

Klasifikasi Sentiment Review Aplikasi MyPertamina Menggunakan Word Embedding FastText dan SVM (Support Vector Machine)

Mustasaruddin, Elvia Budianita*, M Fikry, Febi Yanto

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11651103375@students.uin-suska.ac.id, ^{2*}elvia.budianita@uin-suska.ac.id, ³mfikry1980@yahoo.com, ⁴febiyanto@uin-suska.ac.id

Email Penulis Korespondensi: elvia.budianita@uin-suska.ac.id

Submitted: 07/02/2023; Accepted: 31/03/2023; Published: 31/03/2023

Abstrak—Aplikasi MyPertamina menjadi syarat untuk membeli Bahan Bakar Minyak (BBM) subsidi yaitu pertalite dan solar, tujuannya supaya pembelian (BBM) subsidi tepat sasaran. Aplikasi MyPertamina banyak mendapat penilaian dan komentar terhadap masyarakat baik dari sisi positif dan negatif, dengan komentar dan penilaian ini diharapkan dapat membantu pemerintah sebagai tolak ukur dalam melaksanakan suatu program. Oleh sebab itu, penelitian kali ini bertujuan menilai aplikasi MyPertamina dengan pengelompokan kelas sentiment 90:10, 80:20 dan 70:30. Dalam penelitian ini, metode yang digunakan Fasttext dan Support Vector Machine (SVM) untuk review aplikasi MyPertamina. Penelitian kali ini menggunakan data sebanyak 8000, data dikelompokkan menjadi tiga porsi data, dengan porsi 90:10, 80:20 dan 70:30. Diperoleh model terbaik SVM dengan porsi data 90:10 dengan jumlah data training 7200 dan 800 data testing, didapat akurasi 80%, recall 50% dan precision 84% tanpa undersampling. Sedangkan jika jumlah data diseimbangkan (undersampling) dengan jumlah data positif 1325, netral 1325 dan negatif 1325, yaitu dengan patokan nilai data terendah dari kelas sentiment diperoleh akurasi sebesar 67%, recall 69% dan precision 57%. Jumlah kelas sentiment yang paling banyak dari porsi data 90:10 adalah negatif yaitu 4300, netral 1575 dan positif 1325, karena banyak ditemukan ulasan pengguna aplikasi MyPertamina yaitu “setelah update aplikasi MyPertamina semakin parah bug nya”.

Kata Kunci: Aplikasi Play Store; MyPertamina; SVM; FastText; Analisa Sentiment

Abstract—The MyPertamina application is a requirement for buying subsidized fuel oil (BBM), namely pertalite and diesel, the goal is that subsidized (BBM) purchases are right on target. The MyPertamina application has received many ratings and comments from the public, both positive and negative, with these comments and ratings expected to help the government as a benchmark in implementing a program. Therefore, this research aims to assess the MyPertamina application by grouping sentiment classes 90:10, 80:20 and 70:30. In this study, the method used is Fasttext and Support Vector Machine (SVM) to review the MyPertamina application. This research uses 8000 data, the data is grouped into three portions of data, with portions of 90:10, 80:20 and 70:30. The best SVM model was obtained with a data portion of 90:10 with a total of 7200 training data and 800 testing data, obtained 80% accuracy, 50% recall and 84% precision without undersampling. Meanwhile, if the amount of data is balanced (undersampling) with the number of positive data 1325, neutral 1325 and negative 1325, that is, with the benchmark of the lowest data value from the sentiment class, an accuracy of 67% is obtained, recall is 69% and precision is 57%. The highest number of sentiment classes from the 90:10 portion of the data is negative, namely 4300, neutral 1575 and positive 1325, because many users found reviews of the MyPertamina application, namely "after updating the MyPertamina application the bugs are getting worse".

Keywords: Play Store App; MyPertamina; SVM; FastText; Sentiment Analysis

1. PENDAHULUAN

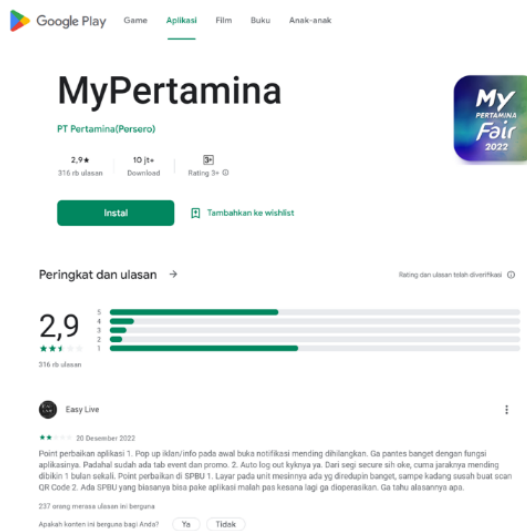
PT.Pertamina (PERSERO) adalah salah satu perusahaan BUMN terbesar di Indonesia. Perusahaan milik negara itu beroperasi di industri perminyakan. Pertamina (PERSERO) bergerak dalam bidang usaha pengolahan dan pemasaran minyak bumi dari hulu ke hilir. Salah satu hasil olahan Pertamina adalah BBM (bahan bakar minyak)[1]. BBM (bahan bakar minyak): merupakan BBM yang terbuat dari penyulingan minyak yang terdapat di dalam bumi. Minyak mentah yang berasal dari dalam bumi pertama kali diproses di kilang untuk menghasilkan produk minyak termasuk bahan bakar. Bahan Bakar Minyak (BBM) merupakan komoditas yang memegang peranan vital dalam semua aktivitas ekonomi[2]. BBM juga menentukan faktor penting dalam menentukan perubahan harga bahan pokok atau inflasi di suatu negara[3]. Pemerintah melihat bisnis minyak dan gas alam sangat berperan penting dalam menghasilkan masukan yang nyata untuk pertumbuhan ekonomi Nasional BBM yang dikelola oleh pemerintah Indonesia dengan UUD No. 22 tahun 2001 tentang minyak dan gas. Pemerintah mencoba pembelian bahan bakar minyak (BBM) bersubsidi yaitu pertalite dan solar dengan mengeluarkan program MyPertamina. Dengan tujuan agar penyaluran pertalite dan solar tepat sasaran agar nantinya aplikasi MyPertamina bisa membantu memudahkan kinerja pemerintah[4].

