

Optimalisasi Metode *Support Vector Machine* (SVM) berbasis Optimized Weight Evolutionary dalam Penentuan Sentimen Komentar

Optimized Weight Evolutionary-based Support Vector Machine (SVM) Optimization for Comment Sentiment

¹Mulyana*, ²Wahyu Utomo

^{1,2}Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

^{1,2}Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, Jakarta, Indonesia

*e-mail: mul@mulyana.net

(*received*: 31 October 2024, *revised*: 2 November 2024, *accepted*: 16 January 2025)

Abstrak

Komentar dalam *feedback* kuesioner memiliki makna sentimen, seperti positif, negatif, atau netral. Setiap komentar ulasan terhadap layanan pelatihan memerlukan tindak lanjut cepat dan akurat untuk meningkatkan kualitas layanan. Namun, klasifikasi sentimen sering kali membutuhkan waktu dan tenaga besar untuk menentukan sentimen secara tepat. Penelitian ini bertujuan meningkatkan efisiensi dan akurasi dalam klasifikasi sentimen pada kuesioner pelatihan. Analisis komparatif dilakukan menggunakan tiga algoritma: Naïve Bayes, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM). Hasil menunjukkan SVM sebagai algoritma tercepat dan paling akurat, dengan waktu pembelajaran 3.067 detik, lebih cepat 100 milidetik dibanding Naïve Bayes dan 45,8 detik dibanding KNN. Akurasi SVM mencapai 60,81%, dengan rata-rata sensitivity sebesar 61%, specificity 80%, dan precision 63%. Selanjutnya, penelitian ini mengintegrasikan metode Optimized Weight Evolutionary untuk meningkatkan akurasi SVM dan menangani pemilihan atribut. Hasil uji menunjukkan peningkatan akurasi SVM sebesar 2,16%, menjadi 63,10%. Proses pembelajaran dilakukan pada dataset 1.153 komentar, dengan 90% data digunakan untuk pelatihan algoritma. Kombinasi SVM dan Optimized Weight Evolutionary terbukti efektif dalam mengklasifikasikan komentar lebih akurat. Hasil penelitian ini memberikan wawasan baru dalam aplikasi klasifikasi sentimen, khususnya untuk *feedback* pelatihan. Optimalisasi algoritma ini dapat membantu perusahaan pelatihan lebih responsif dalam menanggapi komentar dan meningkatkan kualitas layanan secara keseluruhan.

Kata kunci: *analisis sentimen, text classification, support vector machine, feature selection, SVM*

Abstract

Comments in questionnaire feedback carry sentiment meanings, such as positive, negative, or neutral. Each review comment on training services requires prompt and accurate follow-up to improve service quality. However, sentiment classification often demands significant time and effort to determine sentiments accurately. This study aims to enhance efficiency and accuracy in sentiment classification for training questionnaires. A comparative analysis was conducted using three algorithms: Naïve Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The results indicate that SVM is the fastest and most accurate algorithm, with a training time of 3.067 seconds, 100 milliseconds faster than Naïve Bayes and 45.8 seconds faster than KNN. SVM achieved an accuracy of 60.81%, with an average sensitivity of 61%, specificity of 80%, and precision of 63%. Subsequently, this study integrated the Optimized Weight Evolutionary method to enhance SVM's accuracy and address attribute selection. Testing results showed a 2.16% improvement in SVM accuracy, bringing it to 63.10%. The training process was conducted on a dataset of 1,153 comments, with 90% of the data used for algorithm training. The combination of SVM and Optimized Weight Evolutionary proved effective in achieving more accurate sentiment classification. This study provides

<http://sistemasi.ftik.unisi.ac.id>

new insights into the application of sentiment classification, particularly for training feedback. Optimizing the algorithm can help training companies respond more effectively to comments and improve overall service quality.

Keywords: *sentiment analysis, text classification, Support Vector Machine, feature selection, SVM*

1 Pendahuluan

Di era digital saat ini, data dari berbagai sumber seperti media sosial, situs web, sensor, dan transaksi bisnis terus meningkat pesat, membentuk "Big Data" [1]. Big Data memiliki potensi besar untuk memberikan wawasan berharga di berbagai bidang seperti bisnis, sains, pemerintahan, dan kesehatan. Namun, tantangan utama adalah mengelola, menganalisis, dan mengekstrak informasi dari data yang kompleks (baik terstruktur maupun tidak terstruktur). Di sinilah teknik text mining memainkan peran penting dalam menganalisis teks dalam Big Data. Text mining, juga dikenal sebagai data mining teks, adalah proses penggalian informasi dari teks tidak terstruktur dengan mengubahnya menjadi data semi-terstruktur [2]. Teknik ini melibatkan metode dan algoritma untuk mengekstrak informasi dan pola tersembunyi dalam teks, membantu mengidentifikasi tren, sentimen, topik, dan entitas dalam data teks yang besar.

Sentimen dalam teks merujuk pada sikap emosional dalam suatu teks[3]. Sikap ini dapat bersifat positif, negatif, atau netral. Sentimen ini sering ditemukan pada data ulasan atau komentar suatu produk atau layanan. PT Brainmatics Indonesia Cendekia menyediakan pelatihan dan mengumpulkan feedback peserta, termasuk komentar terkait materi, instruktur, dan fasilitas. Komentar ini belum diolah otomatis menjadi wawasan bermanfaat, padahal dapat membantu menilai keberhasilan pelatihan dan menyusun rekomendasi cepat. Saat ini, analisis sentimen dilakukan oleh pegawai, namun metode manual ini lambat dan berisiko menurunkan kualitas pelayanan.

Analisis sentimen telah banyak diteliti. Okoye meneliti feedback peserta pendidikan [4], sementara Tao mengkaji ulasan produk makanan [5]. Metode yang umum digunakan adalah Naive Bayes, SVM, Decision Trees, KNN, dan Neural Networks [6]. SVM populer untuk komparasi metode klasifikasi, namun menghadapi masalah dalam generalisasi [7] dan pemilihan fitur [8][9].

Tujuan dari penelitian ini yaitu melakukan komparasi algoritma untuk memilih metode klasifikasi sentimen komentar yang paling cepat, akurat dan melakukan pemilihan atribut yang berpengaruh sebagai teknik untuk meningkatkan akurasi klasifikasi.

2 Tinjauan Literatur

Tinjauan studi dilakukan untuk mengetahui berbagai masalah dan teknik *text classification* yang digunakan. Metodologi pencarian studi terdahulu studi mengadopsi artikel Kumar [10]. Pencarian studi terkait dilakukan dengan memasukkan istilah "text classification" dan "sentimen analysis". Pencarian referensi penelitian terkait menggunakan operator boolean "and" dan "or".

Ringkasan hasil literatur studi kemudian dilakukan klasifikasi berdasarkan pada tujuan tinjauan studi yang meliputi metode *text classification* apa saja yang paling banyak digunakan atau yang telah diteliti, jenis data apa saja yang diolah, berapa banyak data yang diolah dan bagaimana penelitian dilakukan, apakah menerapkan metode yang ada atau melakukan *improvement*, bagaimana teknik optimalisasi yang dilakukan, dan bagaimana metode pelabelan himpunan data yang dilakukan.

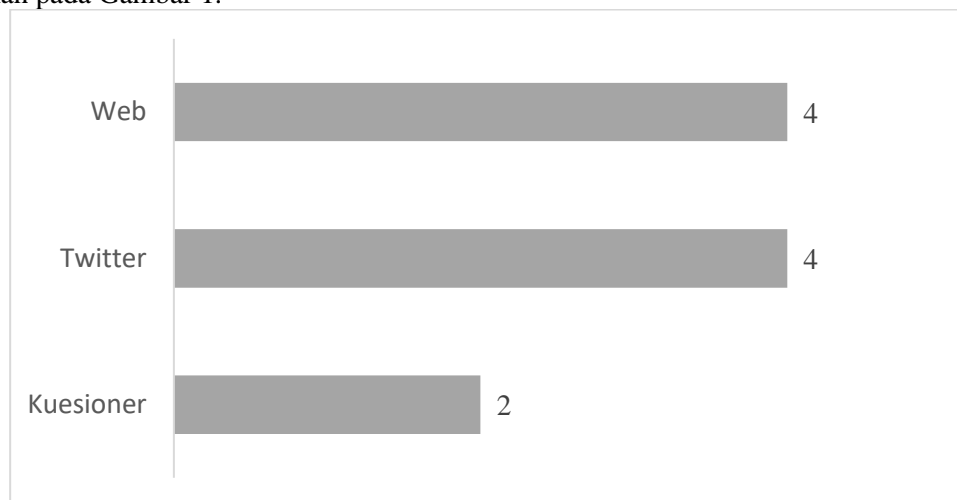
Sebanyak 4 (empat) dari 10 penelitian yang diteliti tidak mencantumkan spesifikasi data yang dianalisis. Berdasarkan 6 paper yang dianalisis, masing-masing menggunakan data mulai dari sebanyak 1.200 *record* hingga terbanyak sebanyak 1.600.000 *record*. Ringkasan jumlah data ditampilkan pada Tabel 1.

Tabel 1. Statistik data pada penelitian terdahulu

Statistik	Jumlah Data
Min	1.200
Maks	1.600.000
Rata-rata	346.851.72

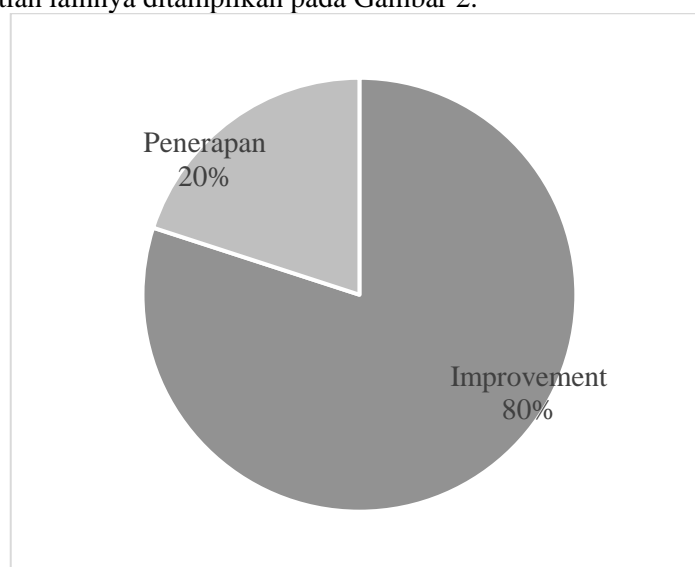
Ragam data yang digunakan dalam analisis *text mining* terdapat 3 jenis data. Data terbanyak yang diolah merupakan data terkait web (berita, website atau aplikasi berbasis web) dan data Twitter. Data

yang paling sedikit diolah untuk *text classification* atau sentimen yaitu terkait data Kuesioner, ditampilkan pada Gambar 1.



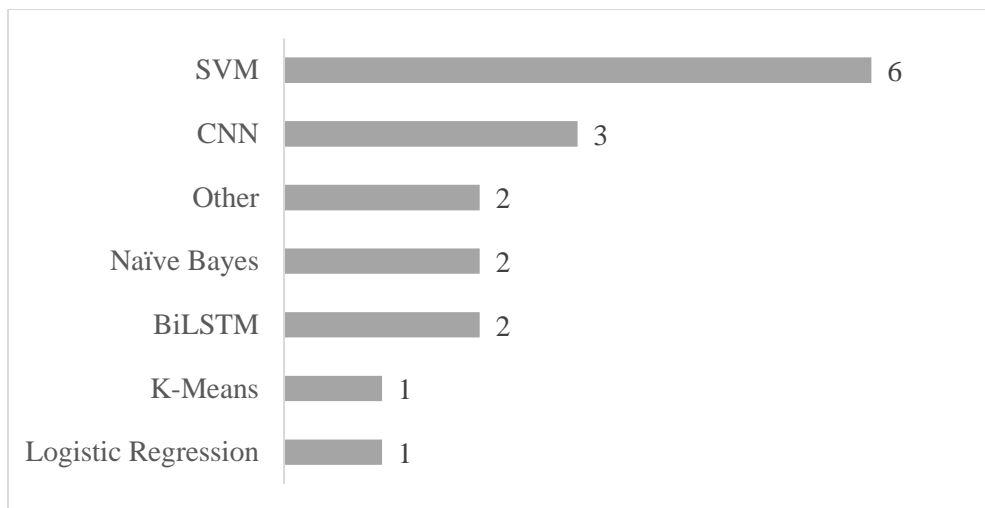
Gambar 1. Jenis data yang digunakan pada penelitian terdahulu

Bentuk penelitian yang dilakukan diklasifikasikan ke dalam 2 (dua) jenis penelitian. Penelitian pertama terkait *improvement* suatu metode atau algoritma *text classification*. Sedangkan jenis penelitian kedua yaitu terkait penerapan algoritma *text classification* atau *improvement* metode non *text mining*. *Improvement text classification* paling banyak dilakukan yaitu sebanyak 8 penelitian atau 80%. Nilai jenis penelitian lainnya ditampilkan pada Gambar 2.



