Klasifikasi Kategori Pengaduan Masyarakat Melalui Kanal LAPOR! Menggunakan *Artificial Neural Network*

Mochamad Ihsan Ananto, Wiwiek Setya Winahju dan Kartika Fithriasari
Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: *ananto15@mhs.statistika.its.ac.id, wiwiek@statistika.its.ac.id, kartika_f@statistika.its.ac.id*

***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 masya-rakat. 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 setiap kategori aduan dengan menggunakan *word cloud*.**

Kata Kunci*— *Artificial Neural Network, LAPOR!, SMOTE, Text Mining, Word Cloud

I. PENDAHULUAN

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. Tak hanya bagi individu, tetapi berbagai sektor seperti sektor pemerintahan telah memanfaatkan internet untuk berbagai tujuan.

Salah satu bentuk teknologi informasi yang mendukung pemanfaat internet di bidang pemerintahan adalah penggunaan aplikasi *electronic government* (*e-government*). *E-government* bertujuan untuk membentuk sistem pelayanan pemerintahan yang lebih mudah dan cepat dengan menggunakan situs internet. 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, pengaduan masyarakat mengenai pelayanan publik diharapkan dapat ditangani dengan cepat, transparan, dan akuntabel

SP4N diwujudkan melalui kanal LAPOR! yaitu sarana aspirasi dan pengaduan berbasis media sosial yang 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. Masyarakat dapat melakukan pengaduan via 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! untuk selanjutnya diteruskan kepada intansi atau dinas terkait.

Laporan pengaduan masyarakat yang masuk melalui LAPOR! tersebut hadir dalam bentuk teks, sehingga dapat diselesaikan dengan cara *text mining*. *Text mining* merupakan proses menggali informasi secara intensif dimana pengguna berhadapan dengan sekumpulan dokumen menggunakan *tools* analisis *data mining* [1]. Proses yang dapat dilakukan dengan *text mining* di antaranya adalah *text clustering* dan *text classification*.

Berdasarkan permasalahan yang ada dalam LAPOR!, maka klasifikasi teks merupakan proses yang tepat karena data aduan masyarakat yang masuk telah terklasifikasikan dalam kategori tertentu. 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 [2] dengan menggunakan algoritma SVM. Peneltian tentang metode ANN pernah dilakukan oleh Reyhana & Fithriasari [3] tentang analisis sentimen perkembangan infrastruktur di Kota Surabaya. Pada penelitian tersebut didapatkan hasil performa klasifikasi yang dihasilkan melalui metode ANN lebih baik dari SVM..

Oleh karena itu pada penelitian ini akan dilakukan *text mining* dengan metode klasifikasi menggunakan *Artificial Neural Networks* (ANN) karena mampu mempelajari model non-linier dan data yang besar. Nantinya akan didapatkan karakteristik aduan masyarakat yang masuk via LAPOR!, berapa tingkat ketepatan klasifikasi yang didapatkan, dan kata kunci apa yang harus menjadi perhatian dinas terkait di tiap kategorinya. Melalui penelitian ini diharapkan dapat memberikan masukan tambahan bagi pihak LAPOR! terkait kategori klasifikasi aduan yang masuk skala prioritas nasional sehingga dapat mempercepat penanganan aduan oleh instansi terkait.

II. TINJAUAN PUSTAKA

A. Text Mining

Text mining bertujuan untuk mengekstrak informasi yang berguna dari dokumen berupa teks yang tidak terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu *text classification* dan *text clustering* dengan menggunakan *tools* analisis yang berhubungan dengan *data mining* [1]. *Text classification* merupakan proses untuk membentuk kelas-kelas dari dokumen berdasarkan kelas kelompok yang sudah diketahui sebelumnya atau *supervised learning*.

Text pre-processing merupakan tahap awal dalam *text mining* sebelum analisis dilakukan. Data teks harus melalui tahap *pre-processing* terlebih dahulu agar data teks yang tidak terstruktur menjadi lebih terstruktur [1]. Tahapan *pre-processing* yang dilakukan yakni diantaranya adalah *data cleaning*, *case folding*, *stemming*, *tokenizing,* dan *stopwords removal*.