Aplikasi adalah perangkat lunak yang dikembangkan untuk melakukan tugas tertentu yang bertujuan untuk memudahkan kinerja atau transaksi manusia dengan manusia lainnya[5]. Mobile Payment merupakan hasil dari perkembangan teknologi, dimana mobile payment diciptakan untuk proses pembayaran dari tunai menjadi non-tunai dengan menggunakan perangkat handphone atau smartphone pembayarannya menggunakan QR Code (quick response code) OTP (one time password). dan NFC (near field communication)[6]. Efek dari kemajuan teknologi sangat berdampak bagi kehidupan manusia, yang dari awalnya semua aktivitas transaksi dilakukan secara

antarmuka, kini sudah bisa dilakukan dengan jarak yang jauh dan kapanpun selagi semua akses internet terhubung dari satu dengan yang lainnya.

MyPertamina adalah aplikasi layanan digital yang dikembangkan oleh perusahaan Pertamina, melalui aplikasi MyPertamina diharapkan memudahkan masyarakat dalam pembelian BBM di Stasiun Pengisian Bahan Bakar Umum (SPBU) secara cashless. Aplikasi MyPertamina dirancang untuk membantu kebutuhan masyarakat agar transaksi menjadi lebih mudah[7]. Peluncuran atau uji coba aplikasi MyPertamina pertama kali pada tanggal 1 Juli 2022 di beberapa wilayah dan kabupaten di Indonesia. Adapun wilayah dan kabupaten yang termasuk dalam uji coba aplikasi My Pertamina yaitu : Kota Bukittinggi, Kabupaten Agam, Kota Padang Panjang, Kabupaten Tanah Datar, Kota Banjarmasin, Kota Bandung, Kota Tasikmalaya, Kabupaten Ciamis, Kota Manado, Kota Yogyakarta dan Kota Sukabumi[8].

Berdasarkan penilaian yang dilakukan oleh pengguna aplikasi MyPertamina didasarkan dari pengalaman pengguna dalam menggunakan aplikasi MyPertamina belum memuaskan. Dilihat dari rating pada Google Play Store yang memiliki rentang 1.0 sampai dengan 5.0, rating yang didapatkan aplikasi MyPertamina adalah 2,9 dengan ulasan tiga ratus empat belas ribu, rating tersebut didapatkan berdasarkan review dari pengalaman pengguna aplikasi MyPertamina setelah menggunakan aplikasi tersebut[9].



Gambar 1. Ulasan dan rating aplikasi MyPertamina di Google Play

Ada beberapa kendala yang dirasakan pengguna aplikasi MyPertamina adalah sering error pada aplikasi, sering terjadi bug yang kemudian membuat pengguna menjadi malas menggunakan aplikasi MyPertamina dan juga kendalanya adalah petugas di lapangan kurang memahami penerapan penggunaan aplikasi MyPertamina [7]. Dari opini dan sentiment pengguna review aplikasi MyPertamina sangat membantu sebagai evaluasi program MyPertamina agar nantinya lebih baik dari sebelumnya, dikarenakan program MyPertamina harus mempertimbangkan berbagai masukan, diantaranya dengan melihat respon dan opini masyarakat.

Analisis sentiment adalah pendapat atau opini seseorang terhadap suatu barang atau objek yang biasanya hasil dari pendapat itu bisa berupa pendapat positif, negatif ataupun netral[10]. Analisis sentiment bisa diartikan sebagai pendapat dari seseorang atau lebih terkait objek atau topik. Tujuannya agar nantinya bisa dijadikan sebagai keputusan untuk mengetahui apakah seseorang suka atau tidak suka dengan suatu objek. Biasanya nilai yang diberikan terkait analisis sentiment adalah negatif dan positif[11]. Terdapat dua cara pendekatan dalam melakukan sentiment analysis, yaitu dengan lexicon-based (pendekatan berbasis leksikal) dan learning-based (pendekatan dengan machine learning). Pendekatan lexicon-based bergantung pada kamus opini (lexicon) untuk melakukan klasifikasi sedangkan learning-based menggunakan dataset yang telah diklasifikasi secara manual sebelumnya untuk sebagai data latihan untuk memperoleh klasifikasi teks opini secara otomatis[12].

