

Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di *Google Playstore* Menggunakan Metode *Support Vector Machine*

Author:

Mohammad Rezza Fahlevi¹

Affiliation:

Institut Pemerintahan Dalam Negeri (IPDN Jatinangor), Indonesia ¹

e-Mail:

rezza@ipdn.ac.id¹

Abstract

The Information and Documentation Management Officer (PPID) application was built as an application to meet the needs of information management and services by Public Bodies for the implementation of Law No. 14 of 2008 about Public Information Disclosure (UU KIP). This application assists public bodies in documenting and serving requests for information to the public. With the launch of the PPIPD application on the Google Play Store, it raises many opinions or responses from application users through the review feature. The reviews are quite many and unstructured and contain opinions from users about their satisfaction with the application. The feedback obtained from users is not only positive but also negative. Users often make complaints about applications that have been used or provide suggestions for features in the application. User reviews are of great interest to the owners of the application to decide the future. Sentiment analysis is an activity used to analyze a person's opinion or opinion on a topic. The Support Vector Machine (SVM) method is a text mining method that includes a classification method and the term Frequency-Inverse Document Frequency (TF-IDF) is a character weighting method. SVM and TF-IDF can be used to analyze sentiment based on user reviews of PPID apps on the Google Play Store. The purpose of this study was to classify user reviews of PPID apps on the Google Play Store using sentiment analysis that has been collected and filtered. The reviews totaled 700 data with labels of 85 positive and 615 negative. And the results of the analysis using SVM produced an average k-fold of 88%, precision of 94%, recall of 100%, f-measure of 97%, and accuracy of 97%.

Keywords: *PPID; Sentiment Analysis, Support Vector Machine, TF-IDF*

Abstrak

Aplikasi Pejabat Pengelola Informasi dan Dokumentasi (PPID) dibangun sebagai sebuah aplikasi untuk memenuhi kebutuhan pengelolaan dan pelayanan informasi oleh Badan Publik untuk penerapan Undang-Undang No 14 Tahun 2008 tentang Keterbukaan Informasi Publik (UU KIP). Aplikasi ini membantu badan publik dalam melakukan pendokumentasian serta melayani permohonan informasi kepada masyarakat. Dengan di luncurkannya aplikasi PPIPD di *Google Play Store* sehingga memunculkan banyak pendapat atau tanggapan dari pengguna aplikasi melalui fitur ulasan. Ulasan yang ada cukup banyak dan tidak terstruktur serta mengandung opini dari para pengguna tentang kepuasan mereka terhadap aplikasi tersebut. Umpan balik yang didapatkan

dari pengguna tidak hanya yang bersifat positif, melainkan juga bersifat negatif. Pengguna seringkali memberikan keluhan-keluhan terhadap aplikasi yang telah digunakan atau memberikan usulan-usulan terhadap fitur dalam aplikasi tersebut. Ulasan pengguna sangat menarik bagi pemilik aplikasi untuk mengambil keputusan di masa depan. Analisis sentimen adalah kegiatan yang digunakan untuk menganalisis pendapat atau opini seseorang tentang suatu topik. Metode *Support Vector Machine* (SVM) adalah metode *text mining* yang mencakup metode klasifikasi dan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode pembobotan karakter. SVM dan TF-IDF dapat digunakan untuk menganalisis sentimen berdasarkan ulasan pengguna aplikasi PPID di *Google Play Store*. Tujuan dari penelitian ini adalah untuk mengklasifikasikan ulasan pengguna aplikasi PPID di *Google Play Store* menggunakan analisis sentimen yang telah dikumpulkan dan disaring. Ulasan tersebut berjumlah 700 data dengan label 85 positif dan 615 negatif. Dan hasil analisis menggunakan SVM menghasilkan rata-rata *k-fold* sebesar 88%, *precision* 94%, *recall* 100%, *f-measure* 97%, dan *accuracy* 97%.

Kata kunci: PPID; Analisis Sentimen, *Support Vector Machine*, TF-IDF

1. PENDAHULUAN

Aplikasi Pejabat Pengelola Informasi dan Dokumentasi (PPID) Kementerian Dalam Negeri (Kemendagri) dibangun sebagai sebuah aplikasi untuk memenuhi kebutuhan pengelolaan dan pelayanan informasi oleh Badan Publik untuk penerapan Undang-undang No 14 Tahun 2008 Tentang Keterbukaan Informasi Publik (UU KIP). Aplikasi ini membantu badan publik dalam melakukan pendokumentasian serta melayani permohonan informasi kepada masyarakat. Dengan di luncurkannya aplikasi PPIPD di *Google Play Store* sehingga memunculkan banyak pendapat atau tanggapan dari pengguna aplikasi melalui fitur ulasan. Ulasan yang ada cukup banyak dan tidak terstruktur serta mengandung opini dari para pengguna tentang kepuasan mereka terhadap aplikasi tersebut. Umpan balik yang didapatkan dari pengguna tidak hanya yang bersifat positif, melainkan juga bersifat negatif. Pengguna seringkali memberikan keluhan-keluhan terhadap aplikasi yang telah digunakan atau memberikan usulan-usulan terhadap fitur dalam aplikasi tersebut. Ulasan pengguna sangat menarik bagi pemilik aplikasi untuk mengambil keputusan di masa depan.