1. *Data Cleaning,* yaitu membersihkan data teks dari kata yang tidak diperlukan untuk mengurangi *noise*. Kata yang

dihilangkan dalam dokumen teks antara lain karakter HTML, *emoticons*, *hashtag* (#), dan URL.
2. *Case Folding,* yaitu merupakan proses untuk mengubah semua karakter teks menjadi non kapital serta menghilangkan tanda baca dan angka [4].
3. *Stemming,* yakni proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, serta kombinasi dari awalan dan akhiran.
4. *Stopwords Removal,* merupakan proses untuk menghapus kata-kata yang tidak memiliki arti yang relevan. 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 [5]
5. *Tokenizing,* yakni proses untuk memecah keseluruhan data teks yang sebelumnya berupa kalimat menjadi kata per kata.

B. *Term Frequency Inverse Document Frequency*

Term Frequency Inverse Document Frequency (TF-IDF) merupakan metode pembobotan yang dilakukan untuk ekstraksi data teks. Tujuan dari TF-IDF adalah menemukan jumlah kata yang diketahui (*tf*) setelah dikalikan dengan frekuensi aduan dimana suatu kata tersebut muncul (*idf*). Metode TF-IDF dilakukan dengan menghitung bobot dengan cara integrasi antara *term frequency* (*tf*) dan *inverse document frequency* (*idf*) [6]. Berikut merupakan rumus untuk memperoleh nilai TF-IDF.

$$w_{ij} = tf_{ij} \times idf_j \,, \tag{1}$$

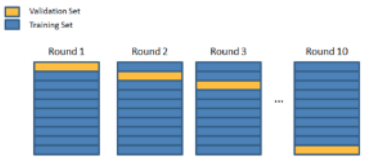
$$idf_i = log\left(\frac{N}{df_i}\right), \tag{2}$$

keterangan:

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

C. *K-fold Cross Validation*

Metode *K-fold cross validation* digunakan untuk mengevaluasi kinerja klasifikasi. Metode ini digunakan untuk mengurangi bias terkait pengambilan sampel dari data. *K-fold cross validation* membagi data kedalam sejumlah bagian yang telah ditentukan secara acak, yang disebut *fold* [7]. 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 1.



Gambar 1. Ilustrasi Pembagian Data

Salah satu variasi dari metode ini adalah *Stratified K-Fold Cross Validation, yakni* untuk menjaga agar setiap *fold* mengandung persentase sampel yang sama dari setiap kelas.

D. *Feature Selection*

Feature selection merupakan proses pemilihan variabel yang relevan untuk digunakan dalam pembentukan model. *Feature selection dapat* mempercepat proses sehingga nantinya akan didapatkan performa yang lebih tinggi [8]. *Feature selection* yang digunakan dalam penelitian ini adalah *Chi-Square* (χ^2) karena menghasilkan performa yang bagus terutama untuk data *multiclass* [9], hipotesis yang digunakan adalah 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)

$$\chi^2 = \sum_{k=1}^{r}\sum_{l=1}^{c}\frac{(n_{kl}-\hat{\mu}_{kl})^2}{\hat{\mu}_{kl}}, \tag{3}$$

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 $\chi^2 > \chi^2_{\alpha}$ dengan (df = n - 1), maka tolak H_0 pada tingkat signifikansi yang digunakan. Jika sebaliknya, maka gagal tolak H_0. [9].

E. *Synthetic Minority Oversampling Technique (SMOTE)*

SMOTE merupakan salah satu metode untuk mengatasi data *imbalance* yang diusulkan oleh Chawla dkk [10]. 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 sintetis berdasarkan tetangga terdekat *k-nearest neighbour* dimana tetangga terdekat dipilih berdasarkan jarak *euclidean* antara kedua data [10].

Pembangkitan data sintetis untuk kelas minor dilakukan dengan menggunakan persamaan berikut:

$$\mathbf{x}_{syn} = \mathbf{x}_i + (\mathbf{x}_{knn} - \mathbf{x}_i)\gamma, \tag{4}$$

keterangan:

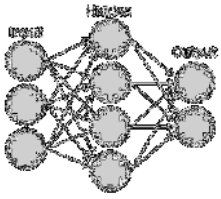
$\mathbf{x}_{syn}$ = data sintetis,
$\mathbf{x}_i$ = data ke-i dari kelas minor,
$\mathbf{x}_{knn}$ = data dengan jarak terdekat dari data yang akan direplikasi,
γ = bilangan random antara 0 – 1.

F. *Artificial Neural Network*

Artificial Neural Network (ANN) merupakan salah satu sistem pemrosesan informasi dengan desain menirukan cara kerja otak manusia. Pembuatan ANN terinspirasi dari kesadaran atas *complex learning system* pada otak yang terdiri dari set-set neuron yang saling berhubungan secara dekat [11].