Penelitian mengenai analisis sentiment yang dilakukan oleh R.Mahendrajaya dkk [13] menggunakan metode Lexicon Based dan Support Vector Machine (SVM) dan TF-IDF untuk mengetahui tingkat akurasi dua kernel dari Support Vector Machine. Pelabelan menggunakan lexicon based dengan menghitung jumlah atau skor sentiment, proses selanjutnya kata-kata yang telah diketahui mengandung nilai positif, negatif dan netral yang terdapat di dalam suatu kalimat, dihitung berdasarkan jumlah nilai opini, untuk sentiment bernilai 1 positif, jika netral bernilai 0, untuk negatif bernilai -1. Hasil dari penelitian ini diperoleh akurasi 89,17%, presisi sebesar 91,5% dan recall 97,83%, namun kernel polynomial memperoleh akurasi 84,38%, presisi 84,23% dan recall 97,83%.

Penelitian selanjutnya dilakukan oleh [14] menggunakan algoritma Convolutional Neural Network (CNN) untuk mencari performa Word Embedding Word2vec, GloVe, Dan FastText Pada Klasifikasi Teks. Hasil dari penelitian ini algoritma Convolutional Neural Network (CNN) dengan word2vec mendapatkan performa 0.925 dengan dataset Reuters News 0.694, GloVe mendapatkan performa 0.958 dengan dataset 0.688 dan FastText

mendapatkan performa 0.979 dengan dataset 0.715. Kesimpulannya adalah performa terbaik didapatkan word embedding FastText karena FastText bisa diandalkan untuk permasalahan out of vocabulary.

Penelitian selanjutnya dilakukan oleh [15] menggunakan Wordembedding pada algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) untuk membandingkan kinerja terbaik dari kedua algoritma tersebut. Hasil yang didapatkan dari penelitian ini SVM sangat mendominasi dibandingkan (NKN), dengan akurasi untuk Algoritma Support Vector Machine (SVM) 81,90%, presisi 81,91% dan recall 81,90% sedangkan untuk Algoritma K-Nearest Neighbor (KNN) akurasi 65,60%, presisi 65.75%, dan recall 65,54%.

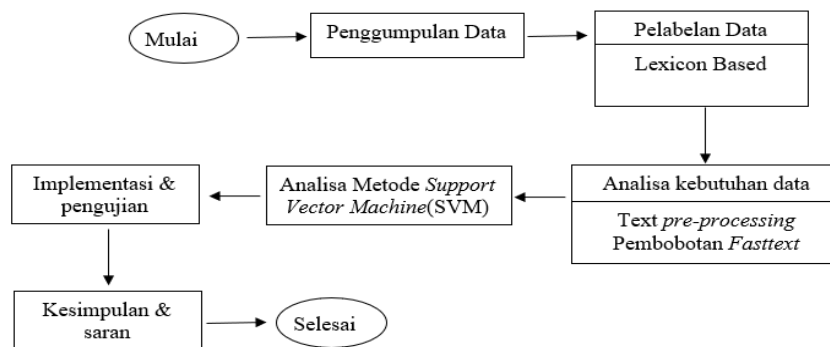
Penelitian selanjutnya dilakukan oleh [16] dengan menggunakan metode support Vector Machine dan TF-IDF terkait vaksin covid-19 untuk menghitung akurasi metode Support Vector Machine dalam mengklasifikasi sentiment terhadap vaksin covid-19 pada media sosial Twitter, kelas sentiment dibagi menjadi 3 yaitu positif, negatif dan netral, jumlah data dalam penelitian ini yaitu 9178 diantaranya 8000 data training, 400 data testing dan 778 validasi. Hasil dari penelitian didapatkan akurasi pelatihan 99%, akurasi saat validasi 71% dan akurasi saat pengujian sebesar 65%, recall 55%, precision 61% dan f1-score 57%.

Berdasarkan penelitian terkait maka penelitian ini akan melakukan analisis sentiment mengenai review aplikasi MyPertamina pada google play store dengan metode wordembedding FastText dan SVM, kelas sentiment dibagi dalam 3 kelompok diantaranya kelompok kelas positif, negatif dan netral. Batasan dalam penelitian hanya menggunakan dataset terkait review pengguna aplikasi MyPertamina. Data yang dipakai dalam penelitian ini sebanyak 20.000 data. Pelabelan data dilakukan secara otomatis menggunakan kamus vader lexicon. SVM melakukan Tunning parameter menggunakan GridSearch untuk mendapatkan model dari setiap parameter yang ada pada SVM, sedangkan wordembedding FastText sebagai kamus vektor untuk menghasilkan performa akurasi dari dataset review aplikasi MyPertamina. Penelitian ini menggunakan bahasa pemrograman python dan confusion matrix untuk melakukan pengujian akurasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan penelitian

Tahapan penelitian ini dibagi dalam beberapa tahapan yaitu, tahap pengumpulan data, pelabelan data, text pre-processing, pembobotan kata FastText, metode support vector machine, implementasi dan pengujian. Tahapan penelitian dapat dilihat pada gambar 2.