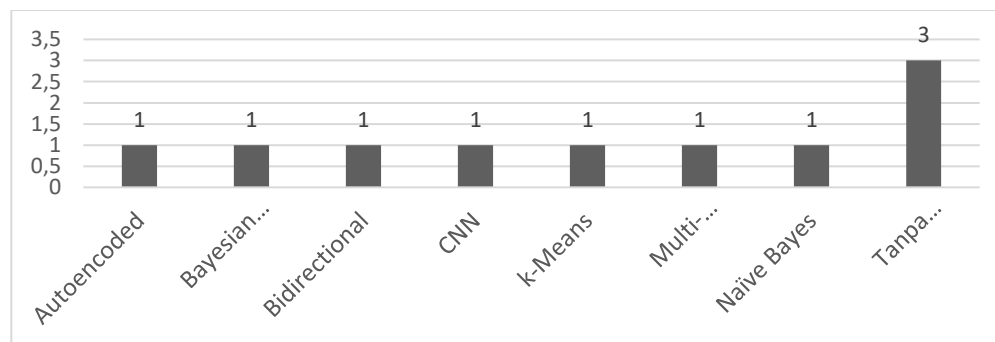
Gambar 2. Jenis penelitian terdahulu pada text mining

Penerapan maupun *improvement* algoritma *text classification* atau sentimen tidak lepas dari metode atau algoritma yang menjadi acuan pada penelitian. Hasil studi menunjukkan banyaknya variasi algoritma yang digunakan. Algoritma yang digunakan adalah untuk mengklasifikasi sentimen. Algoritma yang paling populer digunakan yaitu Support Vector Machine (SVM). Sedangkan yang paling sedikit adalah K-Means dan Logistic Regression. Statistik penggunaan algoritma pada penelitian ditampilkan pada Gambar 3.



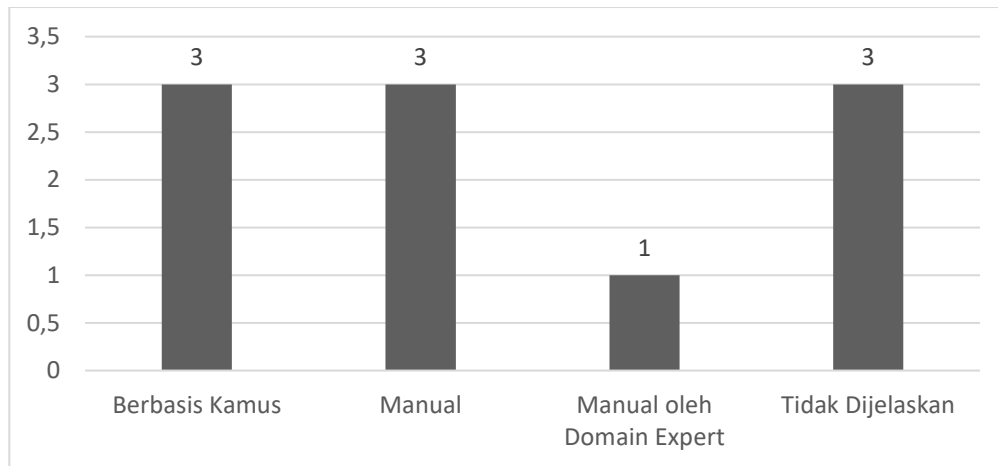
Gambar 3. Penggunaan algoritma pada penelitian terdahulu

Setiap penelitian sebelumnya menggunakan teknik pembobotan atau atau *feature selection* yang berbeda-beda dalam melakukan optimalisasi. Terdapat tiga penelitian yang tidak menggunakan teknik pembobotan. Pada penelitian sebelumnya tidak ditemukan adanya penggunaan teknik Optimized Weights Evolutionary atau Genetic Algorithm dalam *text mining*. Ringkasan metode yang dipakai ditampilkan pada Gambar 4.



Gambar 4. Teknik optimalisasi penelitian sebelumnya

Adapun teknik pelabelan yang digunakan pada dataset untuk menandai sentimen dari suatu kalimat secara umum terdapat 3 cara, seperti ditampilkan pada Gambar 5. Cara pertama yaitu menggunakan kamus, kemudian menggunakan teknik manual dan menggunakan manual secara spesifik melibatkan *domain expert*. Terdapat tiga penelitian tidak menyebutkan teknik apa yang digunakan. Sehingga cara manual merupakan teknik yang paling banyak digunakan (termasuk melibatkan *domain expert*).



Gambar 5. Teknik pelabelan penelitian sebelumnya

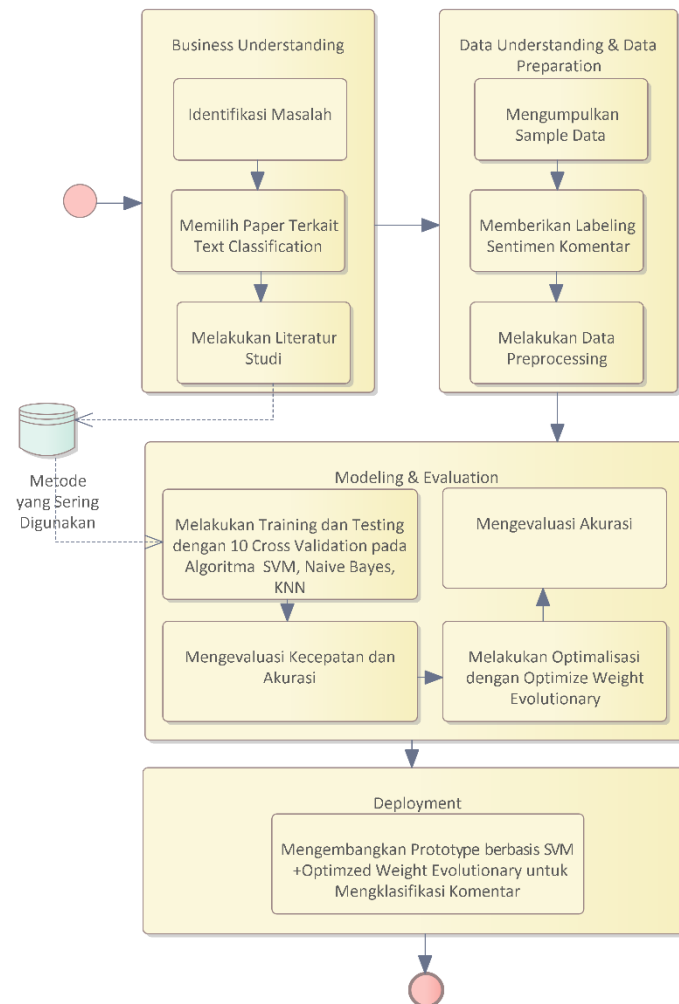
Berdasarkan tinjauan studi yang dilakukan pada sepuluh penelitian terdahulu, metode yang digunakan dalam analisis sentimen menghadapi masalah dalam pemilihan atribut untuk mengklasifikasi sentimen. Metode pembobotan atau pemilihan atribut dilakukan untuk meningkatkan akurasi. Teknik ini di antaranya menggunakan CNN, k-Means dan Naïve Bayes dan *autoencoded*.

Sedangkan pelabelan dataset yang disiapkan untuk pembentukan pola, terdapat dua teknik yaitu dilakukan secara manual dan dilakukan dengan basis kamus. Teknik manual dilakukan satu per satu atau melibatkan tim ahli. Sedangkan berbasis kamus yaitu ekstraksi sentimen dari kamus berisi kumpulan teks emosional.

Oleh karena itu, diajukan teknik pembobotan atau pemilihan atribut berbasis Optimized Weight yang belum pernah dilakukan pada penelitian sebelumnya. Adapun teknik pelabelan data mengacu pada kebanyakan penelitian sebelumnya yaitu dilakukan secara manual dan melibatkan tim ahli.

3 Metode Penelitian

Metode penelitian menggunakan metodologi Cross-Industry Process for Data Mining (CRISP-DM). CRISP-DM merupakan standar *de-facto* dalam proyek analisis, *data science* dan data mining [11]. Metodologi CRISP-DM terdiri dari fase business understanding, data understanding, data preparation, modeling, evaluation dan deployment, ditampilkan pada Gambar 6.



Gambar 6. Tahapan penelitian

3.1 Data Understanding dan Data Preparation

Pengumpulan data dilakukan dengan mengekspor data hasil kuesioner dari sistem pelatihan perusahaan (LMS) yang dikelola oleh Direktorat Learning. Adapun data yang dikelola terpisah pada *file spreadsheet* yang belum dikelola sistem, data kemudian digabungkan ke dalam *file* hasil ekspor dari sistem. Hasil penggabungan data kemudian disimpan dalam satu *file spreadsheet* (<http://s.id/dataset-komentar>).

Data yang digunakan dalam penelitian menggunakan data populasi komentar selama 8 tahun terakhir yaitu periode 2014 sampai dengan 2022. Namun demikian, karena adanya COVID-19, maka data tahun 2019 dan 2020 tidak termasuk. Jumlah data populasi adalah 1.153. Data kemudian dilakukan pelabelan yaitu penentuan sentimen komentar apakah positif, negatif atau netral. Pemilihan status sentimen mengacu pada penelitian Okoye [3]. Tiga sentimen juga dapat menggambarkan ketidaksukaan, kesukaan dan netral atau biasa-biasa saja [12]. Penerapan dalam perusahaan, sentimen negatif perlu ditindaklanjuti dengan melakukan perbaikan layanan. Sentimen positif akan dijadikan sebagai media *marketing* berupa testimonial peserta. Sedangkan sentimen netral tidak memerlukan tindakan khusus atau prioritas yang harus dilakukan perusahaan.

Proses pelabelan dilakukan secara manual yaitu menggunakan peran ahli yang memahami terkait konteks objek yang dibicarakan oleh peserta. Ahli dalam hal ini akan melibatkan para pejabat di perusahaan yang terdiri dari Direktur dan para manajer baik dari Direktorat Learning maupun Direktorat Marketing.

Data kemudian dilakukan sampling untuk menyamakan jumlah kelas. Setelah data terdistribusi secara seimbang, kemudian diambil sebesar 90% untuk proses *training* atau pembentukan pola.

Sedangkan 10% sisanya akan digunakan untuk pengujian. Data terdiri dari 6 atribut yaitu id, nama pelatihan, tanggal mulai, instruktur, komentar dan sentimen. Contoh data ditampilkan pada Tabel 2. Namun demikian, data yang akan diolah untuk analisis sentimen akan fokus pada atribut komentar dan sentimen untuk mengolah terkait sentimen yang terkandung dalam komentar.

Tabel 2. Contoh dataset

ID	Nama Pelatihan	Tanggal Mulai	Instruktur	Komentar	Sentimen
1	CCNA Exploration 4	30/1/2014	Ika A. S	Router dan Switch sangat lambat, perbarui	Negatif
3	CCNA Exploration 3	31/1/2014	Joko P	PC lambat, harusnya diupgrade	Negatif
8	CCNA Exploration 2	13/1/2014	Joko P	Perangkat praktik seperti Router dan Switch tidak kompatibel, perlu diupgrade	Negatif
12	Java Fundamental Enterprise Edition	9/4/2014	Hendro S	Pengunduran kursus beberapa periode karena waiting list, jadi mesti nunggu lama banget	Negatif
2	CCNA Exploration 4	30/1/2014	Ika A. S	Menu makan dibuat lebih bervariasi dan penambahan coffee break dengan sirop	Netral
4	System Analysis and Design	24/8/2014	Hendro S	Modul sebaiknya menggunakan versi Bahasa Indonesia	Netral
7	System Analysis and Design	24/8/2014	Hendro S	Kalau bisa ada parkir gratis	Netral
85	Adobe Flash CS4 - Brainmatics	18/8/2014	Irvan H. T	Bila ada promosi pelatihan kursus bisa di email ke saya	Netral
9	CCNA Exploration 3	15/2/2014	Ika A. S	Instruktur yang qualified	Positif
78	CCNA Exploration 4	14/6/2014	Ika A. S	Good	Positif
79	CCNA Exploration 4	14/6/2014	Ika A. S	Sudah baik	Positif
80	CCNA Exploration 4	14/6/2014	Ika A. S	Pelatihannya bagus, pertahankan dan lanjutkan	Positif

Instrumentasi yang digunakan dalam mendukung penelitian menggunakan tools Microsoft Excel dan Rapidminer. Microsoft excel digunakan untuk mengelola file dataset. Excel juga digunakan untuk melakukan *preprocessing* yaitu dengan memanfaatkan fitur Power Query dalam menggabungkan beberapa dataset. Rapidminer pada penelitian digunakan untuk melakukan tahapan *data preparation*

khususnya melakukan *preprocessing* dan juga pembagian 10 fold cross validation, perbandingan dan juga optimalisasi algoritma.