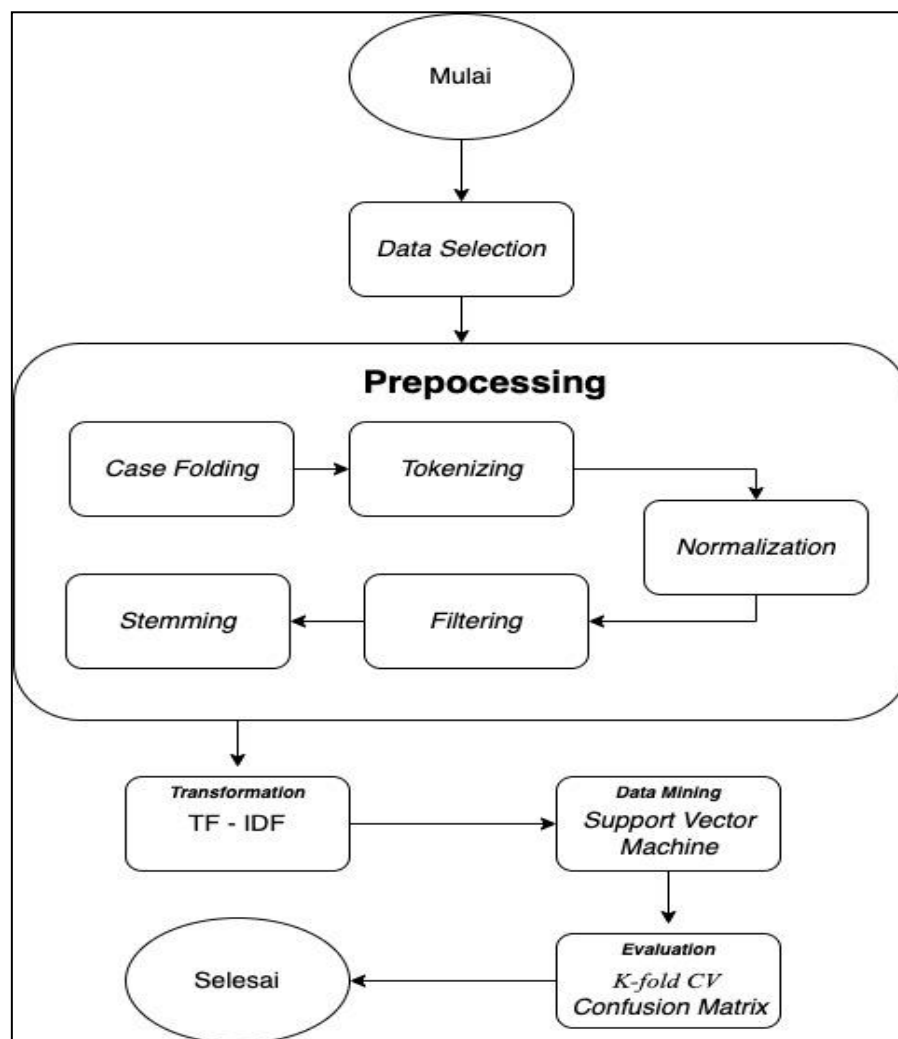
Analisis sentimen adalah kegiatan yang digunakan untuk menganalisis pendapat atau opini seseorang tentang suatu topik. Tugas dasar Analisis sentimen adalah mengklasifikasikan beberapa teks dari dokumen, kalimat atau fitur, kalimat dari fitur tersebut bisa bersifat positif, negatif dan netral. Dalam melakukan analisis sentimen diperlukan metode yang menunjang klasifikasi. Metode yang digunakan adalah *Support Vector Machine* (SVM), berdasarkan hasil penelitian analisis sentimen sebelumnya yang dilakukan oleh Putranti dan Winarko. Pada penelitian tersebut dilakukan analisis sentimen pada objek *twitter* dengan mengimplementasikan metode *Maximum Entropy* dan SVM dan menghasilkan nilai akurasi sebesar 86,81% (Putranti & Winarko, 2014). Penelitian analisis sentimen lainnya dilakukan dengan membandingkan hasil akurasi metode *Naïve Bayes* dan SVM. Berdasarkan penelitian tersebut didapatkan SVM

adalah metode klasifikasi yang baik untuk analisis sentimen dengan hasil akurasi 98% (K, S. T. & Shetty, J., 2017.).

Dari berbagai referensi penelitian yang telah dilakukan, metode SVM menjadi salah satu pilihan metode yang akan digunakan peneliti dalam analisis sentimen. Metode *Support Vector Machine* (SVM) adalah metode *text mining* yang mencakup metode klasifikasi dan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode pembobotan karakter. SVM dan TF-IDF dapat digunakan untuk menganalisis sentimen terkait ulasan pengguna aplikasi PPID di *Google Play Store*.

