySQL)
5	Informasi dataset <i>open data</i> berbasis CKAN (MongoDB)	Kategorisasi dan penilaian dataset <i>open data</i> berbasis CKAN	Nilai dataset <i>open data</i> berbasis CKAN (MySQL)
6	Nilai dataset BPS dan <i>open data</i> berbasis CKAN (MySQL)	Penilaian penerapan <i>open data</i> milik pemerintah daerah	Nilai penerapan <i>open data</i> setiap pemerintah daerah (MySQL)
7	Nilai penerapan <i>open data</i> setiap pemerintah daerah (MySQL)	Visualisasi penilaian <i>open data</i> setiap pemerintah daerah	Hasil visual nilai penerapan <i>open data</i> milik pemerintah daerah (Situs Egovbench)

a. Data yang disimpan pada basis data MongoDB

Pada basis data MongoDB, data yang disimpan merupakan data hasil akuisisi data yang berasal dari situs BPS dan open data berbasis CKAN milik pemerintah daerah. Berikut adalah *collection* yang dibuat untuk menyimpan data tersebut:

Tabel 4.3 Document Key pada MongoDB

Collection	Atribut	Keterangan
bps_kesehatan	id_pemda	Id pemerintah daerah pemilik dataset
	title	Informasi berupa judul dari dataset yang diambil dari situs BPS
	update	Informasi terkait pembaruan dataset berdasarkan 4 angka terakhir pada judul dataset
	format	Informasi terkait format dokumen dataset yang dapat diakses
	license	Informasi terkait lisensi dataset yang tersedia
	crawlMonth	Informasi bulan dan tahun dilakukannya pengambilan data
ckan_pemda	id_pemda	Id pemerintah daerah pemilik dataset
	package_id	Array yang berisi daftar id <i>packages</i> yang dimiliki pemerintah daerah pada situs portal data berbasis CKAN
	crawlMonth	Informasi bulan dan tahun dilakukannya pengambilan data
ckan_package	id_pemda	Id pemerintah daerah pemilik dataset
	title	Informasi berupa judul dari dataset yang diambil dari metadata "title"
	update	Informasi terkait pembaruan dataset yang diambil dari metadata "created" didalam metadata "resources"
	format	Informasi terkait format dokumen dataset yang diambil dari metadata

		“format” didalam metadata “resources”
	license	Informasi terkait lisensi dataset yang diambil dari metadata “license_title”
	crawlMonth	Informasi bulan dan tahun dilakukanya pengambilan data

b. Data yang disimpan pada basis data MySQL

Pada basis data MySQL, data yang disimpan meliputi daftar situs pemerintah daerah, daftar situs BPS dan CKAN, daftar kata *parser*, dan hasil penilaian penerapan *open data*. Berikut adalah tabel yang dibuat untuk menyimpan data tersebut:

Tabel 4.4 Tabel dan Atributnya pada MySQL

Tabel	Atribut	Tipe Data	Keterangan
pemda	id_pemda	Int(11)	Id yang digunakan untuk mengidentifikasi pemerintah daerah
	nama_pemda	Varchar(50)	Nama dari pemerintah daerah di Indonesia
	bps_pemda	Varchar(50)	Url dari situs BPS milik pemerintah daerah
	ckan_pemda	Varchar(50)	Url dari situs portal open data milik pemerintah daerah
kategori_opendata	id_kategori	Varchar(4)	Kode dari kategori yang ditentukan oleh Open Data Barometer
	nama_kategori	Varchar(30)	Nama dari kategori yang ditentukan oleh Open Data Barometer
	klaster_kategori	Varchar(30)	Pengelompokan klaster kategori yang ditentukan oleh Open Data Barometer yaitu <i>innovation, social policy, dan accountability</i>
Parser_opendata	id_parser	Int(5)	Id yang digunakan untuk mengidentifikasi subkategori yang dilakukan <i>parsing</i>
	id_kategori	Varchar(4)	Id yang digunakan untuk mengidentifikasi kategori yang dilakukan <i>parsing</i>
	nama_subkategori	Varchar(30)	Nama subkategori yang dilakukan <i>parsing</i>

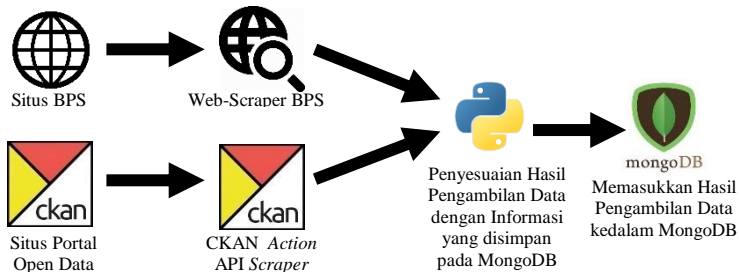
	kata_parser	Varchar(20)	Kata <i>parser</i> yang digunakan dalam menentukan kategori dataset
	timestamp	timestamp	Catatan waktu pembuatan/perubahan dari baris untuk mengidentifikasi kapan waktu parser dibuat
bps_scored_dataset	date	Date	Catatan waktu tanggal penilaian terhadap dataset
	id_pemda	Int(11)	Id pemerintah daerah pemilik dataset
	id_dataset	Varchar(30)	Id yang digunakan untuk mengidentifikasi dataset dari situs BPS yang dinilai
	id_kategori	Varchar(4)	Id kategori dari dataset yang dinilai
	subkategori	Varchar(30)	Nama subkategori dataset yang dinilai
	judul_dataset	Varchar(255)	Judul dari dataset yang dinilai
	aExist	Int(11)	Jawaban pertanyaan “ <i>Does the data exist?</i> ”
	bAvailable	Int(11)	Jawaban pertanyaan “ <i>Is it available online from government in any form?</i> ”
	cMachineReadable	Int(11)	Jawaban pertanyaan “ <i>Is the dataset provided in machine readable format?</i> ”
	dBulk	Int(11)	Jawaban pertanyaan “ <i>Is the machine readable data available in bulk?</i> ”
	eFree	Int(11)	Jawaban pertanyaan “ <i>Is the dataset available free of charge?</i> ”
	fLicense	Int(11)	Jawaban pertanyaan “ <i>Is the data openly licensed?</i> ”
	gUpdated	Int(11)	Jawaban pertanyaan “ <i>Is the dataset up to date?</i> ”
	hSustainable	Int(11)	Jawaban pertanyaan “ <i>Is the publication of the dataset sustainable?</i> ”
iDiscoverable	Int(11)	Jawaban pertanyaan “ <i>Was it easy to find information on this dataset?</i> ”	
jLinked	Int(11)	Jawaban pertanyaan “ <i>Are (linked) data URIs provided for key elements in the dataset?</i> ”	
	date	Date	Catatan waktu tanggal penilaian terhadap dataset

ckan_score_dataset	id_pemda	Int(11)	Id pemerintah daerah pemilik dataset
	id_dataset	Varchar(30)	Id yang digunakan untuk mengidentifikasi dataset dari situs BPS yang dinilai
	id_kategori	Varchar(4)	Id kategori dari dataset yang dinilai
	subkategori	Varchar(30)	Nama subkategori dataset yang dinilai
	judul_dataset	Varchar(255)	Judul dari dataset yang dinilai
	aExist	Int(11)	Jawaban pertanyaan <i>“Does the data exist?”</i>
	bAvailable	Int(11)	Jawaban pertanyaan <i>“Is it available online from government in any form?”</i>
	cMachineReadable	Int(11)	Jawaban pertanyaan <i>“Is the dataset provided in machine readable format?”</i>
	dBulk	Int(11)	Jawaban pertanyaan <i>“Is the machine readable data available in bulk?”</i>
	eFree	Int(11)	Jawaban pertanyaan <i>“Is the dataset available free of charge?”</i>
	fLicense	Int(11)	Jawaban pertanyaan <i>“Is the data openly licensed?”</i>
	gUpdated	Int(11)	Jawaban pertanyaan <i>“Is the dataset up to date?”</i>
	hSustainable	Int(11)	Jawaban pertanyaan <i>“Is the publication of the dataset sustainable?”</i>
	iDiscoverable	Int(11)	Jawaban pertanyaan <i>“Was it easy to find information on this dataset?”</i>
jLinked	Int(11)	Jawaban pertanyaan <i>“Are (linked) data URIs provided for key elements in the dataset?”</i>	
result_opendata	date	Date	Catatan tanggal penilaian penerapan open data milik pemerintah daerah
	id_pemda	Int(11)	Id pemerintah daerah dari hasil penilaian penerapan open data
	id_kategori	Varchar(4)	Id kategori dari penerapan open data yang dinilai
	aResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan aExist

	bResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan bAvailable
	cResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan cMachineReadable
	dResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan dBulk
	eResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan eFree
	fResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan fLicense
	gResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan gUpdated
	hResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan hSustainable
	iResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan iDiscoverable
	jResult	Int(11)	Hasil penghitungan nilai penerapan open data untuk pertanyaan jLinked
	totalScore	Int(11)	Hasil akhir penilaian penerapan open data yang didapatkan dari penjumlahan aResult hingga jResult.

4.1.4 Pengambilan Data

Dalam melakukan pengambilan data, dibuat kode program *scraper* yang berfungsi untuk mendapatkan informasi yang dibutuhkan dalam melakukan penilaian penerapan *open data* milik pemerintah daerah. Pengambilan data dilakukan dengan mendapatkan informasi terkait dataset yang dipublikasikan oleh pemerintah daerah melalui situs BPS dan juga *open data* berbasis CKAN. Kode program *scraper* melakukan akses kedalam basis data MySQL untuk mendapatkan daftar situs yang telah dikumpulkan pada proses sebelumnya, daftar situs ini menjadi input program untuk melakukan proses *scraping*. Setelah proses *scraping* dilakukan, didapatkan informasi yang dibutuhkan untuk selanjutnya dimasukkan kedalam basis data MongoDB.



Gambar 4.4 Alur Pengambilan Data

a. Pembuatan BPS Scraper

Dalam membuat *web-scraper* untuk dataset yang bersumber dari situs BPS, dilakukan analisis secara langsung terhadap struktur halaman kategori pada situs BPS. Analisis dilakukan dengan menentukan data yang harus diambil dari halaman tersebut, setelah data yang dibutuhkan diketahui, maka dilakukan akses terhadap struktur HTML dari halaman BPS tersebut. Halaman yang dicek adalah “<https://bps.go.id/subject/30/kesehatan.html>”. Pada halaman tersebut, diketahui bahwa seluruh informasi yang dibutuhkan untuk diambil berada didalam suatu tabel yang berisi judul dataset kategori kesehatan. Sehingga dalam struktur HTML, dicari tag HTML berupa *table*, ditemukan tag *table* dengan id bernama “listTabel1”. Tag tersebut menjadi masukan pada kode program *web-scraper python* yang dibuat menggunakan *library BeautifulSoup*. Berdasarkan judul dataset yang tersedia pada tabel ini, kita mengambil atribut berupa judul dan tanggal terakhir pembaruan isi dari dataset berdasarkan 4 digit terakhir dari judul dataset. Informasi terkait tipe dokumen bisa didapatkan ketika kita membuka salah satu halaman dengan memilih salah satu judul dataset. Saat ini, seluruh informasi pada dataset tersedia dalam bentuk format file XLS, sehingga seluruh atribut format dataset yang bersumber dari BPS diberikan isi BPS. Informasi terkait lisensi dari dataset tidak tersedia sehingga atribut lisensi dikosongi. Terakhir, informasi pemilik dataset didapatkan dengan atribut `id_pemda`

berdasarkan sub-domain dari BPS yang sedang diakses dari hasil pengumpulan daftar situs. Informasi yang telah dikumpulkan tersebut disimpan kedalam *Collection* *bps_kesehatan* pada basis data MongoDB.

Misalkan pada alamat situs web BPS <https://jakarta.bps.go.id/subject/30/kesehatan.html>, dengan judul dataset “Jumlah Fasilitas Kesehatan, 2010 – 2013”, data yang disimpan oleh *scraper* adalah:

Tabel 4.5 Data yang disimpan oleh Web-Scraper BPS

<i>Key</i>	<i>Value</i>
id_pemda	“162”
title	“Jumlah Fasilitas Kesehatan, 2010 – 2013”
update	“2013”
format	“xls”
license	“”
crawlMonth	“2018-05”

b. Pembuatan CKAN Crawler

Dalam membuat *crawler* untuk dataset yang bersumber pada portal *open data* berbasis CKAN, dilakukan pengecekan ketersediaan penggunaan API dengan cara mengakses situs dan menambahkan “/api/action/package_list” pada url situs. Situs yang memberikan response berupa JSON merupakan situs benar. Pembuatan *crawler* dibagi menjadi 2 tahap, tahap pertama untuk mendapatkan daftar seluruh id dari *packages* atau dataset yang tersedia, tahap kedua untuk mendapatkan metadata dari seluruh *packages* atau dataset yang tersedia. Menggunakan *library urllib*, kode program *python* dibuat untuk mendapatkan hasil *response* dari situs terhadap akses API yang dilakukan. Tahap pertama, penggunaan alamat situs dengan tambahan “/api/action/package_list” memberikan *response* yang memberikan daftar id setiap dataset yang tersedia, informasi ini disimpan pada *Collection* *ckan_pemda* pada basis data MongoDB. Misalkan pada alamat situs portal open data

milik pemerintah daerah Kota Bogor yaitu http://data.kotabogor.go.id/api/action/package_list, data yang disimpan terkait hasil *response* adalah:

Tabel 4.6 Data Daftar Dataset yang disimpan CKAN Crawler

<i>Key</i>	<i>Value</i>
id_pemda	“187”
package_id	[0] a-penetapan-kinerja-tujuan-pembangunan-daerah [1] administrasi-dan-manajemen-umum [2] administrasi-legislasi [3] alamafacebook [4] analisa-alternatif-program-kegiatan-kelurahan-katulampa [5] anggaran-kesehatan-kabupaten-kota [6] anggaran-pendapatan-dan-belanja-daerah-tahun-anggaran-2017-dengan-rahmat-tuhan-yang-maha-esa [7] angka-kematian-pasien [8] angka-kesembuhan-dan-pengobatan-lengkap-tb-paru-bta-serta-keberhasilan-pengobatan [9] aplikasi-atau-sistem-informasi-prioritas-1

Tahap kedua, penggunaan hasil daftar id dataset yang telah disimpan pada *Collection* `ckan_pemda` menjadi masukan dalam melakukan akses terhadap situs portal open data dengan tambahan “`api/action/package_show?id=<package_id >`” yang memberikan *response* JSON berisi metadata dari dataset dengan id tersebut. Metadata yang diambil merupakan metadata yang dibutuhkan dalam melakukan penilaian, sesuai yang telah dijelaskan pada Tabel 4.3. metadata yang dibutuhkan tersebut kemudian disimpan dalam *Collection* `ckan_package` pada basis data MongoDB. Misalkan pada dataset `annakarenina` yang dimiliki oleh pemerintah daerah Salatiga diakses dengan http://data.kotabogor.go.id/api/action/package_show?id=angka-kematian-pasien, data yang disimpan oleh *crawler* adalah:

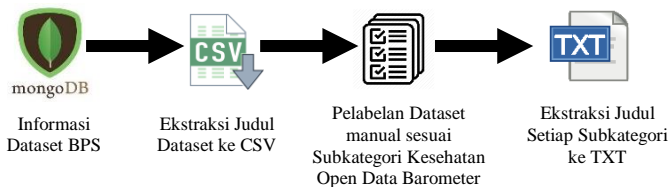
Tabel 4.7 Data Metadata Dataset yang disimpan CKAN *Crawler*

<i>Key</i>	<i>Value</i>
id_pemda	“187”
title	“ANGKA KEMATIAN PASIEN”
update	“2017-08-29”
format	“XLSX”
license	“Creative Commons Attribution”
crawlMonth	“2018-05”

4.2 Kategorisasi dan Penilaian *Open Data*

4.2.1 Pelabelan Subkategori terhadap Dataset

Dalam melakukan kategorisasi, dilakukan pemberian label terhadap setiap dataset yang bersumber dari situs BPS. Setiap dataset yang berasal dari situs BPS dikategorikan berdasarkan subkategori pada dataset kategori kesehatan menurut Open Data Barometer. Pengecekan juga dilakukan terhadap contoh yang diberikan pada Open Data Barometer dan memiliki hasil yang tidak jauh berbeda dibandingkan dataset yang berasal dari BPS.