Pada tahapan *preprocessing* menggunakan text processing, agar teks dapat direpresentasikan dalam bentuk vektor atau matriks yang memungkinkan analisis lebih lanjut [13]. *Text mining* adalah teknik yang digunakan untuk menggali informasi berharga dari teks yang tidak terstruktur. Proses ekstraksi informasi meliputi langkah-langkah seperti pre-processing, tokenisasi, *stemming*, dan *stop word removal* [14]. Proses *feature extraction* melibatkan representasi teks dalam bentuk yang dapat digunakan dalam algoritma *text classification*, seperti model Bag of Words (BoW) dan metode *term frequency-inverse document frequency* (TF-IDF) [12]. Formula TF-IDF ditampilkan pada persamaan 1. N merupakan jumlah dokumen yang diproses.

Pada penelitian ini, dibandingkan juga antara penggunaan *stemming* dan tanpa *stemming*. Proses penyederhanaan kata sering meningkatkan kemampuan mengekstraksi dokumen yang relevan, terkadang juga dapat sedikit mengurangi keakuratan [15]. Terdapat banyak metode *stemming* yang diusulkan dalam Bahasa Indonesia, di antaranya adalah Nazief-Adriani, Arifin-Setiono Algorithm dan Sastrawi [16]. Pada studi perbandingan yang dilakukan, Sastrawi mampu melakukan *stemming* lebih baik dibandingkan metode lainnya. Pada penelitian menggunakan teknik *stemming* Sastrawi.

3.2 Modeling dan Evaluation

Tahapan *modeling* dan *evaluation* dilakukan proses *training* dan *testing* pada tiga algoritma untuk dikomparasi. Tiga algoritma yang dikomparasi yaitu SVM, Naïve Bayes dan KNN. Hasil *training* dan *testing* dikomparasi berdasarkan kecepatan dan akurasi. Hasil terbaik kemudian ditingkatkan akurasinya melalui teknik pemilihan atribut berupa pembobotandan kemudian dievaluasi kembali.

Proses *training* dan *testing* untuk menguji kecepatan proses algoritma diukur berdasarkan *training time* dan *inference time*. Pengukuran *training time* merupakan selisih antara waktu *training* dimulai dengan waktu *training* selesai, ditampilkan pada Persamaan 1. Sedangkan mengukur kecepatan prediksi atau *inference time* merupakan selisih antara waktu *testing* mulai dengan waktu *testing* selesai, sebagaimana ditampilkan pada Persamaan 2.

$$\text{Training time} = \text{Waktu Mulai-Waktu Training Selesai} \quad (1)$$

$$\text{Inference time} = \text{Waktu Testing Mulai-Waktu Testing Selesai} \quad (2)$$

Proses *training* dan *testing* untuk menguji akurasi dilakukan berbasis *cross validation*. *Cross validation* dapat memastikan bahwa model tersebar ke sebagian besar data selama pelatihan, berpotensi menghasilkan pembelajaran yang lebih baik dan evaluasi yang lebih akurat [16]. Metode *cross-validation* digunakan untuk menghindari *overlapping* pada data *testing*. *Cross-validation* disebut juga dengan *k-fold cross-validation*. *K-fold cross validation* seringkali subset dibuat *stratified* (bertingkat). Tahapan *cross-validation* terdiri dari:

1. Bagi data menjadi k subset yg berukuran sama
2. Gunakan setiap subset untuk data *testing* dan sisanya untuk data *training*

Metode evaluasi standar yang paling terkenal adalah *10-fold cross-validation*. Hasil dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa *10-fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat. *10-fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian.

Evaluasi dilakukan terhadap Evaluasi model klasifikasi terdapat beberapa ukuran yang dapat digunakan [23], di antaranya adalah

1. *Accuracy, Accuracy, Recognition Rate*, ditampilkan pada Persamaan 3
2. *Error Rate, Misclassification Rate*, ditampilkan pada Persamaan 4
3. *Sensitivity, True Positive Rate, Recall*, ditampilkan pada Persamaan 5
4. *Specificity, True Negative Rate* ditampilkan pada Persamaan 6
5. *Precision*, ditampilkan pada Persamaan 7
6. *F, F1, F-Score, Harmonic Mean of Precision and Recall*, ditampilkan pada Persamaan 8
7. $F\beta$

Ada empat istilah yang digunakan dalam merepresentasikan pengukuran model klasifikasi, yaitu sebagai berikut:

1. *True Positives* (TP) yaitu jumlah data positif yang ditandai benar oleh *classifier*
2. *True Negatives* (TN) yaitu jumlah data negatif yang ditandai benar oleh *classifier*
3. *False Positives* (FP) yaitu jumlah data negatif yang salah ditandai oleh *classifier*
4. *False Negatives* (FN) yaitu jumlah data positif yang salah ditandai oleh *classifier*

Empat istilah ini digambarkan dengan menggunakan *confusion matrix* dalam menginterpretasikan hasil pengukuran, seperti ditampilkan pada Tabel 3.

Tabel 3. Confusion matrix

		Predicted Class		
		Positive	Negative	Total
Actual Class	Positive	TP	FN	P
	Negative	FP	TN	N
	Total	P'	N'	P+N

$$Accuracy = \frac{TP+TN}{P+N} \quad (3)$$

$$Error Rate = \frac{FP+FN}{P+N} \quad (4)$$

$$Sensitivity = \frac{TP}{P} \quad (5)$$

$$Specificity = \frac{TN}{N} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$F, F1, F - score = \frac{2 \times precision \times recall}{precision+recall} \quad (8)$$

Setelah dilakukan optimalisasi, Uji-t dilakukan untuk melihat signifikansi hipotesis. Uji-t berpasangan (*paired t-test*) merupakan teknik statistik untuk analisis hipotesis yang digunakan ketika sampel datanya tidak independen atau berkaitan [17]. Karakteristik umum dalam situasi berkaitan ini adalah ketika subjek penelitian (atau objek studi) mengalami dua jenis intervensi yang berbeda. Meskipun subjek yang diuji adalah sama, penelitian ini menghasilkan dua jenis data sampel, yaitu hasil dari intervensi pertama dan kedua. Persamaan *paired t-test* ditampilkan pada Persamaan 9. Sebelum menghitung *paired t-test*, perlu diketahui terlebih dahulu nilai standar deviasi dan variansinya, seperti ditampilkan persamaannya pada 10 dan 11.

$$t_{hit} = \frac{\bar{D}}{\frac{SD}{\sqrt{n}}} \quad (9)$$

$$SD = \sqrt{var} \dots \quad (10)$$

$$var(s^2) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (11)$$

t = nilai t hitung

\bar{D} = rata-rata selisih pengukuran 1 dan 2

SD = standar deviasi selisih pengukuran 1 dan 2

n = jumlah sample

Interpretasi dari hasil perhitungan yaitu a. Untuk menginterpretasikan uji t-test terlebih dahulu harus ditentukan Nilai signifikansi α dan Df (*degree of freedom*) = N-k, khusus untuk paired sample t-test df = N-1. Selanjutnya adalah membandingkan t_{hit} dengan $t_{tab=\alpha; n-1}$. Apabila $t_{hit} >$

t_{tab} artinya berbeda secara signifikan (H_0 ditolak). Apabila $t_{hit} < t_{tab}$ artinya tidak berbeda secara signifikan (H_0 diterima).

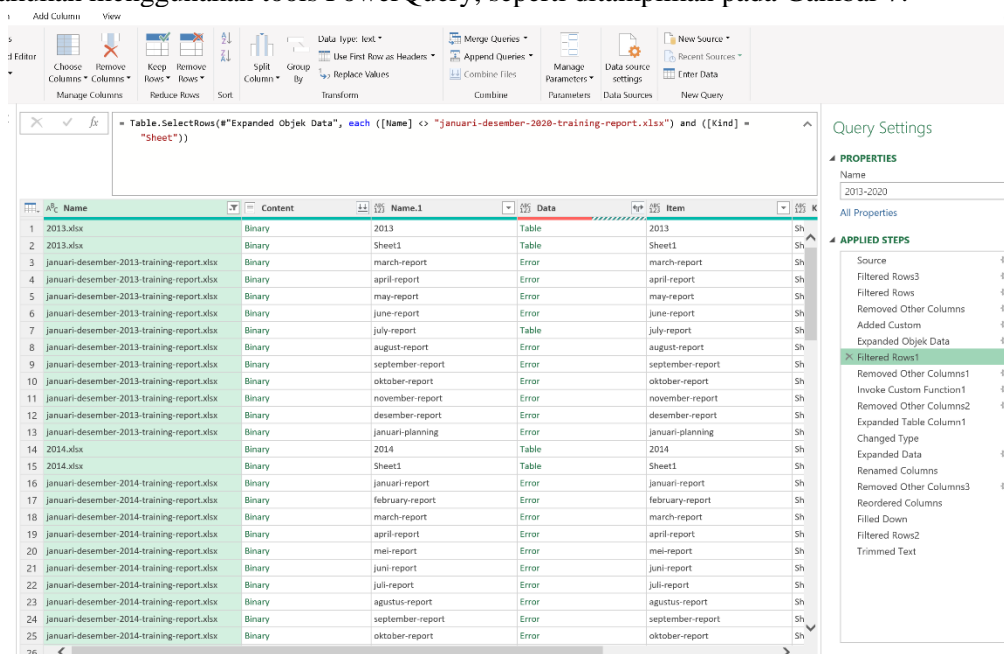
3.3 Deployment

Setelah didapatkan akurasi terbaik, dilakukan *deployment* di mana model diterapkan pada dataset baru yang belum pernah dilakukan proses *training*. Hasil optimalisasi akan dilakukan pengujian *performance* untuk mengetahui kecepatan dan akurasi. Deploymen juga diujicobakan pada prototipe sistem untuk mengotomasi pengklasifikasian sentimen komentar.

4 Hasil dan Pembahasan

4.1 Data Understanding dan Data Preparation

Data digabungkan dari berbagai *file spreadsheet* baik hasil ekspor dari LMS maupun *file-file spreadsheet* terdahulu sebelum menggunakan LMS. Penggabungan data dan juga pembersihan awal data dilakukan menggunakan tools PowerQuery, seperti ditampilkan pada Gambar 7.



Gambar 7. Penggabungan data dengan powerquery

Data hasil penggabungan terdiri dari 4 atribut sebelum dilakukan pelabelan. Sedangkan data yang telah diberikan label menjadi 5 atribut. Penjelasan atribut data ditampilkan pada Tabel 4.

Tabel 4. Penjelasan atribut data

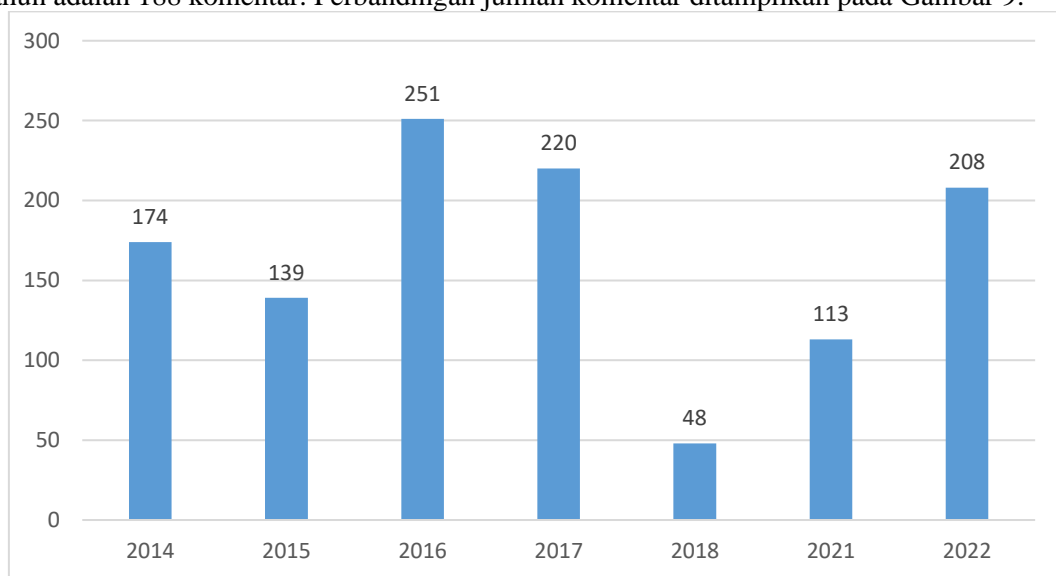
No	Nama Atribut	Penjelasan
1	No	Unik identifier atau nomor urutan data
2	Nama <i>Training</i>	Nama produk <i>course</i> yang telah diselenggarakan
3	Tanggal	Merupakan tanggal pelaksanaan <i>training</i> dimulai
4	Komentar	Tanggapan peserta <i>training</i> setelah mengikuti <i>training</i>
5	Sentimen	Hasil pelabelan apakah komentar negatif, positif atau netral

Sebelum data diolah, data dieksplorasi terlebih dahulu dengan menampilkan statistik data tersebut. Statistik ini memastikan apakah terdapat data yang kosong atau tidak, termasuk melakukan eksplorasi terkait berapa jumlah komentar di setiap tahunnya, *training* apa yang paling banyak mendapatkan komentar dalam rentang 10 *training* terbanyak, serta berapa jumlah komentar di masing-masing label sentimen. Berdasarkan Gambar 8, tidak ditemukan adanya data yang kosong atau *missing*.



