



TUGAS AKHIR - KS184822

**KLASIFIKASI KATEGORI PENGADUAN
MASYARAKAT MELALUI KANAL LAPOR!
MENGUNAKAN *ARTIFICIAL NEURAL NETWORK*
(ANN)**

**MOCHAMAD IHSAN ANANTO
NRP 062115 4000 0123**

**Dosen Pembimbing
Dra. Wiwiek Setya Winahju, M.S.
Dr. Kartika Fithriasari, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**



TUGAS AKHIR - KS184822

**KLASIFIKASI KATEGORI PENGADUAN
MASYARAKAT MELALUI KANAL *LAPOR!*
MENGUNAKAN *ARTIFICIAL NEURAL NETWORK*
(ANN)**

**MOCHAMAD IHSAN ANANTO
NRP 062115 4000 0123**

**Dosen Pembimbing
Dra. Wiwiek Setya Winahju, M.S.
Dr. Kartika Fithriasari, M.Si.**

**PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**



FINAL PROJECT - KS184822

**Classification of Public Complaints Categories
Through LAPOR! CANAL using Artificial Neural
Network (ANN)**

**MOCHAMAD IHSAN ANANTO
NRP 062115 4000 0123**

Supervisors

Dra. Wiwiek Setya Winahju, M.S.

Dr. Kartika Fithriasari, M.Si.

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019**

(Halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN

KLASIFIKASI KATEGORI PENGADUAN MASYARAKAT MELALUI KANAL LAPOR! MENGUNAKAN ARTIFICIAL NEURAL NETWORK (ANN)

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh :

Mochamad Ihsan Ananto
NRP. 062115 4000 0123

Disetujui oleh Pembimbing:

Dra. Wiwiek Setya Winahju, M.S.

NIP. 19560424 198303 2 001

Dr. Kartika Fithriasari, M.Si.

NIP. 19691212 199303 2 002

(*Styawa*)
(*KF*)



Mengetahui,
Kepala Departemen Statistika

Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

(Halaman ini sengaja dikosongkan)

KLASIFIKASI KATEGORI PENGADUAN MASYARAKAT MELALUI KANAL LAPOR! MENGUNAKAN ARTIFICIAL NEURAL NETWORK (ANN)

Nama Mahasiswa : Mochamad Ihsan Ananto
NRP : 062115 4000 0123
Departemen : Statistika
Dosen Pembimbing : Dra. Wiwiek Setya Winahju, M.S.
Dr. Kartika Fithriasari, M.Si.

Abstrak

LAPOR! merupakan sarana aspirasi dan pengaduan masyarakat terkait kinerja pemerintah berbasis media sosial. Oleh karena laporan pengaduan masyarakat yang masuk tersebut berbentuk teks, maka dapat diselesaikan dengan cara text mining. Sehingga dilakukan analisis klasifikasi teks menggunakan Artificial Neural Network serta SMOTE untuk mengatasi data imbalance dan Chi-Square untuk proses seleksi variabel. Data yang digunakan adalah data historis aduan masyarakat melalui kanal LAPOR! tahun 2015. Melalui proses seleksi variabel, didapatkan sejumlah 428 term atau kata yang memberikan pengaruh terhadap kategori aduan masyarakat. Ketepatan klasifikasi yang dihasilkan melalui metode Artificial Neural Network dengan feature selection dan 3 nodes hidden layer adalah precision 0,794, sensitivity 0,818 dan F1-Score 0,800. Selain itu didapatkan topik permasalahan yang patut mendapatkan perhatian lebih pada kategori aduan energi, pangan dan maritim adalah kata raskin. Untuk kategori infrastruktur dan transportasi adalah aduan mengenai jalan. Pada kategori kesehatan, yakni kata BPJS. Untuk kategori pendidikan adalah kata terima yakni berkaitan dengan pembagian KIP. Lalu untuk reformasi birokrasi adalah KTP. Sedangkan untuk pariwisata dan lingkungan hidup adalah kata imigrasi.

Kata kunci: Artificial Neural Network, LAPOR!, SMOTE, Text Mining, Word Cloud

(Halaman ini sengaja dikosongkan)

CLASSIFICATION OF PUBLIC COMPLAINTS CATEGORIES THROUGH LAPOR! CANAL USING ARTIFICIAL NEURAL NETWORK (ANN)

Name : Mochamad Ihsan Ananto
Student Number : 062115 4000 0123
Department : Statistics
Supervisors : Dra. Wiwiek Setya Winahju, M.S.
Dr. Kartika Fithriasari, M.Si.

Abstract

LAPOR! is a media of aspirations and public complaints related to the performance of government based on social media. Because of public complaints received in the form of text, it can be solved by text mining. Therefore, the text classification analysis is used with application of Artificial Neural Network and SMOTE to overcome imbalance data and Chi-Square for the variable selection process. The data used is the historical data of public complaints through LAPOR! canal in 2015. Through variable selection process, obtained 428 terms or words which give effect to the category of public complaints. The performance measure of classification produced through the Artificial Neural Network method with a feature selection and 3 hidden layer nodes precision 0,794, sensitivity 0,818 dan F1-Score 0,800. In addition, the topic of the issue that deserves more attention in the category of energy, food and maritime complaints is the raskin word. For the category of infrastructure and transportation, there is a complaint regarding the road. In the health category, BPJS. For the education category, the word accept is related to the distribution of KIP. Then for bureaucratic reform is KTP. While for tourism and the environment is the word immigration

Keywords: *Artificial Neural Network, LAPOR!, SMOTE, Text Mining, Word Cloud*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan atas rahmat dan hidayah yang diberikan Allah SWT sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul “Klasifikasi Kategori Pengaduan Masyarakat Melalui Kanal LAPOR! menggunakan *Artificial Neural Network* (ANN)” dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Kedua orang tua yakni Bapak Sugih Nugroho, Ibu Sri Noviandri dan adik penulis Adinda Anugerah atas segala do'a, nasehat, kasih sayang, dan dukungan yang diberikan kepada penulis demi kesuksesan dan kebahagiaan penulis.
2. Bapak Dr. Suhartono selaku Ketua Departemen Statistika dan Dr. Santi Wulan Purnami, S.Si, M.Si. selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
3. Ibu Diaz Fitra Aksioma S.Si, M.Si, dan Ibu Ni Luh Putu Satyaning Pradnya Paramita S.Si., M.Si. selaku dosen-dosen wali selama masa studi yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika.
4. Ibu Dra. Wiwiek Setya Winahju, M.S dan Ibu Dr. Kartika Fithriasari, M.Si. selaku dosen pembimbing yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan serta motivasi selama penyusunan Tugas Akhir.
5. Bapak Prof. Drs. Nur Iriawan, MIKom., Ph.D. dan Bapak Dr. R. Mohamad Atok, S.Si., M.Si. selaku dosen penguji yang sabar dalam memberikan komentar serta masukan dan saran dalam penyelesaian Tugas Akhir.
6. Seluruh dosen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS.

7. Teman-teman Vivacious Statistika ITS Σ 26 angkatan 2015, yang selalu memberikan dukungan serta mengisi kehidupan penulis selama kuliah di Departemen Statistika.
8. Devita Prima Vernanda yang selalu menemani penulis selama masa studi serta memberikan motivasi dan semangat untuk menyelesaikan tugas akhir ini.
9. Teman-teman kos Bu Aisyah terutama Haekal, Ifthor, dan Hafizh dan Alumni SMAN 47 2015 Arya, Eko, Hamka, Ghiyats, Irfan, Basyar, Daffa, Panji, dan Triyoga sebagai teman penulis di awal hidup perantauan.
10. Teman-teman kepanitiaan terutama PR ITS EXPO yang telah memberikan penulis banyak pengalaman.
11. Teman-teman organisasi BEM, HIMASTA-ITS yang telah memberikan penulis banyak pengalaman organisasi
12. Teman-teman Jalanjalanmen dan Keluarga Cemarrah yang memberikan suntikan semangat kepada penulis saat kembali ke kampung halaman.
13. Serta semua teman, relasi dan berbagai pihak yang tidak bisa penulis sebutkan namanya satu persatu yang telah membantu dalam penulisan laporan ini.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Juli 2019

Penulis

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	iii
Abstrak	v
Abstract	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 <i>Text Mining</i>	7
2.2 <i>Term Frequency Inverse Document Frequency</i>	8
2.3 <i>K-fold Cross Validation</i>	8
2.4 <i>Feature Selection</i>	9
2.5 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	10
2.6 <i>Artificial Neural Network</i>	11
2.7 Ketepatan Klasifikasi.....	13
2.8 <i>Word Cloud</i>	15
2.9 LAPOR!.....	15
BAB III METODOLOGI PENELITIAN	17
3.1 Sumber Data.....	17
3.2 Variabel Penelitian & Struktur Data.....	17
3.3 Langkah Analisis.....	18
3.4 Diagram Alir.....	20
BAB IV ANALISIS DAN PEMBAHASAN	23
4.1 Karakteristik Data Aduan Masyarakat.....	23

4.2	Klasifikasi Aduan Masyarakat menggunakan <i>Artificial Neural Network</i> (ANN).....	24
4.2.1	<i>Pre-Processing Data</i>	25
4.2.2	<i>Oversampling Data</i> dengan <i>Synthetic Minority Oversampling Technique</i> (SMOTE)	29
4.2.3	Ketepatan Klasifikasi Menggunakan <i>Artificial Neural Network</i>	30
4.3	Visualisasi <i>Word Cloud</i>	35
BAB V KESIMPULAN DAN SARAN		43
5.1	Kesimpulan	43
5.2	Saran	44
DAFTAR PUSTAKA		45
LAMPIRAN		49

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Pembagian Data	9
Gambar 2.2 Arsitektur ANN.....	11
Gambar 2.3 Visualisasi Data Teks dengan Word Cloud	15
Gambar 2.4 Alur Kerja LAPOR!	16
Gambar 3.1 Diagram Alir	20
Gambar 4.1 Jumlah Aduan Masyarakat Tiap Kategori.....	23
Gambar 4.2 Persentase Status Aduan Masyarakat.....	24
Gambar 4.3 Jaringan ANN 428 feature dan 3 nodes hidden layer	33
Gambar 4.4 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Energi Pangan dan Maritim.....	36
Gambar 4.5 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Infrastruktur dan Transportasi.....	37
Gambar 4.6 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Kesehatan	38
Gambar 4.7 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Pendidikan.....	39
Gambar 4.8 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Reformasi Birokrasi	40
Gambar 4.9 (a) <i>Word Cloud</i> (b) <i>Pareto Chart</i> Kategori Aduan Pariwisata dan Lingkungan Hidup	41

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> Untuk <i>Binary Classification</i>	14
Tabel 2.2 <i>Confusion Matrix</i>	14
Tabel 3.1 Variabel Penelitian	17
Tabel 3.2 Struktur Data Awal	18
Tabel 4.1 Contoh Data Sebelum dan Sesudah <i>Case Folding</i> dan <i>Cleaning</i>	25
Tabel 4.2 Contoh Data Sebelum dan Sesudah Tahap <i>Stemming</i>	26
Tabel 4.3 Contoh Data Sebelum dan Sesudah Tahap <i>Stopwords Removal</i>	27
Tabel 4.4 Contoh Data Sebelum dan Sesudah Tahap <i>Tokenizing</i>	27
Tabel 4.5 <i>Count Vectorizer</i> kata dalam aduan	28
Tabel 4.6 Nilai χ^2 untuk setiap kata	29
Tabel 4.7 Jumlah Data <i>Training</i>	30
Tabel 4.8 Hasil Ketepatan Klasifikasi dengan ANN	31
Tabel 4.9 <i>Confusion Matrix</i>	32
Tabel 4.10 Ketepatan Klasifikasi Tiap Kategori	32
Tabel 4.11 Ketepatan Klasifikasi dengan <i>k-fold</i>	34

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1. Data aduan masyarakat via kanal LAPOR! 2015..	49
Lampiran 2. TF-IDF dari Kata Kunci.....	52
Lampiran 3. Data hasil <i>tokenizing</i>	53
Lampiran 4. Syntax <i>pre-processing</i>	54
Lampiran 5. Syntax <i>Count Vectorized</i> dan TF-IDF.....	57
Lampiran 6. Syntax <i>Splitting Data</i> dan SMOTE.....	57
Lampiran 7. Syntax <i>Artificial Neural Network</i>	58
Lampiran 8. Syntax <i>Artificial Neural Network</i> dengan <i>K-fold</i> ..	59
Lampiran 9. Syntax <i>Word Cloud</i>	60
Lampiran 10. Output <i>Weight and Bias</i>	60
Lampiran 11. Hasil <i>Chi Square</i>	61
Lampiran 12. Surat Keterangan Pengambilan Data	62

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi terutama internet, telah berkembang sangat pesat. Berdasarkan laporan Wearesocial jumlah pengguna internet dunia saat ini telah mencapai 4,021 miliar orang yang berarti setengah dari total manusia di bumi telah menggunakan internet. Bahkan menurut Kementerian Komunikasi dan Informatika Republik Indonesia, pengguna internet di Indonesia pada tahun 2016 mencapai 132,7 juta orang dari total penduduk Indonesia sebanyak 256,2 juta. Hal ini menunjukkan bahwa perkembangan teknologi informasi khususnya internet memberikan manfaat bagi manusia karena fungsinya sebagai sumber informasi serta sarana bersosialisasi satu sama lain yang dilakukan secara *online*, sehingga memungkinkan manusia untuk saling berinteraksi. Manusia dapat berbagi informasi, berbagi pengalaman, mempromosikan sesuatu, dan berbagai hal lainnya satu sama lain meskipun berada di lokasi yang berjauhan ataupun perbedaan waktu sekalipun. Tak hanya bagi individu, banyak pihak di berbagai sektor telah memanfaatkan internet untuk berbagai tujuan tidak terkecuali sektor pemerintahan.

Sektor pemerintahan memanfaatkan internet sebagai upaya untuk meningkatkan kualitas layanan publik bagi masyarakat. Perkembangan teknologi informasi mendesak pemerintah untuk memberikan kebebasan bagi masyarakat dalam mendapatkan informasi terkait kebijakan serta kinerja pemerintah. Salah satu bentuk teknologi informasi yang mendukung pemanfaatan internet adalah penggunaan aplikasi *electronic government (e-government)*. *E-government* bertujuan untuk membentuk sistem pelayanan pemerintahan yang lebih mudah dan cepat dengan menggunakan situs internet. Berbagai negara di belahan dunia saat ini telah mengimplementasikan *e-government* dengan strategi yang disesuaikan dengan kondisi sosial politik serta geografisnya masing-masing. Berdasarkan Survey *e-government* PBB Tahun 2018, Denmark menduduki pe-

ringkat pertama sebagai negara dengan Indeks *e-government* tertinggi, sedangkan Indonesia berada di peringkat 107 dari 193 negara anggota PBB. Penerapan *e-government* sendiri dapat memudahkan penyebaran informasi sekaligus mendorong masyarakat untuk dapat berpartisipasi aktif dalam memberikan masukan untuk terwujudnya pelayanan publik yang prima sebagaimana diatur dalam UU No. 14 Tahun 2008 tentang Keterbukaan Informasi Publik (KIP). Untuk mewujudkan hal tersebut maka sesuai UU No. 3 Tahun 2015 pemerintah Indonesia membentuk Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) yakni integrasi pengelolaan pengaduan pelayanan publik secara berjenjang. Melalui SP4N ini pengaduan masyarakat mengenai pelayanan publik diharapkan dapat ditangani dengan cepat, transparan, dan akuntabel sesuai dengan kewenangan instansi terkait sehingga dapat mendorong peningkatan kinerja pengelolaan pengaduan masyarakat terkait pelayanan publik.