Gambar 2. Tahapan penelitian review aplikasi My pertamina menggunakan SVM

a. Identifikasi Masalah

Identifikasi masalah pada penelitian ini yaitu mengidentifikasi sentiment review aplikasi MyPertamina di google play dengan menerapkan metode Support Vector Machine (SVM) dan pembobotan kata fitur FastText untuk mendapatkan tingkat akurasi pada analisis sentiment review aplikasi Mypertamina. Berikut penjelasan dan tahapan dalam penelitian ini.

b. Pengumpulan Data

Pengambilan data dalam penelitian ini menggunakan crawling terkait data review pengguna aplikasi MyPertamina sebanyak 20.000 data dengan menggunakan bahasa pemrograman python di google collab dengan kata kunci 'com.dafturn.mypertamina'. Data yang didapat dari crawling dari tanggal 1 Oktober 2022 kemudian data diberikan label untuk setiap review yang mengandung sentiment. tahap pelabelan data review aplikasi MyPertamina dilabel dengan metode Lexicon Based. Setelah dilakukan pemilahan ulang data, data valid yang digunakan dalam penelitian ini berjumlah 8000 data, kemudian data tersebut ditranslate ke bahasa inggris agar nantinya proses pelabelan sesuai dengan kamus lexicon base vader yang berbahasa inggris, selanjutnya hasil dari translate akan dipindahkan kembali ke data yang berbahasa indonesia. kemudian data dibagi ke dalam 2 kelompok data yaitu data latih atau data training 7200 data dan data uji atau data test berjumlah 800 data.. Hasil analisis dalam penelitian ini mendapatkan jumlah respon yang paling banyak dari review pengguna aplikasi MyPertamina adalah negatif, karena banyak ditemukan ulasan pengguna aplikasi MyPertamina yaitu "setelah update aplikasi MyPertamina semakin parah bug nya".

c. Lexicon Vader

VADER (Valence Aware Dictionary and Sentiment Reasoner) adalah salah satu metode analisis lexicon based. Cara kerja kerjanya dengan menganalisis text berdasarkan kamus laxicon yang akan menghasilkan class sentiment. Dalam jurnal ini (Hutto, C.J. & Gilbert, E.E. 2014), penelitian analisis sentiment pada VADER dapat dibagi menjadi empat bagian: positif, negatif, netral, dan majemuk. Vader Sentiment Lexicon adalah salah satu kamus Lexicon, kamus ini memiliki 7500 token yang berisi g kata bahasa Inggris dan emosi serta pendekatan yang berupa gabungan huruf dan suku kata [17]. Skor dictionary based atau laxicon based vader dihitung menggunakan rumus :

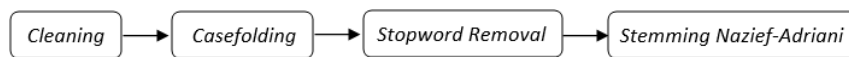
if $\sum k \text{ Score } (k) > 0$ then positive (1)

if $\sum k \text{ Score } (k) < 0$ then Negative (2)

if $\sum k \text{ Score } (k) = 0$ then Netral (3)

d. Text Pre-processing

Text pre-processing adalah salah satu proses dalam pengolahan data, terutama untuk membersihkan data yang nantinya dipakai sebagai data penelitian, tujuannya adalah berfungsi menghapus kata-kata yang dianggap tidak penting sehingga membuat data menjadi bersih dan terstruktur, contoh kata yang dimaksudkan seperti simbol, nomor hashtag, dll. Agar nantinya perhitungan menjadi optimal[18]. Dalam penelitian ini akan dilakukan 4 tahapan text preprocessing bisa dilihat pada gambar 3



Gambar 3. Tahapan text pre-processing

1. **Cleaning**

Tahapn ini menghilangkan emoticon, dan karakter yang tidak memiliki makna dalam text seperti : hastag (#), tanda baca (“”) dan spasi yang berlebihan.

2. **Casefolding**

Tahapan ini merubah semua kata yang huruf dasarnya kapital menjadi huruf kecil.

3. **Stopword Removal**

Tahapan ini adalah proses meghilangkan kata yang termasuk ke dalam katagori stopwords. Stopword sendiri ialah kata yang sering muncul namun tidak memiliki arti.

4. **Stemming Nazif-Adriani**

Stemming Nazief-Adriani adalah algoritma stemming untuk text berbahasa indonesia. Stemming adalah proses menemukan kata dasar dengan menghilangkan imbuhan yang menempel pada kata.

e. FastText

FastText adalah model Wordembedding yang dikembangkan oleh Facebook. FastText adalah evolusi dari model wordembedding Word2Vec. FastText mampu mencapai kinerja luar biasa dalam representasi kata dan klasifikasi kalimat dengan mengeksplotasi informasi tingkat karakter, terutama dalam kasus kata-kata langka [18]. Cara kerja FastText yaitu memakai banyak kata untuk diolah, tidak hanya satu kata saja. FastText memakai n-gram untuk pengolahan kata. Contoh aplikasi n-gram pada kata “program”, FastText bekerja menggunakan trigram (n=3) dimana kata akan di bilah menjadi “pro”,”rog”,”ogr”,”gra”,”ram”. FastText terbukti efektif dalam klasifikasi teks seperti sentiment analysis[19].