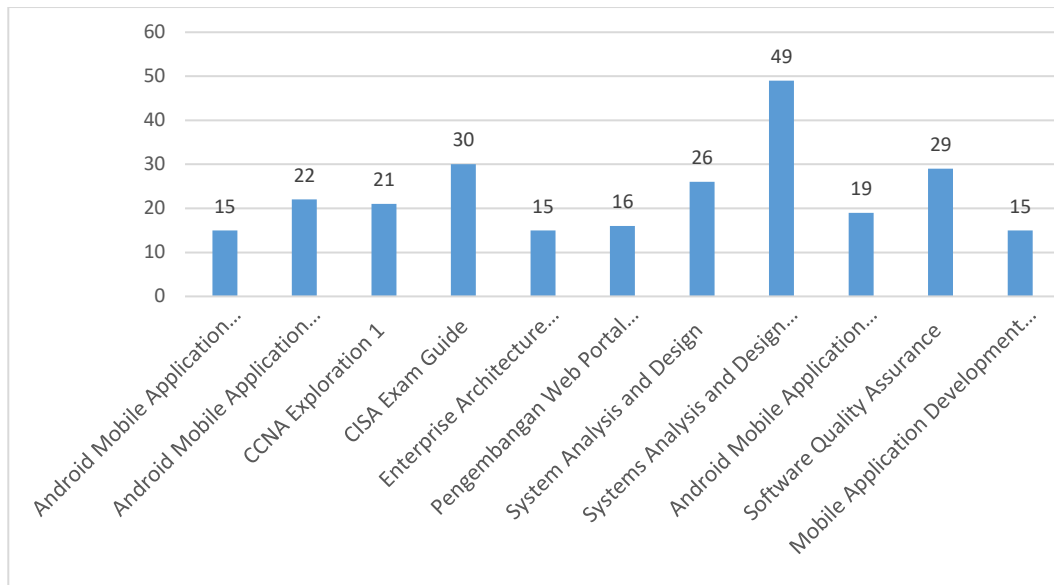
Gambar 8. Ringkasan kualitas data

Berdasarkan jumlah komentar per tahun, 2016 memiliki komentar paling banyak yaitu terdapat 251 komentar. Sedangkan komentar paling sedikit yaitu tahun 2018. Adapun nilai rata-rata komentar per tahun adalah 188 komentar. Perbandingan jumlah komentar ditampilkan pada Gambar 9.



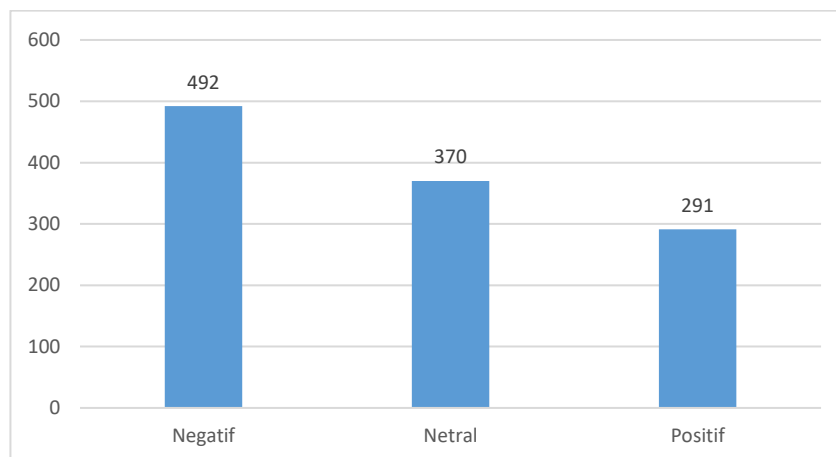
Gambar 9. Jumlah komentar per tahun

Jumlah komentar paling banyak berdasarkan 10 *training* teratas ditampilkan pada Gambar 10. Berdasarkan data yang dikumpulkan, jumlah komentar terbanyak berada pada *training* Systems Analysis and Design. Sedangkan paling sedikit di antara sepuluh *training* teratas yaitu *training* Mobile Application Development, Android Mobile Application Development dan Enterprise Architecture.



Gambar 10. Komentar terbanyak pada 10 training teratas

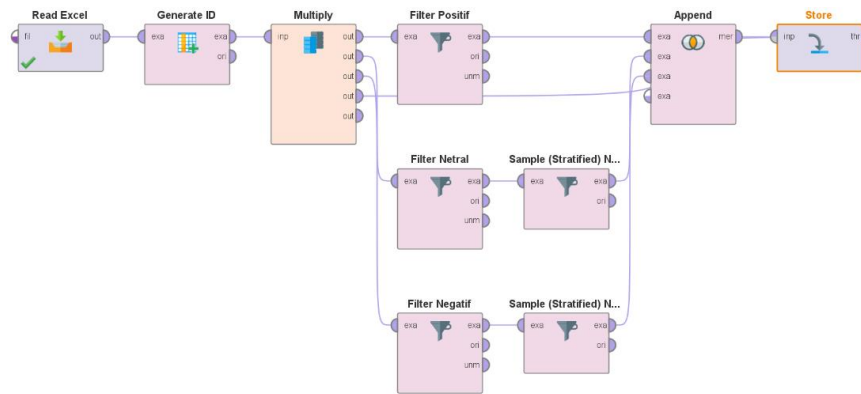
Adapun jumlah komentar di masing-masing label sentimen ditampilkan pada Gambar 11. Jumlah label pada masing-masing label tidak sama atau seimbang. Komentar negatif lebih dominan yaitu sebanyak 492. Sedangkan komentar positif merupakan label paling sedikit yang berjumlah 291. Oleh karena itu, perlu dilakukan persiapan data dengan melakukan sampling agar jumlah kelas pada data menjadi seimbang.



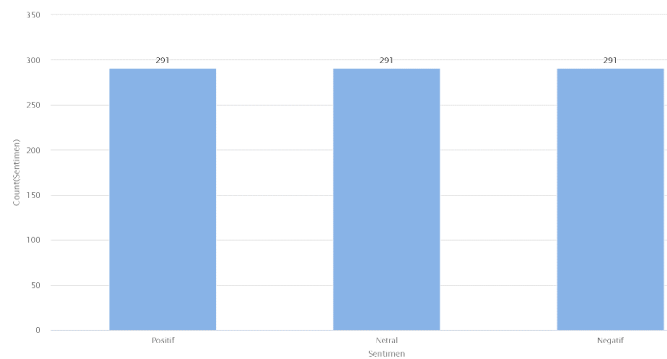
Gambar 11. Jumlah tiap label

Terdapat dua metode sampling paling populer untuk menangani data tidak seimbang yaitu *oversampling* dan *undersampling* [18]. *Undersampling* dipilih karena secara waktu lebih efektif, tidak memerlukan waktu lama untuk melakukan training label yang diperlukan seperti pada *oversampling*. Selain itu, *undersampling* juga mengurangi duplikasi data, tidak seperti yang terjadi pada *oversampling*.

Sampling dilakukan dengan menerapkan *undersampling*, di mana jumlah label terkecil menjadi jumlah target di setiap label. Dalam hal ini jumlah label terkecil yaitu jumlah positif. Pengambilan sample juga dilakukan secara *stratified*. Proses *undersampling* ditampilkan pada Gambar 12. Data hasil sampling kemudian digabung (*append*) dalam satu dataset sehingga jumlah dataset sampling menjadi 873 komentar. Statistik hasil akhir *undersampling* ditampilkan pada Gambar 13. Jumlah semua label menjadi sama yaitu sebanyak 291 di masing-masing label.



Gambar 12. Proses undersampling data

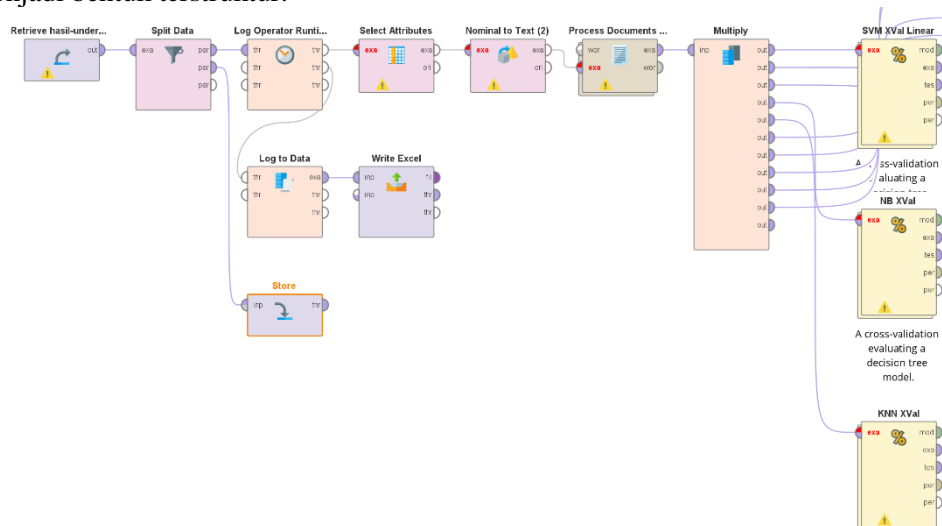


Gambar 13. Jumlah data tiap label hasil sampling

4.2 Modeling dan Evaluation

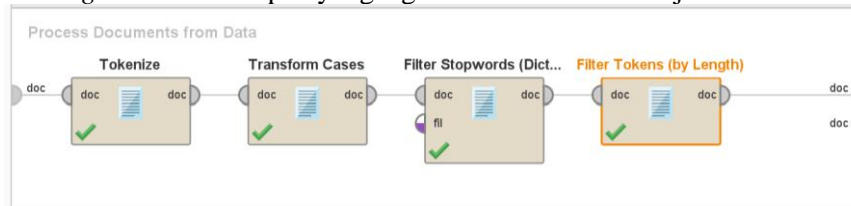
Tahapan ini meliputi pembahasan hasil pengukuran perbandingan kecepatan algoritma dan juga pengukuran kinerja. Proses pengukuran kecepatan dilakukan bersamaan dengan proses *training* pembentukan pola untuk mengklasifikasi sentimen komentar. Proses ini sekaligus membandingkan kinerja masing-masing algoritma yang dibandingkan. Proses pengukuran kecepatan dan kinerja ditampilkan pada Gambar 14.

Tahapan dimulai dengan mengambil data yang telah dilakukan sampling. Data kemudian dibagi menjadi dua bagian sebesar 90:10. 90% akan digunakan untuk pembentukan pola. Sedangkan 10% akan digunakan untuk pengujian di tahapan akhir pada *deployment*. Log proses terkait waktu yang diperlukan untuk menjalankan setiap proses ditempatkan untuk memperoleh informasi kecepatan pemrosesan. Pada tahapan ini juga dilakukan proses *text processing*, di mana mengubah data tidak terstruktur menjadi bentuk terstruktur.



Gambar 14. Proses modeling

Text processing pada tahapan ini terdiri dari tokenisasi, *transform case*, *stop word removal* (menggunakan *dictionary* berbahasa Indonesia), dan *filter token by length*. Urutan tahapan ini ditampilkan pada Gambar 15. Di *modelling* awal ini tidak menggunakan *stemming* terlebih dahulu. *Stemming* akan dilakukan setelah mendapatkan kinerja yang terbaik, sehingga dapat diketahui apakah penggunaan *stemming* memiliki dampak yang signifikan terhadap kinerja atau tidak.



Gambar 15. Tahapan text processing

Hasil pengukuran kecepatan juga dilakukan terhadap tiga algoritma terpilih yaitu Naïve Bayes, KNN dan SVM. Hasil pengukuran waktu menggunakan alat bantu Rapidminer. Hasil perbandingan waktu yang diperlukan selama proses training dengan masing-masing 10 kali percobaan (10-fold-cross validation) ditampilkan pada Tabel 5. Satuan waktu dalam perbandingan ini adalah *millisecond*.