Saat ini SP4N hadir dalam bentuk LAPOR! yakni sarana aspirasi dan pengaduan berbasis media sosial. LAPOR! dikelola dan dikembangkan oleh Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (PAN-RB) bersama Kementerian Dalam Negeri, Kantor Staf Presiden, dan Ombudsman Republik Indonesia sejak Maret 2016. LAPOR! hadir sebagai saluran partisipasi masyarakat untuk pengawasan pembangunan dan pelayanan publik di Indonesia. Dimana masyarakat dapat melakukan pengaduan dengan layanan LAPOR! melalui website www.lapor.go.id, SMS di 1708, aplikasi LAPOR! dan twitter @LAPOR1708. Aduan dari masyarakat yang berupa laporan tersebut kemudian akan diverifikasi terlebih dahulu oleh administrator LAPOR! dan selanjutnya diteruskan ke instansi terkait paling lambat 3 hari kerja setelah laporan dilakukan. Nantinya setiap aduan dari masyarakat akan dipublikasikan pada situs LAPOR! dan dilengkapi dengan *Tracking ID* yang dapat digunakan pengguna untuk melakukan penelusuran atas suatu laporan sudah sejauh apa telah ditindaklanjuti. Hingga Maret 2017, LAPOR! telah digunakan lebih dari 560.000 pengguna dan menerima rata-rata lebih dari 800 laporan masyarakat per

harinya dan telah terpadu dengan 34 Kementerian, 24 Pemprov, 94 Pemkab, 33 Pemkot, 149 Pemda, 128 PTN dan Kopertis, 116 BUMN, serta 131 Perwakilan RI di luar negeri (LAPOR!, 2017). Hal ini membuat LAPOR! telah menjadi suatu sarana efektif dan tepat guna bagi masyarakat untuk menyampaikan segala bentuk aspirasi dan keluhan karena dapat dibaca serta ditangan secara langsung oleh pembuat kebijakan terkait. Nantinya pengaduan masyarakat tersebut akan terkategori menjadi beberapa kategori yang dapat ditindaklanjuti oleh instansi terkait.

Saat ini LAPOR! masih belum mampu untuk mengetahui kategori laporan yang diprioritaskan atau laporan yang paling mendapat sorotan dari masyarakat via SMS. Hal ini berdampak terhadap tindak lanjut laporan yang tidak dapat dilakukan secara cepat dan efisien. Oleh karena laporan pengaduan masyarakat yang masuk tersebut berbentuk teks, maka dapat diselesaikan dengan cara *text mining*. *Text mining* dapat diartikan sebagai suatu proses untuk menggali informasi secara intensif dimana pengguna berhadapan dengan sekumpulan dokumen menggunakan *tools* analisis *data mining* (Feldman & Sanger, 2007). Proses yang dapat dilakukan dengan *text mining* di antaranya adalah *text clustering* dan *text classification*. *Text clustering* berhubungan dengan penemuan struktur kelompok yang belum terlihat atau *unsupervised learning* dari sekumpulan dokumen. Sedangkan *text classification* merupakan proses untuk membentuk kelas-kelas dari dokumen berdasarkan pada kelas kelompok yang sudah diketahui sebelumnya atau *supervised learning*. (Darujati & Gumelar, 2012). Sebelum dilakukan analisis, teks harus melalui *preprocessing data* terlebih dahulu yakni dengan melakukan *data cleaning*, *case folding* (mengubah teks menjadi non kapital), *stemming* (menghilangkan imbuhan), *tokenizing* (memecah keseluruhan teks menjadi kata per kata), dan terakhir mendapatkan kata kunci untuk dilanjutkan ke analisis.

Berdasarkan permasalahan yang ada dalam LAPOR!, maka klasifikasi teks merupakan proses yang tepat karena data aduan masyarakat yang masuk akan terklasifikasikan dalam kategori ter-

tentu. Beberapa metode yang sering digunakan dalam klasifikasi teks diantaranya adalah *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Neural Networks*. Penelitian sebelumnya mengenai klasifikasi teks lewat aduan LAPOR! pernah dilakukan oleh Megawati (2015) dengan menggunakan algoritma SVM. Penelitian tentang metode ANN pernah dilakukan oleh Reyhana & Fithriasari (2018) tentang analisis sentimen perkembangan infrastruktur di Kota Surabaya. Pada penelitian tersebut didapatkan hasil performa klasifikasi melalui metode ANN lebih baik dari SVM.

1.2 Rumusan Masalah

Melalui kanal LAPOR! masyarakat dapat melakukan pelaporan berupa pengaduan, saran, hingga rekomendasi terkait kebijakan dan kinerja pemerintah di lingkungan mereka. Tentunya laporan-laporan tersebut harus dengan cepat ditangani sehingga upaya perbaikan pelayanan publik dapat terwujud. Sehingga perlu adanya cara untuk mengklasifikasikan laporan masyarakat yang berupa dokumen teks secara otomatis. Oleh karena itu pada penelitian ini akan dilakukan *text mining* dengan salah satu metode klasifikasi teks dari *Neural Networks* yakni *Artificial Neural Networks*. Maka permasalahan utama yang akan dibahas dalam penelitian ini adalah bagaimana karakteristik dari data aduan masyarakat yang masuk via LAPOR!, berapa tingkat ketepatan klasifikasi dari model berdasarkan *data training* dengan menggunakan *Artificial Neural Networks* (ANN), dan kata kunci apa yang paling sering dilaporkan masyarakat.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan karakteristik aduan masyarakat melalui kanal LAPOR!
2. Mendapatkan hasil ketepatan klasifikasi aduan masyarakat menggunakan *Artificial Neural Network* (ANN)

3. Mendapatkan kata kunci aduan masyarakat di masing-masing kategori yang divisualisasikan dengan *word cloud*.

1.4 Manfaat Penelitian

Melalui penelitian ini diharapkan dapat memberikan bahan masukan tambahan bagi pihak LAPOR! terkait kategori klasifikasi aduan yang masuk skala prioritas nasional. Nantinya hasil dari klasifikasi aduan masyarakat diharapkan dapat mempercepat aduan oleh instansi terkait karena telah mendapatkan model dari *data training*. Sehingga pemerintah dapat memberikan pelayanan yang lebih baik dan efisien.

1.5 Batasan Masalah

Dalam penelitian ini, batasan masalah yang digunakan adalah kategori aduan masyarakat yang masuk via LAPOR! Dikategorikan ulang kedalam kategori aduan prioritas nasional yakni energi pangan dan maritim, infrastruktur dan transportasi, kesehatan, pendidikan, reformasi birokrasi, serta pariwisata dan lingkungan hidup. Data yang digunakan adalah data historis aduan masyarakat melalui kanal LAPOR! tahun 2015 yang diunggah oleh Kantor Staf Kepresidenan di situs data.go.id.

(Halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

2.1 *Text Mining*

Text mining dapat diartikan sebagai suatu proses menggali informasi secara intensif dimana pengguna berhadapan dengan sekumpulan dokumen teks dengan menggunakan *tools* analisis yang berhubungan dengan *data mining*. *Text mining* bertujuan untuk mengekstrak informasi yang berguna dokumen berupa teks yang tidak terstruktur melalui indentifikasi dan eksplorasi suatu pola menarik. Adapun tugas khusus dari *text mining* antara lain yaitu klasifikasi teks (*text classification*) dan pengelompokan teks (*text clustering*) (Feldman & Sanger, 2007).

Text clustering berhubungan dengan penemuan struktur kelompok yang belum terlihat atau *unsupervised learning* dari sekumpulan dokumen. Sedangkan *text classification* merupakan proses untuk membentuk kelas-kelas dari dokumen berdasarkan kelas kelompok yang sudah diketahui sebelumnya atau *supervised learning* (Darujati & Gumelar, 2012).

Text pre-processing merupakan tahap awal dalam *text mining* sebelum analisis lebih lanjut akan dilakukan. Data teks harus melalui *pre-processing* terlebih dahulu agar data yang tidak terstruktur menjadi lebih terstruktur (Feldman & Sanger, 2007). Beberapa *pre-processing* untuk data teks yakni dengan melakukan *data cleaning*, *case folding*, *stemming*, *tokenizing*, dan *stopwords removal*.

1. *Case Folding*, yaitu merupakan proses untuk mengubah semua karakter teks menjadi non kapital serta menghilangkan tanda baca dan angka (Weiss, 2010).
2. *Data Cleaning*, yaitu membersihkan data teks dari kata yang tidak diperlukan untuk mengurangi *noise*. Kata yang dihilangkan dalam dokumen teks antara lain karakter HTML, *emojis*, *hashtag* (#), dan URL.

3. *Stemming*, yakni suatu proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, serta kombinasi dari awalan dan akhiran.
4. *Tokenizing*, yakni proses untuk memecah keseluruhan teks yang sebelumnya berupa kalimat menjadi kata per kata
5. *Stopwords Removal*, merupakan proses untuk menghapus kata-kata yang tidak memiliki arti yang relevan yang berakhir sebagai pemisahan kata per kata. Sehingga akan ada kata yang tidak memiliki arti sesuai yang dihilangkan untuk mendapatkan ciri dari dokumen. Contoh kata yang perlu dihilangkan yakni seperti “ini, itu, dan, atau” dan banyak lagi kata-kata sejenis.

2.2 Term Frequency Inverse Document Frequency

Term Frequency Inverse Document Frequency (TF-IDF) merupakan metode pembobotan yang dilakukan untuk ekstraksi data teks. Tujuan TF-IDF adalah menemukan jumlah kata yang diketahui (tf) setelah dikalikan dengan frekuensi aduan dimana suatu kata tersebut muncul (idf). Metode TF-IDF dilakukan melalui perhitungan bobot dengan cara integrasi antara *term frequency* (tf) dan *inverse document frequency* (idf) (Rahman, 2018). Berikut merupakan rumus untuk memperoleh nilai TF-IDF.

$$w_{ij} = tf_{ij} \times idf_j, \quad (2.1)$$

$$idf_i = \log \left(\frac{N}{df_i} \right), \quad (2.2)$$

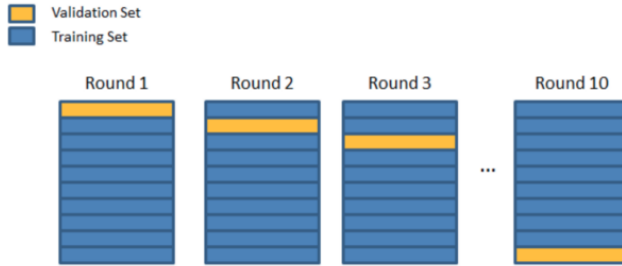
keterangan:

- N = jumlah keseluruhan aduan,
 tf_{ij} = jumlah munculnya kata i pada ulasan j ,
 df_i = banyaknya aduan yang mengandung kata I ,

2.3 K-fold Cross Validation

Metode *K-fold cross validation* sudah sering digunakan untuk mengevaluasi kinerja klasifikasi. *K-fold cross validation* digunakan untuk mengurangi bias terkait pengambilan sampel dari data dengan secara acak membagi data kedalam sejumlah bagian yang

telah ditentukan yang disebut *fold* (Gokgoz & Subasi, 2015). K merupakan besar angka partisi data yang digunakan untuk pembagian *data training* dan *data testing*. Ilustrasi pembagian data menggunakan *K-fold cross validation* terdapat pada Gambar 2.1



Gambar 2.1 Ilustrasi Pembagian Data
(Sumber: towardsdatascience.com)

Dalam beberapa kasus, variasi dalam *K-fold cross validation* dapat dilakukan. Salah satunya adalah menjaga agar setiap *fold* mengandung persentase sampel yang sama dari setiap kelas. Variasi ini juga dikenal sebagai *Stratified K Fold Cross Validation*.

2.4 Feature Selection

Feature selection atau seleksi variabel merupakan proses pemilihan variabel yang relevan untuk digunakan dalam pembentukan model. Keuntungan dari pemilihan atribut adalah untuk mempercepat proses sehingga nantinya akan didapatkan performa yang lebih tinggi (Meesad, Boonrawd, & Nui pian, 2011). *Feature selection* yang digunakan dalam penelitian ini adalah *Chi-Square* (χ^2) karena menghasilkan performa yang bagus terutama untuk data *multiclass* (Bachri, Kusnadi, & Nurhayati, 2017), dengan hipotesis sebagai berikut.

$H_0: \pi_{kl} = \pi_{k+}\pi_{+l}$ (Tidak ada hubungan antar variabel)

$H_1: \pi_{kl} \neq \pi_{k+}\pi_{+l}$ (Ada hubungan antar variabel)

$$\chi^2 = \sum_{k=1}^r \sum_{l=1}^c \frac{(n_{kl} - \hat{\mu}_{kl})^2}{\hat{\mu}_{kl}}, \quad (2.3)$$

keterangan:

- n_{kl} = Frekuensi pada sel baris ke- k dan kolom ke- l ,
 $\hat{\mu}_{kl}$ = Frekuensi harapan pada sel baris ke- k dan kolom ke- l ,
 π_{k+} = Total pada baris ke- k ,
 π_{+l} = Total pada kolom ke- l ,
 r = jumlah baris,
 c = jumlah kolom.

Jumlah derajat bebas yang digunakan diperoleh dengan mengurangi jumlah kelas target dengan 1 ($df = n - 1$). Jika $\chi^2 > \chi_{\alpha}^2$ dengan ($df = n - 1$), maka tolak H_0 pada tingkat signifikansi yang digunakan. Jika sebaliknya, maka gagal tolak H_0 . (Bachri, Kusnadi, & Nurhayati, 2017).

2.5 *Synthetic Minority Over-sampling Technique (SMOTE)*

Synthetic Minority Over-sampling Technique (SMOTE) merupakan salah satu metode untuk mengatasi data *imbalance* yang diusulkan oleh (Chawla, 2002). Ide dasar dari SMOTE yaitu menambah jumlah sampel pada kelas minor agar memiliki jumlah data yang setara dengan kelas mayor. Hal ini dilakukan dengan cara membangkitkan data sintesis berdasarkan tetangga terdekat *k-nearest neighbour* dimana tetangga terdekat dipilih berdasarkan jarak *euclidean* antara kedua data (Chawla, 2002). Pembangkitan data sintesis dilakukan dengan menggunakan persamaan berikut:

$$\mathbf{x}_{syn} = \mathbf{x}_i + (\mathbf{x}_{knn} - \mathbf{x}_i)\gamma \quad , \quad (2.4)$$

keterangan:

- \mathbf{x}_{syn} = data sintesis,
 \mathbf{x}_i = data ke- i dari kelas minor,
 \mathbf{x}_{knn} = data dengan jarak terdekat dari data yang akan direplikasi,
 γ = bilangan acak antara 0 – 1.