Gambar 4.5 Alur Pelabelan Data

Tahapannya dilakukan dengan cara *export* judul dataset yang berasal dari *Collection* *ckan_package* kedalam bentuk *csv* dengan menggunakan kode program *python*. Setelah judul dataset BPS berhasil dimasukkan kedalam *CSV*, maka penulis sebagai *annotator* tunggal melakukan pelabelan dataset secara manual sesuai dengan subkategori dataset kesehatan pada Open Data Barometer yaitu:

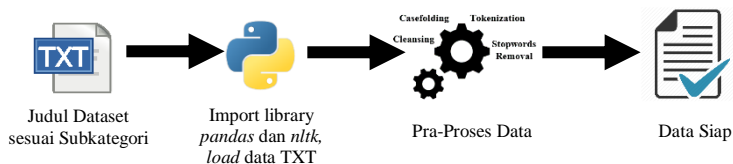
1. *Mortality and Survival Rates*
2. *Levels of Vaccination*
3. *Levels of Access to Health Care*

4. *Health Care Outcomes for Particular Groups*
5. *Patient Satisfaction with Health Services*
6. *Waiting Times for Medical Treatment*
7. *Spend per Admission*

Setelah pelabelan manual yang dilakukan oleh penulis selesai, setiap judul dataset pada setiap subkategori disimpan dalam file yang berbeda sesuai dengan subkategori masing-masing.

4.2.2 Pra-Proses Data

Pada proses ini, data yang telah dikategorisasikan berdasarkan subkategori pada Open Data Barometer diolah untuk menghilangkan tanda baca, karakter tertentu, mengubah menjadi huruf kecil, dan pemecahan judul menjadi per kata dengan memanfaatkan *library* Python seperti *pandas* dan *nltk*. Tahapan yang dilalui meliputi:



Gambar 4.6 Alur Pra-Proses Data

a. Cleansing

Pada tahap ini, judul dataset yang telah dimuat diproses untuk menghilangkan karakter-karakter berupa tanda baca, simbol dan angka yang masih dapat ditemukan pada judul dataset pada saat pengambilan data. Misalkan:

Sebelum <i>Cleansing</i>		Setelah <i>Cleansing</i>	
Jumlah	Fasilitas	Jumlah	Fasilitas
Kesehatan	Kota	Sehat	Kota Bandung
Bandung,	2016		

b. Casefolding

Pada tahap ini, judul dataset yang telah bersih dari simbol diubah ke bentuk huruf kecil seluruhnya untuk menyeragamkan struktur katanya. Misalkan:

Sebelum <i>Casefolding</i>	Setelah <i>Cleansing</i>
Jumlah Fasilitas Sehat Kota Bandung 2016	jumlah fasilitas sehat kota bandung

c. Tokenizing

Pada tahap ini, judul dataset yang telah diseragamkan dipisahkan menjadi setiap kata penyusun judul dataset tersebut. Misalkan:

Sebelum <i>Tokenizing</i>	Setelah <i>Cleansing</i>
Jumlah Fasilitas Sehat Kota Bandung 2016	([jumlah], [fasilitas], [sehat], [kota], [bandung])

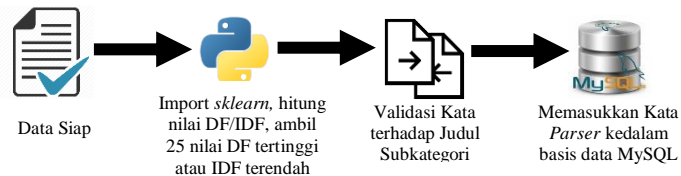
d. Stopwords Removal

Pada tahap ini, judul dataset yang telah dipisah menjadi setiap kata tersebut diidentifikasi untuk menghilangkan kata yang sering muncul pada dokumen namun tidak bermakna penting. Daftar kata yang digunakan merujuk pada penelitian Fadillah Z. Tala[43] dan beberapa tambahan kata oleh penulis yang dinilai dapat menimbulkan bias terhadap penilaian IDF dokumen pada studi kasus judul dataset ini, daftar kata tambahan penulis dapat dilihat pada Lampiran B.

Sebelum <i>Stopwords Removal</i>	Setelah <i>Stopwords Removal</i>
([jumlah], [fasilitas], [sehat], [kota], [bandung])	([jumlah], [fasilitas], [sehat], [bandung])

4.2.3 Penentuan Kata *Parser*

Penentuan kata *parser* merupakan proses menentukan kata yang digunakan sebagai *filter* untuk memastikan bahwa dataset yang diambil merupakan dataset kategori kesehatan. Tahap ini dilakukan dengan melakukan percobaan penghitungan DF (*Document Frequency*) maupun IDF (*Inverse Document Frequency*) terhadap hasil pra-proses data.



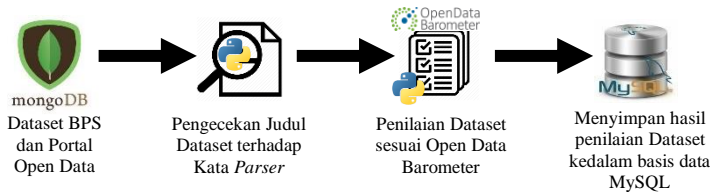
Gambar 4.7 Alur Penentuan Kata *Parser*

Data judul yang telah siap untuk setiap subkategori dihitung nilai DF/IDF-nya menggunakan *library sklearn*. Setelah mendapatkan nilai DF dan IDF untuk setiap kata yang terdapat pada seluruh dokumen yang dimiliki, dokumen diurutkan berdasarkan nilai yang didapatkan oleh setiap kata. DF diurutkan dari yang terbesar sebagaimana nilai DF menunjukkan jumlah frekuensi dokumen yang mengandung kata tersebut. IDF diurutkan dari yang terkecil sebagaimana nilai IDF yang semakin kecil menandakan bahwa semakin banyak dokumen yang mengandung kata tersebut. Setelah diurutkan, maka diambil 25 kata urutan teratas sebagai pertimbangan penentuan kata *parser*.

Metodologi penentuan kata *parser* yang digunakan dalam tugas akhir ini dilakukan oleh penulis dengan mencari kata dari urutan teratas yang relevan terhadap judul subkategori berdasarkan tahap pelabelan yang telah dilakukan oleh penulis. Penulis dapat memilih lebih dari satu kata *parser* untuk setiap kategorinya sesuai dengan relevansi kata terhadap judul dataset. Setelah beberapa kata yang sesuai dipilih, maka *parser* tersebut dimasukkan kedalam basis data MySQL.

4.2.4 Penilaian terhadap setiap Dataset

Penilaian terhadap setiap dataset merupakan proses menentukan jawaban dari setiap pertanyaan yang diajukan pada *Open Data Barometer* terhadap setiap dataset yang telah diambil pada proses akuisisi data. Tahap ini dilakukan dengan memberikan nilai 1 atau 0 terhadap setiap pertanyaan penilaian *open data*.



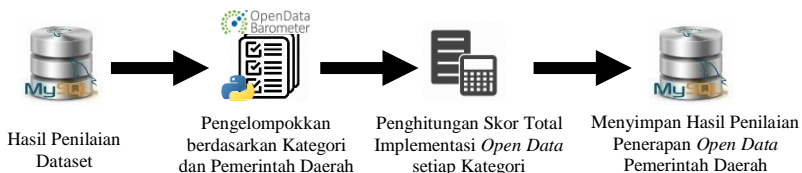
Gambar 4.8 Alur Penilaian Dataset

Informasi terkait dataset yang telah dikumpulkan pada basis data MongoDB yang berasal dari situs BPS dan situs Portal Open Data diambil. Setiap dataset dilakukan pengecekan judul terhadap kata *parser* yang telah ditentukan, apabila judul dataset tidak mengandung kata *parser* yang telah ditentukan maka tidak dilakukan penilaian terhadap dataset tersebut, sedangkan judul dataset yang mengandung kata *parser* dikategorikan sesuai dengan kata *parser* yang ditemukan pada judul dataset tersebut. Dataset yang mengandung kata *parser* tersebut selanjutnya dinilai dengan menjawab pertanyaan sesuai dengan informasi dataset yang diambil. Hasil penilaian untuk setiap dataset kemudian disimpan kedalam basis data MySQL.

4.2.5 Penilaian Open Data setiap Pemerintah Daerah

Penilaian Open Data setiap Pemerintah Daerah merupakan proses untuk memberikan nilai kepada setiap Pemerintah Daerah terhadap dataset yang dimilikinya. Nilai dari setiap dataset yang telah dihitung pada proses sebelumnya dikelompokkan berdasarkan Pemerintah Daerah pemilik dataset. Nilai terbaik dari gabungan seluruh dataset yang dimiliki oleh Pemerintah Daerah dari setiap kategori dataset menjadi nilai akhir *Open Data* setiap Pemerintah Daerah (pada penelitian ini kategori yang dilakukan penilaian hanya dataset

kategori kesehatan). Nilai total untuk setiap kategori dihitung menggunakan logika kondisi sesuai dengan penjelasan pada Tabel 3.2. Hasil penilaian *open data* setiap Pemerintah Daerah ini disimpan kedalam basis data MySQL.



Gambar 4.9 Alur Penilaian Penerapan *Open Data* Pemerintah Daerah

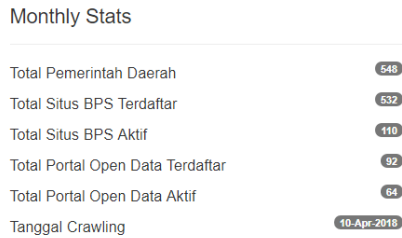
4.3 Visualisasi Aplikasi

4.3.1 Pembuatan Aplikasi Visualisasi

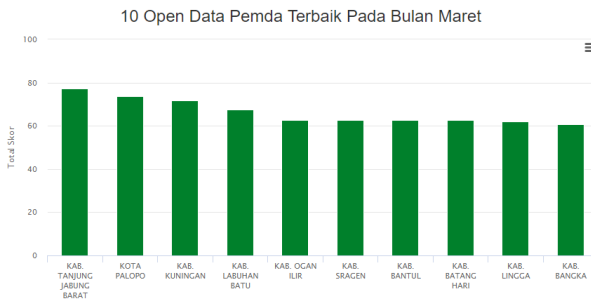
Pembuatan aplikasi visualisasi merupakan proses membuat aplikasi yang berfungsi untuk menampilkan hasil penilaian *open data* pemerintah daerah yang telah dilakukan kepada pengguna. Aplikasi ini dibuat dengan menggunakan bahasa pemrograman PHP versi 7 dengan memanfaatkan *framework* berbasis *Model – View – Controller* (MVC) yaitu Laravel. Pada bagian Model, dibuat kode untuk mengakses basis data MySQL yang telah menyimpan informasi terkait pemerintah daerah dan penilaian terhadap *Open Data*. Pada bagian Controller, dibuat kode untuk mengambil data dari basis data sesuai dengan yang dibutuhkan dan membuat grafik yang ingin divisualisasikan dengan data yang telah diambil. Grafik dibuat memanfaatkan *Highcharts Javascript* yang bisa menampilkan berbagai jenis grafik seperti *Bar Chart*, *Pie Chart*, dan *Line Chart*. Pada bagian View, dibuat kode untuk membuat tampilan halaman situs yang digunakan untuk menampilkan grafik dan tabel hasil penilaian *open data*. Halaman yang dibuat seperti halaman utama penilaian *open data*, halaman daftar pemerintah daerah yang berisi nilai penerapan *open data*, dan halaman detail untuk menunjukkan perkembangan nilai penerapan *open data* pemerintah daerah setiap bulannya.

4.3.2 Desain Antarmuka

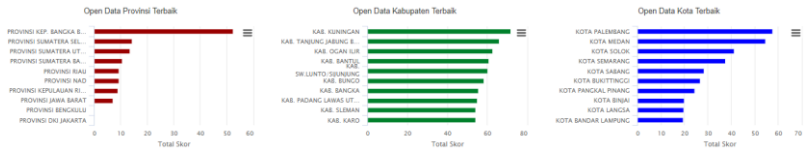
Desain antarmuka merupakan rancangan tampilan aplikasi yang dapat diakses oleh pengguna. Aplikasi menampilkan grafik-grafik terkait penerapan *open data* pemerintah daerah, dan tabel berisi informasi terkait pemerintah daerah. Pengguna dapat mengakses situs menggunakan peramban untuk menampilkan grafik dan tabel. Desain antarmuka yang dibuat mengacu pada tampilan antarmuka yang telah diimplementasikan pada situs <http://egovbench.addi.is.its.ac.id> sebelumnya. Berikut beberapa desain antarmuka untuk grafik dan tabel yang ditampilkan pada situs:



Gambar 4.10 Rancangan Visualisasi Statistik Bulanan



Gambar 4.11 Rancangan Visualisasi Nilai *Open Data* Terbaik



Gambar 4.12 Rancangan Visualisasi Nilai *Open Data* menurut Jenis Pemerintah Daerah

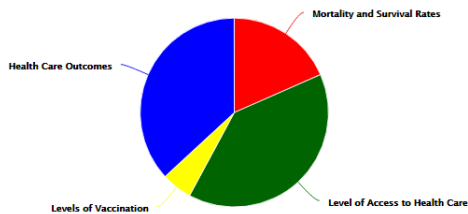
Search:

#	Nama Pemda	Link BPS	Link Portal Open Data	Jumlah Dataset	Total Skor	Detail
1	PROVINSI NAD	Link BPS	Link CKAN	102	40.00	Click Here
2	KAB. ACEH SELATAN	Link BPS	Link CKAN	142	50.00	Click Here
3	KAB. ACEH TENGGARA	Link BPS	Tidak Ada	12	30.00	Click Here
4	KAB. ACEH TIMUR	Link BPS	Link CKAN	12	55.00	Click Here
5	KAB. ACEH TENGAH	Link BPS	Link CKAN	70	55.00	Click Here
6	KAB. ACEH BARAT	Link BPS	Tidak Ada	0	0.00	Click Here
7	KAB. ACEH BESAR	Link BPS	Link CKAN	3	45.00	Click Here
8	KAB. PIDIE	Link BPS	Tidak Ada	47	70.00	Click Here
9	KAB. ACEH UTARA	Link BPS	Link CKAN	13	60.00	Click Here
10	KAB. SIMEULUE	Link BPS	Tidak Ada	25	60.00	Click Here

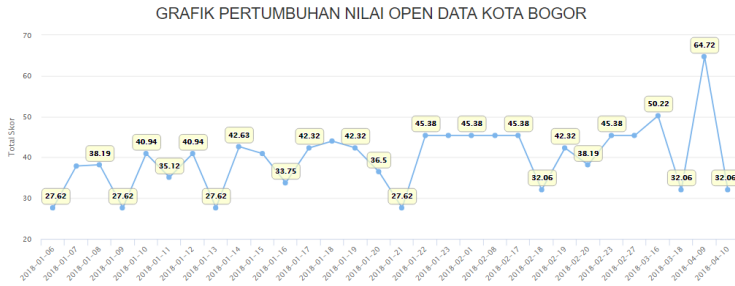
Previous | 1 | 2 | 3 | 4 | 5 | ... | 53 | Next

Gambar 4.13 Rancangan Visualisasi Informasi terkait Pemerintah Daerah

Jumlah Dataset berdasarkan Subkategori pada Kategori Kesehatan



Gambar 4.14 Rancangan Visualisasi berdasarkan Subkategori Dataset



Gambar 4.15 Rancangan Visualisasi Perkembangan Nilai *Open Data*

Pertanyaan	Map Data	Land Ownership	National Statistics	Government Budget	Government Spend	Company Register	Legislation	Public Transport	International Trade	Health Performance	Education Performance	Crime Statistics	Environment Data
Data Existence	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Online Availability	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Machine Readable	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Bulk Availability	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Free of Charge	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Openly Licensed	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Up to Date	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Update Frequency	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Information Clarity	✓	✗	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✗
Linked Data	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Gambar 4.16 Rancangan Visualisasi Detail Nilai *Open Data* Pemerintah Daerah

BAB V IMPLEMENTASI

Pada bab ini dijelaskan hasil dari implementasi perancangan yang telah dijelaskan pada bab perancangan. Penjelasan dalam bagian implementasi meliputi penjelasan lingkungan implementasi, implementasi kode dalam setiap tahapan, dan hasil pembuatan visualisasi.