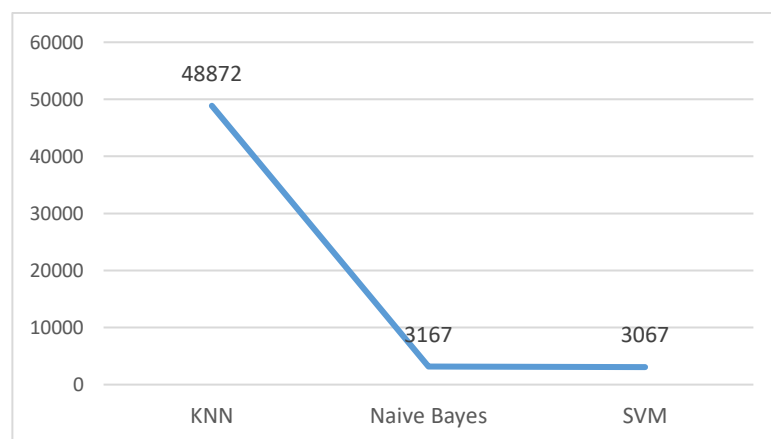
Total waktu yang diperlukan untuk melakukan proses training paling lama adalah metode KNN. KNN memerlukan waktu total 48872 *millisecond* atau sekitar 48.87 detik. KNN memerlukan waktu rata-rata 48.872 *millisecond* untuk setiap setiap iterasi.

Tabel 5. Perbandingan kecepatan algoritma

Uji Coba Ke-	Waktu (Milisecond)		
	KNN	Naïve Bayes	SVM
1	4702	182	290
2	4824	183	300
3	4833	185	301
4	4861	187	297
5	4803	191	303
6	4912	197	312
7	4942	198	308
8	4954	201	313
9	4991	202	324
10	5050	1984	319
Total Waktu	48872	3167	3067
Rata-rata	48.872	3.167	3.067

Berdasarkan Tabel 5, waktu yang diperlukan oleh Naïve Bayes adalah selama 3167 *millisecond* atau 3.167 detik untuk melakukan proses *training*. Naïve Bayes memerlukan waktu rata-rata 3.167 *millisecond* untuk setiap iterasi. Waktu proses ini jauh lebih kecil dibandingkan dengan KNN.

SVM memerlukan waktu total *training* 100 *millisecond* lebih cepat dibandingkan dengan Naïve Bayes. Rata-rata waktu yang diperlukan untuk satu iterasi adalah selama 3.067 *millisecond*. Jika dilihat berdasarkan grafik perbandingan pada Tabel 5, SVM unggul lebih cepat dibandingkan dengan algoritma lainnya.



Gambar 16. Perbandingan waktu proses

<http://sistemasi.ftik.unisi.ac.id>

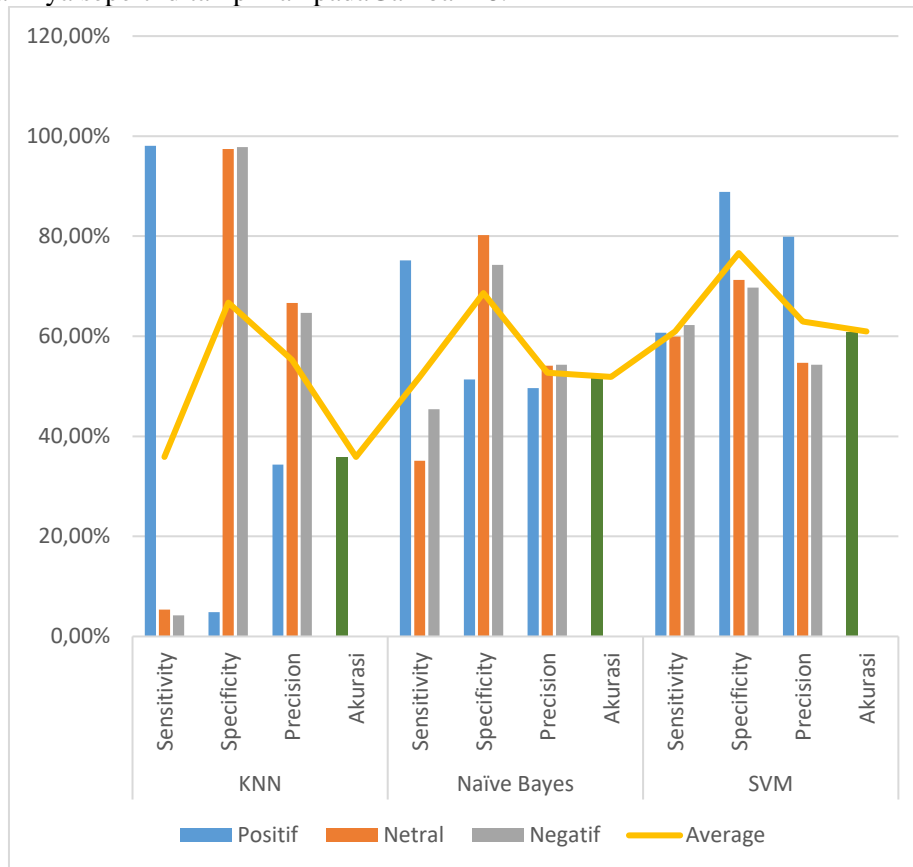
Komparasi kinerja juga dilakukan untuk memperoleh akurasi yang paling baik di antara ketiga algoritma tersebut yaitu yaitu KNN, dan Naïve Bayes dan SVM. Kinerja yang terbaik dianalisis berdasarkan akurasi, *sensitivity* (*recall*) dan *specificity*.

Hasil komparasi kinerja dibandingkan bersamaan ditampilkan pada Tabel 6. Nilai rata-rata di semua label masing-masing kinerja dihitung. Berdasarkan nilai rata-rata kinerja di tiap label, algoritma SVM memiliki nilai *sensitivity*, *specificity* dan *precision* yang lebih tinggi dibandingkan dengan algoritma lainnya.

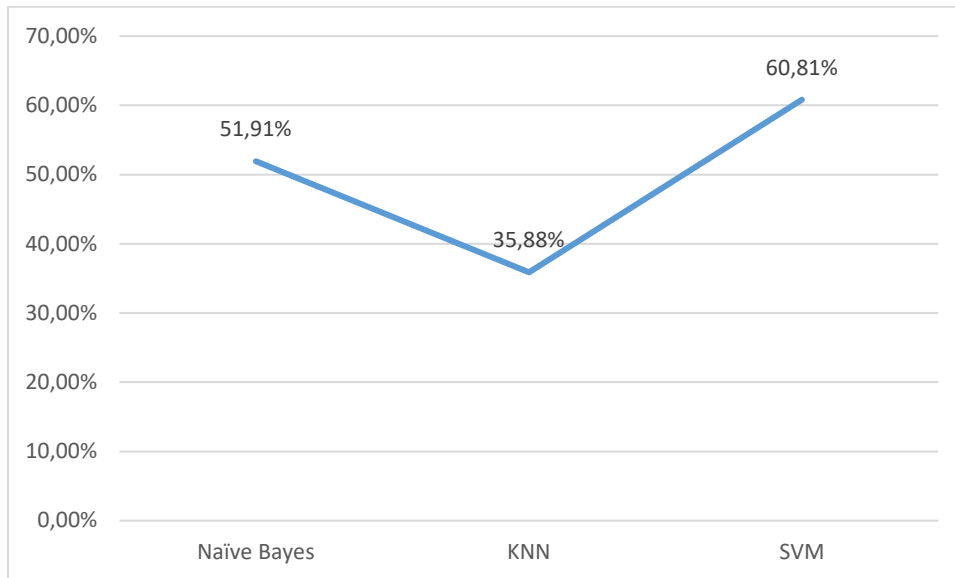
Tabel 6. Perbandingan kinerja algoritma

Algoritma	Kinerja	Positif	Netral	Negatif	Average
KNN	Sensitivity	98%	5%	4%	36%
	Specificity	5%	97%	98%	67%
	Precision	34%	67%	65%	55%
	Akurasi				35.88%
Naïve Bayes	Sensitivity	75%	35%	45%	52%
	Specificity	51%	80%	74%	69%
	Precision	50%	54%	54%	53%
	Akurasi				51.91%
SVM	Sensitivity	63%	57%	62%	61%
	Specificity	90%	80%	69%	80%
	Precision	83%	53%	53%	63%
	Akurasi				60.94%

Berdasarkan Gambar 17, SVM secara umum lebih unggul. Demikian apabila dilihat berdasarkan kinerja pada label positif, *specificity* dan *precision* lebih tinggi pada SVM dibandingkan dengan KNN dan Naïve Bayes. Berdasarkan Tingkat akurasi, SVM juga lebih unggul dibandingkan dengan algoritma lainnya seperti ditampilkan pada Gambar 18.

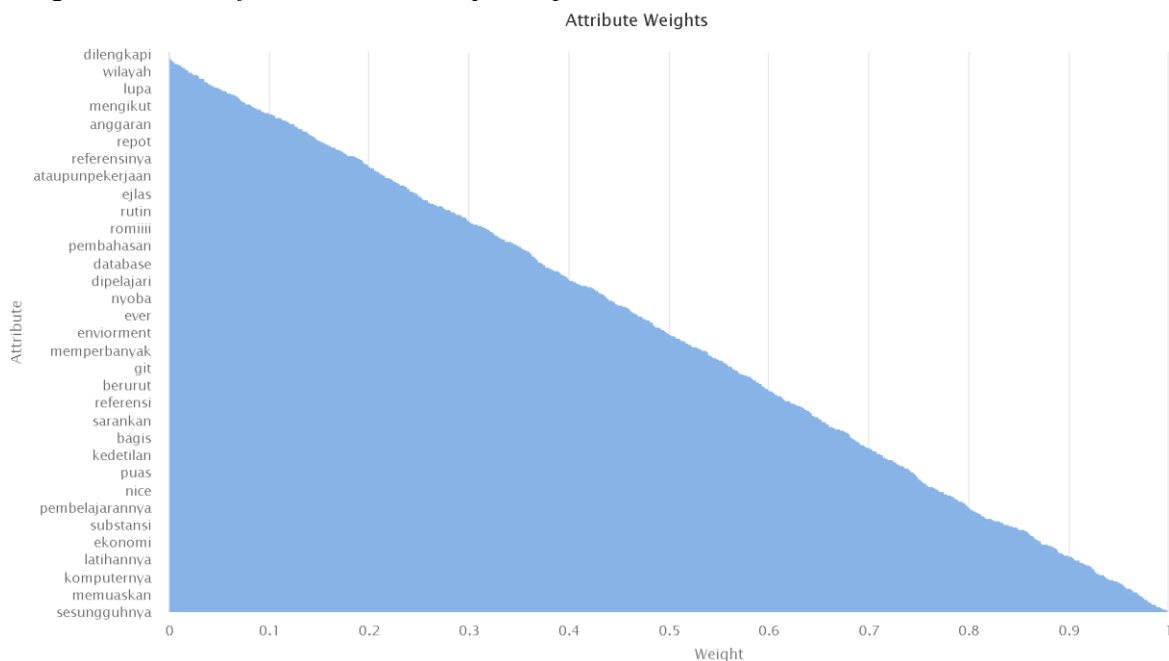


Gambar 17. Perbandingan sensitivity, specificity dan precision



Gambar 18. Perbandingan akurasi

Berdasarkan hasil pengukuran kecepatan dan kinerja, maka algoritma yang terbaik adalah SVM. Pada tahapan ini dilakukan optimasi SVM dengan teknik *feature selection* berbasis pembobotan dengan metode Optimized Weight Evolutionary (OWE). OWE yang digunakan dalam *feature selection* berbasis skema *non dominated sorting*. Nilai bobot atribut ditampilkan Gambar 19. Sedangkan atribut *confusion matrix* ditampilkan pada Tabel 7.



Gambar 19. Grafik bobot atribut

Berdasarkan hasil pembobotan, dari 1638 atribut terdapat 30 atribut yang tidak memiliki bobot. Adapun 20 atribut yang memiliki bobot cukup tinggi di antaranya adalah atribut atau kata “*sebenarnya*”, “*hadapi*”, “*memerlukan*”, “*mockup*”, “*implementasi*”, “*kerja*”, “*mengatasi*”, “*pray*”, “*baru*”, “*tercantum*”, “*sample*”, “*pembelajaran*”, “*hadiah*”, “*berasal*”, “*parking*”, “*tertarik*”, “*dikasih*”, “*selingan*”, “*terjawab*” dan “*matang*”.