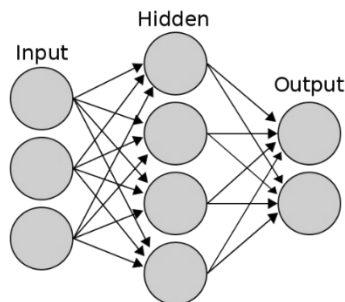
Pada prosedur *k-fold cross validation*, SMOTE dilakukan pada masing-masing data *training* di tiap *fold*, untuk menghindari

kemungkinan terdapat data replikasi atau data dengan pola yang sama pada data *training* dan *testing* (Santos, 2018)

2.6 Artificial Neural Network

Artificial Neural Network (ANN) atau bisa disebut sebagai Jaringan Syaraf Tiruan (JST) merupakan cabang dari ilmu kecerdasan buatan (*artificial intelligence*). ANN merupakan salah satu sistem pemrosesan informasi dengan desain menirukan cara kerja otak manusia. Pembuatan ANN terinspirasi dari *complex learning system* pada otak yang terdiri dari set-set neuron yang saling berhubungan secara dekat (Meinanda dalam Damayanti, 2018)

Kelebihan dari pemodelan ANN diantaranya; tidak memerlukan banyak asumsi, dapat memodelkan sistem secara non-linier dengan baik, dan mampu memberikan model yang mendekati sistem nyata (Astuti, 2017). Arsitektur *Neural Network* yang tergambar melalui Gambar 2.2, terdiri atas 3 bagian yakni *Input Layer* yaitu bagian yang menerima masukan langsung dari lingkungan, *Hidden Layer* yaitu bagian tersembunyi, dan *Output Layer* yaitu bagian yang menghasilkan *output* akhir dari jaringan ANN. Arsitektur tersebut disebut juga *Multi Layer Perceptron* (MLP) atau *Fully-Connected Layer*.



Gambar 2.2 Arsitektur ANN

(Sumber: Wikipedia)

Neural network adalah kumpulan dari *input/output* yang terhubung dimana setiap sambungan mempunyai bobot (*weight*). Jaringan tersebut menyesuaikan bobot sehingga dapat memprediksi

label kelas yang benar dari *input*. Algoritma backpropagation menunjukkan pembelajaran *multilayer feed-forward neural network* yang terdiri dari lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output (Han & Kamber, 2006).

Setiap *input* yang terhubung ke setiap *nodes* pada *hidden layer* maupun *output layer* masing-masing memiliki bias dan *weight*. Nantinya akan ada perhitungan pada *nodes hidden layer* tanpa fungsi aktivasi sebagaimana persamaan (2.5) dan setelah fungsi aktivasi yang dijelaskan melalui persamaan (2.6). Selanjutnya dilakukan perhitungan untuk mendapatkan nilai prediksi yang dijelaskan melalui persamaan (2.7) dan (2.8).

$$p_j = a_j + \sum_{i=1}^{n_x} w_{ij}x_i, \quad (2.5)$$

$$q_j = f(p_j), \quad (2.6)$$

$$r_l = b_l + \sum_{j=1}^{n_h} v_{jl}q_j, \quad (2.7)$$

$$s_l = f(r_l), \quad (2.8)$$

dengan :

x_i = variabel input,

a_j = nilai bias pada *hidden layer*,

w_{ij} = nilai pembobot pada *hidden layer*,

b_l = nilai bias pada *output layer*,

v_{jl} = nilai pembobot pada *output layer*,

n_x = banyak variabel *input* yang digunakan,

n_h = banyak *nodes* pada *hidden layer*,

p_j = *output* setiap *nodes hidden layer* tanpa fungsi aktivasi,

q_j = *output* setiap *nodes hidden layer* setelah dimasukkan dalam fungsi aktivasi,

p_j = *output* setiap *nodes output layer* tanpa fungsi aktivasi,

q_j = *output* setiap *nodes output layer* setelah dimasukkan dalam fungsi aktivasi.

Fungsi Aktivasi merupakan fungsi yang digunakan dalam *neural networks* untuk menghitung *weight* dan *bias*. Fungsi aktivasi juga menggambarkan hubungan antara *input* untuk mengeluarkan nilai *output* yang dapat berbentuk *linear* ataupun *non-linear*. (Nwankpa dkk, 2018). Beberapa fungsi aktivasi yang digunakan yakni *Rectified Linear Unit* (ReLU) dan *Softmax*.

ReLU merupakan fungsi aktivasi yang sering digunakan. ReLU memaksa elemen input yang kurang dari 0 ke nilai 0 yang ditunjukkan melalui persamaan (2.9).

$$q_j = f(p_j) = \max(0, p_j) = \begin{cases} p_j, & \text{jika } p_j \geq 0, \\ 0, & \text{jika } p_j < 0, \end{cases} \quad (2.9)$$

dengan:

p_j = *input* pada setiap *nodes* pada persamaan (2.5),

q_j = *output* pada setiap *nodes* setelah dimasukkan fungsi aktivasi,

j = banyak *nodes* pada *layer*.

Fungsi aktivasi *Softmax* merupakan tipe lain dari fungsi aktivasi yang menghasilkan *output* dengan nilai di antara 0 dan 1. *Softmax* digunakan dalam kasus *multiclass* dengan hasil merupakan probabilitas untuk setiap kelasnya dengan kelas target mempunyai probabilitas tertinggi.

$$s_j = f(r_j) = \frac{\exp(r_j)}{\sum_{i=1}^k \exp(r_i)}, \quad (2.10)$$

dengan:

r_l = *input* pada setiap *nodes* pada persamaan (2.7),

s_l = *output* pada setiap *nodes* setelah dimasukkan fungsi aktivasi,

k = banyak kelas yang digunakan.

2.7 Ketepatan Klasifikasi

Pengukuran ketepatan klasifikasi dilakukan untuk melihat performa klasifikasi yang telah dilakukan. Data aktual dan data hasil prediksi dari model klasifikasi disajikan dengan menggunakan *Confusion matrix* yang mengandung informasi tentang kelas data

yang aktual direpresentasikan pada baris matriks dan kelas data hasil prediksi pada kolom (Khaulasari, 2016).

Tabel 2.1 *Confusion Matrix Untuk Binary Classification*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	<i>tp</i>	<i>fn</i>
Negatif	<i>fp</i>	<i>tn</i>

Ketepatan klasifikasi dapat dievaluasi dengan menghitung jumlah kelas positif yang terklasifikasi dengan benar (*true positive*), jumlah kelas negatif yang terklasifikasi dengan benar (*true negative*), jumlah kelas negatif yang salah terklasifikasi ke dalam kelas positif (*false positive*) atau jumlah kelas positif yang salah terklasifikasi ke dalam kelas negatif (*false negative*). Keempat perhitungan ini dapat dilihat melalui *confusion matrix* untuk kasus *binary classification* yang terdapat pada Tabel 2.1 (Sokolova & Lapalme, 2009). Sedangkan untuk kasus *multiclass classification* seperti pada penelitian ini dapat dilihat melalui Tabel 2.2.

Tabel 2.2 *Confusion Matrix*

Kelas Aktual	Kelas Prediksi						Total
	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	
C ₁	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	n ₁
C ₂	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅	X ₂₆	n ₂
C ₃	X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅	X ₃₆	n ₃
C ₄	X ₄₁	X ₄₂	X ₄₃	X ₄₄	X ₄₅	X ₄₆	n ₄
C ₅	X ₅₁	X ₅₁	X ₅₁	X ₅₁	X ₅₁	X ₅₆	n ₅
C ₆	X ₆₁	X ₆₂	X ₆₃	X ₆₄	X ₆₅	X ₆₆	n ₃₆
Total	n ₁	n ₂	n ₃	n ₄	n ₅	n ₆	N _{total}

Penilaian ketepatan klasifikasi untuk *multiclass classification* didefinisikan perkelas secara individual. Kualitas keseluruhan klasifikasi dapat dinilai dengan dua cara yakni melalui rata-rata dari tiap kriteria yang dihitung untuk setiap kelasnya (*macro-averaging*) atau jumlah penghitungan untuk mendapatkan kumulatif *tp*; *fn*; *tn*; *fp* (*micro-averaging*) (Sokolova & Lapalme, 2009). Berikut merupakan beberapa kriteria untuk menilai ketepatan klasifikasi.

$$\text{Precision} = \frac{\sum_{a=1}^6 tp_a}{\sum_{a=1}^6 (tp_a + fp_a)}, \quad (2.11)$$

$$\text{Sensitivity} = \frac{\sum_{a=1}^6 tp_a}{\sum_{a=1}^6 (tp_a + fn_a)}, \quad (2.12)$$

$$F1 \text{ Score} = 2 \times \left(\frac{(\text{Precision} \times \text{Sensitivity})}{(\text{Precision} + \text{Sensitivity})} \right). \quad (2.13)$$

2.8 Word Cloud

Word cloud merupakan salah satu metode visualisasi dokumen teks yang sering digunakan. *Word cloud* merupakan representasi grafis dari dokumen teks dengan melakukan *plotting* kata-kata yang sering muncul kedalam ruang dua dimensi. Melalui *word cloud*, dapat diketahui seberapa besar frekuensi dari kata yang muncul melalui besar kecilnya ukuran huruf kata tersebut. Semakin besar ukuran kata, maka semakin besar frekuensi kata tersebut muncul dalam dokumen. (Castella & Sutton, 2013). Contoh dari visualisasi dokumen teks dengan *word cloud* ditunjukkan pada Gambar 2.3.



Gambar 2.3 Visualisasi Data Teks dengan Word Cloud
(Sumber: Castella & Sutton, 2013)

2.9 LAPOR!

Layanan Aspirasi dan Pengaduan Online Rakyat (LAPOR!) merupakan sarana aspirasi dan pengaduan berbasis media sosial. LAPOR! dikelola dan dikembangkan oleh Kementerian Pendaya-

gunaan Aparatur Negara dan Reformasi Birokrasi (PAN-RB) bersama Kementerian Dalam Negeri, Kantor Staf Presiden, dan Ombudsman Republik Indonesia sejak Maret 2016.

Berdirinya LAPOR! merupakan amanat dari UU No. 3 Tahun 2015 untuk membentuk Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) yakni integrasi pengelolaan pengaduan pelayanan publik secara berjenjang. Melalui SP4N pengaduan masyarakat mengenai pelayanan publik diharapkan dapat ditangani dengan cepat, transparan, dan akuntabel sesuai dengan kewenangan instansi terkait. Masyarakat dapat menyampaikan aduan via LAPOR! Melalui website www.lapor.go.id, SMS 1708, aplikasi LAPOR! dan twitter @LAPOR1708. Alur kerja dari LAPOR! dijelaskan melalui Gambar 2.4.



Gambar 2.4 Alur Kerja LAPOR!
(Sumber: lapor.go.id)

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapat melalui situs data.go.id yang diunggah oleh admin dari LAPOR!. Data merupakan rekapitulasi aduan masyarakat terhadap kinerja pemerintah di berbagai daerah pada tahun 2015 yang dilaporkan melalui kanal LAPOR!. Total jumlah data yang digunakan dalam penelitian ini adalah sebanyak 6916 aduan yang terbagi kedalam enam kategori yakni energi pangan dan maritim, infrastruktur dan transportasi, kesehatan, pendidikan, reformasi birokrasi, serta pariwisata dan lingkungan hidup sebagaimana terlampir pada Lampiran 1.

3.2 Variabel Penelitian & Struktur Data

Pada penelitian ini terdapat dua variabel yang digunakan setelah tahap *text pre-processing*. Yakni terdiri dari variabel prediktor (x) yaitu bobot dari kata dasar pada setiap aduan dan variabel respon (y) yaitu klasifikasi kategori aduan sebagaimana dijelaskan melalui Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Skala
	Kategori aduan	
	0 = energi pangan dan maritim	
	1 = infrastruktur dan transportasi	
y	2 = kesehatan	Nominal
	3 = pendidikan	
	4 = reformasi birokrasi	
	5 = pariwisata dan lingkungan hidup	
x	Bobot kata ke-j yang muncul pada aduan	Rasio

Struktur data awal yang memuat isi laporan dan kategori aduan sebelum dilakukan tahapan *text pre processing* ditunjukkan melalui Tabel 3.2.

Tabel 3.2 Struktur Data Awal

TrackingID	Isi Aduan	Kategori
1301625	Ada 2 lubang tengah jalan bahayakan pengendara bermotor di jl.k.h.ahmad dahlan	Infrastruktur
1302327	Saya ingin memberi masukan mengenai gunungan sampah yang menumpuk di Pasar Induk Kramat Jati	Pariwisata & Lingkungan Hidup
:	:	:
1300415	Saya ingin melaporkan gas elpiji 3 kg di daerah kami Kec. Kayen langka sejak 5 bulan lalu	Energi pangan dan maritim

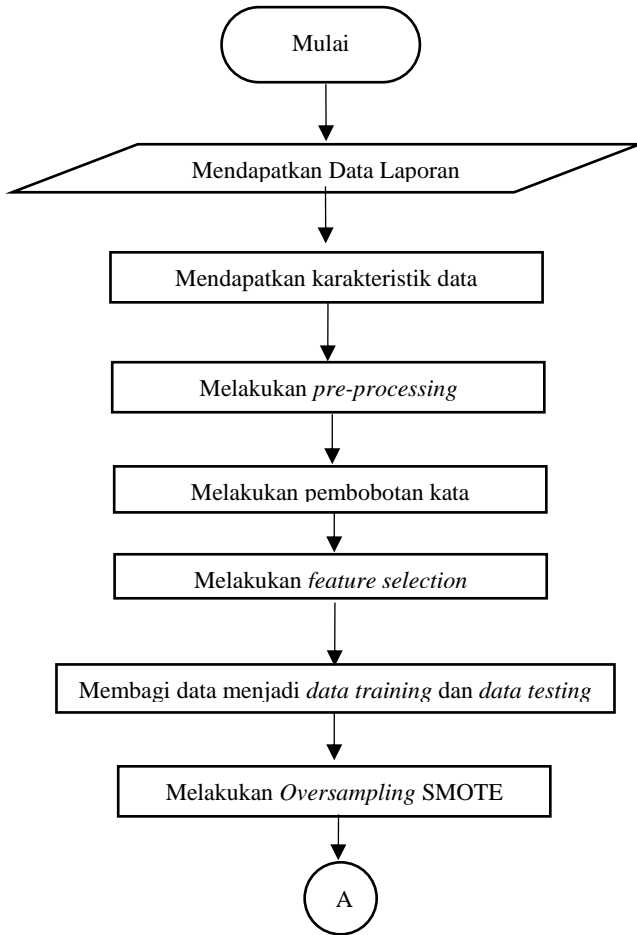
3.3 Langkah Analisis

Langkah analisis yang akan dilakukan pada penelitian ini adalah sebagai berikut.