2. METODE PENELITIAN

Penelitian ini menggunakan metode KDD (*Knowledge Discovery in Database*) karena memiliki keunggulan proses pengidentifikasian pola yang terorganisir dari se kumpulan data yang kompleks, sehingga datanya mudah untuk dipahami (Ramos et al., 2021; Rizki et al., 2021). Gambar 1 menunjukkan alur penelitian yang menerapkan metode KDD.



Gambar 1. Alur Penelitian

Alur penelitian pada gambar 1 dijelaskan sebagai berikut.

2.1. Data Selection

Data *selection* merupakan tahapan pertama dalam KDD, pada tahap ini akan dilakukan pengumpulan, seleksi, dan pelabelan data. Data dikumpulkan dari situs *Google Play Store* menggunakan *Web Scrapping (WS)* dengan nama aplikasi PPID Kemendagri sebanyak 700 data komentar. WS atau yang dikenal sebagai ekstraksi web adalah teknik untuk mengekstraksi data dari *World Wide Web (WWW)* dan menyimpannya ke file sistem atau basis data untuk dijadikan analisis data. WS dapat dilakukan baik secara manual oleh seorang pengguna atau secara otomatis oleh *bot* atau *crawler web*. Proses WS dari Internet dapat dibagi menjadi dua langkah berurutan, yaitu mengakuisisi sumber daya web dan kemudian mengekstraksi informasi yang diinginkan dari data yang diperoleh. Secara khusus, program WS dimulai dengan meminta HTTP untuk memperoleh sumber daya dari yang ditargetkan oleh situs web. Permintaan ini dapat diformat kedalam URL yang berisi permintaan *GET* atau HTTP yang berisi *POST*. Setelah permintaan berhasil diterima dan diproses oleh situs web yang ditargetkan, sumber daya yang diminta akan diambil dari situs web dan kemudian dikirim kembali ke program WS. Sumber daya ini bisa dalam berbagai format, seperti halaman web yang dibangun dengan HTML, XML atau JSON, atau data multimedia seperti gambar, audio, atau video. Terdapat dua modul penting dari WS - modul untuk menulis permintaan HTTP, seperti *Urllib2* atau *selenium* dan satu lagi untuk parsing dan mengekstraksi informasi dari kode HTML mentah, seperti *BeautifulSup* atau *Pyquery* (B. Zhao, 2017).

2.2. Preprocessing

Pada tahap *preprocessing* ini, data mentah yang telah terkumpul akan diolah menjadi data yang dapat digunakan pada tahap selanjutnya. *Preprocessing* merupakan proses awal yang akan mentransformasikan data masukan menjadi data dengan format yang sesuai dan siap untuk diproses. Beberapa contoh hal yang dilakukan dalam *preprocessing* meliputi berbagai proses yang diperlukan antara lain : penggabungan, perubahan bentuk, ataupun pentransformasian data sebagai cara untuk membersihkan, mengintegrasikan, mereduksi dan mendiskritisasi. Lebih lanjut proses yang ada dalam *preprocessing* dapat terdiri dari salah satu kegiatan proses ataupun gabungan dari beberapa proses diatas. Proses yang ada tergantung dari tujuan yang akan dicapai dalam *preprocessing* tersebut (Karthick & Malathi, 2015). Tahap *preprocessing* terdiri dari 5 proses yaitu:

- a. *Case Folding* merupakan proses *preprocessing* yang bertujuan untuk mengubah huruf kapital menjadi huruf kecil (*lowercase*) agar lebih terstruktur dan mudah dalam pemrosesan lebih lanjut.
- b. *Tokenization* atau tokenisasi bertujuan untuk menghilangkan tanda baca, *tag*, *emoticon* dan memotong setiap kalimat ke dalam sekumpulan kata (kata tunggal). Tahapan tokenisasi disesuaikan dengan kebutuhan pemrosesan.

- c. *Normalization* atau Normalisasi bertujuan untuk memperbaiki penulisan kata yang salah ketik (*typo*) atau yang tidak baku seperti perulangan untuk penekanan kata, misalnya: “sangat” yang ditulis “sangaaaat”.
- d. *Filtering* bertujuan untuk pemilihan kata-kata penting dari hasil tokenisasi dan menghapus kata-kata yang tidak digunakan dalam proses identifikasi, seperti kata yang, kecuali, tetapi, dan, dapat, dan lain-lain.
- e. *Stemming* merupakan proses untuk mengembalikan kata ke dalam bentuk dasarnya (kata dasar).