f. SVM (Support Vector Machine)

SVM (Support Vector Machine) termasuk machine learning (supervised learning) yang bisa memprediksi setiap kelas dengan melihat dari hasil proses pelatihan. SVM bisa diartikan sebagai garis pemisah atau pembatas yang umumnya disebut dengan hyperplane, dimana garis ini bertujuan untuk memisahkan review dengan sentiment positif, negatif dan netral[18]. Adapun tahapan pada metode SVM untuk penelitian yang akan dilakukan yaitu :

1. Review pada data train dan data testing diubah menjadi vektor dengan menggunakan model FastText.
2. Data dibagi menjadi 3 vorsni data yaitu 90:10, 80:20, dan 70:30
3. Data latih (training) dipakai untuk pelatihan model SVM.
4. Tuning parameter SVM menggunakan metode GridSearchCV untuk mencari hasil maxsimal kernel, nilai C dan gamma. Pada penelitian ini kernel yang digunakan adalah {Sigmoid, Polynominal, RBF} beserta nilai C = {0.1, 1, dan 10} dan nilai gamma = {10, 1, 0.1, 0.01}.
5. Memilih satu model terbaik dari hasil tuning parameter, setelah mendapatkan model terbaik, kemudian data diuji coba dengan setiap nilai C dan gamma, untuk melihat perbandingan tingkat akurasi dari 3 vorsni data yang telah ditetapkan.
6. Model terbaik digunakan untuk klasifikasi data uji (test).

g. Pengujian Confusion Matrix

Hasil dari pengujian ini akan dievaluasi dengan confusion matrix untuk melihat hasil sistem yang telah didapatkan. Confusion matrix adalah sebuah teknik atau cara yang digunakan untuk mengevaluasi klasifikasi model untuk mengukur objek yang benar atau salah[20]. Pada Confusion matrix akurasi adalah gambaran keakuratan model

dalam mengklasifikasikan, presisi adalah gambaran tingkat akurasi yang diminta dengan data yang dihasilkan model, sedangkan recall adalah gambaran keberhasilan model dalam menemukan informasi. Pengukuran akurasi menggunakan confusion matrix dapat dilihat pada tabel 1.

Tabel 1. Confusion Matrix

	Positif	Negatif	Netral
Positif	Tpos	FPosNeg	FposNet
Negatif	FnegPos	Tneg	FnegNet
Netral	FNetPos	FnetNeg	TNet

$$Akurasi = \frac{TP+TN}{TP++TN+FP+FN} \times 100\% \tag{4}$$

$$Presisi = \frac{TP}{FP+TP} \times 100\% \tag{5}$$

$$Recall = \frac{TP}{FN+TP} \times 100\% \tag{6}$$

Dimana :

TP = True Positif, data positif yang diprediksi bernilai positif atau benar

FN = Fals Negatif, data positif namun diprediksi sebagai data negatif atau salah

TN = True Negatif, data negatif yang diprediksi positif atau benar

FP = False Negatif, data positif namun diprediksi sebagai data positif atau salah

3. HASIL DAN PEMBAHASAN

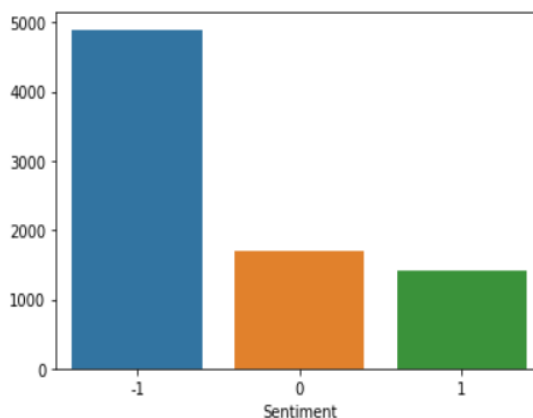
Tahapan ini adalah hasil dan pembahasan di setiap proses di penelitian ini, yaitu tentang analisa sentiment mengenai review aplikasi MyPertamina pada google play store menggunakan Fasttext dan SVM (Support Vector Machine).

3.1 Translate dan Pelabelan

Sebelum data di label menggunakan Laxicon based, data diterjemahkan yang awalnya bahasa indonesia menjadi bahasa inggris, disebabkan pada proses pelabelan data yang digunakan adalah kamus vader sentiment. Dataset yang berjumlah 8000 data dari review aplikasi MyPertamina setelah dilakukan proses pelabelan menghasilkan sebanyak 1405 kelas positif, 1698 kelas netral dan 4897 kelas negatif. Hasil translate dan pelabelan serta visualisasi data bisa dilihat pada tabel 2 dan gambar 4. Tabel 2 adalah hasil translate dan pelabelan, gambar 4. visualisasi jumlah kelas sentiment setelah dilakukan pelabelan. Warna biru pada gambar 3 melambangkan kelas sentiment negatif, warna oren melambangkan kelas sentiment netral dan warna hijau adalah kelas sentiment positif.