Kelebihan dari pemodelan ANN diantaranya yakni; tidak memerlukan banyak asumsi, dapat memodelkan sistem secara non-linier dengan baik, dan mampu memberikan model yang mendekati sistem nyata [12]. Arsitektur *Neural Network* yang tergambarkan melalui Gambar 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*.

Setiap *input* yang terhubung ke tiap *nodes* pada *hidden layer* maupun *output layer* masing-masing memiliki *bias* dan *weight*.



Gambar 2. Arsitektur ANN

Nantinya akan ada perhitungan pada *nodes hidden layer* tanpa fungsi aktivasi sebagaimana persamaan (5) dan setelah fungsi aktivasi yang dijelaskan melalui persamaan (6). Selanjut-nya dilakukan perhitungan untuk mendapatkan nilai prediksi yang dijelaskan melalui persamaan (7) dan (8).

$$p_j = a_j + \sum_{j=1}^{n_x} w_{ij} x_i, \tag{5}$$

$$q_j = f(p_j), \tag{6}$$

$$r_l = b_l + \sum_{l=1}^{n_h} v_{jl} q_j, \tag{7}$$

$$s_l = f(r_l), \tag{8}$$

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*,
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* [13]. 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 (9).

$$q_j = f(p_j) = \max(0, p_j) = \begin{cases} p_j, & \text{jika } p_j \geq 0, \\ 0, & \text{jika } p_j < 0, \end{cases} \tag{9}$$

dengan:
p_j = *input* pada setiap *nodes*,
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 merupa-kan probabilitas untuk setiap kelasnya dengan kelas target mempunyai probabilitas tertinggi.

$$s_j = f(r_j) = \frac{\exp(r_j)}{\sum_{i=1}^{k} \exp(r_i)}, \tag{10}$$

dengan:
r_l = *input* pada setiap *nodes*,
s_l = *output* pada setiap *nodes* setelah dimasukkan fungsi aktivasi,
k = banyak kelas yang digunakan.

G. Ketepatan Klasifikasi

Untuk melihat performa klasifikasi yang telah dilakukan, maka dilakukan pengukuran ketepatan klasifikasi. 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 [14].

Tabel 1. *Confusion Matrix* Untuk *Binary Classification*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	*tp*	*Fn*
Negatif	*fp*	*tn*

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 penghitungan ini dapat dilihat melalui *confusion matrix* untuk kasus *binary classification* yang terdapat pada Tabel 1 sesuai dengan penelitian Sokolova dan Lapalme [15]. Sedangkan untuk kasus *multiclass classification* seperti pada penelitian ini dapat dilihat melalui Tabel 2.

Tabel 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. 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*) [15]. Berikut merupakan beberapa kriteria untuk menilai ketepatan klasifikasi.

$$Precision = \frac{\sum_{a=1}^{6} tp_a}{\sum_{a=1}^{6} (tp_a + fpa)}, \tag{11}$$

$$Sensitivity = \frac{\sum_{a=1}^{6} tp_a}{\sum_{a=1}^{6} (tp_a + fna)}, \tag{12}$$

$$F1\ Score = 2 \times \left(\frac{(Precision \times Sensitivity)}{(Precision + Sensitivity)} \right). \tag{13}$$

H. Word Cloud

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 ukuran huruf kata tersebut. Semakin besar ukuran

kata, maka semakin besar frekuensi kata tersebut muncul dalam dokumen. [16].

I. LAPOR!

Layanan Aspirasi dan Pengaduan Online Rakyat merupakan 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. Berdirinya LAPOR! merupakan amanat dari UU No. 3 Tahun 2015 untuk membentuk Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N) yang diharapkan dapat ditangani dengan cepat, transparan, dan akuntabel sesuai dengan kewenangan instansi terkait. Masyarakat dapat melakukan pengaduan via LAPOR! melalui *website* www.-lapor.go.id, SMS 1708, aplikasi dan twitter @LAPOR1708.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder dari situs data.go.id yang diunggah oleh admin dari LAPOR!. Data yang digunakan merupakan rekapitulasi dari aduan masyarakat melalui LAPOR! terkait kinerja pemerintah di berbagai daerah pada tahun 2015. Total data yang digunakan yakni sebanyak 6916 aduan yang terbagi kedalam 6 kategori yaitu energi pangan dan maritim, infrastruktur dan transportasi, kesehatan, pendidikan, reformasi birokrasi, serta pariwisata dan lingkungan hidup.

B. Variabel Penelitian

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.

Tabel 3 Variabel Penelitian

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

Struktur data awal yang memuat isi laporan dan kategori aduan sebelum dilakukan *text pre-processing* ditunjukkan melalui Tabel 4.