Tabel 7. Confusion matrix, sensitivity dan precision OWE SVM

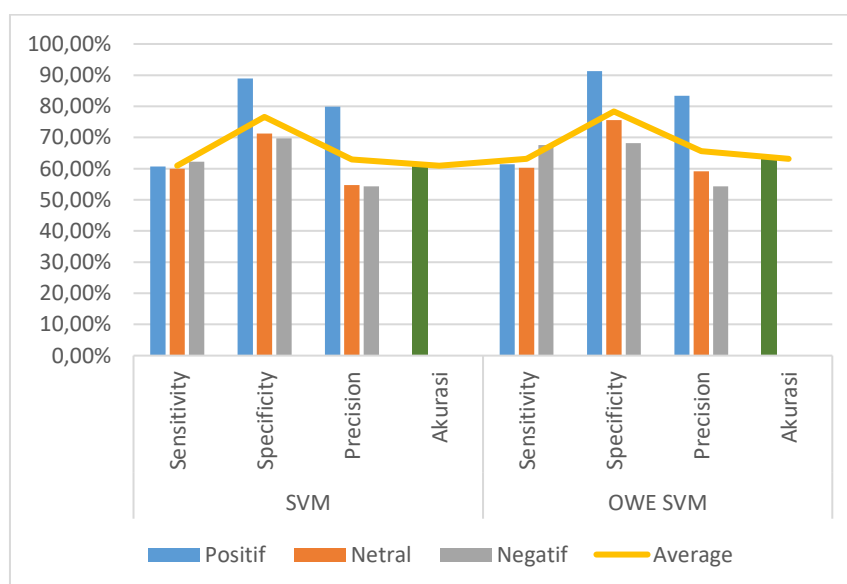
	true Positif	true Netral	true Negatif	class precision
pred. Positif	161	20	12	83.42%
pred. Netral	36	158	73	59.18%
pred. Negatif	65	84	177	54.29%
class recall	61.45%	60.31%	67.56%	

Berdasarkan hasil pengukuran pada Tabel 8, *Feature selection* dengan OWE untuk mengoptimalkan kinerja SVM (diperkenalkan dengan istilah OWE SVM), berhasil meningkatkan nilai akurasi sebesar 2.16%, meningkat dari 60.81% menjadi 63.10%. Nilai *sensitivity* pada OWE SVM meningkat dibandingkan dengan SVM biasa, baik untuk label positif, Netral maupun negatif, sebagaimana ditampilkan pada Tabel 8. Sedangkan *specificity* dan *precision* pada OWE SVM meningkat untuk label positif dan netral.

Tabel 8. Perbandingan kinerja optimalisasi

Algoritma	Kinerja	Positif	Netral	Negatif	Average
SVM	Sensitivity	60.69%	59.92%	62.21%	60.94%
	Specificity	88.89%	71.24%	69.76%	76.63%
	Precision	79.90%	54.70%	54.33%	62.98%
	Akurasi		60.94%		60.94%
OWE SVM	Sensitivity	61.45%	60.31%	67.56%	63.10%
	Specificity	91.28%	75.62%	68.16%	78.35%
	Precision	83.42%	59.18%	54.29%	65.63%
	Akurasi		63.10%		63.10%

Oleh karena itu rata-rata nilai *sensitivity*, *specificity* dan *precision* pada OWE SVM lebih tinggi jika dibandingkan dengan SVM biasa. Perbandingan kinerja dalam bentuk *chart* ditampilkan pada Gambar 20.



Gambar 20. Perbandingan sensitivity, specificity dan precision SVM dan OWE SVM

Proses pembelajaran pada OWE SVM ini belum menerapkan teknik *stemming* pada tahapan *text processing*. Meskipun proses *stemming* biasanya meningkatkan keberhasilan dalam mengekstraksi informasi pada teks atau dokumen, tetapi juga dapat sedikit menurunkan keakuratan. Namun demikian, *stemming* umumnya memungkinkan hasil yang lebih berkualitas dalam penerapan *text mining*. Oleh karena, *stemming* diujicobakan dengan metode OWE SVM. *Confusion matrix* hasil penerapan ditampilkan pada Tabel 9.

Tabel 9. Confusion matrix hasil penerapan stemming pada OWE SVM

	true Positif	true Netral	true Negatif	class precision
pred. Positif	180	19	15	84.11%
pred. Netral	41	159	83	56.18%
pred. Negatif	70	113	193	51.33%
class recall	61.86%	54.64%	66.32%	

Nilai *precision* dan akurasi hasil penerapan *stemming* pada OWE SVM lebih rendah bila dibandingkan dengan OWE SVM tanpa *stemming*. *Stemming* pada dataset ini hanya mampu meningkatkan nilai *sensitivity* pada label positif sebesar 0.41% dan *specificity* netral sebesar 3.41%. Rata-rata kinerja hasil *stemming* pada OWE SVM yaitu 67.73%. Sedangkan rata-rata hasil OWE SVM 1.30% sedikit lebih tinggi, atau sebesar 69.03%. Oleh karena itu, penggunaan *stemming* pada proses pembelajaran pada penelitian ini tidak digunakan.

Pengujian signifikansi juga dilakukan untuk mengetahui apakah terdapat perbedaan signifikan atau tidak antara penggunaan metode yang diusulkan yaitu OWE SVM dengan SVM biasa dan *stemming* OWE SVM. Data hasil pengukuran kinerja ditampilkan seperti pada Tabel 10 untuk memudahkan perhitungan uji beda.

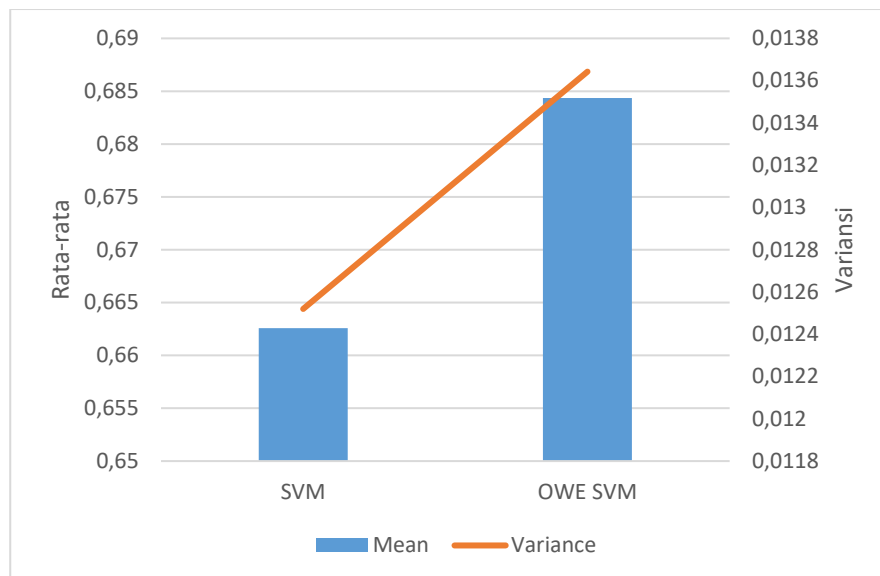
Tabel 10. Kinerja SVM, OWE SVM dan stemming OWE SVM

Kinerja	SVM	OWE SVM	Stemming OWE SVM
Sensitivity Positif	60.69%	61.45%	61.86%
Sensitivity Netral	59.92%	60.31%	54.64%
Sensitivity Negatif	62.21%	67.56%	66.32%
Specificity Positif	88.89%	91.28%	91.19%
Specificity Netral	71.24%	75.62%	0.7903
Specificity Negatif	69.76%	68.16%	0.6494
Precision Positif	79.90%	83.42%	84.11%
Precision Netral	54.70%	59.18%	56.18%
Precision Negatif	54.33%	54.29%	51.33%
Akurasi	60.94%	63.10%	60.94%

Hasil perhitungan rata-rata kinerja SVM dan OWE SVM, variansi dan jumlah sampel yang diuji beda ditampilkan pada Tabel 11. Nilai rata-rata OWE SVM sedikit lebih tinggi dibandingkan dengan SVM seperti dilihat pada Gambar 21. Demikian juga dengan nilai variansinya menunjukkan bahwa OWE SVM sedikit lebih tinggi dari pada SVM.

Tabel 11. Rata-rata dan variansi SVM dan OWE SVM

	SVM	OWE SVM
Mean	0.662588	0.684366
Variance	0.01252	0.013643
Observations	10	10



Gambar 21. Perbandingan rata-rata dan variansi SVM dan OWE SVM

Berdasarkan Tabel 11, variansi dari SVM adalah 0.01252, dan OWE SVM adalah 0.01364. Hal ini menunjukkan bahwa skor OWE SVM sedikit lebih bervariasi dibandingkan dengan SVM. Nilai statistik t adalah -3.0350. Nilai negatif menunjukkan bahwa rata-rata grup pertama (SVM) lebih rendah daripada grup kedua (OWE SVM). Probabilitas untuk mendapatkan nilai t yang lebih ekstrem dalam arah negatif dalam uji satu ekor adalah 0.0071, yang kurang dari 0.05, sehingga menunjukkan signifikansi statistik.

Nilai t kritis untuk uji satu ekor adalah 1.8331, nilai t harus lebih kecil dari -1.8331 untuk menolak hipotesis nol pada tingkat signifikansi 5% dalam uji satu ekor. Probabilitas untuk mendapatkan nilai t yang lebih ekstrem di kedua arah dalam uji dua ekor adalah 0.0141, yang juga kurang dari 0.05, menunjukkan signifikansi statistik. Nilai t kritis untuk uji dua ekor adalah 2.2622, nilai t harus lebih kecil dari -2.2622 atau lebih besar dari 2.2622 untuk menolak hipotesis nol pada tingkat signifikansi 5% dalam uji dua ekor.

Tabel 12. Hasil uji beda (paired t-test)

Parameter	Nilai
Pearson Correlation	0.981224
Hypothesized Mean Difference	0
df	9
t Stat	-3.03505
P(T<=t) one-tail	0.007065
t Critical one-tail	1.833113
P(T<=t) two-tail	0.014131
t Critical two-tail	2.262157

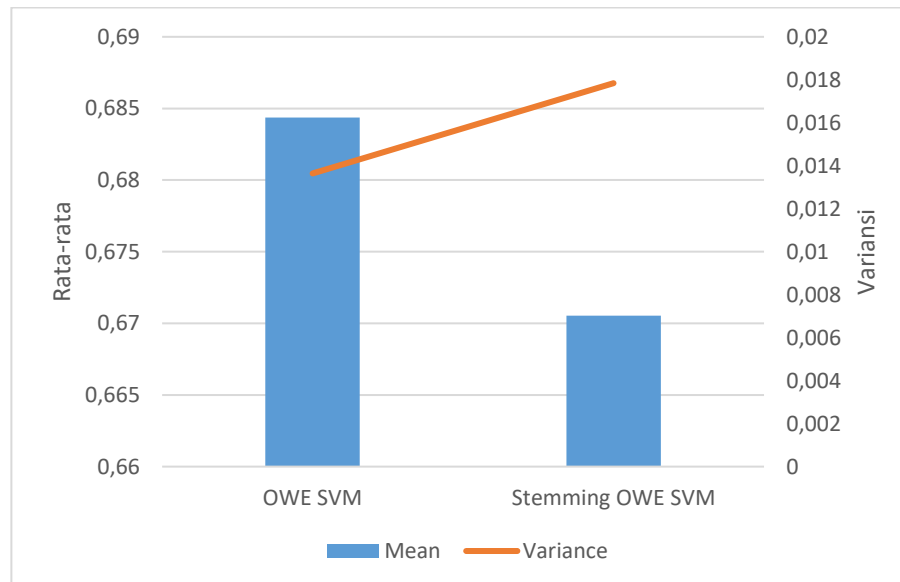
Berdasarkan pada Tabel 12 disimpulkan bahwa ada perbedaan yang signifikan secara statistik antara rata-rata SVM dan OWE SVM. Perbedaan signifikan juga dapat dilihat berdasarkan analisis satu arah maupun dua arah, karena nilai t yang dihitung (-3.035047147) adalah lebih rendah dari nilai t kritis untuk *one-tail* dan *two-tail*, dan p -value (0.007065371 untuk *one-tail* dan 0.014130742 untuk *two-tail*) adalah kurang dari 0.05 sebagaimana ditampilkan pada Tabel 12. Hal ini menunjukkan bahwa hipotesis nol (tidak ada perbedaan nyata antara *rata-rata* kedua variabel) dapat ditolak, sehingga penggunaan OWE SVM terdapat perbedaan signifikan dibandingkan dengan SVM biasa.

Sedangkan hasil perhitungan rata-rata kinerja OWE SVM dan *stemming* OWE SVM, variansi dan jumlah sample yang diuji beda ditampilkan pada Tabel 13. Nilai rata-rata OWE SVM sedikit

lebih tinggi dibandingkan dengan *stemming* OWE SVM seperti dilihat pada Gambar 22. Namun nilai variansi OWE SVM sedikit lebih rendah dari pada *stemming* OWE SVM.