Tabel 2. Hasil translate dan pelabelan

Content	Content	Vader Sentiment	Sentiment
Pertamina makin di depan	Pertamina is getting ahead	0.1000	Positive
Sangat membantu	Very helpful	0.4754	Positive
Terima kasih	Thank you	0.3612	positive



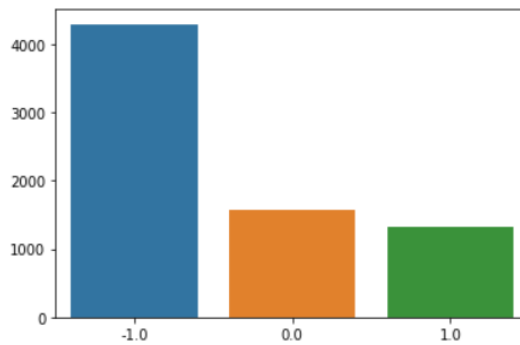
Gambar 4. Visualisasi data hasil pelabelan review aplikasi MyPertamina

Setelah data di translate dan di label, selanjutnya data dibagi menjadi 3 porsi data dengan perbandingan jumlah data latih (Training) dan data uji (Test) yaitu 90%:10, 80%:20% , dan 70%:30% . Pembagian jumlah data bisa dilihat pada tabel 3.

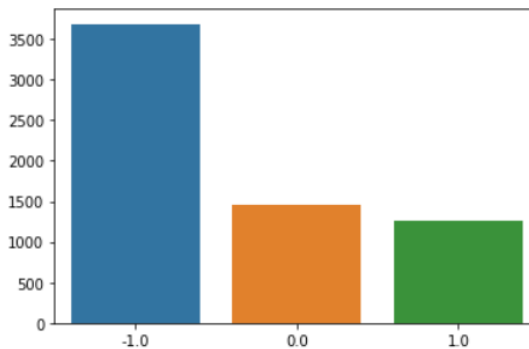
Tabel 3. Pembagian jumlah data

Dataset	Jumlah Data			
	Latih(Training)	Uji(Test)	Latih(Training)	Uji(Test)
Dataset A	90%	10%	7200	800
Dataset B	80%	20%	6400	1600
Dataset C	70%	30%	5600	2400

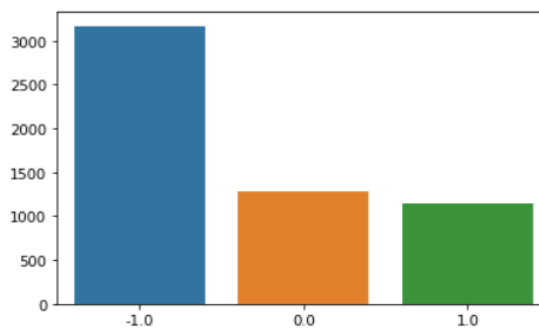
Tahapan berikutnya pembagian jumlah data menjadi 3 porsi data, dari setiap dataset memiliki jumlah kelas sentiment yang berbeda, dataset A memiliki jumlah kelas sentiment positif berjumlah 1325, netral berjumlah 1575 dan negatif berjumlah 4300, dataset B memiliki jumlah kelas sentiment positif 1252, netral 1461 dan negatif 3687, dan dataset C memiliki jumlah kelas sentiment positif 1146, netral 1286 dan negatif 3136. Visualisasi jumlah kelas sentiment pada tiga Dataset A, B dan C bisa dilihat pada gambar 5, 6 dan 7.



Gambar 5. Visualisasi jumlah kelas sentiment pada Dataset A



Gambar 6. Visualisasi jumlah kelas sentiment pada Dataset B



Gambar 7. Visualisasi jumlah kelas sentiment pada Dataset C

3.2 Text Pre-processing

Text pre-processing adalah tahap penting dalam melakukan analisis klasifikasi. Contoh kalimat pada proses text pre-processing yaitu “Aplikasinya bagus dan Membantu.,” hasil dari text pre-processing pada penelitian ini bisa dilihat pada tabel 4.

Tabel 4. Text pre-processing

Pre-processing	Hasil
Kalimat dasar	Aplikasinya bagus dan Membantu.,

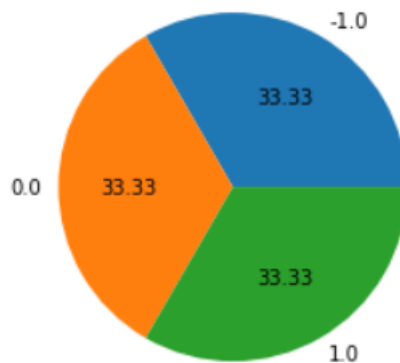
Pre-processing	Hasil
Celaning	Aplikasinya bagus dan Membantu
case Folding	aplikasinya bagus dan membantu
Stopwords Removal	aplikasinya bagus membantu
Stemming	aplikasi bagus bantu

3.3 Undersampling

Undersampling adalah proses penyetaraan data dengan mengacu pada data kelas sentiment terendah dari 3 kelas sentiment. untuk dataset A dengan porsi data 90%:10% kelas sentiment terendah adalah 1325. Untuk lebih jelas terkait data undersampling bisa dilihat pada tabel 5 dan gambar 8 visualisasi data undersampling .