2.3. Transformation

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode untuk melakukan pembobotan kata dari proses ekstraksi kata dengan menerapkan perhitungan kata umum di *information retrieval*. Metode pembobotan ini merupakan penggabungan antara *term frequency* dan *inverse document frequency*. *Term frequency* merupakan jumlah kemunculan sebuah *term* pada sebuah dokumen. Besarnya jumlah term yang muncul berbanding lurus dengan pembobotan yang diberikan. Sedangkan *inverse document frequency* adalah proses untuk mengukur seberapa penting kata dalam suatu dokumen (Akbari et al., 2012). Nilai TF-IDF dapat diperoleh menggunakan persamaan (1), (2), dan (3).

$$W_{t,d} = \begin{cases} 1 + \log_{10} t_{t,d}, & \text{if } t_{t,d} > 0 \\ 0, & \text{if } t_{t,d} = 0 \end{cases} \quad (1)$$

$$idf_t = \log_{10} \frac{N}{Df_t} \quad (2)$$

$$W_{t,d} = W_{t,d} \cdot idf_t \quad (3)$$

Keterangan:

$W_{t,d}$ = Bobot kata di setiap dokumen

$t_{t,d}$ = Jumlah kemunculan term dalam dokumen

N = Jumlah keseluruhan dokumen

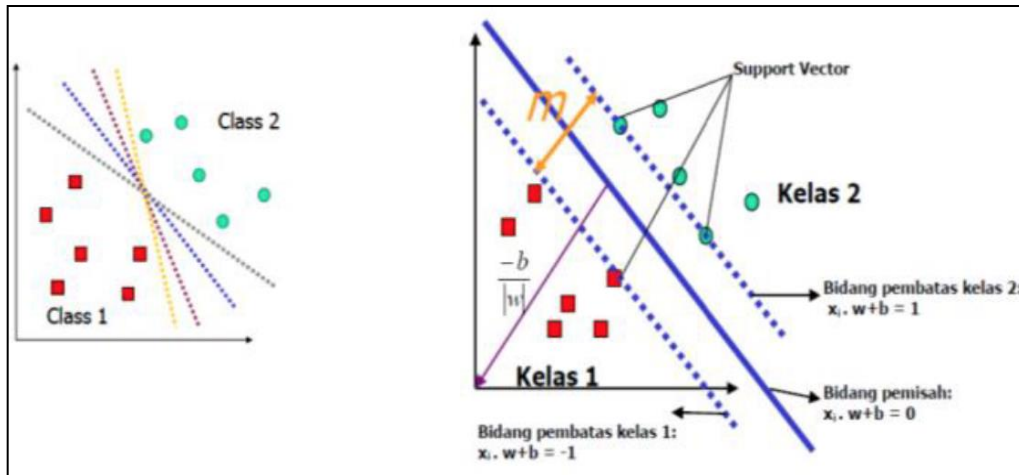
Df = Jumlah dokumen yang mengandung term

idf = Bobot *inverse* dalam nilai df

$W_{t,d}$ = Pembobotan TF-IDF

2.4. Data Mining

Pada tahap data *mining* akan dilakukan proses klasifikasi sentimen pada data ulasan dengan menggunakan algoritma *Support Vector Machine* (SVM). SVM merupakan satu di antara banyak algoritma yang digunakan untuk klasifikasi dan termasuk dalam kategori *supervised learning*. Konsep kerja *Support Vector Machine* yaitu dengan mencari *hyperplane* atau garis pembatas paling optimal yang berfungsi untuk memisahkan dua kelas (Lidya et al., 2015). Konsep kerja dari algoritma *Support Vector Machine* (SVM) dapat dilihat pada Gambar 2.



Gambar 2. Hyperplane Support Vector Machine (Lidya et al., 2015)

Untuk memperoleh garis *hyperplane* yang paling optimal dalam memisahkan data ke dua buah kelas tersebut, maka digunakan perhitungan *margin hyperplane* dan menemukan titik maksimal. Dalam memperoleh *hyperplane* pada SVM, dapat menggunakan persamaan (4).

$$(w \cdot xi) + b = 0 \tag{4}$$

Di dalam data xi , yang termasuk pada kelas -1 dapat dirumuskan pada persamaan (5).

$$(w \cdot xi + b) \leq 1, yi = -1 \tag{5}$$

Sedangkan data xi yang termasuk pada kelas +1 dapat dirumuskan seperti pada persamaan (6).

$$(w \cdot xi + b) \geq 1, yi = 1 \tag{6}$$

Radial Basis Function (RBF) Kernel

RBF kernel merupakan fungsi kernel yang biasa digunakan dalam analisis ketika data tidak terpisah secara linear. RBF kernel memiliki dua parameter yaitu *Gamma* dan *complexity*. Parameter *complexity* atau biasa disebut sebagai C merupakan parameter yang bekerja sebagai pengoptimalan SVM untuk menghindari misklasifikasi di setiap sampel dalam training dataset. Parameter *Gamma* menentukan seberapa jauh pengaruh dari satu sampel training dataset dengan nilai rendah berarti “jauh”, dan nilai tinggi berarti “dekat”. Dengan *gamma* yang rendah, titik yang berada jauh dari garis pemisah yang masuk akal dipertimbangkan dalam perhitungan untuk garis pemisah. Ketika *gamma* tinggi berarti titik – titik berada di sekitar garis yang masuk akal akan dipertimbangkan dalam perhitungan (Patel, 2017). Berikut merupakan persamaan dari RBF kernel (7).