5.1 Lingkungan Implementasi

Pada bagian ini, dipaparkan spesifikasi perangkat yang digunakan dalam pembuatan dan implementasi penelitian. Perangkat yang dimaksud meliputi perangkat keras dan perangkat lunak pengembangan aplikasi. Berikut merupakan daftar perangkat yang digunakan dalam penelitian ini:

Tabel 5.1 Spesifikasi Perangkat Lunak

Hardware	Spesifikasi
Jenis	Toshiba Satellite C40-A
Processor	Intel Core i5-4200M
Memory	DDR3L 8GB
Storage	- 128GB SSD - 1TB HDD

Tabel 5.2 Perangkat Lunak Pengembangan Aplikasi

Fungsi	Software / Teknologi
Sistem Operasi	Windows 10 Education 64-Bit
Webserver	PHP Build-in Webserver
Bahasa Pemrograman	- PHP 7.1.8 x86 - Python 3.6.4 x64

IDE	- Spyder 3.2.8 - Visual Studio Code 1.24.1
Text Editor	Notepad++ 7.5.3
Relational DBMS	MariaDB 10.1.26
Non Relational DBMS	MongoDB 3.6.3
Data Labeling	Microsoft Excel 365
Web Browser	- Google Chrome version 67 - Firefox version 60 - Microsoft Edge version 42
Python Library	- mysql.connector 8.0.11 - pymongo 3.4.0 - beautifulsoup4 (bs4) 4.6.0 - pandas 0.22.0 - sklearn 0.19.1 - nltk 3.2.5

5.2 Implementasi Akuisisi Data

5.2.1 Pengumpulan Daftar Situs

Berdasarkan tahapan yang telah dilakukan untuk mengumpulkan daftar situs BPS dan *open data* milik pemerintah daerah Indonesia secara manual dengan menggunakan mesin pencari Google dengan kata kunci yang ditetapkan, pada tanggal 28 Maret 2018 didapatkan total daftar situs sebagai berikut:

Tabel 5.3 Jumlah Daftar Situs yang Terkumpul

No.	Situs	Jumlah
1	BPS	548 Situs
2	Open Data berbasis CKAN	15 Situs

Jumlah pemerintah daerah di Indonesia saat ini adalah 548 pemerintah daerah sehingga menurut tahap ini, seluruh pemerintah daerah memiliki situs BPS yang dapat diakses, sedangkan hanya 15 pemerintah daerah yang memiliki situs Portal Open Data berbasis CKAN.

5.2.2 Penyimpanan Data

Dalam mempersiapkan penilaian terhadap *open data* yang dilakukan, dibuatlah penyimpanan data sesuai rancangan yang telah dibuat. Basis data yang digunakan dalam melakukan penyimpanan dibedakan menjadi 2 yaitu MongoDB dan MySQL. MongoDB berfokus pada penyimpanan hasil akuisisi data, sedangkan MySQL berfokus pada penyimpanan informasi milik pemerintah daerah, kategori *open data*, daftar kata *parser*, dan penilaian *open data* terhadap hasil akuisisi data.

a. Pembuatan Basis Data MongoDB

Pembuatan basis data MongoDB, dilakukan dengan menentukan nama basis data dan *Collection* dari basis data tersebut. *Command* dijalankan pada Mongo Shell untuk membuat basis data dan *Collection* sesuai rancangan yang dibuat.

b. Pembuatan Basis Data MySQL

Pembuatan basis data MySQL tidak dibuat dari awal melainkan pengembangan dari basis data Egovbench yang telah ada sebelumnya, sehingga beberapa hal yang dilakukan adalah *update* terhadap atribut pada tabel yang sudah ada dalam basis data dan membuat tabel beserta atribut yang belum ada dalam basis data.

5.2.3 Implementasi Pembuatan *Crawler*

Implementasi pembuatan *crawler* merupakan tahapan pembuatan kode berbasis python untuk mengambil data yang berasal dari situs BPS dan situs portal *open data* yang dimiliki oleh pemerintah daerah. *Input* yang dibutuhkan dalam menjalankan kode ini adalah alamat situs BPS dan alamat situs portal *open data* yang tersimpan pada basis data MySQL. *Output* yang diharapkan adalah hasil pengambilan informasi terkait dataset yang dapat disimpan ke dalam basis data MongoDB.

a. *Web-Scraper* situs BPS

Dalam membuat *web-scraper* untuk mendapatkan informasi dataset yang dimiliki oleh situs BPS, dilakukan request akses ke halaman BPS setiap pemerintah daerah dengan menambahkan “/subject/30/kesehatan.html#subjekViewTab3” pada URL yang telah disimpan dibasis data untuk menampilkan tabel judul dataset dengan kategori kesehatan. Atribut *format* dan *license* yang disimpan pada proses ini independen terhadap proses *scraping* karena telah diasumsikan diawal berdasarkan kondisi situs BPS saat ini. Atribut yang didapatkan pada proses ini meliputi *id_pemda*, *title*, *update*, dan *crawlMonth*. Berikut *pseudocode* yang dijalankan:

```

GET link bps pemda dari basis data MySQL
FOR setiap link bps:
    DO request halaman bps, scrap tabel bps
    FOR setiap baris pada tabel bps
        DO ambil judul dataset, cleanse judul dataset,
        Buat key-value dari judul dataset, masukkan
        hasil akuisisi data kedalam basis data MongoDB
    ENDFOR
ENDFOR

```

Kode 5.1 Pseudocode Web-Scraper Situs BPS

Contoh hasil dokumen yang disimpan dalam basis data MongoDB *collection* *bps_kesehatan*:

Tabel 5.4 Contoh Informasi Hasil *Scraping* Situs BPS

<i>Key</i>	<i>Value</i>
<i>_id</i>	ObjectId("5aea1ef95f5236be5331f04c")
<i>id_pemda</i>	1
<i>title</i>	Jumlah Fasilitas Kesehatan Tahun 1989 2014
<i>update</i>	2014
<i>format</i>	xls
<i>license</i>	
<i>crawlMonth</i>	2018-05

b. CKAN *Crawler*

Pembuatan CKAN *Crawler* dibagi menjadi 2 tahap dengan melakukan pemanggilan fungsi *Action* API milik CKAN untuk setiap tahapnya. Kedua tahap melakukan request API menggunakan alamat situs portal open data dengan tambahan fungsi API. Pada tahap pertama, fungsi API yang digunakan adalah “Get Package List” yang dapat diakses dengan menambahkan “api/action/package_list” pada alamat portal open data milik pemerintah daerah. Hasil yang didapatkan berupa format JSON yang berisi daftar *package/dataset* yang dimiliki oleh portal open data tersebut. Berikut *pseudocode* yang dijalankan:

```

GET link ckan pemda dari basis data MySQL
FOR setiap link ckan pemda:
    DO request Package List API CKAN
    IF request sukses AND response berupa JSON, THEN
        READ JSON, buat key-value dari atribut JSON
        masukkan hasil akuisisi data kedalam basis data
        MongoDB
    ENDFIF
ENDFOR

```

Kode 5.2 Pseudocode Package List Crawler CKAN

Contoh hasil dokumen tahap pertama proses *crawling* yang disimpan dalam basis data MongoDB *collection* ckan_pemda:

Tabel 5.5 Contoh Informasi Hasil *Crawling* Daftar Dataset CKAN

<i>Key</i>	<i>Value</i>
_id	ObjectId("5aea5c7a5f5236be53321d7c")
id_pemda	187
package_id	[0] a-penetapan-kinerja-tujuan-pembangunan-daerah [1] administrasi-dan-manajemen-umum [2] administrasi-legislasi [3] alamafacebook

	[4] analisa-alternatif-program-kegiatan-kelurahan-katulampa [5] anggaran-kesehatan-kabupaten-kota [6] anggaran-pendapatan-dan-belanja-daerah-tahun-anggaran-2017-dengan-rahmat-tuhan-yang-maha-esa [7] angka-kematian-pasien [8] angka-kesembuhan-dan-pengobatan-lengkap-tb-paru-bta-serta-keberhasilan-pengobatan [9] aplikasi-atau-sistem-informasi-prioritas-1
crawlMonth	2018-05

Pada tahap kedua, fungsi API yang digunakan adalah “Get Package Show” yang dapat diakses dengan menambahkan “api/action/package_show?id=<package_id>” pada alamat portal open data milik pemerintah daerah. Hasil yang didapatkan berupa format JSON yang berisi metadata dari dataset yang dimiliki oleh portal open data tersebut. Berikut *pseudocode* yang dijalankan:

```

GET informasi daftar package pemda yang diambil pada
bulan ini dari basis data MongoDB
FOR setiap informasi daftar package milik pemda:
  GET daftar id package
  FOR setiap id package:
    DO request Package Show API CKAN
    IF request sukses AND response berupa JSON,
    THEN READ JSON
      IF ada metadata resource AND not NULL THEN
        Clean metadata, buat key-value dari
        atribut JSON yang dibutuhkan, masukkan
        hasil akuisisi data kedalam basis data
        MongoDB
      ENDIF
    ENDIF
  ENDFOR
ENDFOR

```

Kode 5.3 Pseudocode Informasi Dataset Crawler CKAN

Contoh hasil dokumen tahap kedua proses *crawling* yang disimpan dalam basis data MongoDB *collection* *ckan_package*:

Tabel 5.6 Contoh Informasi Hasil *Crawling* Metadata Dataset CKAN

<i>Key</i>	<i>Value</i>
<i>_id</i>	ObjectId("5b0ae0cf476d335782f36311")
<i>id_pemda</i>	187
<i>title</i>	ANGKA KEMATIAN PASIEN
<i>update</i>	2017-08-29
<i>format</i>	Xlsx
<i>license</i>	creative commons attribution
<i>crawlMonth</i>	2018-05

5.3 Implementasi Kategorisasi dan Penilaian Open Data

5.3.1 Pelabelan Dataset

Dalam melakukan pelabelan dataset, tahapan yang dilakukan meliputi *input* berupa ekstraksi judul dataset BPS dari *Collection* *bps_kesehatan* kedalam bentuk CSV, pelabelan dataset secara manual oleh penulis sebagai *annotator* untuk mencocokkan judul dataset yang dimiliki dengan subkategori yang ada pada dataset kategori kesehatan menurut Open Data Barometer, serta *output* berupa file TXT yang berisi judul dataset yang telah dilabeli.

a. Ekstraksi Judul Dataset ke CSV

Ekstraksi judul dataset ke CSV dilakukan menggunakan kode program python sederhana. Hasil dari ekstraksi file tersebut berisi informasi *id_pemda* pemilik dataset dan judul dataset yang dimilikinya seperti pada tabel berikut:

id_pemda	Judul
1	Persentase Balita Menurut Kabupaten/Kota dan Penolong Persalinan Terakhir 2012
10	Jumlah Penderita Penyakit Menular di Kabupaten Simeulue Tahun 2014
12	Jumlah Kematian Ibu Menurut Puskesmas Dalam Kabupaten Bireuen 2015
13	Jumlah Kelahiran Bayi 2014
24	Angka Harapan Hidup Beberapa Kabupaten/Kota Di Aceh Tahun 2013 - 2015

b. Pelabelan manual Dataset sesuai Subkategori Kesehatan

Pelabelan secara manual dataset dilakukan dengan menggunakan aplikasi Microsoft Excel, pada aplikasi ini penulis memberikan label kepada setiap judul dataset satu persatu. Setiap judul dataset hanya dapat menjadi salah satu dari 7 subkategori yang ada. Berikut merupakan hasil pelabelan manual dataset yang dilakukan terhadap judul dataset BPS:

Tabel 5.7 Hasil Pelabelan Dataset

No.	Subkategori	Jumlah Dataset
1	<i>Mortality and Survival Rates</i>	764
2	<i>Levels of Vaccination</i>	167
3	<i>Levels of Access to Health Care</i>	2058
4	<i>Health Care Outcomes for Particular Groups</i>	641
5	<i>Patient Satisfaction with Health Services</i>	0
6	<i>Waiting Times for Medical Treatment</i>	0
7	<i>Spend per Admission</i>	0
8	Tidak Terkategorisasi	61

5.3.2 Pra-Proses Data

Pada proses pra-proses data, *input* yang dibutuhkan merupakan judul dataset yang telah dikategorisasikan sebelumnya dan dipisahkan dalam format TXT. Proses ini dilakukan menggunakan kode program python dengan memanfaatkan library yaitu *pandas* dan *nlk*. Selain library tersebut, daftar stopwords juga digunakan untuk menghapus kata yang tidak memiliki arti signifikan namun sering muncul dari judul dataset. berikut *psoudocode* yang dijalankan:

```

GET daftar stopwords
OPEN katalog judul dataset
READ katalog judul dataset tiap baris
GET daftar judul dataset dari katalog judul dataset
FOR setiap judul dataset
    DO cleanse, casefold, tokenize, remove stopwords
    terhadap judul dataset, gabungkan kembali kata
    hasil pra-proses menjadi judul dataset
ENDFOR

```

Kode 5.4 Pseudocode Pra-Proses Data

Dari tahapan pra-proses data tersebut, berikut merupakan contoh hasil pra-proses data yang dilakukan:

Tabel 5.8 Contoh Hasil Pra-Proses Data

Judul Sebelum Pra-Proses	Judul Setelah Pra-Proses
Persentase Balita Menurut Kabupaten/Kota dan Penolong Persalinan Terakhir 2012	['persentase', 'balita', 'penolong', 'persalinan']
Persentase Penduduk yang Mempunyai Keluhan Kesehatan Selama Sebulan yang lalu Menurut Kabupaten/Kota 2012	['persentase', 'penduduk', 'keluhan', 'kesehatan', 'sebulan']
Jumlah Angka Kematian Ibu Kunjungan Bayi Bayi Lahir Hidup dan Balita Dirinci Menurut Kecamatan di	['angka', 'kematian', 'kunjungan', 'bayi', 'bayi', 'lahir', 'hidup', 'balita', 'dirinci', 'simeulue']

Kabupaten Simeulue Tahun 2014	
Jumlah Kasus HIV/Aids IMS DBD Diare TB Dan Malaria Menurut Kecamatan Di Kabupaten Bireuen 2015	['hiv', 'aids', 'ims', 'dbd', 'diare', 'tb', 'malaria', 'bireuen']
Perkiraan Angka Harapan Hidup menurut Kabupaten/Kota (tahun) 2012 – 2014	['perkiraan', 'angka', 'harapan', 'hidup']

5.3.3 Teknik Penentuan Kata *Parser*

Dalam menentukan kata *parser*, judul yang telah selesai tahap pra-proses tersebut dimasukkan kedalam list sesuai dengan subkategori masing-masing dan dilakukan penghitungan nilai DF/IDF terhadap list tersebut. Proses ini memanfaatkan *library* sklearn milik python yang berfungsi sebagai penghitung nilai DF/IDF. Setiap subkategori diambil 25 kata dengan nilai DF tertinggi atau IDF terendah sebagai pertimbangan penentuan kata *parser* yang digunakan untuk kategorisasi dataset. Hasil 25 kata yang didapatkan tersebut divalidasi kecocokannya dengan nama subkategori oleh penulis dan ditetapkan sebagai kata *parser* untuk subkategori tersebut. Berikut *pseudocode* yang dijalankan:

```

GET daftar judul hasil pra-proses
CALCULATE nilai DF/IDF untuk setiap kata dalam daftar
judul
SORT kata berdasarkan nilai DF tertinggi / IDF terendah
GET 25 kata teratas hasil pengurutan

```

Kode 5.5 *Pseudocode* penghitungan DF/IDF kata pada judul dataset

Hasil 25 kata DF/IDF untuk setiap subkategori dapat dilihat pada LAMPIRAN C – Daftar 25 kata DF/IDF Terendah setiap Subkategori yang menunjukkan bahwa kata yang berasal dari penghitungan nilai DF dan IDF memiliki hasil yang sama, terdapat beberapa perbedaan urutan yang terjadi akibat nilai DF atau IDF yang sama terhadap kata yang tertukar posisinya.