Tabel 4 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 me-ngenai gunungan sampah yang me-numpuk 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

C. Langkah Analisis

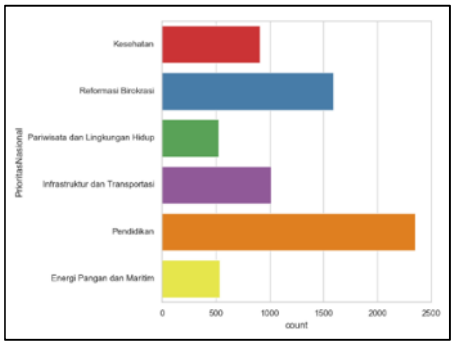
Langkah analisis yang akan dilakukan pada penelitian ini adalah sebagai berikut.

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* dengan langkah sebagai berikut;
   a. *Data cleaning*, yaitu membersihkan data teks dari kata yang tidak diperlukan untuk mengurangi *noise* dengan menghapus tautan berupa URL, menghilangkan angka, spasi ganda, dan tanda baca.
   b. *Case folding,* yaitu proses untuk mengubah kata menjadi non kapital dan menghilangkan tanda baca maupun angka,
   c. *Stemming,* yaitu menghilangkan imbuhan pada setiap kata untuk mendapatkan kata kunci atau kata dasar,
   d. *Stopwords removal,* yakni merupakan proses yang dilakukan untuk menghapus kata-kata yang tidak memiliki arti yang relevan sesuai dengan penelitian F.Z Tala [17] ,
   e. *Tokenizing,* yakni memecah kalimat isi aduan menjadi kata per kata,
   f. Melakukan pembobotan kata dengan TF-IDF,
   g. Melakukan *feature selection* atau seleksi variabel menggunakan χ^2. Yakni dengan membandingkan nilai χ^2 dengan χ^2_α , dimana jika keputusannya adalah gagal tolak H_0 maka kata tersebut tidak berpengaruh terhadap kategori aduan masyarakat.
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 metode *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 (11), (12), dan (13)
7. Melakukan visualisasi kata di tiap hasil klasifikasi dengan melakukan *plotting* kata-kata yang sering muncul menggunakan *word cloud*.

IV. ANALISIS DAN PEMBAHASAN

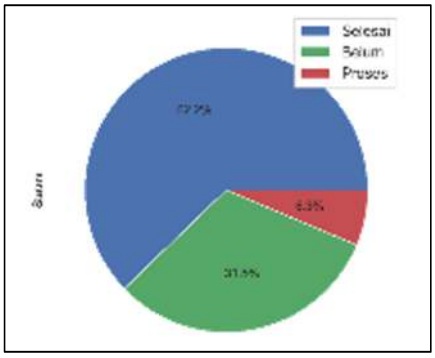
A. Karakteristik Data Aduan Masyarakat

Data yang digunakan merupakan data aduan masyarakat via kanal LAPOR! pada tahun 2015 yang diambil melalui situs data.go.id. Aduan tersebut berjumlah 6916 yang terbagi ke dalam enam kategori aduan. Melalui data tersebut, dapat diketahui bagaimana karakteristik data awal sebelum proses analisis. Data tersebut memuat isi aduan, kategori, dan status tindak lanjut aduan. Terlebih dahulu ditunjukan jumlah aduan masyarakat di tiap kategorinya melalui Gambar 3.



Gambar 3. Jumlah Aduan Masyarakat Tiap Kategori

Gambar 3 menunjukkan jumlah aduan masyarakat yang diterima oleh LAPOR! selama tahun 2015. Diketahui bahwa selama tahun 2015, permasalahan Pendidikan merupakan kategori yang paling sering diadukan oleh masyarakat dengan jumlah aduan sebesar 2348. Hal tersebut disinyalir karena pada tahun 2015 sejalan dengan program pemerintah dalam pembagian Kartu Indonesia Pintar (KIP), sehingga banyak aduan dari masyarakat terkait distribusi KIP yang belum berjalan dengan baik. Sedangkan aduan paling sedikit yang diterima oleh LAPOR! adalah aduan mengenai Pariwisata dan Lingkungan Hidup dengan jumlah aduan hanya sebanyak 526 aduan. Data aduan masyarakat via LAPOR! tersebut selanjutnya diteruskan ke dinas terkait untuk ditindaklanjuti serta dapat dilacak status aduannya oleh pelapor.