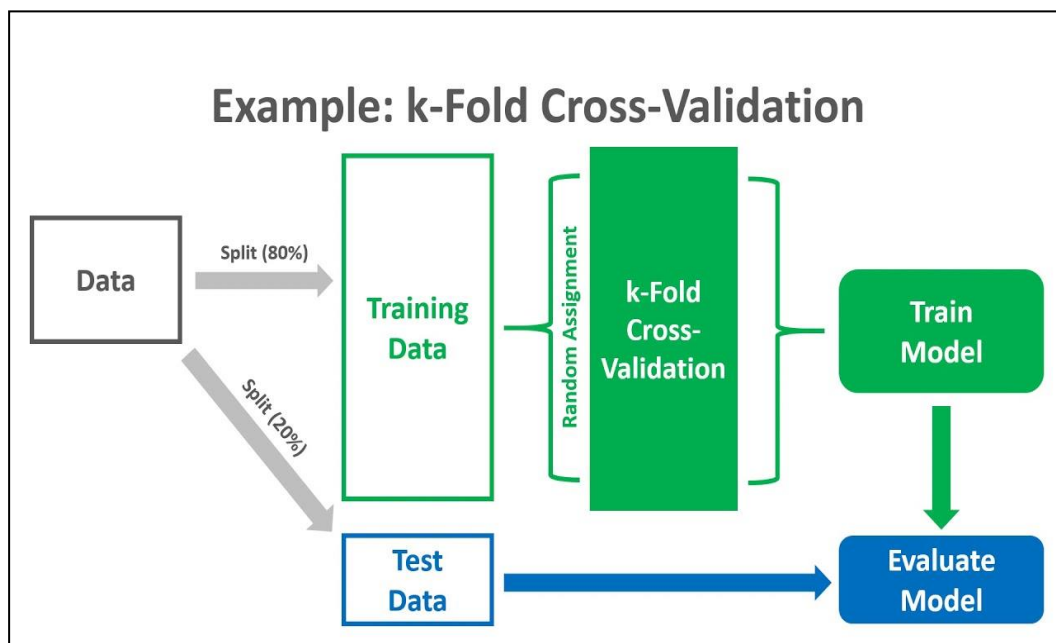
$$K(x,z) = \exp [-\gamma || x - z ||^2] \tag{7}$$

2.5. Evaluation

K-Fold Cross Validation

K-Fold Cross Validation (K-fold CV) adalah metode untuk memperkirakan kesalahan prediksi untuk evaluasi kinerja model. Dalam *cross validation* dikenal sebagai estimasi rotasi, dengan membagi data menjadi himpunan bagian k dengan ukuran yang hampir sama, model dalam klasifikasi dilatih dan diuji sebanyak k. Disetiap pengulangan, salah satu himpunan bagian akan digunakan sebagai data pengujian dan sub kelompok data k lainnya berfungsi sebagai data pelatihan (Nurhayati, et al., 2014).

K-fold CV merupakan metode untuk mengevaluasi kinerja *classifier*, metode ini dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah *instance* tidak banyak). *K-fold CV* adalah suatu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan reduksi dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. *K-fold CV* diawali dengan membagi data sejumlah *n-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam n buah partisi dengan ukuran yang sama variabel Data ke 1, variabel Data ke 2, variabel Data ke 3 .. Dn selanjutnya proses uji dan latih dilakukan sebanyak n kali. Dalam iterasi ke-i partisi Di akan menjadi data uji dan sisanya akan menjadi data latih. Untuk penggunaan jumlah *fold* terbaik untuk uji validitas, dianjurkan menggunakan *10-fold cross validation* dalam model. Contoh mekanisme penerapan *K-fold CV* terdapat pada gambar 3.



Gambar 3. Contoh k-Fold Cross Validation (Caughlin, D.E., 2014)

Pembagian data akan dilakukan dengan membagi data komentar dari hasil ekstraksi fitur menjadi dua bagian, yaitu data latih dan data uji. Bagian pertama berupa data latih yang digunakan untuk membentuk model klasifikasi dengan melakukan

pelatihan menggunakan data latih yang sudah dibagi. Bagian kedua berupa data uji yang digunakan untuk proses evaluasi. Data dari hasil ekstraksi fitur merupakan data dalam bentuk *Bag of Words* (BoW). Pembagian data dalam bentuk BoW ditentukan menggunakan metode *K-fold CV*. Penelitian ini menggunakan *K* sebesar 10- *fold* (bagian) secara merata, pembagian data dengan menggunakan 10-fold CV akan diulang sebanyak 10 kali dengan setiap perulangannya menghasilkan 70 data dengan jumlah total dataset sebanyak 700 data. Rincian pembagian yaitu sembilan bagian dari tiap *fold* digunakan sebagai data latih, dimana setiap *fold* berisi 630 komentar. Sedangkan sisanya yaitu satu *fold* digunakan sebagai data uji yang terdapat 70 komentar.