Tabel 13. Rata-rata dan variansi SVM dan OWE SVM

	OWE SVM	Stemming OWE SVM
Mean	0.68436576	0.670542842
Variance	0.013642669	0.017843426
Observations	10	10



Gambar 22. Perbandingan rata-rata dan variansi OWE SVM dan stemming OWE SVM

Tabel 14 menunjukkan bahwa Rata-rata OWE SVM adalah 0.6844 dan *Stemming* OWE SVM adalah 0.6705. Ini menunjukkan bahwa, secara rata-rata, OWE SVM memiliki skor lebih tinggi dibandingkan dengan *stemming* OWE SVM. Sedangkan variansi untuk OWE SVM adalah 0.01364, sementara *stemming* OWE SVM adalah 0.01784, yang menunjukkan bahwa terdapat sedikit lebih banyak variasi dalam skor *Stemming* OWE SVM dibandingkan dengan OWE SVM. Kedua grup memiliki jumlah observasi yang sama, yaitu 10.

Nilai t Statistik adalah 1.6979, yang merupakan perhitungan dari perbedaan antara rata-rata yang diamati dibagi dengan kesalahan standar dari perbedaan. Nilai p untuk uji satu arah adalah 0.06187, yang berarti ada 6.187% kemungkinan mendapatkan hasil sejauh ini atau lebih jauh dari rata-rata hipotesis jika hipotesis nol benar. Nilai t kritis untuk uji satu arah adalah 1.8331, yang berarti nilai t harus lebih besar dari ini untuk menolak hipotesis nol pada tingkat signifikansi 5%.

Nilai p untuk uji dua arah adalah 0.12375, yang berarti ada 12.375% kemungkinan mendapatkan hasil sejauh ini atau lebih jauh dari rata-rata hipotesis jika hipotesis nol benar. Nilai t kritis untuk uji dua arah adalah 2.2622, yang berarti nilai t harus lebih besar dari ini untuk menolak hipotesis nol pada tingkat signifikansi 5%.

Tabel 14. Hasil uji beda (Paired t-test) OWE SVM dan stemming OWE SVM

Parameter	Nilai
Pearson Correlation	0.987781322
Hypothesized Mean Difference	0
df	9
t Stat	1.697934503
P(T<=t) one-tail	0.061873284
t Critical one-tail	1.833112933
P(T<=t) two-tail	0.123746568

<http://sistemasi.ftik.unisi.ac.id>

t Critical two-tail 2.262157163

Berdasarkan Tabel 14 disimpulkan bahwa tidak ada cukup bukti untuk menolak hipotesis nol pada tingkat signifikansi 5% untuk kedua uji satu dan dua arah, karena nilai t-Stat (1.6979) lebih rendah dari kedua nilai t kritis, dan nilai p untuk uji dua arah (0.12375) lebih tinggi dari 0.05. Hal ini menunjukkan bahwa, meskipun ada perbedaan rata-rata, perbedaan tersebut tidak cukup signifikan secara statistik dengan data yang tersedia dan tingkat signifikansi yang dipilih.

4.3 Deployment

Setelah didapatkan hasil yang paling optimal, model kemudian digunakan untuk melakukan uji coba pada data baru yang tidak pernah digunakan untuk *training*. Data ini sebanyak 10% dari 1153 data atau sebanyak 87 data. Kondisi data uji coba ditampilkan pada Gambar 23. Data yang diuji tidak ada yang kosong atau *missing value*, berasal dari berbagai tahun mulai dari tahun 2014 sampai tahun 2022. Namun demikian, pada data ini masih terdapat label yang sebelumnya sudah diberikan, sehingga dapat diketahui berapa nilai akurasi.

Name	Type	Missing	Statistics	Filter (5 / 5 attributes)
id	Integer	0	Min: 7, Max: 1108, Average: 543.345	
Sentimen	Nominal	0	Least: Positif (29), Most: Negatif (29)	Values: Negatif (29), Netral (29), ...[1 more]
Nama Training	Nominal	0	Least: Zend PHP [...] ation (0), Most: System A [...] sign (3)	Values: System A [...] d Design (3), Andro
Tanggal	Date-time	0	Earliest date: Feb 15, 2014, Latest date: Dec 20, 2021	
Komentar	Nominal	0	Least: yang pas [...] elat2 (0), Most: bagus (3)	

Gambar 23. Data uji beda

Error! Reference source not found. menampilkan beberapa data hasil klasifikasi sentimen dengan menggunakan model OWE SVM. Jika hasil ini diukur, maka tingkat akurasi adalah sebesar 64.37%. Hasil pengukuran ditampilkan pada Tabel 15.

id	Tanggal	Nama Trai...	Komentar ↓	Sentimen	prediction(Sentimen)
159	Nov 26, 2014	Oracle Data...	waktunya kurang	Negatif	Negatif
909	Aug 2, 2021	DMBOK Fun...	waktunya di tambahkan, dikarenakan materi yang begitu banyak	Netral	Negatif
589	Apr 6, 2017	PHP & MyS...	waktu yang terbatas sehingga materi tidak bisa semua dipelajari, semoga esok s...	Netral	Netral
350	Feb 6, 2016	Java Funda...	waktu trining jangan di persingkat, maksudnya tidak di porsir (dipaksakan dalam...	Negatif	Positif
413	Jul 2, 2016	Joomla CMS	waktu pelatihan yang lebih panjang agar materi yang disampaikan lebih mendalam	Negatif	Netral
795	Jan 6, 2018	Android Mob...	untuk pelatihan android bila bisa disediakan handset+kabel untuk test	Negatif	Negatif
1108	Dec 1, 2021	Data Mining ...	untuk disetiap akhir pelatihan baiknya dibuatkan summary pelatihan di hari tsb.,	Netral	Netral
550	Nov 19, 2016	PHP Frame...	tutor ok, tempat ok, fasilitas ok	Positif	Positif
1014	Mar 29, 2021	Enterprise Ar...	training yang sangat bagus materinya,	Positif	Positif
476	Oct 29, 2016	Android Mob...	training membuat game di android	Netral	Netral
472	Oct 15, 2016	Systems Ana...	topiknya dikembangkan ke arah yang lebih luas tidak hanya it secara teknikal, mu...	Netral	Netral
877	Jul 5, 2021	Enterprise Ar...	tolong dapat di persiapkan lebih matang terkait module2nya	Negatif	Netral
864	Apr 12, 2021	CISA Exam ...	time management,	Negatif	Positif
448	Sep 17, 2016	Business Pro...	sudah sangat bagus	Positif	Positif
1016	Jun 7, 2021	PHP Laravel...	sudah baik secara keseluruhan, dipertahankan dan dipertingkatkan	Positif	Positif
905	Jul 26, 2021	Data Mining ...	sudah bagus gabungan antara teori dan praktek	Positif	Positif
195	Mar 7, 2015	Database Ap...	sudah bags, diberi lagi banyak latihan praktik	Positif	Netral

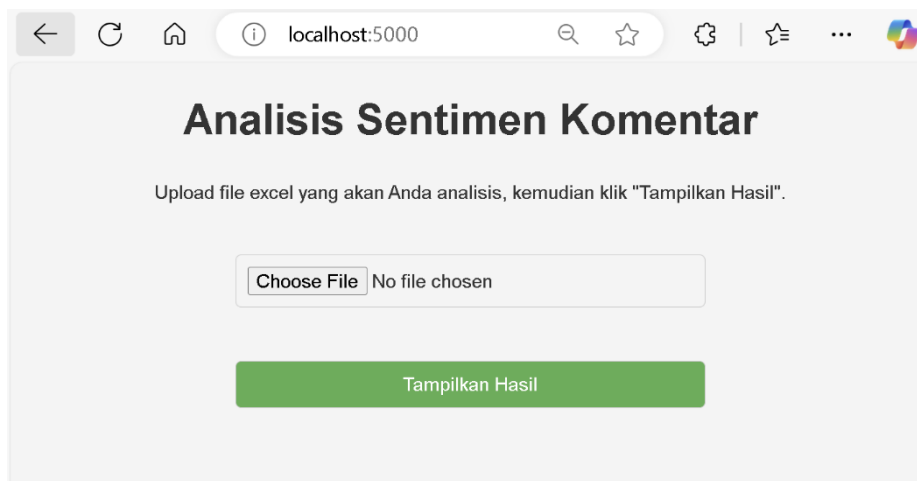
Gambar 24. Data hasil uji coba

Prediksi sentimen positif berjumlah 20 komentar atau sebesar 68.97% dari total 29 jumlah komentar dengan sentimen positif. Sedangkan Jumlah prediksi komentar Netral adalah 19 komentar dari jumlah komentar Netral atau sebesar 65.52%. Adapun Jumlah prediksi komentar negatif adalah 17 komentar dari total 29 komentar atau sebesar 58.62%, sebagaimana di tampilkan pada Tabel 15.

Tabel 15. Confusion Matrix uji coba

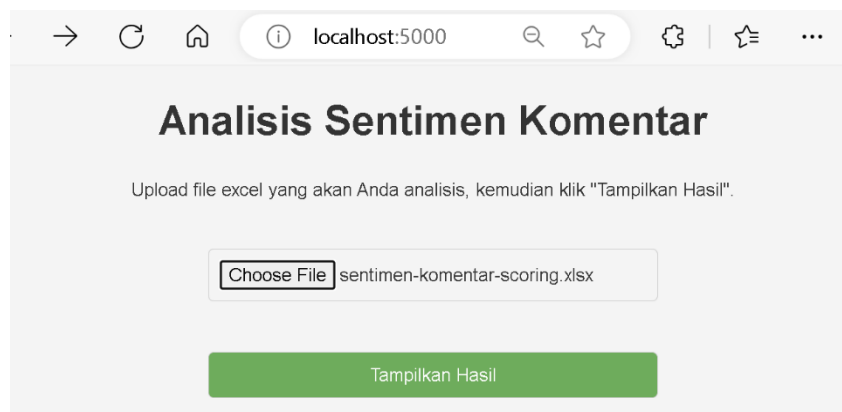
	true Positif	true Netral	true Negatif
pred. Positif	20	2	6
pred. Netral	2	19	6
pred. Negatif	7	8	17

Adapun contoh implementasi model ke dalam aplikasi sehingga aplikasi dapat dengan cepat memprediksi atau mengklasifikasi sentimen suatu komentar yang masuk, maka ditampilkan seperti pada Gambar 25. Pada aplikasi ini dapat melakukan unggah data yang berisi komentar, menampilkan hasil klasifikasi, dan ringkasan hasil klasifikasi serta melakukan unduh *file* hasil klasifikasi sentimen.



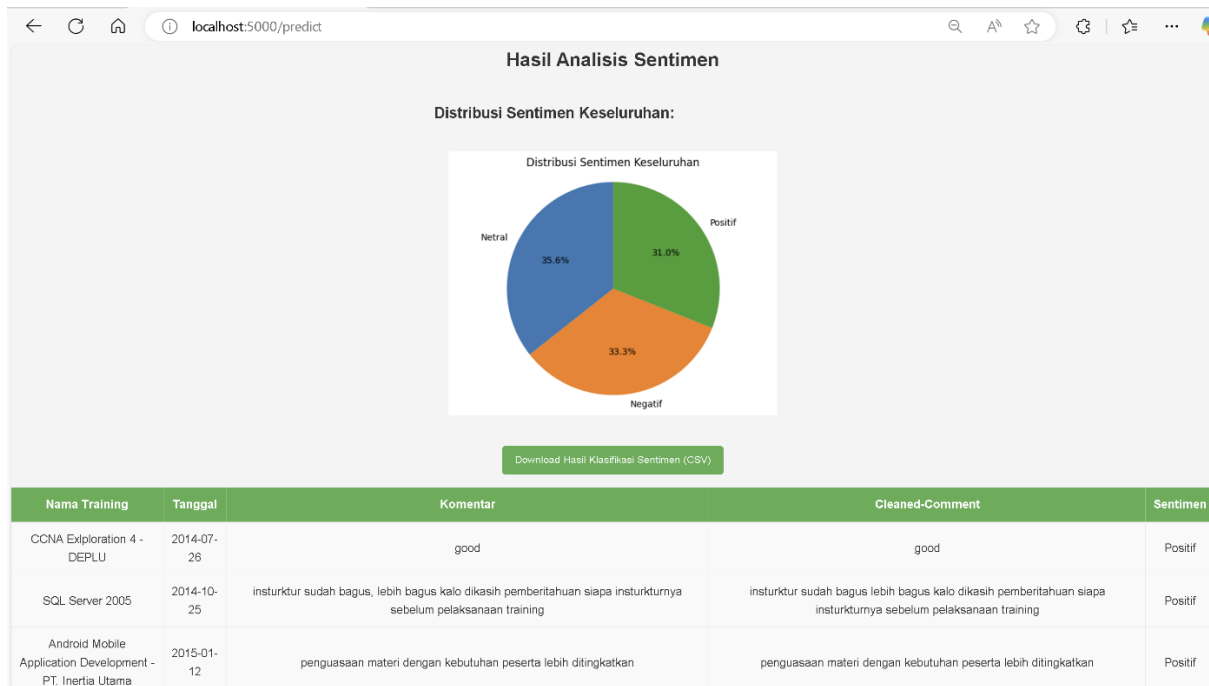
Gambar 25. Implementasi klasifikasi sentimen komentar pada web

Saat menu “Choose File” dan file excel dipilih, maka akan menampilkan informasi nama file yang diunggah seperti ditampilkan pada Gambar 26. Aplikasi akan melakukan klasifikasi komentar setelah pilihan “Tampilan Hasil” diklik. Halaman hasil analisis ditampilkan pada Gambar 27.