Gambar 4. Persentase Status Aduan Masyarakat

Dari Gambar 4 diketahui status tindak lanjut aduan masyarakat oleh dinas terkait. Mayoritas dari data aduan masyarakat yang masuk via LAPOR! yakni sebanyak 4301 aduan telah selesai ditindaklanjuti oleh instansi terkait. Sedangkan untuk status aduan yang masih dalam proses tindak lanjut, diakibatkan karena masih dalam proses tindak lanjut hingga data tersebut diunggah atau tidak adanya pembaruan status aduan dari dinas terkait.

B. *Pre Processing Data*

Sebelum dilanjutkan analisis, dilakukan *pre-processing* data terlebih dahulu. Dalam kasus data teks, beberapa tahapan *pre-processing* yang dilakukan yakni *case folding*, yaitu proses untuk mengubah semua karakter teks menjadi non kapital serta menghilangkan tanda baca dan angka. *Data cleaning*, pembersihan data teks dari kata yang tidak diperlukan dengan cara menghapus tautan berupa URL, menghilangkan angka, spasi ganda, dan tanda baca. *Stemming*, 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 *tokenizing,* dan *stopwords removal*. Sehingga data aduan masyarakat sebelum dan sesudah semua proses *pre-processing* tersebut dijelaskan melalui Tabel 5.

Tabel 5. Contoh Data Sebelum dan Sesudah *Pre-Processing*

Sebelum *Pre-Processing*	Setelah *Pre-Processing*
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	'kps' 'tempat' 'pegang' 'kartu' 'kps' 'tahun' 'sekolah' 'terima' 'uang' 'bantu' 'kartu' 'kps' 'pegang' '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', 'jalan', 'cipedes', 'rumah', 'nomor', 'kota', 'bandung', 'mati', 'gelap', 'butuh', 'biaya'

Hasil dari *pre-processing* yaitu tahap *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 6.

Tabel 6. *Count Vectorizer* kata dalam aduan

No	Kategori	Kata Kunci				
		acara	bpjs	dokter	…	yogyakarta
1	2	0	2	1	…	0
2	4	0	1	0	…	0
⋮	⋮	⋮	⋮		⋮	⋮
253	4	1	0	0	…	0
⋮	⋮	⋮	⋮		⋮	⋮
5063	2		2	1	…	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
6915	3	0	0	0	…	0

Berdasarkan Tabel 6 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 TF-IDF menggunakan persamaan (1) dan (2).

Selanjutnya yakni dilakukan proses seleksi variabel yang relevan untuk digunakan dalam pembentukan model. Keuntungan dari proses seleksi variabel ini adalah untuk mempercepat proses serta menghasilkan 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 7.

Berdasarkan Tabel 7 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' dengan nilai *Chi-Square* sebesar 1321,65 dan P-*value* sebesar 0,000.

Tabel 7. Nilai χ^2 untuk setiap kata

Kata ke-*i*	Nilai Chi-Square	P-Value	Keputusan
1	3.747	0.586	Gagal Tolak H_0
2	11.288	0.046	Tolak H_0
3	6.999	0.221	Gagal Tolak H_0
4	15.592	0.008	Tolak H_0
5	8.122	0.150	Gagal Tolak H_0
⋮	⋮	⋮	⋮
692	1321.655	0.000	Tolak H_0
⋮	⋮	⋮	⋮
949	15.867	0.007	Tolak H_0
950	1.586	0.903	Gagal Tolak H_0

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 seleksi variabel sebanyak 428 kata.

C. *Oversampling Data* dengan SMOTE

Melalui bagian A, 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. Data dibagi menjadi data *training* dan data *testing* dengan perbandingan 90:10. Jumlah keseluruhan data *training* dan data *testing* yang digunakan dalam analisis ditunjukkan melalui Tabel 8.

Tabel 8. Jumlah Data *Training*

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

Berdasarkan Tabel 8, diketahui bahwa kategori 3 yakni kategori Pendidikan merupakan kelas mayoritas. Sehingga perlu dilakukan *oversampling* dengan metode SMOTE pada data *training* agar kategori lainnya 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 jumlah data pada kelas mayor. Sehingga data *training* telah seimbang untuk digunakan dalam analisis klasifikasi dengan menggunakan metode *Artificial Neural Network*.

D. *Klasifikasi Menggunakan Artificial Neural Network*

*Artificial Neural Network (*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 [18], 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* [19]. Sedangkan metode yang digunakan untuk melakukan optimasi parameter *weight* dan *bias* adalah *Stochastic Gradient Descent* (SGD).