Confusion Matrix

Tahap evaluasi akan menggunakan *Confusion Matrix* (CM). CM merupakan suatu instrumen yang digunakan untuk mengevaluasi performa dari model klasifikasi yang telah dihasilkan. Pada CM, hasil kelas prediksi akan dibandingkan dengan kelas data yang sebenarnya (Salekhah, C.I., 2016). Hasil tersebut kemudian akan digunakan untuk menghitung nilai *precision*, *recall*, *f-measure*, dan *accuracy*. Pengukuran evaluasi dengan CM dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

Data Aktual	Data Prediksi		
	<i>True</i>	<i>False</i>	Total
<i>True</i>	TP	FN	P
<i>False</i>	FP	TN	N
Total	P'	N'	P+N

Keterangan:

TP (*True Positive*) = Data positif yang terklasifikasi secara benar.

TN (*True Negative*) = Data negatif yang terklasifikasi secara benar.

FP (*False Positive*) = Data negatif yang terklasifikasi menjadi positif.

FN (*False Negative*) = Data positif yang terklasifikasi menjadi negatif.

Skenario Pengujian

Pengujian ini menggunakan fungsi kernel RBF dengan nilai parameter *complexity* dan *gamma*. Pada proses parameter tuning menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f-measure* serta rata-rata *k-fold* terbaik. Kemudian tahap evaluasi terhadap hasil *feature selection* dengan menggunakan metode klasifikasi SVM dengan berbagai kombinasi nilai parameter yaitu parameter *complexity* (C) dan parameter *gamma* (γ). Adapun nilai untuk masing-masing parameter yang ditentukan sebagai nilai parameter optimal menggunakan metode Taguchi diantaranya adalah (1,1.25,1.50,1.75,2) C = dan $\gamma = (0.2, 0.4, 0.6, 0.8)$.

3. HASIL DAN PEMBAHASAN

3.1. Data Selection

Pengumpulan data ulasan aplikasi Pejabat Pengelola Informasi dan Dokumentasi (PPID) yang dilakukan dengan teknik *scraping* menggunakan *library google play scraper* menghasilkan sebanyak 700 data. Data yang telah terkumpul memiliki 5 atribut yaitu *reviewId*, *userName*, *userImage*, *content*, *at*, *content*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *replyContent* and *repliedAt*. Selanjutnya dilakukan penyeleksian atribut yang digunakan yaitu hanya *content* dan *score*. Kemudian dilakukan pelabelan data ulasan secara manual dan divalidasi oleh ahli Bahasa dan Sastra Indonesia. Pelabelan data tersebut menghasilkan 85 ulasan positif dan 615 ulasan negatif. Hasil pada tahap data selection dapat dilihat pada gambar 4.

	label	content
0	negatif	aplikasi kecewa fungsi sayak kalau memang dala...
1	negatif	bisa masuk periksa email menu sepam terus klik...
2	negatif	baru coba daftar langkah sangat mudah telah da...
3	negatif	saran masuk mohon tim developer laku runcing d...
4	positif	bisa daftar mau masuk cetak dokumen ada bisa k...
...
696	negatif	apak gagal
697	negatif	apak kampang
698	positif	bagus banget
699	negatif	jelas

Gambar 4. Hasil Data Selection

3.2. Preprocessing

Pada tahap ini dilakukan untuk mengubah data agar sesuai dan dapat diolah pada tahapan selanjutnya. Contoh hasil dari 5 proses pada tahap *preprocessing* yang telah dilakukan dapat dilihat pada tabel 2.

Tabel 2. Contoh Hasil Preprocessing

Proses	Hasil
Data Collection	Aplikasinya baguuss, bermanfaat untuk kerjaan saya sekarang
Case Folding	aplikasinya baguuss, bermanfaat untuk kerjaan saya sekarang
Tokenizing	'aplikasinya', 'baguuss', 'bermanfaat', 'untuk', 'kerjaan', 'saya', 'sekarang'
Normalization	'aplikasinya', bagus, 'bermanfaat', 'untuk', 'kerjaan', 'saya', 'sekarang'
Filtering	'aplikasinya', bagus, 'bermanfaat', 'kerjaan'
Stemming	'aplikasi', bagus, 'manfaat', 'kerjaan'

3.3. Transformation

Pada proses ini, dilakukan perhitungan jumlah kemunculan tiap *term* pada keseluruhan data komentar (TF), menghitung jumlah komentar yang mengandung

setiap *term* (DF), menghitung *inverse* dari nilai DF (IDF), dan mengkalikan nilai *tf* dan *idf* untuk memperoleh nilai bobot tiap komentar. Proses tersebut dilakukan menggunakan persamaan (1), (2), (3). Hasil pembobotan pada 3 sampel data komentar dapat dilihat pada tabel 3 sedangkan perhitungan pembobotan dapat dilihat pada tabel 4.