Tabel 5. Undersampling data

Dataset	Positif	Netral	Negatif	Jumlah Data
Dataset A	1325	1325	1325	3975



Gambar 8. Visualisasi data undersampling

3.4 FastText

FastText adalah perubahan kalimat atau kata ke bentuk vektor, dalam penelitian ini fasttext dibangun menggunakan data training dan data test hingga didapatkan kamus vektor. Kamus tersebut nantinya digunakan untuk vektorisasi pada review data training dan data test untuk menjadi inputan pada SVM . adapun cara kerja vektorisasi ini yaitu, dengan menghitung setiap vektor pada kamus vektor dan dilakukan penjumlahan sesuai dengan kalimat pada review. Berikut contoh vektorisasi :

Kamus Word embedding FastText:

Kamus Vektor = ([0.23183146, 0.68667823, 0.11995368,....., 0.6883]
 [0.04239526, -0.18824285, -0.5649203,....., 2.0305202]
 [0.29710793, -0.12721363, -0.29338947,....., -1.1695946]

 [-1.1397922, 0.75067955, -0.84227026 ,.....1.3697175]
 [-0.23304936, -0.02467519, -1.1641237,.....-0.21433264])
 Pertamina = ([0.23183146, 0.68667823, 0.11995368,.....-0.21433264])
 Makin = ([0.4414671, 2.07399, 0.14900331,.....0.36202338])
 Di = ([-3.0183673, 5.504036, 0.22811006,.....-0.7291017])
 Depan = ([-0.24270841, -0.6015855, -0.9359991,.....-0.42745432])
 Pertamina makin di depan = ([-4.20214273e-02, 3.24152201e-01, 2.40940433e-02,.....-1.29352316e-01])

3.5 Tuning Parameter

Adapun tujuan dilakukannya tuning parameter adalah untuk meningkatkan performa machine learning yang akan dibangun. Peningkatan performa ini dibutuhkan untuk mengetahui parameter terbaik untuk model. sedangkan optimasi parameter pada SVM dilakukan menggunakan GridSearchCV. Parameter SVM yang digunakan dalam penelitian ini yaitu :

Kernel = { RBF, Sigmoid, Polynomial }
 C = { 0.1, 1, 10 }
 Gamma = { 10, 1, 0.1, 0.01 }

Pelatihan model SVM menggunakan data Training dengan mencari parameter terbaik yang akan divalidasi menggunakan data test untuk mendapatkan hasil model yang optimal. Berikut ini hasil dari tuning parameter yang dilakukan pada SVM dengan jumlah porsi data 90%:10%, 80%:20% dan 70%:30%. Adapun hasil, akurasi dan training model pada masing-masing porsi data bisa dilihat pada Tabel 6, 7, 8 dan 9

Tabel 6. Hasil akurasi data 90:10

Kernel	C/gamma	Akurasi		
		0.1	1	10
RBF	10	75%	75%	75%
	1	75%	75%	75%
	0.1	75%	76%	75%
	0.01	76%	80%	79%

Tabel 7. Hasil akurasi data 80:20

Kernel	C/gamma	Akurasi		
		0.1	1	10
RBF	10	76%	76%	76%
	1	76%	76%	76%
	0.1	76%	77%	76%
	0.01	77%	80%	78%

Tabel 8. Hasil akurasi data 70:30

Kernel	C/gamma	Akurasi		
		0.1	1	10
RBF	10	72%	72%	72%
	1	72%	73%	73%
	0.1	72%	74%	72%
	0.01	73%	76%	75%

Tabel 9. Training Model Dan Akurasi Tanpa Undersampling

Dataset	Parameter (GridSearch CV)						
	Kernel	C	gamma	Training Model	Akurasi	recall	precision
Dataset A	RBF	1	0.01	67%	80%	50%	84%
Dataset B		1	0.01	65%	80%	50%	79%
Dataset C		1	0.01	64%	76%	48%	74%

Setelah dilakukan pengujian dari 3 Dataset A, B dan C tanpa undersampling data, diperoleh model terbaik dan akurasi tertinggi yang nantinya akan diambil sebagai hasil akhir dalam penelitian ini, dimana pada tabel di atas dapat dilihat, model terbaik dan akurasi tertinggi terdapat pada Dataset A. dengan mendapatkan hasil training model 67%, akurasi 80%, recall 50% dan precision 84%. Dikarenakan jumlah kelas sentiment di pengujian Dataset A mengalami ketimpangan jumlah data yang bisa dilihat pada gambar 4, maka langkah selanjutnya dalam penelitian ini juga akan dilakukan kesetaraan data dengan metode undersampling untuk mengetahui tingkat akurasinya. Hasil perbandingan akurasi dataset A tanpa undersampling dan undersampling bisa dilihat pada tabel 10

Tabel 10. Akurasi sebelum dan sesudah undersampling

Dataset A	SVM		
	Akurasi	recall	precision
Tanpa undersampling	80%	50%	84%
Undersampling	67%	69%	57%