Langkah awal yang dilakukan untuk melakukan klasifikasi data aduan masyarakat dengan ANN adalah dengan melakukan pemilihan semua *feature* atau 428 *feature* serta jumlah *nodes hidden layer* yang menghasilkan ketepatan klasifikasi terbaik. Yakni dengan membandingkan nilai *macro averaging* dari kriteria ketepatan klasifikasi yaitu *precision, sensitivity,* dan *F1-Score*. Hasil ketepatan klasifikasi dari masing-masing pemilihan *feature* dan jumlah *nodes* ditunjukkan pada Tabel 10.

Tabel 10. 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

Berdasarkan Tabel 10, didapatkan ketepatan klasifikasi terbaik yakni dengan menggunakan 428 *feature* sebagai *input* dan menggunakan 3 *nodes hidden layer*. Sehingga dapat disimpulkan bahwa *feature selection* mampu meningkatkan nilai ketepatan klasifikasi. Untuk *confusion matrix* dari hasil klasifikasi, dijelaskan melalui Tabel 11.

Tabel 11. Confusion Matrix pada *fold* ke-6

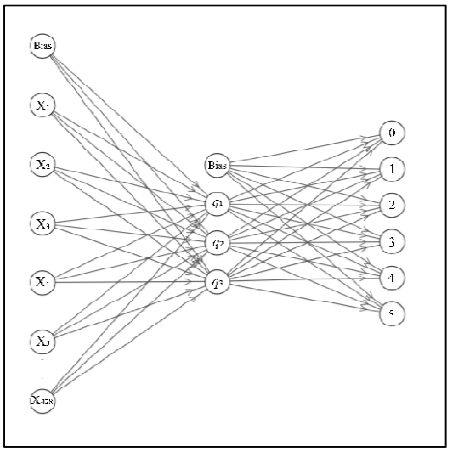
Kelas Aktual	Kelas Prediksi					
	0	1	2	3	4	5
0	42	6	0	1	4	1
1	3	93	0	1	5	4
2	0	0	157	5	2	5
3	0	3	2	50	2	0
4	3	17	30	11	144	15
5	0	8	1	1	8	68

Tabel 11 menunjukkan kinerja klasifikasi dari *Artificial Neural Network*. Dari Tabel 11 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 ke dalam kategori 3, 4 aduan pada kategori 4, dan 1 aduan ke dalam kategori 5. Hal ini juga berlaku untuk kategori aduan lainnya. Sehingga hasil ketepatan klasifikasi yang dihasilkan dijelaskan melalui tabel 12.

Tabel 12. 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* pada *hidden layer* ditunjukkan melalui Gambar 5.



Gambar 5. Jaringan ANN dengan 428 *feature* dan 3 *nodes hidden layer*

Selanjutnya yakni melakukan *K-fold cross validation* untuk mengetahui tingkat kesalahan teknik klasifikasi. 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 13.

Tabel 13. 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 ketapatan 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.

$$p_1 = 0{,}91 + (-0{,}5683\, x_1 - 0{,}6291\, x_2 - 0{,}2100\, x_3 + \ldots - 0{,}1679\, x_{428}),$$
$$p_2 = 0{,}8013 + (-0{,}4805\, x_1 - 0{,}0076\, x_2 + 0{,}3389\, x_3 + \ldots - 0{,}2781\, x_{428}),$$
$$p_3 = 1{,}2489 + (-0{,}8587\, x_1 - 0{,}4181\, x_2 - 0{,}0650\, x_3 + \ldots - 0{,}0382\, x_{428}).$$

Setelah melalui fungsi aktivasi ReLu pada *hidden layer*, maka persamaan pada *output layer* dengan fungsi aktivasi *Softmax* adalah sebagai berikut.

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

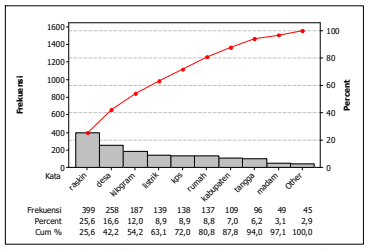
Melalui hasil di atas, akan dihasilkan probabilitas dari suatu aduan masyarakat yang masuk via LAPOR!. Yakni kelas target dari aduan tersebut adalah kategori yang memiliki probabilitas tertinggi. Yakni s_0 untuk kategori energi pangan dan maritim, s_1 untuk infrastruktur dan transportasi hingga s_5 untuk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

E. Visualisasi 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 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 via LAPOR! pada setiap kategori.