Tabel 3. Contoh Data yang Digunakan dalam Pembobotan

Komentar ke -	Komentar
1	Aplikasinya baguuss, bermanfaat untuk kerjaan saya sekarang
2	Tentor keren dalam mengajar
3	Terobosan generasi muda Indonesia dalam dunia pendidikan

Tabel 4. Perhitungan Pembobotan

Term	TF-1	TF-2	TF-3	DF	IDF	TF-IDF1	TF-IDF2	TF-IDF3
Aplikasi	1	0	0	1		0,477	0	0
Keren	1	1	0	2				0
Generasi	1	0	1	2			0	
Muda	1	0	1	2			0	
Indonesia	1	0	1	2			0	
Tentor	0	1	0	1		0		0
Dalam	0		1	1		0	0	
Mengajar	0	1	0	1		0		0
Terobosan	0	0	1	1		0	0	
Dunia	0	0	1	1		0	0	
Pendidikan	0	0	1	1		0	0	

3.4. Data Mining

Proses klasifikasi ditentukan ke dalam 2 kelas yaitu positif dan negatif menggunakan *multi-class SVM one vs. one*. Dataset kemudian dibagi menjadi training set dan testing set menggunakan dengan ratio 0,1. Artinya 90 % data digunakan untuk proses training, dan 10% digunakan untuk proses testing. Algoritma *Support Vector Machine* (SVM) sendiri memiliki fungsi kernel dan parameter yang dapat digunakan ketika proses klasifikasi. Parameter yang digunakan dapat ditentukan nilainya dan perbedaan nilai pada parameter dapat mempengaruhi performa dari model. Untuk itu, dilakukan proses pengujian berupa parameter tuning pada fungsi kernel sehingga menghasilkan *machine learning* model yang memiliki performa optimal.

Pengujian dilakukan dengan memberikan input nilai pada masing-masing parameter yang digunakan seperti *complexity* dan *gamma*. Proses tersebut dilakukan pada fungsi kernel RBF dari algoritma *Support Vector Machine* (SVM) sesuai dengan parameter yang dimiliki. Adapun nilai untuk masing-masing parameter yang ditentukan sebagai nilai parameter optimal menggunakan metode Taguchi diantaranya adalah (1,1.25,1.50,1.75,2) C = dan $\gamma = (0.2, 0.4, 0.6, 0.8)$.

Radial Basis Function (RBF) Kernel

RBF kernel merupakan fungsi kernel yang biasa digunakan dalam analisis ketika data tidak terpisah secara linear. RBF kernel memiliki dua parameter yaitu *Gamma* dan *complexity*. Parameter *complexity* atau biasa disebut sebagai C merupakan parameter yang bekerja sebagai pengoptimalan SVM untuk menghindari misklasifikasi di setiap sampel dalam *training* dataset. Parameter *gamma* menentukan seberapa jauh pengaruh dari satu sampel *training*. Pada kernel RBF, dilakukan pengujian terhadap parameter *complexity* dan *gamma*. Dengan menggunakan nilai parameter *complexity* antara 1 hingga 2, nilai testing *accuracy* terus mengalami peningkatan hingga mencapai yang terbaik pada nilai *complexity* = (1.50, 1.75, 2) dan *gamma* = (0.6, 0.8) yaitu sebesar 97%. Selain *accuracy*, kernel RBF juga menghasilkan nilai *precision*, *recall*, dan *f-measure* yang terbaik dengan penggunaan nilai parameter yang sama.

Hasil Pengujian Cross Validation

Untuk melakukan validasi tingkat *accuracy* dari suatu model, maka dilakukan pengujian dengan metode *10-fold cross validation*. Pengujian ini menggunakan fungsi kernel RBF dengan nilai parameter *complexity* dan *gamma*. Pada proses parameter tuning menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f-measure* serta rata-rata *k-fold* terbaik.

Tahapan seleksi *feature* pada dataset telah dilakukan kemudian tahap evaluasi terhadap hasil *feature selection* dengan menggunakan metode klasifikasi SVM dengan berbagai kombinasi nilai parameter yaitu parameter *complexity* (C) dan parameter *gamma* (γ). Adapun nilai untuk masing-masing parameter yang ditentukan sebagai nilai parameter optimal diantaranya adalah (1,1.25,1.50,1.75,2) C = dan γ = (0.2, 0.4,0.6,0.8). Berikut hasil evaluasi berdasarkan parameter parameter *complexity* (C) dan parameter *gamma* (γ) dapat dilihat pada tabel 5.