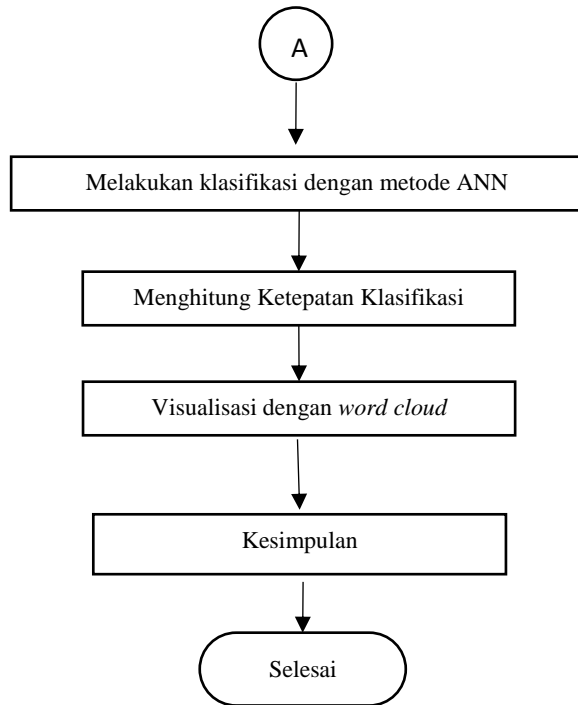
1. Mengambil data aduan masyarakat melalui LAPOR! dari situs data.go.id dan disimpan ke dalam *database*. Lalu dilihat karakteristik dari data tersebut.
2. Melakukan *pre-processing data* yakni dengan melakukan;
 - a. *Data cleaning*, yaitu membersihkan data teks dari kata yang tidak diperlukan untuk mengurangi *noise*. Yakni dengan cara menghapus tautan berupa URL, menghilangkan angka, spasi ganda, dan tanda baca.
 - b. *Case folding*, yaitu proses untuk mengubah semua karakter teks menjadi non kapital serta menghilangkan tanda baca dan angka,
 - c. *Stemming*, yaitu menghilangkan imbuhan pada setiap kata untuk mendapatkan kata kunci atau kata dasar. Imbuhan yang dihilangkan yakni awalan, akhiran, sisipan, serta kombinasi dari awalan dan akhiran.
 - d. *Stopwords removal*, yakni merupakan proses yang dilakukan untuk menghapus kata-kata yang tidak memiliki arti yang relevan. Daftar *stopwords* didapat dari penelitian F. Z. Tala yang berjudul “*A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*” (Tala, 2003),

- e. *Tokenizing*, yaitu memecah kalimat isi aduan menjadi kata per kata,
 - f. Melakukan pembobotan kata dengan TF-IDF menggunakan persamaan (2.1) dan (2.2).
 - g. Melakukan *feature selection* atau seleksi variabel menggunakan χ^2 sesuai dengan persamaan (2.3). Yakni dengan membandingkan nilai χ^2 dengan χ^2_α , dimana jika keputusannya adalah gagal tolak H_0 maka kata tersebut tidak berpengaruh terhadap kelas yang bersangkutan.
3. Melakukan *oversampling* dengan metode *Synthetic Minority Oversampling Technique* (SMOTE).
 4. Membagi data aduan ke dalam *data training* dan *data testing* menggunakan *Stratified K-fold Cross Validation* agar semua *fold* mengandung persentase sampel yang sama dari setiap kelas.
 5. Melakukan klasifikasi dengan *Artificial Neural Network* (ANN)
 - a. Menentukan besar inisiasi bobot,
 - b. Mempropagasikan data input ke depan,
 - c. Menghitung input unit ke-j dengan memperhatikan lapisan sebelumnya dan menghitung output setiap unit ke-j,
 - d. Menghitung koreksi bobot,
 - e. Melakukan iterasi,
 - f. Memilih inisiasi bobot yang menghasilkan solusi optimum.
 6. Melakukan evaluasi hasil klasifikasi dengan melihat hasil ketepatan klasifikasi menggunakan persamaan (2.11), (2.12), dan (2.13)
 7. Melakukan visualisasi kata di tiap hasil klasifikasi dengan melakukan *plotting* kata-kata yang sering muncul menggunakan *word cloud*.

3.4 Diagram Alir



Gambar 3.1 Diagram Alir



Gambar 3.1 Diagram Alir (lanjutan)

(Halaman ini sengaja dikosongkan)

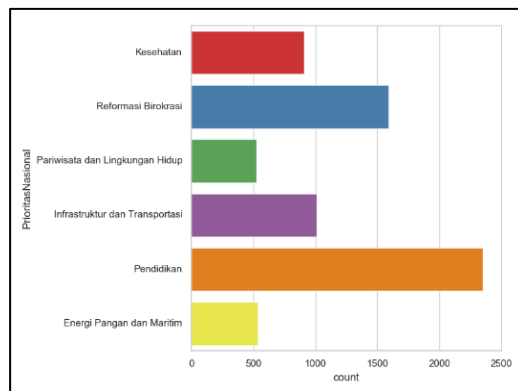
BAB IV

ANALISIS DAN PEMBAHASAN

Pada penelitian ini dilakukan klasifikasi kategori pengaduan masyarakat menggunakan *Artificial Neural Network*. data yang digunakan adalah data aduan masyarakat tahun 2015 dengan jumlah aduan sebanyak 6916 yang terbagi dalam 6 kategori aduan masyarakat dengan jumlah data di setiap kategorinya berbeda-beda. Sehingga sebelum klasifikasi dilakukan *oversampling* untuk mengatasi *imbalance* di tiap kategori dengan menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Kemudian akan dilakukan pengukuran ketepatan klasifikasi dengan menggunakan *Precision*, *Sensitivity*, dan *F1-Score*.

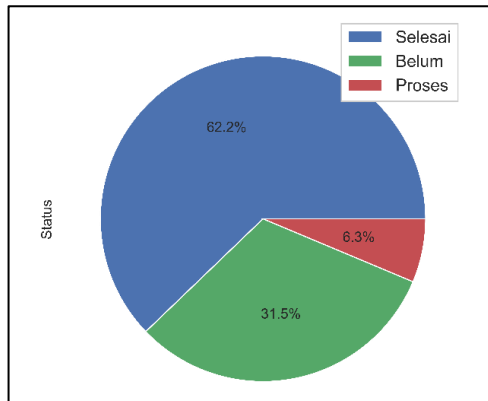
4.1 Karakteristik Data Aduan Masyarakat

LAPOR! merupakan sarana aspirasi dan pengaduan masyarakat terkait kinerja pemerintah berbasis media sosial dan elektronik. Data aduan masyarakat via LAPOR! tersebut kemudian terbagi dalam 6 kategori laporan. Dari data tersebut, dapat diketahui bagaimana karakteristik data awal sebelum dilakukan proses analisis. Data tersebut memuat isi aduan, kategori, dan status tindak lanjut aduan. Terlebih dahulu akan ditunjukkan jumlah aduan masyarakat di tiap kategorinya melalui Gambar 4.1.



Gambar 4.1 Jumlah Aduan Masyarakat Tiap Kategori

Gambar 4.1 menunjukkan jumlah aduan masyarakat yang diterima oleh LAPOR! selama tahun 2015. Diketahui bahwa aduan masyarakat terbanyak terdapat pada kategori Pendidikan dengan jumlah aduan sebesar 2348. Hal tersebut disinyalir karena pada tahun 2015 sejalan dengan program pemerintah dalam pembagian Kartu Indonesia Pintar (KIP), sehingga masyarakat kerap mengajukan distribusi KIP. Sedangkan aduan paling sedikit yang diterima adalah kategori aduan Pariwisata dan Lingkungan Hidup dengan jumlah aduan sebesar 526. Data aduan masyarakat via LAPOR! tersebut selanjutnya diteruskan ke dinas terkait untuk ditindaklanjuti dan dapat dilacak status aduannya oleh pelapor.



Gambar 4.2 Persentase Status Aduan Masyarakat

Dari Gambar 4.2 diketahui status tindak lanjut aduan masyarakat oleh dinas terkait. Sebanyak 4301 aduan yang masuk via LAPOR! telah selesai ditindaklanjuti oleh dinas terkait. Sedangkan untuk status aduan yang masih dalam proses tindak lanjut, diakibatkan karena masih dalam proses lanjutan hingga data ini diterbitkan atau tidak adanya pembaruan status aduan dari dinas terkait.

4.2 Klasifikasi Aduan Masyarakat menggunakan *Artificial Neural Network* (ANN)

Proses selanjutnya yakni melakukan klasifikasi aduan masyarakat menggunakan *Artificial Neural Network*. Tahapan-tahapan

yang dilakukan yakni dengan melakukan *pre processing*, *oversampling* SMOTE, serta mendapatkan ketepatan klasifikasi menggunakan *Artificial Neural Network* dengan membandingkan semua *feature* dan sebagian *feature*.

4.2.1 *Pre-Processing Data*

Sebelum dilakukan analisis, akan dilakukan *pre-processing* data terlebih dahulu. Dalam kasus data teks, beberapa tahapan *pre-processing* yang dilakukan yakni *case folding*, *data cleaning*, *stemming*, *tokenizing*, dan *stopwords removal*. Contoh data sebelum dan setelah *pre-processing* di setiap tahapnya dijelaskan sebagai berikut.

1. *Case Folding dan Data Cleaning*

Case folding merupakan proses untuk mengubah semua karakter teks menjadi non kapital serta menghilangkan tanda baca dan angka. Sedangkan *Data Cleaning* merupakan pembersihan data teks dari kata yang tidak diperlukan dengan cara menghapus tautan berupa URL, menghilangkan angka, spasi ganda, dan tanda baca.

Tabel 4.1 Contoh Data Sebelum dan Sesudah *Case Folding* dan *Cleaning*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
KPD KPS DI TMPAT saya PEMEGANG KARTU KPS Dari thn 2013 sampai skrang anak saya sekolah tdak pernah menerima uang bantuan dari kartu KPS saya pemegang kartu KPS	kpd kps di tmpat saya pemegang kartu kps dari thn sampai skrang anak saya sekolah tdak pernah menerima uang bantuan dari kartu kps saya pemegang kartu kps
⋮	⋮
Lampu PJU di jalan. Cipedes Atas depan rumah No. 39 Kota Bandung mati sekitarnya gelap, kita sangat membutuhkannya jika memang ada biaya utk perbaikan tersebut, kami siap.....	lampu pju di jalan cipedes atas depan rumah no kota bandung mati sekitarnya gelap kita sangat membutuhkannya jika memang ada biaya utk perbaikan tersebut kami siap

Setelah proses *Case Folding* dan *Data Cleaning* pada data teks, tahap *pre-processing* selanjutnya yang akan dilakukan adalah *Stemming*.

2. *Stemming*

Stemming merupakan proses untuk menghilangkan imbuhan pada setiap kata untuk mendapatkan kata kunci atau kata dasar. Imbuhan yang dihilangkan yakni awalan, akhiran, sisipan, serta kombinasi dari awalan dan akhiran. Contoh data teks sebelum dan sesudah melalui tahap *stemming* ditampilkan pada Tabel 4.2.

Tabel 4.2 Contoh Data Sebelum dan Sesudah Tahap *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
kpd kps di tmpat saya pemegang kartu kps dari thn sampai skrang anak saya sekolah tdak pernah menerima uang bantuan dari kartu kps saya pemegang kartu kps	kpd kps di tmpat saya pegang kartu kps dari thn sampai skrang anak saya sekolah tdak pernah terima uang bantu dari kartu kps saya pegang kartu kps
⋮	⋮
lampu pju di jalan cipedes atas depan rumah no kota bandung mati disekitarnya gelap kita sangat membutuhkannya jika memang ada biaya utk perbaikan tersebut kami siap	lampu pju di jalan cipedes atas depan rumah no kota bandung mati sekitar gelap kita sangat butuh jika memang ada biaya utk baik sebut kami siap

Setelah proses *stemming* telah dilakukan, langkah selanjutnya adalah melakukan proses *stopwords removal*.

3. *Stopwords Removal*

Stopwords Removal yakni merupakan proses yang dilakukan untuk menghapus kata-kata yang tidak memiliki arti yang relevan. Daftar *stopwords* yang digunakan merupakan *stopwords* Bahasa Indonesia yang disusun berdasarkan penelitian Fadillah Z Tala pada tahun 2003 serta ditambah oleh kata-kata tidak bermakna yang terkait dengan topik penelitian. Contoh data teks sebelum dan sesudah melalui tahap *Stopwords Removal* ditampilkan pada Tabel 4.3.

Tabel 4.3 Contoh Data Sebelum dan Sesudah Tahap *Stopwords Removal*

Sebelum <i>Stopwords Removal</i>	Setelah <i>Stopwords Removal</i>
kpd kps di tmpat saya pegang kartu kps dari thn sampai skrang anak saya sekolah tdak pernah terima uang bantu dari kartu kps saya pegang kartu kps	kps tempat pegang kartu kps tahun sekolah terima uang bantu kartu kps pegang kartu kps
⋮	⋮
lampu pju di jalan cipedes atas depan rumah no kota bandung mati sekitar gelap kita sangat butuh jika memang ada biaya utk baik sebut kami siap	lampu pju jalan cipedes rumah nomor kota bandung mati gelap butuh biaya

Langkah terakhir dari *pre-processing* data teks yakni melakukan proses *Tokenizing*.

4. *Tokenizing*

Tokenizing merupakan proses untuk memecah keseluruhan teks yang sebelumnya berupa kalimat menjadi kata per kata. Sehingga nantinya didapatkan kata kunci dari setiap aduan untuk nantinya dihitung besaran frekuensinya. Contoh data teks sebelum dan sesudah tahap *Tokenizing* ditampilkan melalui Tabel 4.4.

Tabel 4.4 Contoh Data Sebelum dan Sesudah Tahap *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
kps tempat pegang kartu kps tahun sekolah terima uang bantu kartu kps pegang kartu kps	'kps' 'tempat' 'pegang' 'kartu' 'kps' 'tahun' 'sekolah' 'terima' 'uang' 'bantu' 'kartu' 'kps' 'pegang' 'kartu' 'kps'
⋮	⋮
lampu pju jalan cipedes rumah nomor kota bandung mati gelap butuh biaya	'lampu', 'pju', 'jalan', 'cipedes', 'rumah', 'nomor', 'kota', 'bandung', 'mati', 'gelap', 'butuh', 'biaya'

Hasil dari proses *tokenizing* tersebut dipakai sebagai kata kunci dari data aduan masyarakat. Lalu akan terbentuk struktur data baru dengan masing-masing kata kunci tersebut menjadi variabelnya

dan diketahui frekuensi kemunculannya di tiap aduan sebagaimana ditunjukkan melalui Tabel 4.5

Tabel 4.5 *Count Vectorizer* kata dalam aduan

No	Kategori	Kata Kunci						
		acara	...	bpjs	dokter	ganti	...	yogyakarta
1	2	0	...	2	1	1	...	0
2	4	0	...	1	0	0	...	0
3	5	0	...	0	0	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
253	4	1	...	0	0	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5063	2	0	...	2	1	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6916	3	0	...	0	0	0	...	0

Tabel 4.5 menunjukkan perhitungan frekuensi kata kunci. Pada pengaduan pertama didapatkan bahwa kata “bpjs” disebutkan sebanyak 2 kali, tetapi tidak disebutkan pada aduan ke-3. Pada pengaduan ke-253 kata “acara” disebutkan sebanyak 1 kali, tetapi tidak disebutkan pada aduan ke-n. Sedangkan kata “wilayah” tidak disebutkan pada aduan pertama dan seterusnya. Selanjutnya dari *count vectorizer* tersebut, akan dilakukan pembobotan untuk masing-masing kata menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF) menggunakan persamaan (2.1) dan (2.2). Hasil dari TF-IDF ditunjukkan dalam Lampiran 2.