Penulis selanjutnya menggunakan IDF dikarenakan nilai dari IDF telah memiliki pembobotan. Setelah dilakukan pemilihan kata *parser* oleh penulis berdasarkan judul subkategori dan dataset yang telah diberikan label, berikut merupakan setiap kata *parser* yang digunakan sebagai *filter* untuk menentukan dataset dengan kategori kesehatan:

Tabel 5.9 Daftar Kata *Parser* yang Terpilih

Subkategori	Kata Parser
<i>Mortality and Survival Rates</i>	Bayi
	Balita
	Kelahiran
	Kematian
<i>Levels of Vaccination</i>	Imunisasi
	Vaksinasi
<i>Levels of Access to Health Care</i>	Kesehatan
	Puskesmas
	Medis
	Posyandu
	Klinik
<i>Health Care Outcomes for Particular Groups</i>	Penyakit
	Penderita
	Rawat
	Sakit
	Pasien
	Berobat

5.3.4 Pembuatan Kode Penilaian Dataset

Pembuatan kode untuk penelitian dataset dilakukan menggunakan kode program python. *Input* yang dibutuhkan dalam menjalankan program ini adalah hasil akuisisi data yang disimpan pada basis data mongoDB dan kata *parser* pada basis data MySQL yang digunakan untuk menentukan kategori dataset. *Output* dari proses ini adalah subkategori untuk setiap dataset yang dikumpulkan dan nilai *open data* yang dimilikinya.

Berikut *pseudocode* yang dijalankan untuk kedua sumber dataset yaitu BPS dan portal open data:

```

GET daftar dataset hasil akuisisi data bulan ini dari
basis data MongoDB
GET daftar kata parser dari basis data MySQL
FOR setiap dataset hasil akuisisi data
  GET judul dataset
  FOR setiap kata parser
    IF judul dataset mengandung kata parser THEN
      SET atribut berdasarkan informasi dataset,
      COMPARE informasi dataset dengan kondisi
      penilaian berdasarkan Open Data Barometer,
      SET nilai setiap pertanyaan dari dataset,
      Masukkan hasil penilaian dataset kedalam
      basis data MySQL
    ENDIF
  ENDFOR
ENDFOR

```

Kode 5.6 Pseudocode Penilaian Dataset

5.3.5 Pembuatan Kode Penilaian Penerapan *Open Data* Pemerintah Daerah

Pembuatan kode untuk penelitian dataset dilakukan menggunakan kode program python. *Input* yang dibutuhkan dalam menjalankan program ini adalah hasil nilai setiap dataset yang telah didapatkan pada proses sebelumnya. Nilai dari setiap dataset tersebut dikelompokkan berdasarkan pemerintah daerah pemilik dataset dan kategori dari dataset tersebut. Kombinasi nilai terbaik setiap dataset dalam satu kategori milik suatu pemerintah daerah menjadi *output* berupa nilai *open data* akhir untuk pemerintah daerah pada kategori tersebut. Berikut *pseudocode* yang dijalankan untuk mendapatkan nilai penerapan *open data* pemerintah daerah:

```

GET id pemda dari basis data MySQL
GET daftar kategori open data dari basis data MySQL
FOR setiap id pemda
  FOR setiap kategori open data
    GET nilai dataset dari pemda dan kategori open
    data yang dinilai pada bulan ini dari basis data
    MYSQL
    IF terdapat nilai dataset THEN
      Lakukan penilaian open data setiap dataset
      yang didapatkan dari dataset BPS dan CKAN,
      CALCULATE total skor, masukkan hasil
      penilaian open data kedalam basis data MySQL
    ENDFOR
  ENDFOR
ENDFOR

```

Kode 5.7 Pseudocode Penilaian Penerapan Open Data

5.4 Pembuatan *Bash Script Scheduler*

Pembuatan *bash script scheduler* ini digunakan sebagai alat untuk menjalankan seluruh proses penilaian *open data* mulai dari proses pengambilan data sampai dihasilkan nilai akhir penerapan *open data* untuk setiap pemerintah daerah. Hal ini dilakukan sehingga setiap kode python yang telah dibuat dapat dijalankan hanya dengan perintah tunggal. Bash script yang telah dibuat kemudian dijadwalkan untuk dijalankan setiap awal bulan menggunakan fitur Cronjob milik Linux. Berikut merupakan *bash script* yang dibuat dengan nama `opendata.sh`:

```

#!/bin/bash
current=$(dirname $0)
cd $current #Go to current directory

python3 bps_scraper.py #Run BPS Scraping
python3 ckan_package_grabber.py #Run CKAN Package List Grabber
python3 ckan_crawler.py #Run CKAN Dataset Crawler
python3 bps_scoring.py #Run BPS Dataset Scoring
python3 ckan_scoring.py #Run CKAN Dataset Scoring
python3 result_opendata.py #Run Open Data Scoring Pemda

```

Kode 5.8 Bash Script `opendata.sh`

```
@monthly /home/egovbench/opendata_script/opendata.sh
```

Kode 5.9 Crontab untuk menjalankan `opendata.sh`

5.5 Implementasi Pembuatan Aplikasi Visualisasi

Pembuatan aplikasi visualisasi bertujuan untuk menyajikan hasil penilaian *open data* kepada pengguna. Aplikasi visualisasi yang dibuat merupakan aplikasi berbasis web menggunakan *framework laravel*. Informasi yang ditampilkan kepada pengguna merupakan grafik dan tabel yang menunjukkan hal-hal terkait *open data* yang telah dinilai. Grafik dibuat dengan menggunakan *library Highchart* dengan tampilan yang disesuaikan terhadap informasi yang ingin disajikan. Berikut merupakan beberapa kode yang digunakan dalam mengambil data dari basis data MySQL dan hasil visualisasi yang ditampilkan dengan menggunakan data yang diambil.

5.5.1 Statistik Penilaian Bulanan

```

1. date_default_timezone_set("Asia/Jakarta");
2. $pemda = pemdaModel::all()->count();
3. $total_bps_pemda = pemdaModel::where('bps_pemda',
   '!=', "")->count('bps_pemda');
4. $active_bps_pemda = bpsOpendataModel::distinct('id
   _pemda')->whereYear('date', '=', date('Y'))-
   >whereMonth('date', '=', date('m'))-
   >count('id_pemda');
5. $total_scored_bps = bpsOpendataModel::whereYear('d
   ate', '=', date('Y'))-
   >whereMonth('date', '=', date('m'))->count();
6. $total_ckan_pemda = pemdaModel::where('ckan_pemda'
   , '!=', "")->count('ckan_pemda');
7. $active_ckan_pemda = ckanOpendataModel::distinct('
   id_pemda')->whereYear('date', '=', date('Y'))-
   >whereMonth('date', '=', date('m'))-
   >count('id_pemda');
8. $total_scored_ckan = ckanOpenDataModel::whereYear(
   'date', '=', date('Y'))-
   >whereMonth('date', '=', date('m'))->count();
9. $last_result_date = resultOpendataModel::select('d
   ate')->orderByDesc('date')->first()

```

Kode 5.10 Query Statistik Bulanan Open Data

Input dalam membuat tampilan statistik bulanan terkait penilaian *open data* meliputi jumlah situs terdaftar, situs aktif, jumlah dataset yang dinilai, dan tanggal penilaian terakhir dilakukan. Berikut merupakan hasil visualisasi yang ditampilkan:

Monthly Stats	
This Month	June 2018
Total Pemerintah Daerah	548
Total BPS Pemda	548
Total BPS Aktif	393
Total Dataset BPS	2989
Total Portal Open Data Pemda	15
Total Portal Open Data Aktif	9
Total Dataset Portal Open Data	2358
Tanggal Penilaian Terakhir	2018-06-10

Gambar 5.1 Visualisasi Statistik Bulanan Open Data

5.5.2 Nilai Open Data Terbaik Pemerintah Daerah

```

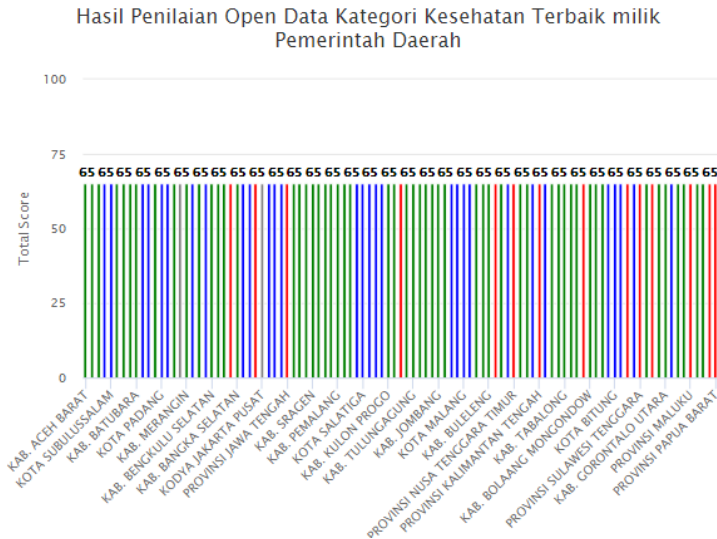
1. $highestResultOpendata = resultOpendataModel::select('totalScore')-
  >where('id_kategori', '=', "D11")-
  >whereYear('date', '=', date('Y'))-
  >whereMonth('date', '=', date('m'))-
  >orderByDesc('totalScore')->first();
2. $top10resultOpendata = resultOpendataModel::join('pemda', 'result_opendata.id_pemda', '=', 'pemda.id_pemda')-
  >select('pemda.nama_pemda', 'pemda.tipe', 'pemda.bps_pemda', 'pemda.ckan_pemda', 'result_opendata.id_pemda', 'result_opendata.id_kategori', 'result_opendata.totalScore')-
  >where('id_kategori', '=', "D11")-
  >where('totalScore', '=', $highestResultOpendata['totalScore'])->whereYear('date', '=', date('Y'))-
  >whereMonth('date', '=', date('m'))-

```

```
>orderByDesc('totalScore')->orderBy('id_pemda')-
>get();
```

Kode 5.11 Query Nilai Open Data Terbaik

Input dalam membuat grafik nilai open data terbaik meliputi nilai open data tertinggi dan pemerintah daerah yang mendapatkan nilai tertinggi tersebut. Berikut merupakan hasil visualisasi yang ditampilkan:



Gambar 5.2 Visualisasi Pemerintah Daerah dengan Nilai Open Data Terbaik

5.5.3 Informasi Terkait Open Data Pemerintah Daerah

```
1. $listPemda = pemdaModel::all()-
  >sortBy('id_pemda');
2. foreach ($listPemda as $lp) {
3.     $idPemda = $lp['id_pemda'];
4.     $lp['bps_pemda'] .= "/subject/30/kesehatan.htm
  l#subjekViewTab3";
5.     $lp['bps_dataset'] = bpsOpendataModel::where('
  id_pemda', '=', $idPemda)-
```

```

>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
6.     $lp['ckan_dataset'] = ckanOpendataModel::where
('id_pemda', '=', $idPemda)-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
7.     $resultScore = resultOpendataModel::select('to
talScore')->where('id_pemda', '=', $idPemda)-
>where('id_kategori', '=', 'D11')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))-
>orderByDesc('date')->first();
8.     $lp['totalScore'] = $resultScore['totalScore']
;
9.     }

```

Kode 5.12 Query Informasi Open Data Pemerintah Daerah

Input dalam membuat tabel informasi pemerintah daerah berupa data pemerintah daerah, alamat situs, dan jumlah dataset yang dimiliki dari BPS dan portal open data setiap pemerintah daerah. Visualisasi ini bertujuan untuk memperlihatkan jumlah dataset dan nilai *open data* yang didapatkan oleh setiap pemerintah daerah. Berikut merupakan hasil visualisasi yang ditampilkan:

🏠 Opendata

Show 10 - entries Search:

ID	Nama Pemerintah Daerah	Dataset BPS	Situs BPS	Dataset CKAN	Situs Portal Data	Total Skor	Detail
1	PROVINSI NAD	12	Link BPS	0	Link Portal CKAN	45	Click Here
2	KAB. ACEH SELATAN	0	Link BPS	0	No Link	0	Click Here
3	KAB. ACEH TENGGARA	0	Link BPS	0	No Link	0	Click Here
4	KAB. ACEH TIMUR	1	Link BPS	0	No Link	45	Click Here
5	KAB. ACEH TENGAH	6	Link BPS	0	No Link	45	Click Here
6	KAB. ACEH BARAT	12	Link BPS	0	No Link	65	Click Here
7	KAB. ACEH BESAR	0	Link BPS	0	No Link	0	Click Here
8	KAB. PIDIE	1	Link BPS	0	No Link	65	Click Here
9	KAB. ACEH UTARA	0	Link BPS	0	No Link	0	Click Here
10	KAB. SIMEULUE	10	Link BPS	0	No Link	45	Click Here

Showing 1 to 10 of 548 entries Previous 1 2 3 4 5 ... 55 Next

Gambar 5.3 Visualisasi Informasi Open Data Pemerintah Daerah

5.5.4 Jumlah Dataset berdasarkan Subkategori

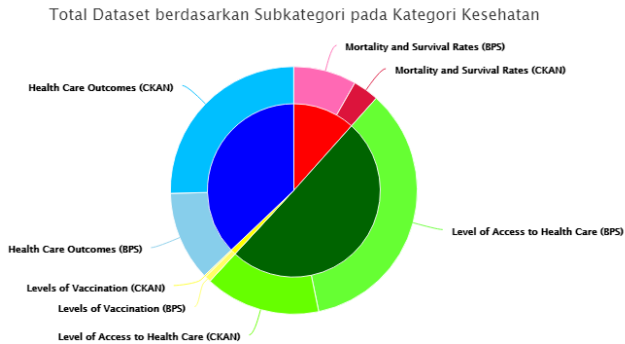
```

1. $countLifeBPS = bpsOpendataModel::where('subkatego
ri', '=', 'Mortality and Survival Rates')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
2. $countAccessBPS = bpsOpendataModel::where('subkate
gori', '=', 'Level of Access to Health Care')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
3. $countVaccineBPS = bpsOpendataModel::where('subkat
egori', '=', 'Levels of Vaccination')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
4. $countOutcomesBPS = bpsOpendataModel::where('subka
tegori', '=', 'Health Care Outcomes')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
5. $countLifeCKAN = ckanOpendataModel::where('subkate
gori', '=', 'Mortality and Survival Rates')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
6. $countAccessCKAN = ckanOpendataModel::where('subka
tegori', '=', 'Level of Access to Health Care')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
7. $countVaccineCKAN = ckanOpendataModel::where('subk
ategori', '=', 'Levels of Vaccination')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();
8. $countOutcomesCKAN = ckanOpendataModel::where('sub
kategori', '=', 'Health Care Outcomes')-
>whereYear('date', '=', date('Y'))-
>whereMonth('date', '=', date('m'))->count();

```

Kode 5.13 Query Jumlah Dataset tiap Subkategori

Input dalam membuat visualisasi jumlah dataset berdasarkan subkategori merupakan hasil penilaian terhadap dataset yang berasal dari BPS dan portal open data. Visualisasi ini bertujuan untuk memperlihatkan banyaknya dataset yang dimiliki seluruh pemerintah daerah berdasarkan subkategori. Berikut merupakan hasil visualisasi yang ditampilkan:



Gambar 5.4 Visualisasi Jumlah Dataset berdasarkan Subkategori

5.5.5 Tren Nilai *Open Data* suatu Pemerintah Daerah

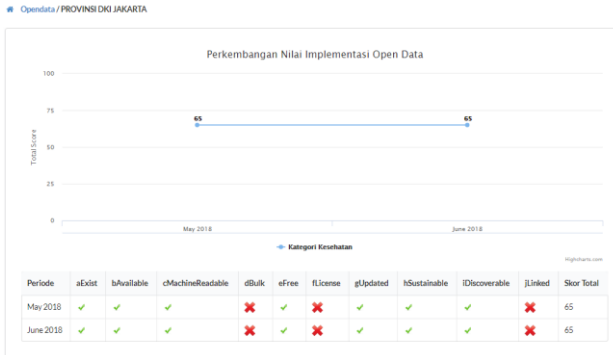
```

1. $pemda = pemdaModel::where('id_pemda', '=', $id)->first();
2. $resultOpendata = resultOpendataModel::where('id_pemda', $id)->where('id_kategori', 'D11')->orderByDesc('date')->take(12)->get()->all();

```

Kode 5.14 *Query Mengambil Nilai Open Data Pemerintah Daerah*

Input dalam membuat visualisasi tren nilai *open data* pemerintah daerah berupa nilai akhir penerapan *open data* yang dimiliki pemerintah daerah. Visualisasi ini bertujuan untuk memperlihatkan tren nilai *open data* yang didapatkan suatu pemerintah daerah. Berikut merupakan hasil visualisasi yang ditampilkan:



Gambar 5.5 Visualisasi Tren Nilai *Open Data* Pemerintah Daerah

5.5.6 Pengelolaan Kata *Parser*

```
1. $parsers = parserOpenDataModel::all();
2. $kategoris = kategoriOpenDataModel::all();
```

Kode 5.15 *Query* Pengambilan Kata *Parser*

Input dalam membuat halaman pengelolaan kata *parser open data* pemerintah daerah berupa daftar kata *parser* yang telah ditentukan dan subkategori sesuai dengan *Open Data Barometer*. Halaman ini bertujuan untuk mengelola kata *parser* dengan fitur dapat menambahkan, mengubah, dan menghapus kata *parser* yang digunakan dalam melakukan kategorisasi dataset. Berikut merupakan hasil visualisasi yang ditampilkan:

Admin / Parser Open Data + Tambah Parser Open Data

Show 10 entries Search:

ID Kategori	Subkategori	Kata Parser	Action
D11	Level of Access to Health Care	kesehatan	🔍 ✖
D11	Health Care Outcomes	penyakit	🔍 ✖
D11	Mortality and Survival Rates	kematian	🔍 ✖
D11	Mortality and Survival Rates	kelahiran	🔍 ✖
D11	Levels of Vaccination	imunisasi	🔍 ✖
D11	Mortality and Survival Rates	bayi	🔍 ✖
D11	Mortality and Survival Rates	balita	🔍 ✖

Gambar 5.6 Halaman Pengelolaan Kata *Parser*

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil dan pembahasan berupa analisis dan pengujian aplikasi terhadap implementasi yang telah dilakukan. Penjelasan pada bab ini meliputi analisis *parser*, analisis penilaian open data, pengujian fungsional, dan pengujian non fungsional.

6.1 Analisis Parser

Analisis *parser* dilakukan untuk mengetahui ketepatan kata *parser* yang terpilih terhadap judul dataset yang dikategorisasikan. Dalam membuat kata *parser* ini, *input* yang digunakan berupa data dari situs BPS yang berhasil diakuisisi pada tanggal 3 Mei 2018. Berikut merupakan langkah-langkah yang dilakukan:

6.1.1 Memuat data

Pada proses pengambilan data yang dilakukan pada bulan Mei, didapatkan total data berjumlah 3691 judul dataset yang berasal dari 548 situs BPS milik setiap pemerintah daerah di Indonesia. Data yang telah diambil tersebut diberi label sesuai dengan subkategori pada Open Data Barometer. Berikut merupakan detail pelabelan data manual yang dilakukan penulis:

Tabel 6.1 Distribusi Label berdasarkan Subkategori

No.	Subkategori	Jumlah Dataset
1	<i>Mortality and Survival Rates</i>	764
2	<i>Levels of Vaccination</i>	167
3	<i>Levels of Access to Health Care</i>	2058
4	<i>Health Care Outcomes for Particular Groups</i>	641
5	<i>Patient Satisfaction with Health Services</i>	0
6	<i>Waiting Times for Medical Treatment</i>	0
7	<i>Spend per Admission</i>	0
8	Tidak Terkategorisasi	61

6.1.2 Pra-proses Data

a. Penetapan Data pembentuk Parser

Dalam penelitian yang dilakukan, pembentukan *parser* dibuat berdasarkan judul dataset yang dimiliki pemerintah daerah hanya yang berasal dari situs BPS. Hal ini dilakukan karena situs BPS telah mengkategorikan dataset tersebut kedalam kategori kesehatan, sedangkan portal open data tidak melakukan kategori terhadap dataset yang dimiliki. Berdasarkan hasil pengambilan data yang dilakukan, terdapat 407 pemerintah daerah yang memiliki dataset pada kategori kesehatan dari total 548 pemerintah daerah yang memiliki situs BPS. Data akhir yang digunakan dalam proses pembentukan *parser* adalah 3630 dataset dari total dataset yang dilabeli yaitu 3691 dikurangi dengan data yang tidak terkategoriisasi berjumlah 61. Berikut merupakan contoh data yang tidak terkategoriisasi oleh pelabelan manual:

Tabel 6.2 Contoh Dataset Tidak Terkategoriisasi

No.	Judul Dataset
1	Pengeluaran Rata-rata per Kapita Sebulan Menurut Kelompok Makanan di Kota Tanjungpinang 2016
2	Pendapatan Bersih Bulan Dana Palang Merah Indonesia Menurut Kabupaten/Kota 2013-2016
3	Fasilitas Tempat Olah Raga Menurut Desa dan Jenisnya Kecamatan Sumbergempol 2016
4	Institusi Masyarakat menurut Kecamatan dan Jenis Institusi 2009-2016
5	Sampah dan Pengelolaannya di Kabupaten Madiun Tahun 2011-2013
6	Banyaknya Desa Menurut Jenis Tempat Pembuangan Sampah dan Ketersediaan Tempat Penampungan Sampah Sementara di Kecamatan Sugio 2014
7	Banyaknya Desa Menurut Keberadaan Kelompok Kegiatan Olahraga tahun 2014
8	Jumlah Sarana dan Prasarana Olahraga di Kabupaten Lamongan 2015

b. Hasil Data setelah tahap Pra-Proses

Dataset awal sejumlah 3630 yang memiliki jumlah kata sebanyak 44.426 kata berkurang menjadi 21.058 kata setelah pra-proses data dilakukan sesuai dengan tahapan yang dilakukan pada sub bab 4.3.3. Pengurangan jumlah kata dalam dataset ini terjadi pada proses *cleaning* dan *stopwords removal* terhadap dataset guna membuang kata dan karakter yang tidak penting dalam judul dataset. Berikut merupakan rincian jumlah kata yang tersisa pada setiap tahapan pra-proses data:

Tabel 6.3 Jumlah Kata setiap Tahapan Pra-Proses

Tahap	Jumlah Kata
Data Hasil Labeling	44.426 kata
<i>Cleaning</i>	41.084 kata
<i>Casefolding</i>	41.084 kata
<i>Tokenizing</i>	41.084 kata
<i>Stopwords Removal</i>	21.058 kata

Dalam tahapan *Cleaning*, jumlah kata pada seluruh dataset yang digunakan untuk melakukan pembentukan *parser* berkurang sebanyak 3.342 kata dengan melakukan penghapusan terhadap setiap simbol dan angka yang ditemui pada judul dataset. dalam tahapan *casefolding* dan *tokenizing*, tidak terjadi pengurangan kata pada seluruh dataset yang digunakan karena proses yang dilakukan berupa mengubah seluruh karakter pada judul dataset ke bentuk huruf kecil dan membentuk *token* dengan memisahkan setiap kata pada judul dataset. Pada tahap pra-proses terakhir yaitu *stopwords removal*, terjadi penurunan jumlah kata yang signifikan yaitu sebanyak 20.026 kata dari proses sebelumnya karena kata yang dianggap tidak penting atau tidak memiliki arti dalam proses penentuan *parser* dihapus dari judul dataset. berikut merupakan ringkasan hasil pra-proses data:

Tabel 6.4 Ringkasan Hasil Pra-Proses Data

Jumlah	Sebelum Pra-Proses	Setelah Pra-Proses
Judul Dataset	3630 dataset	3630 dataset
Total Kata	44.426 kata	21.058 kata
Kata Unik	1.922 kata	1.147 kata

6.1.3 Pembuatan kata *parser* berdasarkan nilai IDF

Berdasarkan dataset hasil pra-proses, dilakukan penghitungan IDF terhadap seluruh dataset menggunakan *TfidfVectorizer* milik *sklearn*. Pada proses ini, setelah nilai IDF untuk setiap kata didapatkan, maka diambil 25 kata dari setiap subkategori yang memiliki nilai IDF terendah. Nilai IDF rendah menunjukkan bahwa kata tersebut merupakan kata yang sering muncul pada seluruh dokumen yang dimiliki. Daftar 25 kata terendah dapat dilihat pada Lampiran C. Hasil dari 25 kata terendah tersebut tidak seluruhnya menjadi kata *parser*, melainkan divalidasi oleh penulis untuk membandingkan kecocokan kata yang sering muncul terhadap subkategori dari dataset yang tersebut. Kata yang ditetapkan menjadi *parser* dapat dilihat pada Tabel 5.9. Total kata *parser* berjumlah 17 kata dengan rincian sebagai berikut:

Tabel 6.5 Jumlah Kata Parser setiap Subkategori

Subkategori	Jumlah Kata <i>Parser</i>
<i>Mortality and Survival Rates</i>	4 kata
<i>Levels of Vaccination</i>	2 kata
<i>Levels of Access to Health Care</i>	5 kata
<i>Health Care Outcomes for Particular Groups</i>	6 kata

6.1.4 Hasil Kategorisasi Dataset Menggunakan Kata Parser

Menggunakan hasil kata *parser* yang telah dibentuk, dilakukan kategorisasi menggunakan dataset yang sama. Dalam kasus ini, pengecekan hasil kategorisasi dataset dilakukan menggunakan dataset hasil pengambilan data yang berasal dari situs BPS pada 3 Mei 2018. Dataset yang telah sebelumnya diberikan label manual oleh penulis dikategorisasikan dengan menggunakan kata *parser* yang telah dibentuk. Berikut merupakan hasil percobaan kategorisasi yang dilakukan dengan merubah urutan subkategori yang dilakukan pengecekan terlebih dahulu:

Hasil Percobaan Kategorisasi 1

Pada percobaan kategorisasi ini, urutan subkategori yang dilakukan kategorisasi adalah:

1. *Health Care Outcomes for Particular Groups*
2. *Levels of Access to Health Care*
3. *Levels of Vaccination*
4. *Mortality and Survival Rates*

Berikut merupakan hasil *confusion matrix* yang didapatkan:

Tabel 6.6 Confusion Matrix Percobaan Kategorisasi 1

Predicted Actual	<i>Mortality and Survival Rates</i>	<i>Levels of Access to Health Care</i>	<i>Levels of Vaccination</i>	<i>Health Care Outcomes</i>
<i>Mortality and Survival Rates</i>	236	57	0	13
<i>Levels of Access to Health Care</i>	29	1539	0	224
<i>Levels of Vaccination</i>	117	7	38	1
<i>Health Care Outcomes</i>	1	146	0	354

Dari tabel *confusion matrix* tersebut, berikut merupakan tingkat *Precision*, *Recall*, dan *F-Measure* untuk setiap subkategori:

Tabel 6.7 Evaluasi Performa Percobaan Kategorisasi 1

Subkategori	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Mortality and Survival Rates</i>	61,62%	30,89%	41,15%
<i>Levels of Access to Health Care</i>	87,99%	74,78%	80,85%
<i>Levels of Vaccination</i>	100%	22,75%	37,07%
<i>Health Care Outcomes</i>	59,8%	55,23%	57,42%
Persentase / Total	77,35%	45,91%	54,12%

Dari hasil evaluasi performa tersebut, didapatkan hasil *precision* sebesar 77,35% dan *recall* sebesar 45,91%. Dari hasil tersebut dapat dikatakan hasil kategorisasi masih kurang baik. Terlihat pada subkategori *Level of Vaccination*, nilai *Precision* mencapai 100% namun nilai *Recall* hanya 22,75%, hal itu menandakan bahwa meskipun prediksi akurat, namun dataset aktual yang terkategori dengan benar hanya sebagian kecil, seperti yang terlihat pada *confusion matrix*.

Hasil Percobaan Kategorisasi 2

Pada percobaan kategorisasi ini, urutan subkategori yang dilakukan kategorisasi adalah:

1. *Levels of Vaccination*
2. *Mortality and Survival Rates*
3. *Levels of Access to Health Care*
4. *Health Care Outcomes for Particular Groups*

Berikut merupakan hasil *confusion matrix* yang didapatkan:

Tabel 6.8 Confusion Matrix Percobaan Kategorisasi 2

Predicted Actual	<i>Mortality and Survival Rates</i>	<i>Levels of Access to Health Care</i>	<i>Levels of Vaccination</i>	<i>Health Care Outcomes</i>
<i>Mortality and Survival Rates</i>	264	35	2	5
<i>Levels of Access to Health Care</i>	71	1499	0	222
<i>Levels of Vaccination</i>	2	0	161	0
<i>Health Care Outcomes</i>	10	145	1	345

Dari tabel *confusion matrix* tersebut, berikut merupakan tingkat *Precision*, *Recall*, dan *F-Measure* untuk setiap subkategori:

Tabel 6.9 Evaluasi Performa Percobaan Kategorisasi 2

Subkategori	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Mortality and Survival Rates</i>	76,08%	34,55%	47,52%
<i>Levels of Access to Health Care</i>	89,28%	72,84%	72,84%
<i>Levels of Vaccination</i>	98,17%	96,41%	96,41%
<i>Health Care Outcomes</i>	60,31%	53,82%	53,82%
Persentase / Total	80,96%	64,41%	70,48%

Dari hasil evaluasi performa tersebut, didapatkan hasil *precision* sebesar 80,96% dan *recall* sebesar 64,41%. Dari hasil tersebut dapat dikatakan bahwa percobaan 2 jauh lebih baik hasilnya dibandingkan dengan percobaan 1. Terlihat bahwa setelah dilakukan perubahan urutan kategorisasi, nilai *recall* pada subkategori *Levels of Vaccination* meningkat menjadi

96,41%. Hal ini menandakan bahwa urutan yang berbeda cukup berpengaruh terhadap hasil *precision* dan *recall* pada kategorisasi. Urutan yang berbeda tersebut menyebabkan dataset yang mengandung lebih dari 1 kata *parser* pada judulnya untuk masuk kedalam subkategori yang dikategorisasikan terlebih dahulu.