Tabel 5. Evaluasi Berdasarkan Parameter *Complexity* (C) dan *Gamma* (γ)

Runs	C	γ	Nilai Akurasi SVM (%)				
			Rata-rata K-Fold	akurasi	precision	recall	f-measure
1	1	0.2	0,88	0,5	0,5	1	0,67
2	1	0.4	0,88	0,5	0,5	1	0,67
3	1	0.6	0,88	0,5	0,5	1	0,67
4	1	0.8	0,88	0,62	0,57	1	0,73
5	1,25	0.2	0,88	0,5	0,5	1	0,67
6	1,25	0.4	0,88	0,5	0,5	1	0,67
7	1,25	0.6	0,88	0,69	0,62	1	0,76
8	1,25	0.8	0,88	0,84	0,5	1	0,67
9	1,50	0.2	0,88	0,5	0,5	1	0,67
10	1,50	0.4	0,88	0,62	0,57	1	0,73
11	1,50	0.6	0,88	0,84	0,76	1	0,86

Runs	C	γ	Nilai Akurasi SVM (%)				
			Rata-rata K-Fold	akurasi	precision	recall	f-measure
12	1,50	0.8	0,88	0,97	0,94	1	0,97
13	1,75	0.2	0,88	0,50	0,50	1	0,67
14	1,75	0.4	0,88	0,72	0,64	1	0,78
15	1,75	0.6	0,88	0,94	0,89	1	0,94
16	1,75	0.8	0,88	0,97	0,94	1	0,97
17	2	0.2	0,88	0,50	0,50	1	0,67
18	2	0.4	0,88	0,75	0,67	1	0,80
19	2	0.6	0,88	0,97	0,94	1	0,97
20	2	0.8	0,87	0,97	0,94	1	0,97

4. Kesimpulan

Berdasarkan dari penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen menggunakan metode *Support Vector Machine* (SVM) sebagai metode klasifikasi, *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai metode pembobotan karakter dalam menganalisis sentimen berdasarkan ulasan pengguna aplikasi PPID Kemendagri di *Google Play Store* mampu diterapkan dengan baik dalam melakukan analisis sentimen sehingga dapat digunakan oleh pemilik aplikasi untuk mengambil keputusan di masa depan. Data yang digunakan adalah data komentar yang berasal dari ulasan aplikasi PPID Kemendagri di *Google Play Store* sebanyak 700 data dengan label 85 positif dan 615 negatif. Dalam tahap evaluasi analisis sentimen menghasilkan rata-rata *k-fold* sebesar 88%, *precision* 94%, *recall* 100%, *f-measure* 97%, dan *accuracy* 97%.

5. DAFTAR PUSTAKA

- Akbari, M.I.H.A.D., Novianty, A. & Setianingsih, C.(2012). Analisis Sentimen Menggunakan Metode *Learning Vector Quantization*. *e-Proceeding of Engineering*, Vol. 4(2), Agustus 2017, pp. 2283-2292.
- B. Zhao, "*Web Scraping*," in *Springer International Publishing AG, USA*, 2017.
- Caughlin, D. E. (2017). *An Introduction to Human Resource Analytics Using R . Version 0.1.1: 2022-05-18*.
- Karthick, R. & Malathi, D.A., 2015. *Preprocessing of Various Data Sets Using Different Classification Algorithms for Evolutionary Programming*. *International Journal of Science and Research (IJSR)*, IV(4), pp.2730-33.
- Lidya, S.K., Sitompul, O.S. & Efendi, S. (2015). *Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K- NN)*. Seminar Nasional Teknologi Informasi dan Komunikasi.
- Nurhayati., Iwan K, Hadihardaja., Indratmo Soekarno., & M. Cahyono. 2014. *A Study of Hold-Out and K-Fold Cross Validation for Accuracy of Groundwater Modeling in Tidal Lowland Reclamation Using Extreme Learning Machine*. *2nd International*

Conference on Technology, Informatics, Management, Engineering & Environment. pp : 228 – 233.

- Patel, S. (2017, Mei 3). Chapter 2 : SVM (*Support Vector Machine*)—*Theory*. Dipetik April 11, 2018, dari *Machine Learning 101*:<https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>.
- Putranti, N. D. & Winarko, E., 2014. Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan *Maximum Entropy* dan *Support Vector Machine*. *IJCCS*, Volume 8, pp. 91-100.
- Ramos, S., Soares, J., Cembranel, S. S., Tavares, I., Foroozandeh, Z., Vale, Z., & Fernandes, R. (2021). *Data mining techniques for electricity customer characterization*. *Procedia Computer Science*, 186, 475 – 488. <https://doi.org/10.1016/j.procs.2021.04.168>.
- Rizki, M., Umam, M. I. H., & Hamzah, M. L. (2020). Aplikasi Data *Mining* Dengan Metode CHAID Dalam Menentukan Status Kredit. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 18 (1), 29 - 33.
- Salekhah, C.I. (2016). Implementasi Metode *Multi Class Support Vector Machine* Untuk Klasifikasi Emosi pada Lirik Lagu Bahasa Indonesia. Universitas Komputer Indonesia (Unikom).