5. *Feature Selection*

Merupakan proses seleksi variabel-variabel yang relevan untuk digunakan dalam pembentukan model. Keuntungan dari seleksi variabel ini adalah untuk mempercepat proses sehingga nantinya akan didapatkan performa yang lebih tinggi. *Feature selection* yang digunakan dalam penelitian ini adalah *Chi-Square* (χ^2) dengan H_0 , tidak ada hubungan antara *term* atau kata dan kategori aduan masyarakat. H_0 akan ditolak jika nilai χ^2 lebih besar dari $\chi^2_{(0,05;5)}$ yaitu sebesar 11,0705 dan *P-value* kurang dari taraf signifikan sebesar 0,05. Hasil Nilai *Chi-square* yang diperoleh untuk setiap *term* atau kata ditunjukkan melalui Tabel 4.5.

Tabel 4.6 Nilai χ^2 untuk setiap kata

Kata ke- <i>i</i>	Nilai Chi-Square	P-Value	Keputusan Berdasarkan Chi-Square	Keputusan Berdasarkan P-Value
1	3,747	0,586	Gagal Tolak H_0	Gagal Tolak H_0
2	11,288	0,046	Tolak H_0	Tolak H_0
3	6,999	0,221	Gagal Tolak H_0	Gagal Tolak H_0
4	15,592	0,008	Tolak H_0	Tolak H_0
5	8,122	0,150	Gagal Tolak H_0	Gagal Tolak H_0
⋮	⋮	⋮	⋮	⋮
692	1321,655	0,000	Tolak H_0	Tolak H_0
⋮	⋮	⋮	⋮	⋮
949	15,867	0,007	Tolak H_0	Tolak H_0
950	1,586	0,903	Gagal Tolak H_0	Gagal Tolak H_0

Melalui Tabel 4.6 dapat diketahui nilai χ^2 dan P-value dari masing-masing kata. *Term* atau kata yang memiliki nilai χ^2 tertinggi adalah *term* ke-692 yakni kata ‘raskin’ yakni sebesar 1321,655 dan P-value sebesar 0,000. Maka keputusan yang diambil adalah Tolak H_0 yang berarti kata ‘raskin’ memberikan pengaruh terhadap kategori aduan masyarakat. Secara keseluruhan, dapat diketahui bahwa proses seleksi variabel mampu mengurangi jumlah *term* sebesar 45%. Sehingga data yang akan digunakan dalam analisis setelah proses *feature selection* adalah sebanyak 428 kata.

4.2.2 Oversampling Data dengan Synthetic Minority Oversampling Technique (SMOTE)

Setelah proses karakteristik data, diketahui bahwa persentase aduan masyarakat di tiap kategori tidak memiliki jumlah yang sama. Kondisi tersebut disebut *imbalanced* data yakni jumlah data suatu kelas melebihi jumlah data kelas lainnya. Oleh karena itu diperlukan penanganan *imbalanced* data dengan menggunakan metode SMOTE. Fungsi dari SMOTE yakni menambah jumlah sampel pada kelas minor agar memiliki jumlah data yang setara dengan kelas mayor. Terlebih dahulu, data dibagi menjadi *data training* dan *data testing* dengan jumlah perbandingan 90:10 agar memiliki

jumlah *data training* besar pada kasus *multiclass*. Jumlah keseluruhan data *training* dan data *testing* yang digunakan dalam analisis ditunjukkan melalui Tabel 4.7.

Tabel 4.7 Jumlah Data *Training* dan Data *Testing*

Kategori	Training	Testing
0	480	54
1	904	106
2	737	169
3	2291	57
4	1372	220
5	440	86

Dari Tabel 4.7 diketahui bahwa kategori 3 atau kategori Pendidikan merupakan kelas mayoritas. Sehingga perlu dilakukan *oversampling* dengan metode SMOTE pada data *training* agar kategori lain memiliki jumlah data yang setara dengan kelas mayoritas. Setelah dilakukan proses SMOTE, maka data *training* telah memiliki jumlah data yang sama di tiap kategorinya yakni sebesar 2291 sesuai dengan data pada kelas mayor. Sehingga data *training* telah seimbang untuk digunakan dalam analisis klasifikasi dengan menggunakan metode *Artificial Neural Network*.

4.2.3 Ketepatan Klasifikasi Menggunakan *Artificial Neural Network* (ANN)

ANN merupakan salah satu sistem pemrosesan informasi dengan desain menirukan cara kerja otak manusia. Penerapan ANN dalam penelitian kali ini karena kemampuannya untuk mempelajari model non-linier dan data yang besar (Mahanta, 2017), dengan menggunakan jaringan *multilayer perceptron*. Algoritma yang digunakan adalah *backpropagation* dengan menggunakan 1 *hidden layer* dengan percobaan jumlah *nodes* yang telah ditentukan sebanyak 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10 *nodes*. Sedangkan fungsi aktivasi yang digunakan adalah *Softmax* karena cocok digunakan dalam data *multiclass* (Zhang dkk, 2019). Sedangkan metode yang digunakan untuk melakukan optimasi parameter *weight* dan *bias* adalah *Stochastic Gradient Descent* (SGD).

Seperti yang telah dijelaskan pada sebelumnya, terdapat 428 variabel yang telah terseleksi. Sehingga langkah awal yang dilakukan untuk melakukan klasifikasi data aduan masyarakat dengan ANN yakni melakukan pemilihan *feature* (semua *feature* atau 428 *feature*) dan jumlah *nodes hidden layer* yang menghasilkan ketepatan klasifikasi terbaik. Yakni dengan membandingkan nilai *precision*, *sensitivity*, dan *F1-Score*. Hasil ketepatan klasifikasi dari masing-masing pemilihan *feature* dan jumlah *nodes* ditunjukkan pada Tabel 4.8.

Tabel 4. 8 Hasil Ketepatan Klasifikasi dengan ANN

Nodes	428 Feature			Semua Feature		
	P	S	Fscore	P	S	Fscore
1	0,607	0,621	0,592	0,541	0,550	0,545
2	0,779	0,779	0,773	0,723	0,739	0,724
3	0,794	0,818	0,800	0,741	0,738	0,737
4	0,767	0,756	0,756	0,724	0,728	0,722
5	0,770	0,778	0,771	0,737	0,746	0,736
6	0,765	0,799	0,776	0,754	0,757	0,753
7	0,745	0,760	0,750	0,744	0,759	0,747
8	0,752	0,781	0,763	0,726	0,733	0,727
9	0,787	0,770	0,777	0,762	0,762	0,760
10	0,766	0,788	0,773	0,765	0,759	0,760

Hasil ketepatan klasifikasi pada Tabel 4.8 merupakan *macro averaging* atau nilai *precision*, *sensitivity* dan *F1-Score* pada setiap kelas, dibagi dengan jumlah kelasnya. Jika menggunakan semua *feature*, jumlah *nodes* yang menghasilkan ketepatan klasifikasi terbaik adalah 10 *nodes* dengan nilai *precision* 0,765, *sensitivity* 0,759, dan *F1-Score* 0,76. Sedangkan dengan menggunakan 428 *feature*, klasifikasi terbaik yang dihasilkan adalah dengan menggunakan 3 *nodes hidden layer* dengan nilai *precision* 0,794, *sensitivity* 0,818 dan *F1-Score* 0,8. Dari hasil tersebut, dapat disimpulkan bahwa *feature selection* mampu meningkatkan nilai ketepatan klasifikasi. Sehingga ketepatan klasifikasi terbaik yakni dengan menggunakan 428 *feature* sebagai *input* serta *hidden layer*

yang digunakan sebanyak 3 *nodes*. Untuk *confusion matrix* dari hasil klasifikasi, dijelaskan melalui Tabel 4.9.

Tabel 4. 9 *Confusion Matrix*

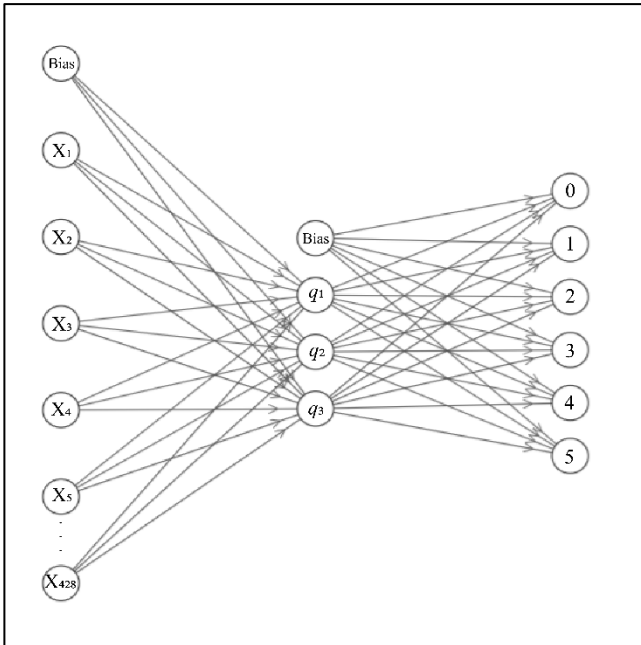
Kelas Aktual	Kelas Prediksi					
	0	1	2	3	4	5
0	42	6	0	1	4	1
1	3	93	0	1	5	4
2	0	0	157	5	2	5
3	0	3	2	50	2	0
4	3	17	30	11	144	15
5	0	8	1	1	8	68

Tabel 4.10 menunjukkan kinerja klasifikasi dari metode klasifikasi *Artificial Neural Network*. Dari tabel 4.10 dapat diketahui bahwa kategori 0 yakni energi pangan dan maritim diklasifikasikan secara benar ke dalam kategori tersebut sebanyak 42 aduan. Sedangkan terdapat 12 aduan salah diklasifikasikan ke dalam kategori lain, yakni 6 aduan ke dalam kategori 1, 1 aduan pada kategori 3, 4 aduan pada kategori 4, dan 1 aduan ke dalam kategori 5. Hal ini berlaku untuk kategori aduan lainnya. Sehingga hasil ketepatan klasifikasi yang dihasilkan dijelaskan melalui tabel 4.10.

Tabel 4. 10 Ketepatan Klasifikasi Tiap Kategori

Kategori	Precision	Sensitivity	F1- score
0	0,875	0,778	0,824
1	0,732	0,877	0,798
2	0,826	0,929	0,875
3	0,725	0,877	0,794
4	0,873	0,655	0,747
5	0,731	0,791	0,760

Ilustrasi jaringan *Artificial Neural Network* dengan 428 *feature* dan 3 *nodes hidden layer* ditunjukkan melalui Gambar 4.3 dengan *weight* dan *bias* terlampir pada Lampiran 11.



Gambar 4.3 Jaringan ANN 428 feature dan 3 nodes hidden layer

Selanjutnya yakni melakukan *K-fold cross validation* untuk mengetahui tingkat kesalahan teknik klasifikasi. Metode ini digunakan untuk mengurangi bias terkait pengambilan sampel dari data dengan membagi data kedalam sejumlah bagian yang telah ditentukan yang disebut *fold*. Partisi data yang digunakan untuk pembagian *data training* dan *data testing* yakni dengan *10-fold cross validation*.

Pada *10-fold cross validation*, data dibagi ke dalam 10 *fold* lalu kemudian dibagi kembali ke dalam data *training* dan data *testing* dengan perbandingan 90:10 dengan metode sampling stratifikasi. Ketepatan klasifikasi dengan menggunakan *10-fold cross validation*, 428 *feature*, dan 3 *nodes hidden layer* ditunjukkan melalui Tabel 4.11. Sama seperti sebelumnya, hasil dari ketepatan klasifikasi merupakan *macro averaging* dari setiap kelasnya.

Tabel 4.11 Ketepatan Klasifikasi dengan *k-fold*

Fold ke-	Ketepatan Klasifikasi		
	Precision	Sensitivity	F1-Score
1	0,731	0,743	0,726
2	0,786	0,811	0,797
3	0,753	0,791	0,769
4	0,749	0,762	0,754
5	0,743	0,783	0,752
6	0,804	0,804	0,8
7	0,745	0,775	0,747
8	0,723	0,737	0,721
9	0,785	0,793	0,783
10	0,785	0,798	0,789
Rata-rata	0,760	0,780	0,764

Rata-rata ketepatan klasifikasi yang dihasilkan melalui 10-fold cross validation yaitu *precision* 0,760, *sensitivity* 0,780, dan *F1-score* 0,764. Model yang dihasilkan melalui metode *Artificial Neural Network* dijelaskan sebagai berikut.

$$p_1 = 0,91 + (-0,5683 x_1 - 0,6291 x_2 - 0,2100 x_3 + \dots - 0,1679 x_{428}),$$

$$p_2 = 0,8013 + (-0,4805 x_1 - 0,0076 x_2 + 0,3389 x_3 + \dots - 0,2781 x_{428}),$$

$$p_3 = 1,2489 + (-0,8587 x_1 - 0,4181 x_2 - 0,0650 x_3 + \dots - 0,0382 x_{428}).$$

Sehingga persamaan untuk setiap *nodes* pada *hidden layer* dengan fungsi aktivasi ReLU yakni sebagai berikut.

$$q_1 = f(p_1) = \max(0, p_1) = \begin{cases} p_1, & \text{jika } p_1 \geq 0, \\ 0, & \text{jika } p_1 < 0, \end{cases}$$

$$q_2 = f(p_2) = \max(0, p_2) = \begin{cases} p_2, & \text{jika } p_2 \geq 0, \\ 0, & \text{jika } p_2 < 0, \end{cases}$$

$$q_3 = f(p_3) = \max(0, p_3) = \begin{cases} p_3, & \text{jika } p_3 \geq 0, \\ 0, & \text{jika } p_3 < 0. \end{cases}$$

Untuk persamaan pada *output layer* sebelum dimasukkan fungsi aktivasi adalah sebagai berikut.

$$\begin{aligned} r_0 &= 9,0164 - 4,5338 q_1 - 4,4940 q_2 - 3,8883 q_3, \\ r_1 &= 1,7051 - 0,9764 q_1 - 6,4906 q_2 - 3,9972 q_3, \\ &\vdots \\ r_5 &= -9,0688 + 6,4526 q_1 + 0,2700 q_2 + 2,5935 q_3. \end{aligned}$$

Sehingga persamaan untuk setiap *nodes* pada *output layer* atau masing-masing kategori dengan menggunakan fungsi aktivasi *Softmax* adalah sebagai berikut.