Hasil Percobaan Kategorisasi 3

Pada percobaan kategorisasi ini, urutan subkategori yang dilakukan kategorisasi adalah:

1. *Levels of Vaccination*
2. *Mortality and Survival Rates*
3. *Health Care Outcomes for Particular Groups*
4. *Levels of Access to Health Care*

Berikut merupakan hasil *confusion matrix* yang didapatkan:

Tabel 6.10 Confusion Matrix Percobaan Kategorisasi 3

Predicted Actual	<i>Mortality and Survival Rates</i>	<i>Levels of Access to Health Care</i>	<i>Levels of Vaccination</i>	<i>Health Care Outcomes</i>
<i>Mortality and Survival Rates</i>	264	35	2	5
<i>Levels of Access to Health Care</i>	71	1374	0	347
<i>Levels of Vaccination</i>	2	0	161	0
<i>Health Care Outcomes</i>	10	87	1	403

Dari tabel *confusion matrix* tersebut, berikut merupakan tingkat *Precision*, *Recall*, dan *F-Measure* untuk setiap subkategori:

Tabel 6.11 Evaluasi Performa Percobaan Kategorisasi 3

Subkategori	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Mortality and Survival Rates</i>	76,08%	34,55%	47,52%
<i>Levels of Access to Health Care</i>	91,84%	66,76%	66,76%
<i>Levels of Vaccination</i>	98,17%	96,41%	96,41%
<i>Health Care Outcomes</i>	53,38%	62,87%	62,87%
Persentase / Total	79,87%	65,15%	69,97%

Dari hasil evaluasi performa tersebut, didapatkan hasil *precision* sebesar 79,87% dan *recall* sebesar 65,15%. Terjadi perubahan yang tidak signifikan yaitu penurunan *precision* dan peningkatan *recall* yang tidak melebihi 2%, begitu pula pada nilai F-Measure yang turun tidak lebih dari 1%. Hal ini menunjukkan bahwa percobaan 2 dan percobaan 3 tidak jauh berbeda. Subkategori yang terpengaruh hanya subkategori yang diubah urutannya yaitu *Health Care Outcomes* dan *Level of Access to Health Care*.

Hasil Percobaan Kategorisasi 4

Pada percobaan kategorisasi ini, urutan subkategori yang dilakukan kategorisasi adalah:

1. *Levels of Vaccination*
2. *Levels of Access to Health Care*
3. *Health Care Outcomes for Particular Groups*
4. *Mortality and Survival Rates*

Berikut merupakan hasil *confusion matrix* yang didapatkan:

Tabel 6.12 Confusion Matrix Percobaan Kategorisasi 4

Predicted Actual	<i>Mortality and Survival Rates</i>	<i>Levels of Access to Health Care</i>	<i>Levels of Vaccination</i>	<i>Health Care Outcomes</i>
<i>Mortality and Survival Rates</i>	234	57	2	13
<i>Levels of Access to Health Care</i>	29	1539	0	224
<i>Levels of Vaccination</i>	2	0	161	0
<i>Health Care Outcomes</i>	1	146	1	353

Dari tabel *confusion matrix* tersebut, berikut merupakan tingkat *Precision*, *Recall*, dan *F-Measure* untuk setiap subkategori:

Subkategori	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
<i>Mortality and Survival Rates</i>	87,97%	30,63%	45,44%
<i>Levels of Access to Health Care</i>	88,35%	74,78%	74,78%
<i>Levels of Vaccination</i>	98,17%	96,41%	96,41%
<i>Health Care Outcomes</i>	59,83%	55,07%	55,07%
Persentase / Total	83,58%	64,22%	70,27%

Dari hasil evaluasi performa tersebut, didapatkan hasil *precision* sebesar 83,58% dan *recall* sebesar 64,22%. Nilai yang didapatkan pada percobaan ini mirip seperti percobaan 2 namun dengan nilai *precision* yang lebih tinggi 3%. Hal ini menunjukkan bahwa percobaan 4 memiliki nilai lebih baik dibandingkan percobaan 2 dan percobaan 3. Namun perlu diperhatikan bahwa nilai *recall* pada subkategori *Mortality and*

Survival Rates menurun sehingga pemilihan urutan kategorisasi perlu dipertimbangkan lagi.

Berdasarkan 4 percobaan yang telah dilakukan terhadap nilai *precision*, *recall*, dan *f-measure*, terdapat satu nilai sama yang didapatkan oleh kategorisasi dengan menggunakan kata *parser* terpilih. Nilai tersebut merupakan nilai *coverage*, yang mana menunjukkan jumlah dataset yang berhasil dikategorisasikan kedalam subkategori terhadap jumlah total dataset yang berhasil diakuisisi. Total dataset yang berhasil diakuisisi dari situs BPS pada 3 Mei 2018 berjumlah 3691 dataset, dari jumlah total tersebut terdapat 2762 dataset yang berhasil terkategorisasi menggunakan kata *parser* terpilih. *Coverage* yang dihasilkan mencapai 76,09%. Hal tersebut menunjukkan bahwa kata *parser* yang terpilih cukup baik dalam melakukan kategorisasi dataset kesehatan yang dimiliki oleh pemerintah daerah.

6.2 Analisis Penilaian Open Data

Analisis penilaian *open data* dilakukan untuk mengetahui hasil penerapan *open data* pemerintah daerah terhadap metode penilaian yang dilakukan oleh Open Data Barometer. Penilaian dilakukan terhadap setiap dataset yang berhasil diambil dan judul dataset mengandung kata *parser* yang telah ditetapkan. Hasil penilaian setiap dataset ini menjadi *input* untuk menentukan nilai penerapan *open data* yang didapatkan oleh pemerintah daerah pemilik dataset tersebut. Berikut merupakan beberapa analisis yang dilakukan:

6.2.1 Jumlah Dataset Ternilai

Menggunakan kode program *web-scraping* dan *crawler* yang dibuat, dataset yang berasal dari situs BPS dan portal *open data* disimpan kedalam basis data MongoDB. Hasil penyimpanan informasi dataset ini dicocokkan dengan kata *parser* yang dibuat, penilaian dilakukan dengan menjawab 10 pertanyaan dataset dari Open Data Barometer terhadap informasi dataset yang dimiliki. Berikut jumlah dataset dari BPS dan portal *open data* pada periode pengambilan data Mei dan Juni 2018:

Tabel 6.13 Jumlah Dataset dan Pemerintah Daerah Pemilik Dataset yang Terambil dan Ternilai

Sumber	Periode <i>Crawling</i>	Dataset Terambil		Dataset Ternilai	
		Jumlah Pemda	Jumlah Dataset	Jumlah Pemda	Jumlah Dataset
BPS	Mei 2018	407 Pemda	3691 Dataset	393 Pemda	2762 Dataset
	Juni 2018	406 Pemda	4131 Dataset	393 Pemda	2989 Dataset
Portal Open Data	Mei 2018	10 Pemda	19759 Dataset	8 Pemda	2271 Dataset
	Juni 2018	12 Pemda	20496 Dataset	9 Pemda	2358 Dataset

Pada Tabel 6.13 dari total 548 Situs BPS dan 15 Situs Portal Open Data yang tercatat pada basis data, dapat diketahui bahwa:

1. Tidak seluruh situs memiliki dataset yang dapat diambil. Situs BPS yang memiliki dataset pada kategori kesehatan berjumlah sekitar 74% dari total situs BPS yang ada, sedangkan situs portal open data yang dapat di *crawl* adalah sejumlah 66-80% pada dua periode pengambilan data yang telah dilakukan.
2. Total dataset yang didapatkan dari situs Portal Open Data jauh lebih banyak dibandingkan dataset yang didapatkan dari situs BPS, padahal jumlah pemilik dataset totalnya sangat jauh lebih besar BPS. Hal ini dikarenakan dataset yang diambil dari Portal Open Data belum spesifik merujuk kedalam suatu kategori, sedangkan data yang diambil dari situs BPS sudah dikategorikan menjadi kesehatan menurut BPS.
3. Total dataset yang ternilai dari kedua sumber dataset tidak berbeda jauh setelah dilakukan penilaian dengan mengecek judul dataset terhadap kata *parser* yang dibuat. Namun terlihat pengurangan yang signifikan pada data yang bersumber dari situs

Portal Open Data terkait jumlah dataset terambil dengan jumlah dataset ternilai.

6.2.2 Karakteristik Penilaian Dataset

Berdasarkan penilaian yang telah dilakukan terhadap dataset, dilakukan pengecekan terhadap berbagai nilai yang didapatkan oleh dataset. Pada periode penilaian dataset bulan Juni 2018, didapatkan data sebagai berikut:

Tabel 6.14 Kombinasi Nilai Dataset yang Terjadi Berdasarkan Sumber

Sumber	Kombinasi Nilai Dataset Terhadap Pertanyaan	Jumlah Dataset
BPS	1 1 1 0 1 0 0 0 1 0	2477
	1 1 1 0 1 0 1 1 1 0	512
Portal Open Data	1 1 1 0 1 0 0 0 1 0	31
	1 1 1 0 1 0 1 1 1 0	2327

Pada Tabel 6.14 dari penilaian yang dilakukan berdasarkan jawaban terhadap pertanyaan Open Data Barometer dilakukan *query* SELECT DISTINCT pada jawaban setiap pertanyaan. Nilai 1 dan 0 pada tabel diatas menunjukkan jawaban benar atau tidaknya informasi dataset terhadap pertanyaan yang diajukan. Dari Tabel 6.14 dapat diketahui bahwa:

1. Hanya terdapat 2 jenis nilai yang berbeda pada masing-masing sumber, dan perbedaan kombinasi nilai pada kedua sumber menunjukkan hasil yang sama. Hal ini dikarenakan banyak asumsi yang dilakukan dalam melakukan penilaian sehingga tidak seluruh pertanyaan terjawab dari pengambilan data. Beberapa pertanyaan dijawab menurut hasil pengamatan penulis terhadap kondisi sumber dataset.
2. Perbedaan kombinasi nilai terdapat pada pertanyaan mengenai *update* dan *sustainability* dataset, yang mana pada aplikasi penilaian dataset, nilai dari *sustainability* dianggap mengikuti nilai *update*.

Artinya perbedaan terjadi karena adanya informasi *key update* yang disimpan pada informasi dataset.

3. Dari jumlah dataset yang ternilai berdasarkan kombinasi nilai terhadap pertanyaan. Terlihat bahwa data yang bersumber dari BPS lebih banyak memiliki nilai *update* 0 sedangkan data yang bersumber dari portal *open data* lebih banyak memiliki nilai *update* 1. Artinya data dari BPS lebih jarang diperbarui daripada data dari portal *open data*. Tetapi terdapat kemungkinan bahwa portal *open data* banyak mendapat nilai 1 pada pertanyaan *update* karena implementasi *open data* masih tergolong baru dilakukan.
- 4.

6.2.3 Karakteristik Penilaian Penerapan *Open Data* Pemerintah Daerah

Setelah nilai dari setiap dataset didapatkan, dilakukan penilaian terhadap penerapan *open data* yang dilakukan oleh setiap pemerintah daerah. Hasil dari penilaian pada setiap kategori menjadi nilai penerapan *open data* pemerintah daerah secara keeseluruhan. Pada basis data, penilaian telah dilakukan terhadap 15 kategori dataset menurut Open Data Barometer, namun karena kategori yang diambil datanya pada penelitian ini merupakan kategori kesehatan, nilai penerapan yang tidak bernilai 0 dari setiap pemerintah daerah merupakan pemerintah daerah yang memiliki dataset kategori kesehatan. Pada periode penilaian Juni 2018, didapatkan data sebagai berikut:

Tabel 6.15 Nilai Penerapan *Open Data* yang didapatkan Pemerintah Daerah

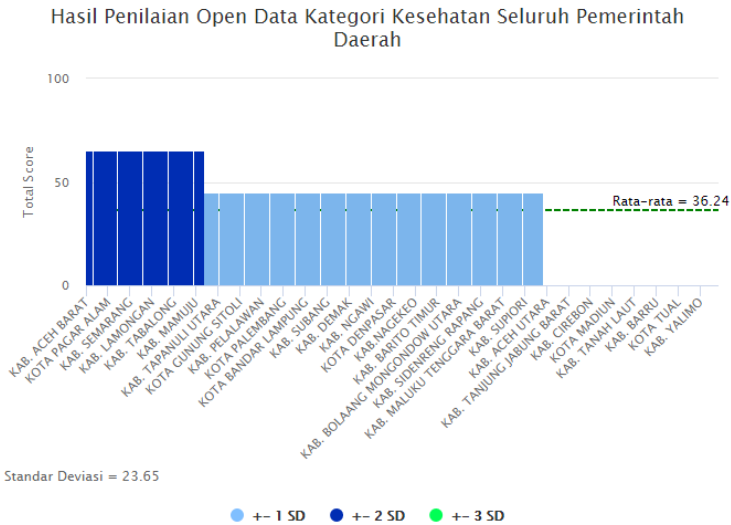
No.	Jumlah Pemda	Kombinasi Nilai Penerapan <i>Open Data</i>	Nilai Akhir Penerapan <i>Open Data</i>
1	102 Pemda	1 1 1 0 1 0 1 1 1 0	65
2	294 Pemda	1 1 1 0 1 0 0 0 1 0	45
3	152 Pemda	0 0 0 0 0 0 0 0 0 0	0

Pada Tabel 6.15 dari penilaian penerapan *open data* yang dilakukan berdasarkan hasil penilaian dataset yang dimiliki oleh pemerintah daerah, dilakukan *query* SELECT DISTINCT pada hasil kategori kesehatan. Nilai 1 dan 0 pada tabel diatas menunjukkan benar atau tidaknya penerapan yang dilakukan pemerintah daerah terhadap Open Data Barometer, nilai ini membentuk nilai total berdasarkan logika penilaian pada Tabel 3.2. Berikut hubungan antara kepemilikan sumber dataset milik pemerintah daerah yang ternilai terhadap nilai akhir penerapan *open data*:

Tabel 6.16 Nilai Akhir Penerapan *Open Data* berdasarkan Kepemilikan Dataset Ternilai

No.	Nilai Akhir Penerapan <i>Open Data</i>	Kepemilikan Dataset Ternilai		
		BPS	Portal Open Data	Kedua Sumber
1	65	93 Pemda	3 Pemda	6 Pemda
2	45	294 Pemda	0 Pemda	0 Pemda
3	0	0 Pemda	0 Pemda	0 Pemda

Berdasarkan nilai yang didapatkan oleh pemerintah daerah pada periode Juni 2018, berikut merupakan visualisasi nilai akhir penerapan *open data* terhadap rata-rata dan standar deviasi penilaian:



Gambar 6.1 Rata-rata dan Standar Deviasi Nilai Peneraan *Open Data* Pemerintah Daerah pada Periode Juni 2018

Dari beberapa informasi statistik yang telah disajikan, dapat diketahui bahwa:

1. Terdapat 3 jenis kombinasi nilai berbeda terhadap hasil peneraan *open data* yang dilakukan oleh pemerintah daerah dari penilaian terhadap dataset. Hal ini dikarenakan penilaian peneraan bersumber pada penilaian dataset pada proses sebelumnya yang mana hanya memiliki 2 jenis kombinasi nilai berbeda. Kombinasi nilai ketiga yang terbentuk adalah nilai 0 untuk seluruh pertanyaan yang menunjukkan bahwa pemerintah daerah tidak memiliki dataset ternilai dari kedua sumber dataset.
2. Berdasarkan sumber kepemilikan dataset ternilai, semua pemerintah daerah yang memiliki dataset ternilai dari portal open data mendapatkan nilai 65. Pemerintah daerah yang memiliki dataset yang berasal dari BPS 24% mendapatkan nilai 65

sedangkan 76% sisanya mendapatkan nilai 45. Hal ini terkait dengan karakteristik penilaian dataset sebelumnya yang telah dibahas bahwa data yang bersumber dari BPS lebih banyak memiliki nilai *update* 0 sedangkan data yang bersumber dari portal *open data* lebih banyak memiliki nilai *update* 1.

3. Jumlah dataset yang dimiliki pemerintah daerah tidak berpengaruh terhadap besar-kecilnya nilai penerapan *open data* yang didapatkan, selama pemerintah daerah minimal memiliki 1 dataset yang ternilai, maka pemerintah daerah dapat memperoleh nilai akhir penerapan *open data* 65 atau 45.
4. Berdasarkan nilai rata-rata dan standar deviasi yang terbentuk terhadap nilai penerapan *open data*. Dapat diketahui bahwa 54% pemerintah daerah yaitu yang mendapatkan nilai 45 berjarak 1 standar deviasi dari nilai rata-rata, sedangkan 46% sisa pemerintah daerah yang mendapatkan nilai 65 dan 0 berjarak 2 standar deviasi dari nilai rata-rata.