(a)

(b)



Gambar 6. (a) *Word Cloud* (b) *Pareto Chart* Kategori Energi Pangan dan Maritim

Melalui Gambar 6.(a) diketahui bahwa tiga kata kunci yang patut mendapatkan perhatian lebih pada kategori aduan energi, pangan dan maritim adalah raskin, desa, kilogram dan listrik. 4 Kata tersebut dipilih karena telah memberikan persentase sebesar 63,1% sebagaimana dijelaskan pada Gambar 6.(b). Kata raskin atau beras miskin menjadi salah satu topik yang paling sering diadukan melalui layanan LAPOR!. Hal ini terkait salah satu program pemerintah dalam penyaluran beras bersubsidi bagi kelompok masyarakat yang memiliki pendapatan rendah [20]. Kata selanjutnya yakni listrik yang terkait dengan aduan masyarakat terhadap pemadaman listrik di suatu daerah. Terakhir, yakni kata desa karena dalam memberikan pengaduan masyarakat menyertakan nama desa.

Gambar 7. *Word Cloud* Kategori Infrastruktur dan Transportasi

Gambar 7 menunjukkan bahwa kata kunci aduan yang menjadi sorotan masyarakat terkait infrastruktur dan transportasi adalah kata jalan dengan kontribusi kata sebesar 40%. Aduan dari masyarakat yang masuk via LAPOR! mengenai kata jalan berhubungan dengan kerusakan jalan, atau aduan lain dengan menyertakan alamat yang bersangkutan.

Gambar 8. *Word Cloud* Kategori Kesehatan

Kata BPJS, KIS, terima, sehat, dan kartu merupakan kata kunci pada kategori kesehatan yang harus ditindaklanjuti oleh dinas terkait karena sering diadukan masyarakat via LAPOR! sebagaimana ditunjukkan melalui Gambar 8. BPJS merupakan badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial. 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. Kata lain yang juga sering diadukan masyarakat adalah KIS atau Kartu Indonesia Sehat yang juga dikelola oleh BPJS Kesehatan. Aduan masyarakat terkait KIS antara lain karena belum tersalurkannya KIS secara baik kepada masyarakat.

Gambar 9. *Word Cloud* Kategori Pendidikan

Melalui Gambar 9 dapat diketahui bahwa tiga kata kunci yang perlu diperhatikan lebih oleh dinas terkait pada kategori pendidikan adalah kata terima, kartu, dan sekolah dengan kontribusi kata sebesar 51,1% jika diambil 10 kata tertinggi. Kata kartu sering diadukan karena kaitannya dengan program pemerintah yakni 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 diadukan karena selaras dengan banyaknya aduan masyarakat yang belum menerima KIP pada saat itu.

Gambar 10. *Word Cloud* Kategori Reformasi Birokrasi

Gambar 10 menunjukkan bahwa kata KTP, Bandung, Jalan, Camat, dan Lurah pada kategori reformasi birokrasi merupakan kata kunci aduan yang paling sering dilaporkan oleh masyarakat dengan persentase kumulatif sebesar 61%. Hal ini terjadi karena pada tahun 2015, Kementrian Dalam Negeri mewajibkan masyarakat untuk melakukan perekaman untuk mengubah KTP mereka menjadi KTP elektronik (*e*-KTP). Masyarakat kerap melaporkan lamanya proses pembuatan *e*-KTP di kantor kecamatan daerah tertentu

Gambar 11. *Word Cloud* 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 dengan kontribusi kata sebesar 67,2% sebagaimana ditunjukkan Gambar 11. Masyarakat kerap memberi aduan terhadap pelayanan kantor imigrasi dalam pendaftaran pembuatan paspor.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, maka diperoleh kesimpulan bahwa kategori dengan jumlah pengaduan terbanyak dari masyarakat via LAPOR! adalah kategori pendidikan dengan jumlah aduan sebanyak 2348 dari total 6916 aduan. 62,2% dari total aduan telah selesai ditindaklanjuti oleh dinas terkait. Melalui seleksi variabel, didapatkan 428 *term* atau kata yang memberikan pengaruh terhadap kategori aduan masyarakat atau dengan kata lain mampu menurunkan jumlah *term* atau kata sebesar 45% yang dapat meningkatkan nilai ketepatan klasifikasi. Sehingga, rata-rata ketepatan klasifikasi yang dihasilkan melalui *Artificial Neural Network* dengan 428 *feature* dan *nodes hidden layer* adalah *precision* 0,794, *sensitivity* 0,818, dan F1-*score* 0,800.