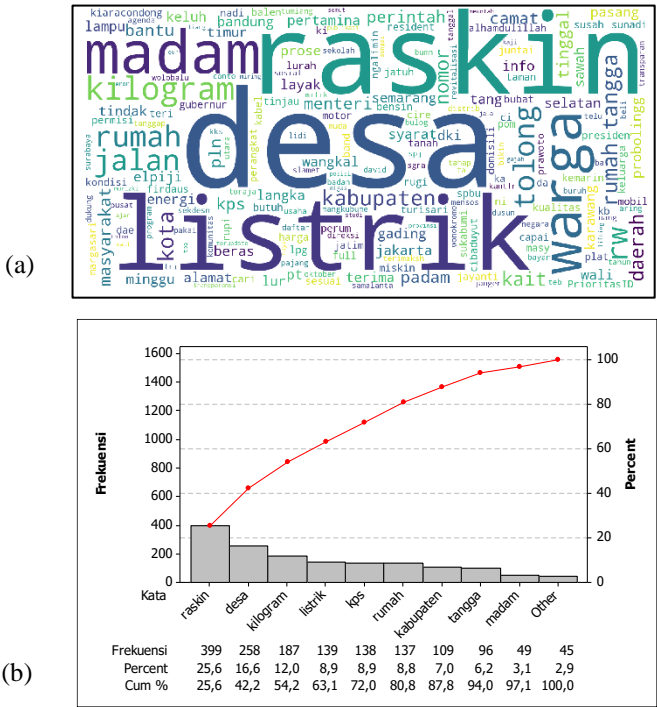
$$\begin{aligned} s_0 &= f(r_1) = \frac{\exp(r_0)}{\exp(r_1) + \exp(r_2) + \exp(r_3) + \exp(r_4) + \exp(r_5) + \exp(r_6)}, \\ s_1 &= f(r_2) = \frac{\exp(r_1)}{\exp(r_1) + \exp(r_2) + \exp(r_3) + \exp(r_4) + \exp(r_5) + \exp(r_6)}, \\ &\vdots \\ s_5 &= f(r_6) = \frac{\exp(r_5)}{\exp(r_1) + \exp(r_2) + \exp(r_3) + \exp(r_4) + \exp(r_5) + \exp(r_6)}. \end{aligned}$$

Melalui persamaan tersebut, akan dihasilkan probabilitas dari suatu aduan masyarakat yang masuk via LAPOR!. Yakni kelas target dari aduan tersebut adalah kategori yang memiliki probabilitas tertinggi. Yakni s_0 untuk kategori energi pangan dan maritim, s_1 untuk infrastruktur dan transportasi hingga s_5 untuk pariwisata dan lingkungan hidup.

4.3 Visualisasi *Word Cloud*

Aduan masyarakat via LAPOR! di tiap kategorinya dapat divisualisasikan melalui *word cloud*. *Word cloud* merupakan representasi grafis dari dokumen teks kedalam ruang dua dimensi dengan melakukan *plotting* kata-kata yang sering muncul. Melalui *word cloud*, akan didapatkan kata kunci yang paling sering diadukan oleh masyarakat melalui layanan LAPOR! di masing-masing kategori prioritas nasional. Semakin besar ukuran kata dalam *word*

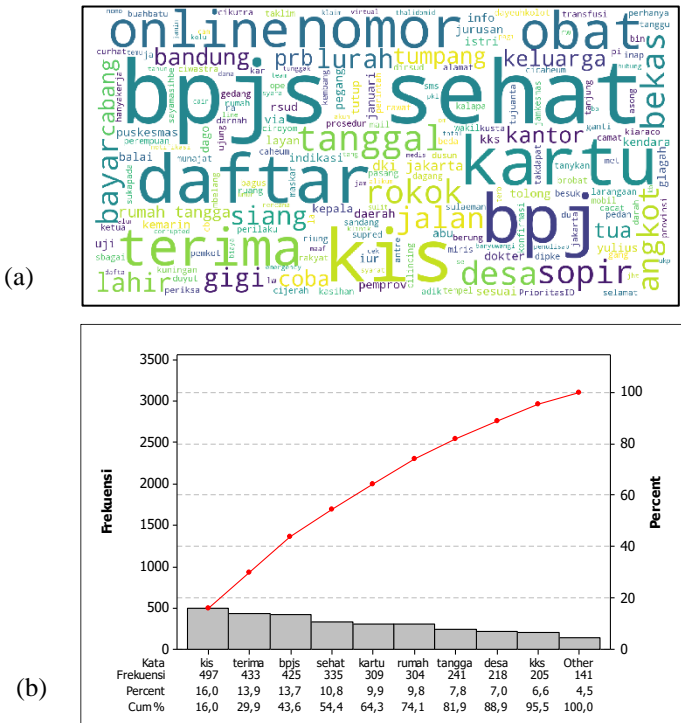
cloud, maka menunjukkan bahwa frekuensi kata tersebut diadukan semakin besar. Sehingga dinas terkait dapat langsung mengetahui permasalahan yang menjadi topik utama di masyarakat pada setiap kategori aduan prioritas nasional. Berikut merupakan visualisasi dari kata-kata yang diadukan melalui LAPOR! pada setiap kategori aduan dengan menggunakan *word cloud*.



Gambar 4.4 (a) *Word Cloud* (b) *Pareto Chart* Kategori Energi Pangan dan Maritim

Gambar 4.4.(a) menunjukkan bahwa kata kunci dari laporan yang patut mendapatkan perhatian lebih pada kategori energi pangan dan maritim adalah raskin, desa, kilogram dan listrik. 4 kata tersebut dipilih karena telah memberikan persentase sebesar 63,1% jika diambil 10 kata dengan frekuensi tertinggi sebagaimana dijelaskan Gambar 4.4.(b). Kata raskin atau beras miskin menjadi

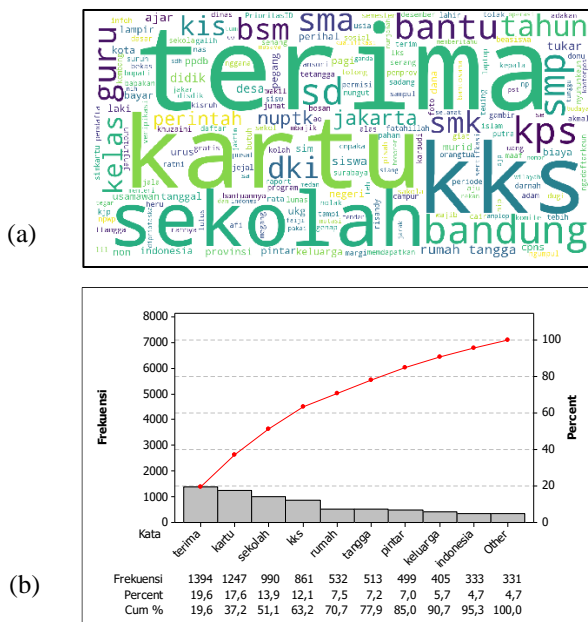
Gambar 4.5.(a) menunjukkan bahwa kata kunci aduan yang menjadi sorotan masyarakat terkait infrastruktur dan transportasi adalah kata jalan. Kata jalan dipilih karena memberikan kontribusi kata sebesar 40% sebagaimana dijelaskan pada Gambar 4.5.(b). Sedangkan kata pendukung lainnya yakni Jakarta dan Bandung. Aduan dari masyarakat yang masuk melalui LAPOR! mengenai kata jalan berhubungan dengan kerusakan jalan, atau aduan lain dengan menyertakan alamat yang bersangkutan. Kota Jakarta dan Bandung kerap menjadi sasaran aduan masyarakat dalam kategori infrastruktur dan transportasi.



Gambar 4.6 (a) Word Cloud (b) Pareto Chart Kategori Kesehatan

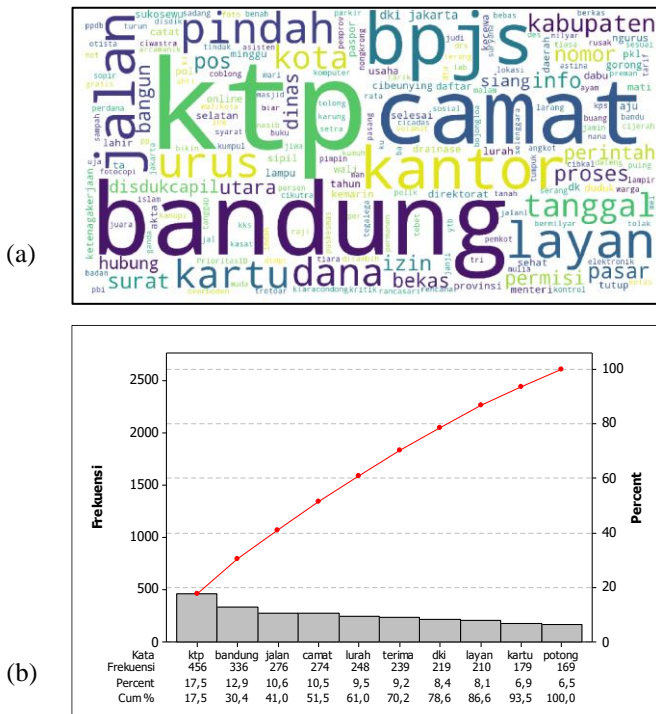
Kata BPJS, KIS, terima, sehat, dan kartu merupakan kata kunci pada kategori kesehatan yang harus ditindaklanjuti oleh di-

nas terkait karena sering diadukan masyarakat via LAPOR! Sebagaimana ditunjukkan melalui Gambar 4.6.(a). 5 Kata tersebut dipilih karena persentase kumulatif yang dihasilkan berdasarkan Gambar 4.6.(b) adalah sebesar 64,3%. BPJS merupakan badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial. BPJS terdiri dari BPJS Kesehatan dan BPJS Ketenagakerjaan. Dalam kasus aduan masyarakat via LAPOR! ini, mayoritas aduan yang masuk adalah terkait dengan BPJS kesehatan. BPJS sering diadukan karena dalam praktik di lapangan, sering ditemukan kendala dalam pelayanan BPJS di berbagai fasilitas kesehatan. Sedangkan kata sehat merupakan kata dasar dari kesehatan sehingga kerap muncul sebagai sebagai penjelas dari kata BPJS yakni BPJS Kesehatan. Kata lain yang juga sering diadukan masyarakat adalah Kartu Indonesia Sehat (KIS). Aduan masyarakat terkait KIS antara lain karena belum tersalurkannya atau belum diterimanya KIS secara baik kepada masyarakat.



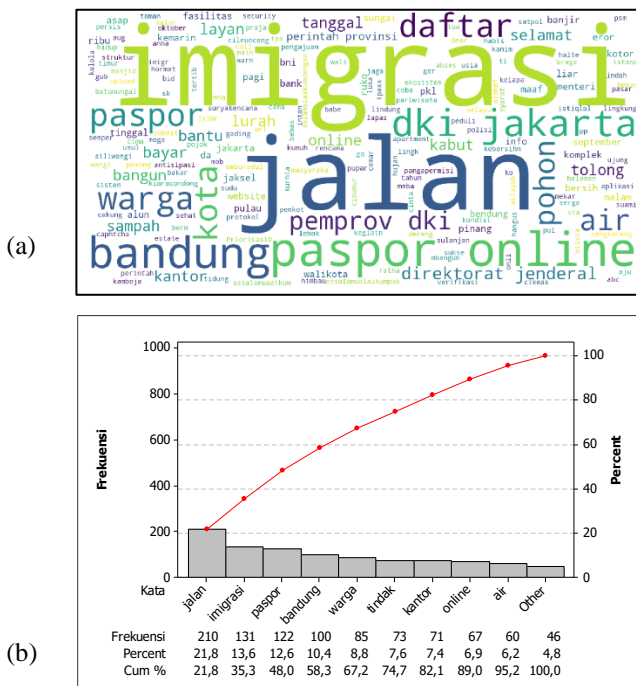
Gambar 4.7 (a) Word Cloud (b) Pareto Chart Kategori Pendidikan

Melalui Gambar 4.7.(a) dapat diketahui bahwa tiga kata kunci yang perlu diperhatikan lebih oleh dinas terkait pada kategori pendidikan adalah kata terima, kartu, dan sekolah. Dipilihnya 3 kata tersebut karena berdasarkan Gambar 4.7.(b) kontribusi kata yang diberikan sebesar 51,1% jika diambil 10 kata tertinggi. Kata kartu sering diadukan karena terkait dengan program pemerintah yaitu Program Indonesia Pintar (PIP). PIP melalui Kartu Indonesia Pintar (KIP) adalah pemberian bantuan tunai pendidikan kepada anak usia sekolah yang berasal dari keluarga miskin, rentan miskin atau pemilik Kartu Keluarga Sejahtera (KKS). Kata terima menjadi kata yang paling sering menjadi topik aduan karena selaras dengan banyaknya aduan masyarakat yang belum menerima KIP pada saat itu.



Gambar 4.8 (a) Word Cloud (b) Pareto Chart Kategori Reformasi Birokrasi

Gambar 4.8.(a) menunjukkan bahwa kata KTP, Bandung, Jalan, Camat, dan Lurah pada kategori reformasi birokrasi merupakan kata kunci aduan yang paling sering dilaporkan oleh masyarakat. 5 Kata tersebut dipilih karena berdasarkan Gambar 4.4.(b) memberikan persentase sebesar 61%. Hal ini terjadi karena pada tahun 2015, Kementerian Dalam Negeri mewajibkan masyarakat untuk melakukan perekaman untuk mengubah KTP mereka menjadi KTP elektronik (*e-KTP*). Sehingga masyarakat kerap melaporkan lamanya proses pembuatan *e-KTP* di kantor kecamatan daerah tertentu, oleh karena itu kata Camat turut dilaporkan oleh masyarakat. Melalui Gambar 4.8.(a) juga diketahui bahwa Bandung merupakan daerah yang sering menjadi sasaran aduan via LAPOR! dalam kategori aduan reformasi birokrasi.



Gambar 4.9 (a) Word Cloud (b) Pareto Chart Kategori Pariwisata dan Lingkungan Hidup

Kata jalan, imigrasi, paspor, bandung, dan warga merupakan kata yang paling sering diadukan oleh masyarakat pada kategori pariwisata dan lingkungan hidup. 5 Kata tersebut memberikan kontribusi sebesar 67,2% sebagaimana ditunjukkan Gambar 4.9. Masyarakat kerap memberi aduan terhadap pelayanan kantor imigrasi dalam pendaftaran pembuatan paspor. Kata kunci lainnya yang kerap diadukan masyarakat yakni kata *online*, juga terkait dengan aduan masyarakat terkait pendaftaran pembuatan paspor secara *online* yang dilakukan oleh pihak imigrasi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan pada bab 4 maka diperoleh kesimpulan sebagai berikut.

1. Aduan terbanyak dari masyarakat via kanal LAPOR! terdapat pada kategori Pendidikan dengan jumlah aduan sebesar 2348 dari total 6916 aduan. Aduan tersebut telah selesai ditindaklanjuti oleh dinas terkait sebanyak 62,2%.
2. Melalui *feature selection* atau seleksi variabel, didapatkan sejumlah 428 *term* atau kata memberikan pengaruh terhadap kategori aduan masyarakat. *Feature selection* mampu menurunkan jumlah *term* atau kata sebesar 45% yang dapat meningkatkan nilai ketepatan klasifikasi. Sehingga, ketepatan klasifikasi yang dihasilkan dengan metode *Artificial Neural Network* dengan *feature selection* dan 3 *nodes hidden layer* di tiap kategori aduan yakni untuk kategori energi, pangan dan maritim adalah *precision* 0,875, *sensitivity* 0,778, dan *F1-Score* 0,824. Untuk kategori infrastruktur dan transportasi adalah *precision* 0,732, *sensitivity* 0,877, dan *F1-Score* 0,798. Untuk kategori kesehatan adalah *precision* 0,826, *sensitivity* 0,929, dan *F1-Score* 0,875. Untuk kategori pendidikan adalah *precision* 0,725, *sensitivity* 0,877, dan *F1-Score* 0,794. Untuk reformasi birokrasi adalah *precision* 0,873, *sensitivity* 0,655, dan *F1-Score* 0,747. Terakhir yakni untuk kategori pariwisata dan lingkungan hidup yaitu *precision* 0,731, *sensitivity* 0,791, dan *F1-Score* 0,760.
3. Berdasarkan visualisasi *word cloud*, didapatkan topik aduan yang patut mendapatkan perhatian lebih dari dinas terkait karena sering diadakan oleh masyarakat melalui kata kunci dari setiap kategori aduan prioritas nasional. Kata kunci aduan yang patut mendapatkan perhatian lebih pada kategori aduan energi, pangan dan maritim adalah kata raskin yakni terkait

pembagian beras miskin oleh pemerintah kepada masyarakat. Untuk kategori infrastruktur dan transportasi adalah aduan mengenai jalan yaitu tentang kerusakan jalan. Pada kategori kesehatan, aduan mengenai BPJS merupakan topik aduan yang kerap dilaporkan oleh masyarakat karena banyaknya masalah yang ditemui pada fasilitas kesehatan. Untuk kategori pendidikan adalah kata terima yakni berkaitan dengan pembagian KIP. Lalu kata KTP merupakan aduan yang paling sering diajukan masyarakat pada kategori reformasi birokrasi karena lamanya proses pembuatan *e*-KTP. Sedangkan kata imigrasi merupakan permasalahan yang kerap diajukan di kategori pariwisata dan lingkungan hidup.