6.3 Pengujian Non Fungsional

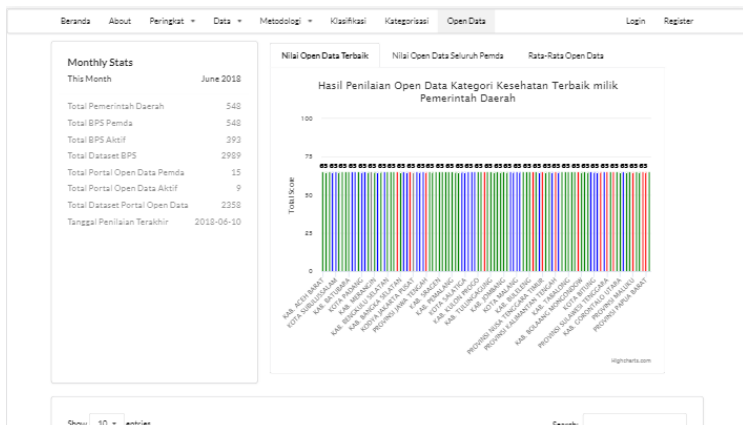
Pengujian non fungsional dilakukan dengan mencoba melakukan akses ke halaman utama aplikasi visualisasi yang telah dibuat terhadap beberapa faktor seperti *load time* situs, peramban, dan perangkat pengakses situs. Hal tersebut bertujuan untuk mengetahui hasil aplikasi visualisasi apabila dihadapkan pada kondisi user yang berbeda. Berikut merupakan beberapa pengujian yang dilakukan:

6.3.1 Pengujian Load Time Situs

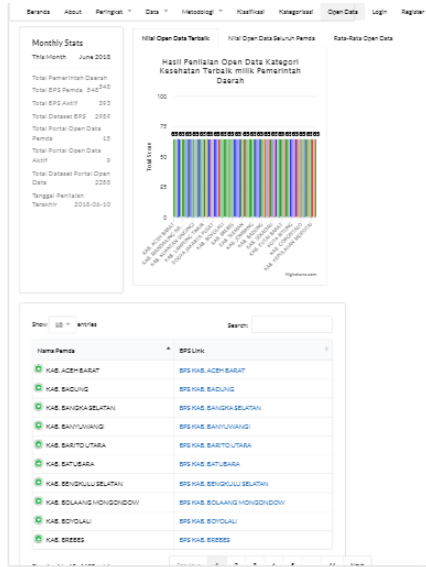
Pengujian terhadap *load time* situs dilakukan menggunakan fitur developer milik Google Chrome. Caranya adalah melakukan penghapusan *cache* setiap percobaan, lalu mengakses situs untuk mendapatkan performa *load time non-cached sites*, dilanjutkan dengan melakukan *refresh* halaman untuk mendapatkan performa *load time cached sites*. Percobaan dilakukan sepuluh kali untuk mendapatkan rata-rata *load time* yang dibutuhkan oleh peramban. Berikut merupakan hasil



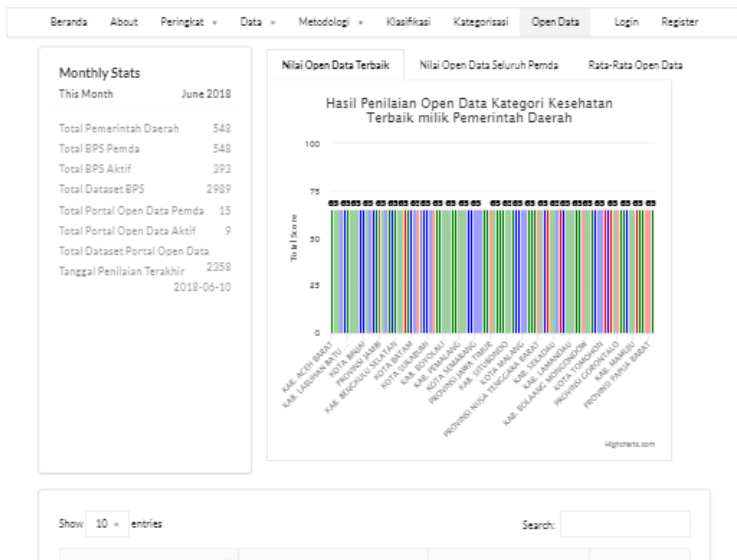
Gambar 6.3 Tampilan pada Mobile dengan Resolusi 720x1280 (Potrait)



Gambar 6.4 Tampilan pada Mobile dengan Resolusi 720x1280 (Landscape)



Gambar 6.5 Tampilan pada Tablet dengan Resolusi 1024x768 (Potrait)



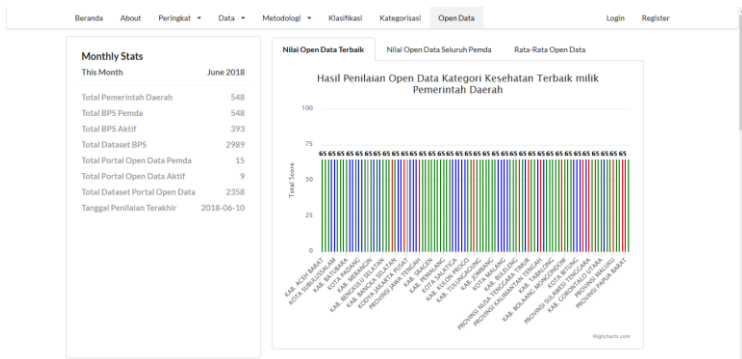
Gambar 6.6 Tampilan pada Tablet dengan Resolusi 1024x768 (Landscape)

6.3.3 Pengujian Akses Peramban

Pengujian akses perangkat situs dilakukan dengan melakukan akses ke situs menggunakan beberapa peramban berbeda yang umum digunakan. Berikut hasil percobaan yang dilakukan terhadap peramban Google Chrome, Mozilla Firefox, dan Microsoft Edge:



Gambar 6.7 Tampilan pada Peramban Google Chrome



Gambar 6.8 Tampilan pada Peramban Mozilla Firefox

BAB VII

KESIMPULAN DAN SARAN

Pada bab ini dipaparkan kesimpulan dari pengerjaan tugas akhir yang telah dilakukan dan saran yang dapat diberikan sebagai pengembangan penelitian yang telah dilakukan.

7.1 Kesimpulan

Berdasarkan pengerjaan tugas akhir berjudul “Rancang Bangun Aplikasi Penilaian Penerapan *Open Data* berdasarkan *Framework* Open Data Barometer (Studi Kasus: Kategori Data Kesehatan)” dapat disimpulkan beberapa hal sebagai berikut:

1. Penelitian ini membuktikan bahwa akuisisi data terhadap dataset kategori kesehatan yang berasal dari situs BPS dapat dilakukan menggunakan *web-scrapers*, sedangkan dataset yang berasal dari situs portal *open data* berbasis CKAN dapat dilakukan menggunakan CKAN *Action API*.
2. Hasil dari akuisisi dataset dengan menggunakan *web-scrapers* dan *crawler* data pada bulan Juni menghasilkan 4131 dataset dari 406 situs BPS pemerintah daerah dan 20496 dataset dari 12 situs portal *open data* berbasis CKAN.
3. Berdasarkan penggunaan kata *parser* saat ini yang terbentuk setelah penghitungan nilai IDF, dataset yang terkategori ke dalam kategori kesehatan pada bulan Juni berjumlah 2989 dataset ternilai yang berasal dari situs BPS dan 2358 dataset ternilai yang berasal dari situs portal *open data* berbasis CKAN.
4. Dalam penelitian yang telah dilakukan, berdasarkan kata *parser* saat ini didapatkan bahwa sejumlah 152 atau sekitar 27,7% pemerintah daerah mendapatkan nilai 0 dikarenakan tidak ada dataset yang ternilai, sedangkan 294 atau sekitar 53,6% pemerintah daerah mendapatkan nilai 45 karena dataset yang

dimiliki tidak diperbarui, dan sisanya sebanyak 102 atau sekitar 18,6% pemerintah daerah mendapatkan nilai 65 karena memiliki dataset yang diperbarui.

7.2 Saran

Berdasarkan hasil pengerjaan tugas akhir yang telah dilakukan, beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya antara lain:

1. Melakukan penilaian *open data* terhadap kategori dataset lain sesuai Open Data Barometer untuk memperkaya ragam nilai penerapan *open data* yang didapatkan oleh pemerintah daerah.
2. Menambahkan sumber dataset yang sesuai dengan kategori yang dinilai berdasarkan ketersediaan data yang dimiliki oleh pemerintah daerah.
3. Melakukan kategorisasi dengan menggunakan metode yang berbeda guna mendapatkan tingkat akurasi yang lebih baik dibandingkan penggunaan nilai IDF.
4. Melibatkan pakar atau seseorang yang ahli dalam bidang sesuai kategori dataset dalam menentukan kata *parser* untuk mengurangi bias terhadap pemilihan kata *parser* yang dilakukan.
5. Melakukan proses *word similarity* dalam kata *parser* yang dibuat sebagai *filter* untuk melakukan kategorisasi sehingga apabila kata mirip yang ditulis berbeda dapat terambil oleh *parser*.
6. Mempertimbangkan peraturan undang-undang tentang penerapan *open data* yang masih dalam tahap pembuatan ketika penelitian ini dilakukan agar kedepannya penilaian *open data* lebih sesuai dengan arahan pemerintah.

DAFTAR PUSTAKA

- [1] Open Government Indonesia, “Exclusive: Why Indonesia is so keen to open up,” 2017. [Online]. Available: <http://www.opengovindonesia.org/news/257/exclusive-why-indonesia-is-so-keen-to-open-up>. [Accessed: 06-Feb-2017].
- [2] KPU, “Visi- Misi - Program Aksi Ir. H. Joko Widodo - Drs. H.M. Jusuf Kala Pemilu Presiden dan Wakil Presiden Tahun 2014.” 2014.
- [3] S. B. Yudhoyono, *Undang-Undang Republik Indonesia Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik*. Indonesia, 2008.
- [4] OGI, “Open Government Indonesia,” 2017. [Online]. Available: <http://www.opengovindonesia.org/about/1/open-government-indonesia>. [Accessed: 06-Feb-2018].
- [5] Satu Data Indonesia, “Tentang Satu Data Indonesia,” 2014. [Online]. Available: <http://data.go.id/tentang/>. [Accessed: 08-Feb-2018].
- [6] World Wide Web Foundation, “Open Data Barometer Global Report (Fourth Edition),” 2016.
- [7] World Wide Web Foundation, “Open Data Barometer 4th Edition - Methodology,” 2015.
- [8] B. H. Handoko, “Rancang Bangun Perangkat Lunak untuk Monitoring Website Pemerintahan dengan Pembobotan Kriteria Berbasis Analytical Hierarchy Process (AHP),” Institut Teknologi Sepuluh Nopember, 2017.
- [9] N. A. Rakhmawati, I. Hafidz, R. P. Kusumawardani, and R. P. Wibowo, “Towards Surabaya as An Open Data

City,” 2016.

- [10] N. Veljković, S. Bogdanović-Dinić, and L. Stoimenov, “Benchmarking open government: An open data perspective,” *Gov. Inf. Q.*, vol. 31, no. 2, pp. 278–290, 2014.
- [11] J. Jansen, S. de Vries, and P. van Schaik, “The Contextual Benchmark Method: Benchmarking e-Government services,” *Gov. Inf. Q.*, vol. 27, no. 3, pp. 213–219, 2010.
- [12] Open Knowledge International, “The Open Definition,” 2014. [Online]. Available: <http://opendefinition.org/>. [Accessed: 29-Jan-2018].
- [13] World Wide Web Foundation, “Open Data Barometer Research Handbook 4th edition,” 2016.
- [14] Open Knowledge International, “What is Open?,” 2014. [Online]. Available: <https://okfn.org/opendata/>. [Accessed: 29-Jan-2018].
- [15] World Wide Web Foundation, “The Open Data Barometer,” 2016. [Online]. Available: <http://opendatabarometer.org/barometer/>. [Accessed: 24-Jan-2018].
- [16] Open Data Charter, “Principles - International Open Data Charter,” 2015. [Online]. Available: <http://opendatabarometer.org/barometer/>. [Accessed: 06-Feb-2018].
- [17] D. Lathrop and L. Ruma, *Open Government: Collaboration, Transparency, and Participation in Practice*, Revised. O’Reilly Media, Inc., 2010.
- [18] White House, “Memorandum on Transparency and Open Government.” 2009.
- [19] B. Ubaldi, “Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives,” *OECD*

Work. Pap. Public Gov., vol. NO.22, no. 22, p. 61, 2013.

- [20] C. Malamud, “Open Government Data Principles,” 2007. [Online]. Available: https://public.resource.org/8_principles.html. [Accessed: 06-Feb-2018].
- [21] A. Ball, “How to License Research Data.” Digital Curation Centre, Edinburgh, 2014.
- [22] Open Knowledge International, “Open Definition 2.1,” 2014. [Online]. Available: <http://opendefinition.org/od/2.1/en/>. [Accessed: 08-Feb-2018].
- [23] Project Open Data, “Open Licenses,” 2015. [Online]. Available: <https://project-open-data.cio.gov/open-licenses/>. [Accessed: 08-Feb-2018].
- [24] Open Data Institute, “Publisher’s Guide to Open Data Licensing,” 2013. [Online]. Available: <https://theodi.org/guides/publishers-guide-open-data-licensing>. [Accessed: 08-Feb-2018].
- [25] Badan Pusat Statistik, “Tentang BPS,” 2017. [Online]. Available: <https://www.bps.go.id/menu/1/informasi-umum.html>. [Accessed: 05-Feb-2018].
- [26] Satu Data Indonesia, “Badan Pusat Statistik,” 2008. [Online]. Available: data.go.id/organization/about/bps. [Accessed: 03-Feb-2018].
- [27] P. Christensson, “Metadata Definition,” 2006. [Online]. Available: <https://techterms.com/definition/metadata>. [Accessed: 06-Feb-2018].
- [28] World Wide Web Consortium, “Metadata and Resource Description,” 2007. [Online]. Available: <https://www.w3.org/Metadata/>. [Accessed: 06-Feb-2018].
- [29] T. Berners-Lee, “Metadata Architecture,” 1997.

- [Online]. Available: <https://www.w3.org/DesignIssues/Metadata.html>. [Accessed: 07-Feb-2018].
- [30] J. Hare, “What is metadata and why is it as important as the data itself?,” 2016. [Online]. Available: <https://www.opendatasoft.com/2016/08/25/what-is-metadata-and-why-is-it-important-data/>. [Accessed: 07-Feb-2018].
- [31] P. Christensson, “Meta Tag Definition,” 2016. [Online]. Available: <https://techterms.com/definition/metatag>. [Accessed: 06-Feb-2018].
- [32] M. Cooper, A. Kirkpatrick, and J. O. Connor, “Understanding WCAG 2.0: A guide to understanding and implementing Web Content Accessibility Guidelines 2.0,” 2016. [Online]. Available: <https://www.w3.org/TR/UNDERSTANDING-WCAG20/>. [Accessed: 07-Feb-2018].
- [33] D. Haynes, *Metadata for Information Management and Retrieval*. London: Facet Publishing, 2004.
- [34] M. Janssen and A. Zuiderwijk, “The potential of metadata for linked open data and its value for users and publishers,” *JeDEM-eJournal eDemocracy Open Gov.*, vol. 4, no. 2, pp. 222–244, 2012.
- [35] S. Zheng, R. Song, J.-R. Wen, and D. Wu, “Joint optimization of wrapper generation and template detection,” *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '07*, no. March, p. 894, 2007.
- [36] Nicolas Poggi, Toni Moreno, Josep Lluís Berral, Ricard Gavaldà, and Jordi Torres, “Web Customer Modeling for Automated Session Prioritization on High Traffic Sites,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, pp. 450–454.