4. KESIMPULAN

Dari penelitian yang telah dilakukan tentang analisis sentiment review aplikasi MyPertamina pada google play store dengan word embedding FastText dan SVM yang dibagi menjadi 3 kelas sentiment dan 3 perbandingan porsi data yang berbeda dilakukan pengujian penyeimbangan data dan tanpa penyeimbangan data menggunakan undersampling, dapat ditarik kesimpulan bahwa porsi data yang memiliki akurasi tertinggi diperoleh pada Dataset A dengan porsi data 90:10 dengan parameter kernel RBF C=1 gamma =0.01 mendapatkan hasil training model 67% serta menghasilkan akurasi 80%, recall 50%, precision 84% untuk data tanpa undersampling. Sedangkan data yang di undersampling mendapatkan akurasi 67% nilai recall 69% dan nilai precision 57%. Penurunan akurasi setelah dilakukannya undersampling untuk porsi data 90:10 disebabkan oleh pengurangan jumlah data training dari yang awalnya berjumlah 7200 data menjadi 3976 data.

REFERENCES

- [1] M. Andriani, "Analisis Pengaruh Kualitas Pelayanan Terhadap keputusan Pembelian pada PT.Pertamina (Studi Pada PT.Pertamina Cemara Asri, Medan)," *J. Stindo Prof.*, vol. VI, no. September, pp. 3–7, 2020.
- [2] D. Yuliani, S. Saryono, D. Apriani, Maghfiroh, and M. Ro, "Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok (Sembako) Di Kecamatan Tambun Selatan Dalam Masa Pandemi," *J. Citizsh. Virtues*, vol. 2, no. 2, pp. 320–326, 2022.
- [3] G. Rozy Hrp, N. Aslami, and P. Studi Manajemen Fakultas Ekonomi Bisnis Islam, "Analisis Damfak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 2, no. 1, pp. 1464–1474, 2022.
- [4] Y. S. N. Lutfi, Ahmad, Syamsir, Aulia Annisa Fitriani, Ira Ramadani, Nabilah Azahra Putri, "Efektivitas Penggunaan Aplikasi My Pertamina di Era Kenaikan BBM Bersubsidi," *J. Pros. Mateandrau*, vol. 1, no. 2, 2022.
- [5] A. Surahmat, "RANCANG BANGUN APLIKASI SISTEM PENJUALAN PADA PERCETAKAN CUBIC ART," *Junal JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 81–86, 2023.
- [6] E. O. Safitri, Y. T. Musityo, and N. H. Wardhani, "Analisis Perilaku Penggunaan Mobile Payment Aplikasi OVO menggunakan Technology Acceptance Model (TAM) Termodifikasi," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, pp. 8184–8189, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] A. A. Sinurat, C. Hendriyani, F. Damayanti, A. Sekretari, and M. Taruna, "Aplikasi MyPertamina Untuk Meningkatkan Keterlibatan Konsumen," *J. Int. Tinj. BISNIS*, vol. 5, no. 1, pp. 65–73, 2022.
- [8] N. K. Hikmawati, "Analisis Kualitas Layanan My Pertamina Menggunakan Pendekatan e-GovQual pada Beberapa Kota Percobaan MyPertamina Service Quality Analysis Using E-GovQual Approach in Several Trial Cities," *J. Manaj. Inform.*, vol. 12, pp. 100–111, 2022.
- [9] "play.google.com," 2022, [Online]. Available: <https://play.google.com/store/apps/details?id=com.dafturn.mypertamina&hl=in&gl=US>
- [10] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektrol dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [11] N. A. F. T. Ardianne luthfika Fairuz, Rima Dias Ramadhani, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial," *J. Data Sci. IOT, Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 10–12, 2021.
- [12] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon: Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021.
- [13] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode SVM," *J. Tek. Univ. Muhammadiyah Ponorogo*, vol. 3, no. 2, pp. 52–63, 2019.
- [14] Z. A. A. Nurdin, Bernadus Anggo Seno Aji, Anugrayai Bustamin, "Perbandingan Kinerja Word Embedding Word2vec, Glove," *J. TEKNOKOMPAK*, vol. 14, no. 2, 2020.
- [15] T. S. Sabrila, V. R. Sari, and A. E. Minarno, "Analisis Sentimen Pada Tweet Tentang Penanganan Covid - 19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K - Nearest Neighbor," *Fountain Informatics J.*, vol. 6, no. 2, 2021.
- [16] M. Rizky, Analisis Sentimen Masyarakat Terhadap Vaksin Covid-19 Menggunakan Metode Support Vector Machine Pada Media Sosial Twitter. 2021.
- [17] P. A. Sumitro, D. I. Mulyana, and W. Saputro, "Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based," *J. Inform. dan Teknol. Komput.*, vol. 02, no. 02, pp. 50–56, 2021.
- [18] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019.
- [19] E. Lim, T. I. Istts, E. I. Setiawan, and T. I. Istts, "Stance Classification Post Kesehatan di Media Sosial Dengan FastText Embedding dan Deep Learning," *J. Intell. Syst. Comput.* 66, pp. 65–73.
- [20] V. S. Ginting, K. Kusrini, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 36–44, 2020.