5.2 Saran

Berdasarkan hasil analisis yang telah dijelaskan, maka saran yang dapat diberikan sebagai pertimbangan penelitian selanjutnya adalah sebagai berikut.

1. Dalam menghadapi data aduan masyarakat, perlu adanya penambahan kata kunci bahasa sehari-hari maupun tafsiran dari berbagai singkatan yang lebih bervariasi agar kata asli dapat terhitung dalam frekuensi kemunculan kata .
2. Mencari data terbaru terkait aduan masyarakat via LAPOR! agar kategori lain yang tidak masuk dalam penelitian ini dapat ditambahkan pada penelitian selanjutnya.
3. Untuk pihak LAPOR! yakni dapat mempertimbangkan hasil klasifikasi dari penelitian ini agar penanganan aduan yang terkait kategori laporan prioritas nasional dapat diselesaikan dengan lebih cepat.

DAFTAR PUSTAKA

- Agresti, A. (2007). *An Introduction To Categorical Data Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Astuti, N. D. (2017). *Klasifikasi Penyakit Gagal Jantung Kongestif Menggunakan Artificial Neural Network*. Surabaya: Program Studi Sarjana Departemen Sistem Informasi Fakultas Teknologi Informasi dan Komunikasi ITS.
- Bachri, O. S., Kusnadi, M. H., & Nurhayati, O. D. (2017). Feature Selection Based On Chi Square In Artificial Neural Network To Predict The Accuracy of Student Study Period. *International Journal of Engineering and Technology*, 8(8), 731-739.
- Castella, Q., & Sutton, C. (2013). *Word Storm: Multiples of Word Clouds for Visual Comparison of Documents*. New York: Cornell University.
- Chawla, N. V. (2002). SMOTE Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Damayanti, F. K. (2018). *Analisis Twitter Pelanggan Belanja Online Menggunakan Metode Naive Bayes Classifier (NBC) dan Artificial Neural Network (ANN)*. Surabaya: Program Studi Sarjana Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data ITS.
- Damayanti, N. (2017). *Klasifikasi Penyakit Paru Dengan Metode Artificial Neural Network*. Surabaya: Program Studi Sarjana Departemen Sistem Informasi Fakultas Teknologi Informasi dan Komunikasi ITS.
- Darujati, C., & Gumelar, A. B. (2012). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Berbahasa Indonesia. *Jurnal Link*, 16(1), 1-8.

- Feldman, R., & Sanger, J. (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Gokgoz, E., & Subasi, A. (2015). Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT. *Biomedical Signal Processing and Control*, 18, 138-144.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kaufman Publisher.
- KataData. (2018). *4 Miliar Penduduk Bumi Telah Terkoneksi Internet*. Diakses pada tanggal 20 Februari 2019 dari Katadata:<https://databoks.katadata.co.id/data-publish/2018/07/23/4-miliar-penduduk-bumi-telah-terkoneksi-internet>
- Kemenko Bidang Pembangunan Manusia RI. (2016). *Pedoman Umum Subsidi Beras Pada Masyarakat Berpendapatan Rendah*. Jakarta: Kemenko Bidang Pembangunan Manusia RI.
- Kementerian Komunikasi dan Informatika RI. (2016). *Laporan Tahunan 2016*. Jakarta: Menteri Komunikasi dan Informatika.
- Kementerian Pendidikan dan Kebudayaan RI. (2016). *Program Indonesia Pintar*. Diakses pada tanggal 20 Februari 2019 dari Kemendikbud Republik Indonesia: <https://indonesia-pintar.kemdikbud.go.id/>
- Khaulasari, H. (2016). *Combine Sampling - Least Square Support Vector Machine Untuk Klasifikasi Multi Class Imbalanced Data*. Surabaya: Program Studi Magister Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data ITS.
- LAPOR! (2017). *Laporan*. Diakses pada tanggal 20 Februari 2019 dari Layanan Aspirasi Pengaduan Online Rakyat: www.lapor.go.id

- Mahanta, J. (2017, July 10). *Introduction to Neural Networks, Advantages and Applications*. Diakses pada tanggal 28 Juni 2019 dari Towards Data Science: <https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications>
- Meesad, P., Boonrawd, P., & Nui pian, V. (2011). A Chi-Square-Test for Word Importance Differentiation in Text Classification. *International Conference on Information and Electronics Engineering*, 6, 110-114.
- Megawati, C. (2015). *Analisis Aspirasi dan Pengaduan di Situs LAPOR! Dengan Menggunakan Text Mining*. Depok: Program Studi Sarjana Departemen Teknik Industri Fakultas Teknik Universitas Indonesia.
- Mishu, S. Z., & Rafiuddin, S. (2016). Performance Analysis of Supervised Machine Learning Algorithms for Text Classification. *International Conference on Computer and Information Technology, North South University, Dhaka, Bangladesh*, 19, 409-413.
- Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of Trends in Practice and Research for Deep Learning*. New York: Cornell University.
- Prasanna, L., & Rao, D. (2018). Text Classification Using Neural Networks. *International Journal of Engineering & Technology*, 7, 603-606.
- Rahman, F. (2018). *Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Pada Pengguna Twitter Mengenai Presiden Joko Widodo*. Surabaya: Program Studi Sarjana Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data ITS.
- Reyhana, Z., Fithriasari, K., Atok, M., & Iriawan, N. (2018). Linking Twitter Sentiment Knowledge with Infrastructure

- Development. *Malaysian Journal of Industrial and Applied Mathematics*, 34, 91-102.
- Saha, D. (2011). Web Text Classification Using a Neural Network. *Second International Conference on Emerging Applications of Information Technology*, 2, 57-60.
- Santos, M. S. (2018). Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches. *IEEE Computational Intelligence Magazine*, 13(4), 59-76.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45, 427-437.
- Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Amsterdam: Master of Logic Project, Institute for Logic, Language and Computation, Universiteit van Amsterdam.
- United Nations. (2018). *United Nations E-Government Survey 2018*. New York: United Nations.
- Weiss, S. M. (2010). *Text Mining: Predictive Methods for Analyzing*. New York: Springer.

LAMPIRAN

Lampiran 1. Data aduan masyarakat via kanal LAPOR! 2015.

Data diambil melalui situs <http://www.data.go.id/dataset/data-aspirasi-dan-pengaduan-masyarakat>

id	TrackingID	Isi Laporan	PrioritasNasional	Status
1	1385716	Bagaimana prosedur periksa ke dokter gigi menggunakan bpjs dan apakah jika mengganti gigi palsu di kenakan biaya saya pengguna bpjs kelas 3. terimakasih	Kesehatan	Selesai
2	1388325	Mengapa dalam pengurusan perpindahan kelas dalam kepesertaan bpjs harus melampirkan fotocopi buku tabungan? padahal sudah ada bukti pembayaran.	Reformasi Birokrasi	Selesai
3	1385668	Assalamuaaikum, Pangapunten sak dereng pun Kang Yoto bendung gerak sak meniko ditutup sumber. Sumber sumber sumur tiang2 dusun sami macet tiang2 bade wudu mawon sulit tulung bukak sekedik terima kasih.	Pariwisata dan Lingkungan Hidup	Selesai
4	1321337	Pak gubenur Basuki cahaya purnama yang terhormat tolong stasiun jatinegara ganggu lingkungan di depannya suara peronnya bising sekali. mohon dikecilkan sedikit. dari Sugito warga didepannya	Infrastruktur dan Transportasi	Selesai
5	1348668	pengelolaan lingkungan untuk struktur air kali cakung dr pupar g sampai babek. pekat. banyak sampah dan limbah. bisakah pengelolaanya sama dengan BKT. Terimakasih	Pariwisata dan Lingkungan Hidup	

6	1317242	39LMTQ17414A04 pasangan saya belum menerima KIS	Kesehatan	Selesai
7	1304828	Pak saya lapor:Anak saya nama ADAM MUHAMMAD,SMA 60 Kelas 12 IPA 1, sampai hari ini tgl 12-1-2015 belum menerima BSM.Terima kasih.	Pendidikan	Belum
8	1309288	373Y4410640A08 katanya, KPS ditukar dgn KKS,SIM Card, KIP & KIS, tapi knp hanya 3kartu yang saya terima sedangkan kartu KIP saya tidak menerima .. Trima ksh sebelumnya ..	Pendidikan	Selesai
9	1311898	Tanah PT.KAI yang telah digusur di.Rw.08 kel.Pejaten Timur.Kec.Ps.Minggu. Kembali d tempati oleh pedagang dan dipunguti uang oleh oknum tertentu.	Infrastruktur dan Transportasi	Proses
10	1315310	(1/2)KPD KPS DI TMPAT saya PEMEGANG KARTU KPS Dari thn 2013 sampai skrang anak saya sekolah tdak pernah menerima uang bantuan dari kartu KPS saya pemegang kartu KPS	Pendidikan	Selesai
11	1309614	375CZ11850A07 SANGAT BUTUH KIP (BSM) (penerima KKS)	Pendidikan	Selesai
12	1302575	Pak gmn masalah kip apa yang blm dpt itu pasti akan dpt	Pendidikan	Selesai
13	1301622	Pak saya mau nanya ttangga saya dpt kip knp tidak merata pembagian kip	Pendidikan	Selesai
14	1318586	Saya pemegang kip tapi belum dapat dana kip	Pendidikan	Belum
15	1332981	39mem17000002 saya mau tanya kenapa saya tidak mendapat kartu kip sedangkan saya masih mempunyai anak yang sekolah (penerima kks)	Pendidikan	Selesai

16	1335393	(01/02)Sy warga rt 01/04 sukabumi selatan,ingin menanyakan apa benar harga raskin mencapai 30 rb rupiah sekarang kecil/15 kilogram sepedrti di rw saya ? .	Energi Pangan dan Maritim	Belum
17	1307491	KENAPA NO KKS YG BELAKANGNYA ADA PDTMPT saya LOM KELUAR DANANYA KATA KANTOR POS DAERAH saya. JD TDK BISA DTUKAR dulu KARTU LAMANYA	Reformasi Birokrasi	Belum
:	:	:	:	:
6915	1375257	. Kementerian Dalam Negeri, Saya ingin bertanya, apakah pengurusan kesalahan NIK dapat di lakukan secara kolektif oleh perwakilan perusahaan ke Disdukcapil, saya sebagai staf pada salah satu perusahaan yang berkepentingan untuk mengurus BPJS	Reformasi Birokrasi	Selesai
6916	1383626	Saya mohon bantuan Bapak/Ibu yang berkompeten dalam masalah saya, saya sebagai Guru pada awalnya mengajar di tingkat SMA tepatnya di SMA Negeri Mangkutana (Kabupaten Luwu Timur) yang sudah disertifikasi dengan Guru Mata Pelajaran Fisika. Tetapi saya dimut	Pendidikan	Belum

Lampiran 2. *TF-IDF* dari Kata Kunci.

No	Kategori	Kata Kunci								
		acara	aceh	adik	administrasi	...	jalan	...	wilayah	yogyakarta
0	2	0	0	0	0	...	0	...	0	0
1	4	0	0	0	0	...	0	...	0	0
2	5	0	0	0	0	...	0	...	0	0
3	1	0	0	0	0	...	0	...	0	0
4	5	0	0	0	0	...	0	...	0	0
5	2	0	0	0	0	...	0	...	0	0
6	3	0	0	0	0	...	0	...	0	0
7	3	0	0	0	0	...	0	...	0	0
8	1	0	0	0	0	...	0	...	0	0
9	3	0	0	0	0	...	0	...	0	0
10	3	0	0	0	0	...	0	...	0	0
11	3	0	0	0	0	...	0	...	0	0
12	3	0	0	0	0	...	0	...	0	0
13	3	0	0	0	0	...	0	...	0	0
14	3	0	0	0	0	...	0	...	0	0
15	0	0	0	0	0	...	0	...	0	0
16	4	0	0	0	0	...	0	...	0	0
17	2	0	0	0.279	0	...	0	...	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
60	5	0	0	0	0	...	0	...	0	0
61	1	0	0	0	0	...	0.221	...	0	0
62	1	0	0	0	0	...	0.232	...	0	0
63	4	0	0	0	0	...	0.232	...	0	0
64	4	0	0	0	0	...	0	...	0	0
65	1	0	0	0	0	...	0.559	...	0	0
66	3	0	0	0	0	...	0	...	0	0
67	3	0	0	0	0	...	0	...	0	0
68	4	0	0	0	0	...	0	...	0	0
69	3	0	0	0	0	...	0.168	...	0	0
70	4	0	0	0	0	...	0	...	0	0
71	1	0	0	0	0	...	0.139	...	0	0
72	1	0	0	0	0	...	0.364	...	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
409	2	0	0.339	0	0.320	...	0	...	0	0
410	4	0	0	0	0	...	0	...	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6914	4	1	0	0	0	...	0	...	0	0
6915	3	0	0	0	0	...	0	...	0	0