- [37] A. Josi, L. A. Abdillah, and Suryayusra, “Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah,” *Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma*. 2014.
- [38] V. Mitra, H. Sujaini, and A. B. P. Negara, “Rancang Bangun Aplikasi Web Scraping Untuk Korpus Paralel Indonesia - Inggris Dengan Metode HTML Dom,” *J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–6, 2017.
- [39] J. Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries,” *Proc. first Instr. Conf. Mach. Learn.*, pp. 1–4, 2003.
- [40] CKAN Association, “About CKAN,” 2012. [Online]. Available: <https://ckan.org/about/>. [Accessed: 07-Feb-2018].
- [41] Open Knowledge International, “CKAN,” 2012. [Online]. Available: <https://okfn.org/projects/ckan/>. [Accessed: 27-Jan-2018].
- [42] CKAN Association, “CKAN Features,” 2012. [Online]. Available: <https://ckan.org/features/>. [Accessed: 29-Jan-2018].
- [43] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” pp. 39–46, 2003.

Halaman ini sengaja dikosongkan

LAMPIRAN A – Asumsi terhadap Pertanyaan Dataset

No.	Pertanyaan	Penjelasan Open Data Barometer	Sumber	Asumsi Penulis
1	Does the data exist?	<i>“If the data is not collected by the government in this country at all, or is not aggregated at a federal level in any form, then answer NO to this question and you don’t need to continue with the rest of the survey for this dataset. If the data is collected at the level of government you are answering for in any form, then you should answer YES”</i>	<i>Open Data Barometer 4th Edition Research Handbook, Halaman 54</i>	Pertanyaan ini tidak bisa dijawab melalui teknik <i>scraping</i> , karena ketersediaan secara <i>offline</i> juga merupakan jawaban benar. Untuk itu jawaban dari pertanyaan ini akan selalu sama dengan jawaban pertanyaan kedua yang akan dinilai dari ketersediaan dataset secara <i>online</i> .
2	Is it available online from government in any form?	<i>“Can you find the data online in any digital format? You should only answer YES if a substantial proportion of the data we are looking for is available, If no government source is available, answer NO to this question and</i>	<i>Open Data Barometer 4th Edition Research Handbook, Halaman 55</i>	Pertanyaan ini merupakan pertanyaan yang mendasari pertanyaan-pertanyaan selanjutnya, ketersediaan dataset secara <i>online</i> dapat dilihat dari situs portal open data dan BPS yang dimiliki oleh pemerintah daerah, apabila dataset tersedia pada situs tersebut maka dapat memenuhi kriteria ini. Tetapi apabila pertanyaan ini terjawab dengan

		<i>you don't need to continue with the survey for this dataset."</i>		jawaban tidak, maka seluruh pertanyaan selanjutnya tidak memenuhi kriteria.
3	Is the dataset provided in machine readable format?	<i>"Can you find a copy of the data in a structured and standardized machine readable format which can be opened in appropriate data manipulation and analysis software with respect to the best example you can locate in terms of availability of open data"</i>	<i>Open Data Barometer 4th Edition Research Handbook,</i> Halaman 56	Pertanyaan ini merupakan pertanyaan terkait format dari dataset yang tersedia pada situs. Berdasarkan panduan Open Data Barometer, data yang termasuk dalam format <i>machine readable</i> merupakan data dalam bentuk terstruktur yang dapat dibuka oleh perangkat lunak yang digunakan untuk memanipulasi dan menganalisis data[1, p. 56]. Pada portal open data berbasis CKAN, ada metadata yaitu "format" yang menjelaskan format dataset yang tersedia, apabila format yang tersedia adalah JSON, XML, XLS, RDF, maka dataset dapat memenuhi kriteria ini. Sedangkan pada situs BPS, setelah dilakukan pengecekan pada beberapa pemerintah daerah, data yang tersedia dapat diunduh dalam bentuk XLS sehingga apabila dataset tersedia, maka seluruh dataset dapat memenuhi kriteria ini.
4	Is the machine readable data available in bulk?	<i>"Bulk access to data files allows analysts and developers to more flexibly build upon the data and to integrate it into other products, services and activities. Look for links to bulk downloads. These might be</i>	<i>Open Data Barometer 4th Edition Research Handbook,</i> Halaman 58	Pertanyaan ini merupakan pertanyaan terkait ketersediaan cara pengunduhan data secara keseluruhan/banyak yang tersedia pada satu halaman situs web. Pada portal open data berbasis CKAN, pemerintah daerah Indonesia tidak memberikan pilihan untuk melakukan pengunduhan data dalam jumlah banyak pada satu halaman,

		<i>organised by month or year for time-series data, or broken down into sub-files for very large datasets.”</i>		begitu pula pada situs BPS, untuk melakukan pengunduhan terhadap dataset, harus menuju halaman dataset tersebut terlebih dahulu, tidak dapat dilakukan langsung sekaligus. Sehingga kedua sumber dataset tidak dapat memenuhi kriteria ini.
5	Is the dataset available free of charge?	<i>“Sometimes datasets are only available for a charge - either a one-off payment or a subscription fee. For example, when maps are licensed by the mapping agency and only provided for a fee, or when a country's legislation is compiled and provided for sale through a private company.”</i>	<i>Open Data Barometer 4th Edition Research Handbook,</i> Halaman	Pertanyaan ini merupakan pertanyaan terkait biaya dari akses dataset, terkadang dataset tersedia dengan cara membayar atau berlangganan kepada penyedia data. Pada kasus ini, pemerintah daerah memberikan datanya tanpa pemakai harus mengeluarkan biaya dari situs portal open data dan BPS. Sehingga apabila dataset tersedia, maka seluruh dataset dapat memenuhi kriteria ini.
6	Is the data openly licensed?	<i>“Unless the license that a dataset is provided under is explicitly stated, users of the data cannot be sure whether they have permission to reuse it, and under what conditions. This will only count as an open license if:</i>	<i>Open Data Barometer 4th Edition Research Handbook,</i> Halaman 59	Pertanyaan ini merupakan pertanyaan terkait lisensi yang diberikan terhadap data, dalam Open Data Barometer, data termasuk <i>open license</i> apabila ada pernyataan bahwa siapa saja boleh menggunakan data tersebut, dan tidak ada batasan penggunaan kembali terhadap data[1, pp. 59–60]. Pada portal open data berbasis CKAN, lisensi data dapat diketahui dari tag metadata “license” dan semacamnya, apabila memenuhi lisensi open data pada umumnya, maka pertanyaan ini dijawab benar.

		<p>- <i>It clearly states that anyone has permission to reuse it</i></p> <p>- <i>It does not restrict what the data can be re-used for</i></p>		<p>Sedangkan untuk situs BPS, tidak ada pernyataan jelas yang menyatakan bahwa data tersebut boleh digunakan oleh siapapun dan untuk keperluan apapun sehingga tidak dapat memenuhi kriteria ini.</p>
7	Is the dataset up to date?	<p><i>“Many datasets are only useful if they are recent and kept updated. Consider any evidence on what was the last update date and how often it is being updated and how often you would expect data of this form to be updated.”</i></p>	<p><i>Open Data Barometer 4th Edition Research Handbook, Halaman 60</i></p>	<p>Pertanyaan ini merupakan pertanyaan terkait pembaruan dataset yang dimiliki, dataset yang diperbarui akan lebih bermanfaat. Pada portal open data, terdapat metadata yang menjelaskan pembaruan terhadap dataset yaitu “metadata_modified”. Sedangkan pada situs BPS, pada setiap judul dataset terdapat tahun yang menjelaskan kapan informasi pada dataset terakhir tersedia. Menurut daftar informasi publik yang disediakan oleh BPS, informasi terkait statistik kesehatan dikumpulkan dalam periode dua tahunan, dan terakhir dikumpulkan pada 2016. Untuk itu data yang dianggap terbaru merupakan data yang dipublikasikan dalam kurun waktu dua tahun terakhir saat pengambilan dataset dilakukan.</p>
8	Is the publication of the dataset sustainable?	<p><i>“Look for evidence that this was a once-off publication, or for evidence that the dataset is being kept regularly updated (for example using past track-</i></p>	<p><i>Open Data Barometer 4th Edition Research</i></p>	<p>Pertanyaan ini merupakan pertanyaan terkait pembaruan berkala dari dataset yang dimiliki. Pada portal open data, penilaian dilakukan sama seperti pertanyaan sebelumnya yaitu dengan menilai atribut “metadata_modified”. Sedangkan pada situs BPS,</p>

		<i>record of publication). In some cases the evidence available to make a judgement will be quite limited. In these cases, where you cannot locate any evidence that the data will be updated regularly, you can answer NO.”</i>	<i>Handbook, Halaman 62</i>	apabila dataset memenuhi kriteria sebelumnya, maka kriteria ini juga dapat dipenuhi karena sifat dari sensus yang dilakukan oleh pemerintah daerah secara berkala.
9	Was it easy to find information on this dataset?	<i>“Answer YES if you think a regular Internet user with a degree level education would be able to locate and find out the information required about this dataset if they were looking for it. Answer NO if you struggled to locate the dataset, or if you think a degree educated Internet user in your country would face considerable struggles in trying to find and find out about this dataset.”</i>	<i>Open Data Barometer 4th Edition Research Handbook, Halaman 63</i>	Pertanyaan ini merupakan pertanyaan terkait kemudahan dalam menemukan informasi pada dataset, dalam panduan Open Data Barometer, dijelaskan bahwa apabila pengguna internet dinegara tersebut pada umumnya dapat menemukan informasi dengan mudah, maka dapat memenuhi kriteria ini [1, p. 63]. Ketersediaan informasi melalui portal open data dan situs BPS sama-sama mudah diakses menggunakan URL yang tersedia, oleh karena itu apabila dataset tersedia, maka dapat memenuhi kriteria ini.
10	Are (linked) data URIs provided for	<i>“First, check to see whether a ‘Linked Data’ version of the dataset or a ‘SPARQL endpoint’ is advertised. This may be provided in the ‘RDF (Resource</i>	<i>Open Data Barometer 4th Edition Research</i>	Pertanyaan ini merupakan pertanyaan terkait ketersediaan <i>linked data</i> pada dataset yang dimiliki oleh pemerintah daerah. Pertanyaan ini dapat dijawab dengan membuktikan adanya versi <i>Linked Data</i> dari dataset yang biasanya didapatkan pada SPARQL

	key elements in the dataset?	<p><i>Descriptor Format)</i> file format, or as a file with the .ttl, .rdf, or .n3 file extension. To answer yes to this question these formats are not required, but if you find these formats then it is likely that the answer is YES.</p> <p><i>Secondly, if you don't find these formats apparent, look to see if there are unique web pages with specific URIs (web addresses) provided for the key entities in the dataset and those are being published by the organisation or agency responsible for the dataset then you should answer YES</i></p> <p><i>Thirdly, if you don't find any of the above, look to see if there are any ID codes or fields associated with the key entities in the dataset, looking either at the associated documentation or the data itself. If that's the</i></p>	Handbook, Halaman 64	<p><i>Endpoints.</i> Implementasi open data situs BPS tidak memiliki SPARQL <i>Endpoints</i>, sama seperti portal open data berbasis CKAN yang dimiliki oleh pemerintah daerah juga tidak memiliki SPARQL <i>Endpoints</i>. Sehingga tidak dapat memenuhi kriteria ini.</p>
--	------------------------------	---	----------------------	---

		<p><i>case then you should answer YES.</i></p> <p><i>If queries for the dataset name and IDs, RDF, Linked Data or SPARQL provides no clear pointers to available data, answer NO.”</i></p>		
--	--	--	--	--

Halaman ini sengaja dikosongkan

LAMPIRAN B – Daftar Stopwords Tambahan

provinsi	sulawesi
kabupaten	papua
kecamatan	maluku
kota	nusa
madya	tenggara
utara	aceh
barat	riau
timur	jambi
selatan	bengkulu
kepulauan	lampung
sumatera	banten
sumatra	jakarta
jawa	yogyakarta
kalimantan	jogjakarta
bali	gorontalo

B-2

Halaman ini sengaja dikosongkan

**LAMPIRAN C – Daftar 25 kata DF/IDF Terendah
setiap Subkategori**

No.	Mortality and Survival Rates	Level of Vaccination	Level of Access to Health Care	Health Care Outcomes for Particular Groups
1	kb	imunisasi	kesehatan	penyakit
2	persentase	jenis	tenaga	jenis
3	usia	balita	fasilitas	penduduk
4	peserta	persentase	sarana	persentase
5	bayi	bayi	desa	kesehatan
6	aktif	kelamin	puskesmas	penderita
7	akseptor	diimunisasi	jenis	kelamin
8	kawin	lengkap	rumah	jalan
9	subur	penduduk	sakit	rawat
10	pasangan	hamil	pelayanan	sakit
11	lahir	dirinci	dokter	sebulan
12	balita	vaksinasi	medis	pasien
13	kontrasepsi	hasil	keluarga	keluhan
14	kelahiran	umur	kerja	setahun
15	anak	pencapaian	ketersediaan	malaria
16	hidup	target	unit	diare
17	jenis	rembang	posyandu	hiv
18	perempuan	cakupan	klinik	dbd
19	alat	kartu	tidur	aids
20	penolong	sasaran	kunjungan	wabah
21	penduduk	realisasi	bayi	tb
22	buruk	tt	berencana	ims
23	keluarga	banggai	dukun	berobat
24	kematian	bitung	gigi	dirinci
25	pus	puskesmas	pembantu	rsud

Keterangan : Kata yang diblok warna merupakan kata parser terpilih

No.	Mortality and Survival Rates	Level of Vaccination	Level of Access to Health Care	Health Care Outcomes for Particular Groups
1	kb	imunisasi	kesehatan	penyakit
2	persentase	jenis	tenaga	jenis
3	usia	balita	fasilitas	penduduk
4	peserta	persentase	sarana	persentase
5	bayi	bayi	desa	kesehatan
6	aktif	kelamin	puskesmas	penderita
7	akseptor	diimunisasi	jenis	kelamin
8	kawin	lengkap	rumah	jalan
9	subur	hamil	sakit	rawat
10	pasangan	penduduk	pelayanan	sakit
11	balita	dirinci	dokter	sebulan
12	lahir	hasil	medis	pasien
13	kontrasepsi	umur	keluarga	keluhan
14	kelahiran	vaksinasi	kerja	setahun
15	anak	pencapaian	ketersediaan	diare
16	hidup	target	unit	hiv
17	jenis	cakupan	posyandu	malaria
18	perempuan	kartu	klirik	aids
19	alat	realisasi	tidur	dbd
20	penolong	rembang	bayi	wabah
21	penduduk	sasaran	kunjungan	ims
22	buruk	tt	berencana	tb
23	keluarga	banggai	dukun	berobat
24	kematian	bitung	gigi	dirinci
25	berat	frekuensi	pembantu	rsud

Keterangan : Kata yang diblok warna merupakan kata parser terpilih

BIODATA PENULIS



Penulis bernama Tubagus Hendro Pramono, lahir di Tuban, Jawa Timur pada tanggal 22 Agustus 1997. Merupakan anak kedua dari dua bersaudara. Penulis telah menempuh pendidikan formal di SD Negeri 1 Sugihan, SMP Negeri 1 Jatirogo, dan SMA Negeri 1 Tuban.

Pada tahun 2014, penulis melanjutkan studi ke jenjang pendidikan yang lebih tinggi di Institut Teknologi Sepuluh Nopember sebagai mahasiswa Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi. Penulis terdaftar sebagai mahasiswa dengan nomor induk (NRP) 5214100013. Selama masa perkuliahan, penulis aktif di berbagai organisasi pada berbagai tingkatan di ITS mulai dari staf ahli Kajian Strategis HMSI ditingkat jurusan, staf ahli Student Research Development BEM FTIf ditingkat fakultas, hingga staf ahli Legislasi DPM ITS ditingkat institut. Penulis juga aktif diberbagai kepanitiaan sebagai staf / koordinator bagian keamanan dan perizinan. Selain itu, penulis juga pernah menjadi asisten dosen pada mata kuliah Desain Basis Data dan membantu dalam penelitian Egov-benchmarking Pemerintah Daerah Indonesia. Pada tahun ke-4, penulis menjadi asisten pada laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI) sekaligus mendalami topik *Open Data* sehingga penulis mengambil bidang minat laboratorium Akuisisi Data dan Diseminasi Informasi (ADDI).

Apabila terdapat pertanyaan mengenai Tugas Akhir ini, penulis dapat dihubungi melalui e-mail hendropramono14@gmail.com