Berdasarkan visualisasi *word cloud*, didapatkan topik aduan yang patut mendapatkan perhatian lebih dari dinas terkait karena sering diadukan masyarakat. Kata kunci aduan yang patut mendapatkan perhatian lebih pada kategori aduan energi, pangan dan maritim adalah kata raskin yakni terkait pembagian beras miskin. 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 diadukan masyarakat pada kategori reformasi birokrasi karena lamanya proses

pembuatan *e-ktp*. Sedangkan kata imigrasi merupakan permasalahan yang kerap diadukan di kategori pariwisata dan lingkungan hidup.

B. Saran

Berdasarkan kesimpulan yang diperoleh, maka saran yang dapat diberikan yakni perlu adanya penambahan kata kunci bahasa sehari-hari maupun tafsiran dari berbagai singkatan yang lebih bervariatif agar kata sebenarnya dapat terhitung dalam frekuensi kemunculan kata. Serta untuk penelitian selanjutnya, agar dapat mencari data terbaru terkait aduan masyarakat via LAPOR! agar kategori lain yang tidak masuk dalam penelitian ini dapat ditambahkan pada penelitian selanjutnya. Untuk pihak LAPOR! yakni dapat mempertimbangkan hasil klasifikasi agar penangananan aduan yang terkait kategori laporan prioritas nasional dapat diselesaikan dengan lebih cepat.

DAFTAR PUSTAKA

[1] R. Feldman dan J. Sanger, Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, New York: Cambridge University Press, 2007.

[2] C. Megawati, "Analisis Aspirasi dan Pengaduan di Situs LAPOR! Dengan Menggunakan Text Mining," 2015.

[3] Z. Reyhana, K. Fithriasari, M. Atok dan N. Iriawan, "Linking Twitter Sentiment Knowledge with Infrastructure Development," *Malaysian Journal of Industrial and Applied Mathematics,* vol. 34, pp. 91-102, 2018.

[4] S. M. Weiss, Text Mining: Predictive Methods for Analyzing, New York: Springer, 2010.

[5] G. N. M. Nata dan P. P. Yudiastra, "Preprocessing Text Mining Pada Email Box Berbahasa Indonesia," dalam *Konferensi Nasional Sistem & Informatika*, Bali, 2017.

[6] F. Rahman, "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Pada Pengguna Twitter Mengenai Presiden Joko Widodo," 2018.

[7] E. Gokgoz dan A. Subasi, "Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT," *Biomedical Signal Processing and Control,* pp. 138-144, 2015.

[8] P. Meesad, P. Boonrawd dan V. Nuipian, "A Chi-Square-Test for Word Importance Differentiation in Text Classification," dalam *International Conference on Information and Electronics Engineering*, Singapore, 2011.

[9] O. S. Bachri, M. H. Kusnadi dan O. D. Nurhayati, "Feature Selection Based On Chi Square In Artificial Neural Network To Predict The Accuracy of Student Study Period," *International Journal of Civil Engineering and Technology,* pp. 731-739, 2017.

[10] K. W. P. Chawla, "SMOTE synthetic minority over-sampling technique.," *Journal of artificial intelligence research,* 2002.

[11] F. K. Damayanti, "Analisis Twitter Pelanggan Belanja Online Menggunakan Metode Naive Bayes Classifier (NBC) dan Artificial Neural Network (ANN)," 2018.

[12] N. D. Astuti, "Klasifikasi Penyakit Gagal Jantung Kongestif Menggunakan Artificial Neural Network," 2017.

[13] C. E. Nwankpa, W. Ijomah, A. Gachagan dan S. Marshall, "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," 2018.

[14] H. Khaulasari, "Combine Sampling - Least Square Support Vector Machine Untuk Klasifikasi Multi Class Imblaanced Data," 2016.

[15] M. Sokolova dan G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management 45 ,* pp. 427-437, 2009.

[16] Q. Castella dan C. Sutton, "Word Storm: Multiples of Word Clouds for Visual Comparison of Documents," 2013.

[17] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Master of Logic Project, Institute for Logic, Language and Computation, Universiteit van, Amsterdam, The Netherlands,* 2003.

[18] J. Mahanta, "Introduction to Neural Networks, Advantages and Applications," 10 July 2017. [Online]. Available: https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications.

[19] A. Zhang, Z. Lipton dan S. A. Mu Li, "Dive into Deep Learning," 2019.

[20] R. I. Kementrian Koordinator Bidang Pembangunan Manusia dan Kebudayaan, Pedoman Umum Subsidi Beras Pada Masyarakat Berpendapatan Rendah, Jakarta, 2016.