Lampiran 3. Data hasil tokenizing.

```

[['prosedur', 'periksa', 'dokter', 'gigi', 'bpjs', 'ganti', 'gigi', 'palsu', 'biaya', 'bpjs', 'kelas'], ['urus', 'pindah', 'kelas', 'bpjs', 'lampir', 'fotocopi', 'buku', 'tabung', 'bukti', 'bayar'], ['assalamu'aaikum', 'pangapermisi', 'dereng', 'bendung', 'gerak', 'menik', 'tutup', 'sumber', 'sumber', 'sumber', 'sumur', 'tiang', 'dusun', 'sami', 'macet', 'tiang', 'bade', 'wudu', 'mawon', 'sulit', 'tolong', 'bukak', 'kedik', 'mawon', 'panga', 'permisi', 'enggang', 'katah', 'matur', 'sembah'], ['gubernur', 'basuki', 'cahaya', 'purnama', 'hormat', 'tolong', 'stasiun', 'jatinegara', 'ganggu', 'lingkung', 'suara', 'peron', 'bising', 'sugito', 'warga'], ['kelola', 'lingkung', 'struktur', 'air', 'cakung', 'pupar', 'babek', 'pekat', 'sampah', 'limbah', 'bisakh', 'pengelolaanya', 'bkt'], ['pasang', 'terima', 'kis'], ['adam', 'sma', 'kelas', 'tanggal', 'terima', 'bsm'], ['kps', 'tukar', 'kks', 'sim', 'kartu', 'kis', 'terima', 'kartu', 'terima', 'kasih'], ['tanah', 'pt', 'kai', 'gusur', 'rw', 'pejaten', 'timur', 'ps', 'minggu', 'dagang', 'pungut', 'uang', 'oknum'], ['kps', 'pegang', 'kartu', 'kps', 'tahun', 'sekolah', 'terima', 'uang', 'bantu', 'kartu', 'kps', 'pegang', 'kartu', 'kps'], ['butuh', 'bsm', 'terima', 'kks'], [], ['ttangga', 'rata'], ['pegang', 'dana'], ['kartu', 'sekolah', 'terima', 'kks'], ['warga', 'sukabumi', 'selatan', 'harga', 'raskin', 'capai', 'rupiah', 'karung', 'kilogram', 'sepedrti', 'rw', 'jawabnannya', 'tebus', 'raskin', 'hranya', 'segtu'], ['nomor', 'kks', 'dtmpt', 'lom', 'dana', 'kantor', 'pos', 'daerah', 'dtukar', 'kartu'], ['selamat', 'siang', 'bpjs', 'adik', 'transfusi', 'darah', 'tanggung', 'kantong', 'tanggung', 'bpjs', 'butuh', 'darah', 'kant', ['ratni', 'bantu', 'dana', 'usamawan', 'lil', 'usamawan'], ['rumah', 'tangga', 'desa', 'raskin'], ['terima', 'kks', 'kis', 'simkartu'], ['kecewa', 'kantor', 'pos', 'dana', 'kps', 'kartu', 'kartu', 'perdana', 'cair', 'kasir', 'danany', 'kosong', 'bgt', 'curang', 'pegawai', 'pos', 'mengambil'], ['pagi', 'kartu', 'indonesia', 'pintar', 'lulus'], ..., ['bantu', 'kompeten', 'guru', 'ajar', 'tingkat', 'sma', 'sma', 'negeri', 'mangkuratana', 'kabupaten', 'luwu', 'timur', 'sertifikasi', 'guru', 'ajar', 'fisika', 'dimut']]

```

Lampiran 4. *Syntax* pre-processing.

```

import pandas as pd
import pandas as dataframe
import string
import nltk
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from nltk.tokenize import word_tokenize

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from IPython.display import display
from wordcloud import WordCloud, STOPWORDS,
ImageColorGenerator
from pylab import figure, axes, pie, title, show
%matplotlib inline

data= pd.read_csv("E://Tugas
Akhir/data_final_edit_6kat.csv", engine="python", sep=";")

#data
datalaporan=data['IsiLaporan']
datalaporan=datalaporan.astype(str)
print(datalaporan)

#lowercase
data_lower=[]
for line in datalaporan:
    a=line.lower()
    data_lower.append(a)
print(data_lower)

#hashtag
datanohashtag = []
for line in data_lower:
    result = re.sub(r"3(\w+)", "",line)
    datanohashtag.append(result)
print(datanohashtag)

#username
datanousername = []

```

```

for line in datanohashtag:
    result = re.sub(r"@S+", "", line)
    datanousername.append(result)
print(datanousername)

#punctuation
datanopunctuation=[]
for line in datanousername :
    result=re.sub(r"[^\w\s]", " ", line)
    datanopunctuation.append(result)
print(datanopunctuation)

#Angka
datanonumber=[]
for line in datanopunctuation :
    result=re.sub("\d", " ", line)
    datanonumber.append(result)
print(datanonumber)

#Clear Space Enter
dataclearspace = []
for line in datanonumber:
    result=re.sub(r"\s+", " ", line)
    dataclearspace.append(result)
print(dataclearspace)

#Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()
datastemmed = map(lambda x: stemmer.stem(x),
dataclearspace)
datastemmed = list(datastemmed)

#Sinonim Kata
kata = {"terimakasih":"terima kasih", "trims":"terima
kasih", "trim":"terima kasih", "nuhun":"terima kasih",
"beroprasi":"operasi", "antri":"antre", "boleh":"boleh", "
bis ":" bus ", "cepat":"cepat", "complain":"komplain",
"dtg":"datang", "dmn":"dimana", "diem":"diam",
"gaada":"tidak ada", "gabisa":"tidak bisa", "gatau":"tidak
tahu", "hour":"jam", "jln":"jalan", " ... , "mengantre":"antre", "
nerima":"terima", "bener":"benar", "ngetem":"mengetem", "nomer"
:"nomor", "ektp":"ktp", "propinsi":"provinsi", "mesjid":"masjid

```

```

", "pami": "jika", "impormasi": "informasi", "error": "eror", "tempat": "tempat", "utk": "untuk"}

stopwords=open('E://stopwords1.txt','r').read()

print(stopwords)

from collections import OrderedDict
def replace_all(datastemmed, dic):
    for i, j in dic.items():
        datastemmed = datastemmed.replace(i, j)
    return datastemmed
dic = OrderedDict(kata)

datachange = []
for line in datastemmed:
    result = replace_all(line, dic)
    datachange.append(result)

laporanawal=[]
laporanfinal=[]
df=[]
for line in datachange:
    word_token = word_tokenize(line)
    word_token = [word for word in word_token if not word in
stopwords and not word[0].isdigit()]
    laporanfinal.append(word_token)
    df.append(" ".join(word_token))

```

Lampiran 5. Syntax *Count Vectorized* dan *TF-IDF*.

```
#Count Vectorizer
from pandas import DataFrame
from sklearn.feature_extraction.text import CountVectorizer
vect = CountVectorizer(min_df=10)
X_dtm = vect.fit_transform(df)
X_dtm = X_dtm.toarray()
cv=DataFrame(X_dtm, columns=vect.get_feature_names())
print(cv)

#tf-idf
from sklearn.feature_extraction.text import TfidfTransformer
tfidf = TfidfTransformer(use_idf=True).fit_transform(cv)
tfidf_train = (tfidf.toarray())
print (tfidf_train)
print (tfidf_train.shape)
tf = DataFrame(tfidf.A,
columns=vect.get_feature_names())
print (tf)
```

Lampiran 6. Syntax *Splitting Data* dan *SMOTE*

```
#Splitting Data
from sklearn.model_selection import
train_test_split,cross_val_score

# Random training and testing data
X_train,X_test,Y_train,Y_test=train_test_split(x_new,y,test_
size=0.1,shuffle=False,random_state=0)

#Balancing Data
import imblearn
import numpy
import numpy as np
from imblearn.over_sampling import SMOTE
sm = SMOTE(random_state=0)
X_train_res, y_train_res = sm.fit_sample(X_train, Y_train)
unique, counts = np.unique(y_train_res, return_counts=True)
print(list(zip(unique, counts)))
```

Lampiran 7. *Syntax* Artificial Neural Network

```

from sklearn.neural_network import MLPClassifier
from sklearn.metrics import
classification_report, confusion_matrix
from sklearn import metrics
from sklearn.metrics import accuracy_score

nn=()
nn =
MLPClassifier(hidden_layer_sizes=(4,), activation='relu', rand
om_state=0, max_iter=1000, verbose=True, batch_size=128)
y_score = nn.fit(X_train_res, y_train_res)
y_pred = nn.predict(X_test)
y_pred_proba = nn.predict_proba(X_test)

Weight_1 = nn.coefs_[0]
print(Weight_1)
np.savetxt("Weight_hidden.csv", Weight_1, delimiter=";")
Weight_2 = nn.coefs_[1]
print(Weight_2)
np.savetxt("Weight_output.csv", Weight_2, delimiter=";")
Bias_0 = nn.intercepts_[0]
print(Bias_0)
np.savetxt("Bias_0.csv", Bias_0, delimiter=";")
Bias_1 = nn.intercepts_[1]
print(Bias_1)
np.savetxt("Bias_1.csv", Bias_1, delimiter=";")

print(confusion_matrix(Y_test, y_pred))
print()
print(classification_report(Y_test, y_pred, digits=3))
print("Akurasi Score :
{: .3f}".format(accuracy_score(Y_test, y_pred)))

```


Lampiran 8. *Syntax* Artificial Neural Network dengan *K*-fold

```

from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import StratifiedShuffleSplit
from sklearn.neural_network import MLPClassifier
from sklearn import metrics
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import
classification_report,confusion_matrix
from sklearn import metrics
from sklearn.metrics import accuracy_score

X_baru, y = x_new, y
kf = StratifiedKFold(n_splits=10,
shuffle=False,random_state=0)
kf.get_n_splits(X_baru)
kf.get_n_splits(y)
cl =
MLPClassifier(hidden_layer_sizes=(4,),activation='relu',rand
om_state=0,max_iter=1000)
nb_classes=6

for train, test in kf.split(X_baru, y):
    (X_baru[train],X_baru[test])
    (y[train],y[test])
    X_train_os, y_train_os = sm.fit_sample(X_baru[train],
y[train])
    a = cl.fit(X_train_os, y_train_os)
    b = cl.predict(X_baru[test])
    Yb_train = np_utils.to_categorical(y_train_os,
nb_classes)
    Yb_test = np_utils.to_categorical(y[test], nb_classes)
    Yb_pred = np_utils.to_categorical(b, nb_classes)
    print("Klasifikasi Artificial Neural Network dengan 428
Feature")
    print()
    print (confusion_matrix(y[test],b))
    print()
    print (classification_report(y[test],b,digits=3))

```

Lampiran 9. Syntax Word Cloud

```

from subprocess import check_output
from wordcloud import WordCloud, STOPWORDS
#mpl.rcParams['figure.figsize']=(8.0,6.0)
mpl.rcParams['font.size']=10
mpl.rcParams['savefig.dpi']=300
mpl.rcParams['figure.subplot.bottom']=.1
wordcloud = WordCloud(
    background_color='white',
    stopwords=stopwords,
    max_words=100,
    max_font_size=300,
    random_state=0
).generate(str(aa2[0]))
print(wordcloud)
fig = plt.figure(1)
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
fig.savefig("wCkat2.png", dpi=300)

```

Lampiran 10. Output Weight and Bias

```

Weight_1 = nn.coefs_[0]
[[-0.56834771 -0.48045755 -0.85865837]
 [-0.62908451 -0.00757292 -0.41813109]
 [-0.20997729 0.3388631 -0.06497094]
 ...
 [ 0.43683727 -0.20417856 -0.69532931]
 [-0.09843454 0.02005189 -0.31053023]
 [-0.16789285 -0.27807983 -0.03820482]]

Weight_2 = nn.coefs_[1]

[[-4.53383033 -0.97644753 1.61564209 -3.77597033 1.1059498
 1 6.45256696]
 [-4.49403071 -6.49055598 5.42521898 4.17822871 1.1416528
 1 0.26996954]
 [-3.88827645 3.99722995 -0.78164444 2.01429792 -3.8260908
 8 2.59345557]]
Bias_0 = nn.intercepts_[0]

[0.91030603 0.80127223 1.2497631 ]
Bias_1 = nn.intercepts_[1]

[ 9.01641682 1.70510634 -3.74343162 -0.38220805 4.14212688
-9.06876279]

```

Lampiran 11. Hasil Chi Square

Kata ke- <i>i</i>	Nilai Chi-Square	P-Value	Keputusan Berdasarkan Chi-Square	Keputusan Berdasarkan P-Value
1	3.747	0.586	Gagal Tolak H_0	Gagal Tolak H_0
2	11.288	0.046	Tolak H_0	Tolak H_0
3	6.999	0.221	Gagal Tolak H_0	Gagal Tolak H_0
4	15.592	0.008	Tolak H_0	Tolak H_0
5	8.122	0.150	Gagal Tolak H_0	Gagal Tolak H_0
6	9.9095	0.0778	Gagal Tolak H_0	Gagal Tolak H_0
7	7.5940	0.1801	Gagal Tolak H_0	Gagal Tolak H_0
8	11.4369	0.0434	Tolak H_0	Tolak H_0
9	68.1704	0.0000	Tolak H_0	Tolak H_0
10	33.1871	0.0000	Tolak H_0	Tolak H_0
11	4.2834	0.5094	Gagal Tolak H_0	Gagal Tolak H_0
12	10.4111	0.0644	Gagal Tolak H_0	Gagal Tolak H_0
13	28.7760	0.0000	Tolak H_0	Tolak H_0
14	10.7099	0.0574	Gagal Tolak H_0	Gagal Tolak H_0
15	57.2714	0.0000	Tolak H_0	Tolak H_0
16	17.3813	0.0038	Tolak H_0	Tolak H_0
17	18.8199	0.0021	Tolak H_0	Tolak H_0
18	21.1852	0.0007	Tolak H_0	Tolak H_0
19	4.8902	0.4294	Gagal Tolak H_0	Gagal Tolak H_0
20	4.5676	0.4709	Gagal Tolak H_0	Gagal Tolak H_0
21	16.6497	0.0052	Tolak H_0	Tolak H_0
∴	∴	∴	∴	∴
692	1321.655	0.000	Tolak H_0	Tolak H_0
693	8.4358	0.1338	Gagal Tolak H_0	Gagal Tolak H_0
694	6.6908	0.2447	Gagal Tolak H_0	Gagal Tolak H_0
∴	∴	∴	∴	∴
949	15.867	0.007	Tolak H_0	Tolak H_0
950	1.586	0.903	Gagal Tolak H_0	Gagal Tolak H_0

Lampiran 12. Surat Keterangan Pengambilan Data

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : Mochamad Ihsan Ananto

NRP : 06211540000123

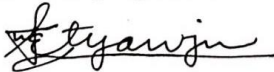
menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini merupakan data sekunder yang diambil dari ~~penelitian / buku/ Tugas Akhir/ Thesis/~~ publikasi lainnya yaitu:

Sumber : <http://www.data.go.id/dataset/data-aspirasi-dan-pengaduan-masyarakat>

Keterangan : Data aspirasi dan pengaduan masyarakat terkait program pembangunan nasional dan pelayanan publik pemerintah yang diperoleh melalui sistem LAPOR!

Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui
Pembimbing Tugas Akhir



(Dra. Wiwiek Setya Winahju, M.S)
NIP. 19560424 198303 2 001

Surabaya, 17 Juni 2019



(Mochamad Ihsan Ananto)
NRP. 06211540000123

*(coret yang tidak perlu)

BIODATA PENULIS



Penulis dilahirkan di Jakarta, 4 Mei 1997 dengan nama lengkap Mochamad Ihsan Ananto dan biasa dipanggil Ihsan atau Ananto. Penulis menempuh pendidikan formal di SD Negeri 2 Palapa Bandar Lampung, SMP Negeri 19 Jakarta, dan SMA Negeri 47 Jakarta, lalu melanjutkan pendidikan formal yakni jenjang perguruan tinggi di Departemen Statistika ITS tahun 2015. Selama masa perkuliahan,

penulis aktif di berbagai organisasi seperti menjadi sebagai Staff Departemen PSDM HIMASTA-ITS pada periode 2016-2017, KOOR SC BCS pada periode 2017-2018, dan Staff Kementerian Komunikasi dan Informatika BEM ITS. Selain itu penulis juga berkesempatan untuk menimba ilmu di berbagai kepanitiaan yakni sebagai Staff *Public Relations* ITS EXPO 2016 dan Staff Ahli *Public Relations* ITS EXPO 2017. Bagi pembaca yang ingin berdiskusi, memberikan saran, maupun kritik mengenai Tugas Akhir ini dapat langsung menghubungi penulis melalui *e-mail* anantosan97@gmail.com atau melalui nomor 082113699323.