Gambar 26. Tampilan informasi unggah file

Berdasarkan penerapan model pada aplikasi web, di mana berhasil melakukan klasifikasi komentar dengan jumlah sentimen positif, netral dan negatif hampir sama. Jumlah komentar yang diprediksi positif ada sebanyak 31%, netral sebesar 35%, dan negatif sebesar 33.3%. Selain ringkasan hasil klasifikasi sentimen dalam bentuk *pie chart*, juga menampilkan dalam bentuk tabel.



Gambar 27. Halaman tampilan hasil analisis sentimen

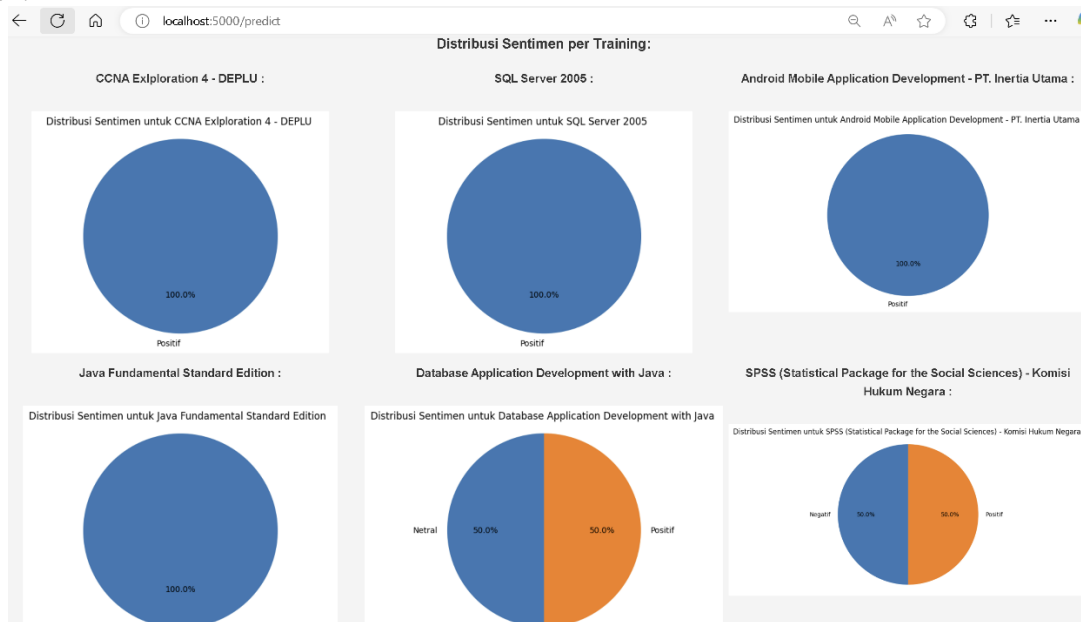
Hasil klasifikasi sentimen dalam bentuk tabel ditampilkan pada Gambar 28. Tabel ini berisi kolom nama *training*, tanggal, komentar, teks hasil *text processing* dan kolom sentimen.

Nama Training	Tanggal	Komentar	Cleaned-Comment	Sentimen
CCNA Exploration 4 - DEPLU	2014-07-26	good	good	Positif
SQL Server 2005	2014-10-25	insturktur sudah bagus, lebih bagus kalo dikasih pemberitahuan siapa insturkturnya sebelum pelaksanaan training	insturktur sudah bagus lebih bagus kalo dikasih pemberitahuan siapa insturkturnya sebelum pelaksanaan training	Positif
Android Mobile Application Development - PT. Inertia Utama	2015-01-12	penguasaan materi dengan kebutuhan peserta lebih ditingkatkan	penguasaan materi dengan kebutuhan peserta lebih ditingkatkan	Positif
Java Fundamental Standard Edition	2015-02-16	bagus	bagus	Positif
Database Application Development with Java	2015-03-07	sudah bags, diberi lagi banyak latihan praktik	sudah bags diberi lagi banyak latihan praktik	Positif
SPSS (Statistical Package for the Social Sciences) - Komisi Hukum Negara	2015-03-25	manstabzzz	manstabzzz	Positif
SPSS (Statistical Package for the Social Sciences) - Komisi Hukum Negara	2015-03-25	lebih aktif lagi	lebih aktif lagi	Negatif
System Analysis and Design	2015-05-16	persiapan perlatan yang mendukung untuk mempresentasikan materi perlu ditingkatkan	persiapan perlatan yang mendukung untuk mempresentasikan materi perlu ditingkatkan	Positif
Android Mobile Application Development	2015-05-23	materi uptodate	materi uptodate	Positif
Android Mobile Application Development	2016-01-09	pengajaran yang lebih baik	pengajaran yang lebih baik	Positif

Gambar 28. Tampilan tabel hasil klasifikasi sentimen

Selain tabel, aplikasi juga menampilkan jumlah sentimen di masing-masing topik pelatihan berdasarkan data yang diunggah. Hasil sentimen di masing-masing pelatihan ditampilkan pada Gambar 29. Berdasarkan cuplikan tampilan pada Gambar 29, komentar di pelatihan Database Application Development with java berjumlah seimbang antara netral dan positif. Sama halnya dengan persentase sentimen di pelatihan SPSS dan Java Fundamentals Standar Edition. Sedangkan di pelatihan Systems Analysis and Design, sentimen negatif lebih dominan dari pada positif. Adapun

pelatihan Android Mobile Application Development secara keseluruhan memperoleh komentar positif.



Gambar 29. Persentase sentimen per pelatihan

Tabel hasil analisis juga dapat diunduh ke dalam komputer. Contoh file CSV yang ditampilkan menggunakan Microsoft Office ditampilkan pada Gambar 30. Kolom yang ada pada file hasil unduhan sama persis dengan yang ditampilkan pada halaman hasil analisis sentimen.

	A	B	C	D	E
1	Nama Training	Tanggal	Komentar	processed comments	Sentimen
2	Windows Server 2008, Configuring Active Directory - Brainmatics (JasaRaharja)	12/12/2011	waktu pelatihan diperpanjang	waktu pelatihan diperpanjang	Negatif
3	Windows Server 2008, Configuring Active Directory - Brainmatics (JasaRaharja)	12/12/2011	bagus! mantap!	bagus mantap	Positif
4	Adobe Photoshop CS4 - Brainmatics	12/12/2011	waktu pelatihan diperpanjang	waktu pelatihan diperpanjang	Negatif
5	Adobe Photoshop CS4 - Brainmatics	12/12/2011	bagus! mantap!	bagus mantap	Positif
6	Microsoft Project 2007 - Brainmatics	12/3/2012	terkonsentrasi	cukup pc trainer yang di beri internet akses, peserta jadi kurang	Netral
7	Microsoft Project 2007 - Brainmatics	12/3/2012	tiap peserta diberi flashdisk/ cd untuk menyimpan file pelatihan	tiap peserta diberi flashdisk cd untuk menyimpan file pelatihan	Netral
8	Microsoft Project 2007 - Brainmatics	12/3/2012	trainer memberi tugas untuk dikerjakan di rumah, karena waktu tersite untuk mengerjakan tugas/quiz pada saat training	trainer memberi tugas untuk dikerjakan di rumah karena waktu	Netral
9	Systems Analysis and Design - Brainmatics	12/3/2012	lebih banyak materi latihan dan panduan dalam implementasi	lebih banyak materi latihan dan panduan dalam implementasi	Netral
10	Systems Analysis and Design - Brainmatics	12/3/2012	sarana dalam ruangan mengajar	sarana dalam ruangan mengajar lebih ditingkatkan kenyamanannya	Positif
11	Systems Analysis and Design - Brainmatics	12/3/2012	memberi kesempatan untuk diskusi antar peserta	memberi kesempatan untuk diskusi antar peserta	Negatif
12	Systems Analysis and Design - Brainmatics	12/3/2012	mengajak peserta untuk saling	mengajak peserta untuk saling mengenal satu sama lain	Netral
13	Systems Analysis and Design - Brainmatics	12/3/2012	praktik sedikit mengenai bpmn	praktik sedikit mengenai bpmn	Netral

Gambar 30. Contoh hasil unduh analisis sentimen

5 Kesimpulan

Keterlambatan dalam menangani masukan atau keluhan pelanggan melalui komentar di LMS dapat merugikan baik perusahaan maupun pelanggan. Mengingat besarnya volume data komentar, pemanfaatan data mining untuk klasifikasi berdasarkan data historis dapat mempercepat respons. Studi komparatif antara tiga algoritma data mining, yaitu KNN, Naïve Bayes, dan SVM,

mengungkapkan bahwa SVM unggul dalam kecepatan dan akurasi. Proses training SVM hanya membutuhkan 3.067 detik, lebih cepat 100 milidetik dibanding Naïve Bayes dan 45.8 detik dari KNN. Selain itu, SVM juga mencatat akurasi tertinggi sebesar 60.81%, yang meningkat menjadi 63.10% dengan penambahan Optimized Weight Evolutionary. Penelitian ini menunjukkan potensi besar integrasi model ini dalam sistem LMS perusahaan untuk meningkatkan efektivitas dan efisiensi dalam menangani komentar dan keluhan pelanggan.

Referensi

- [1] D. Henderson, S. Earley, and L. Sebastian-Coleman, Eds., *DAMA DMBOK: Data Management Body of Knowledge*. New Jersey: Technics Publications, 2017.
- [2] V. Kotu, *Predictive Analytics and Data Mining*. 2015. [Online]. Available: <http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf>
- [3] S. N. Fadhilah and F. S. Utomo, "Algoritma Naïve Bayes untuk Analisis Sentiment Review Blibli.com di Google Play Store," *SISTEMASI: Jurnal Sistem Informasi*, vol. 13, no. 2, pp. 834–4, Mar. 2024, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [4] K. Okoye, S. D. N. Daruich, J. F. E. De La O, R. Castano, J. Escamilla, and S. Hosseini, "A Text Mining and Statistical Approach for Assessment of Pedagogical Impact of Students' Evaluation of Teaching and Learning Outcome in Education," *IEEE Access*, vol. 11, pp. 9577–9596, 2023, doi: 10.1109/ACCESS.2023.3239779.
- [5] D. Tao, P. Yang, and H. Feng, "Utilization of Text Mining as a Big Data Analysis Tool for Food Science and Nutrition," Mar. 01, 2020, *Blackwell Publishing Inc.* doi: 10.1111/1541-4337.12540.
- [6] B. Agarwal and N. Mittal, "Text Classification using Machine Learning Methods," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2014, pp. 701–709. doi: 10.1007/978-81-322-1602-5_75.
- [7] Y. Chen and Z. Yuling, "Research on Text Sentiment Analysis Based on CNNs and SVM," 2018.
- [8] C. Wan, Y. Wang, Y. Liu, J. Ji, and G. Feng, "Composite Feature Extraction and Selection for Text Classification," *IEEE Access*, vol. 7, pp. 35208–35219, 2019, doi: 10.1109/ACCESS.2019.2904602.
- [9] K. Korovkinas, P. Danenas, and G. Garšva, "SVM and K-Means Hybrid Method for Textual Data Sentiment Analysis," *Baltic Journal of Modern Computing*, vol. 7, no. 1, pp. 47–60, 2019, doi: 10.22364/bjmc.2019.7.1.04.
- [10] S. Kumar, A. K. Kar, and P. V. Ilavarasan, "Applications of Text Mining in Services Management: A Systematic Literature Review," Apr. 01, 2021, *Elsevier Ltd.* doi: 10.1016/j.jjime.2021.100008.
- [11] R. E. Krider and D. S. Putler, *Customer and Business Analytics*. Chapman and Hall/CRC, 2012.
- [12] B. Liu, *Sentiment Analysis Mining Opinions, Sentiments, and Emotions, Second Edition*. Cambridge: University Printing House, 2020.
- [13] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge University Press, 2009.
- [14] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.
- [15] C. C. Aggarwal, *Data Mining The Text Book*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [16] M. Jordan, J. Kleinberg, and B. Schölkopf, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [17] Nuryadi, T. D. Astuti, E. S. Utami, and M. Budiantara, *Dasar-dasar Statistik Penelitian*. Yogyakarta: Sibuku Media, 2017. [Online]. Available: www.sibuku.com
- [18] H. Haibo and M. Yunqian, Eds., *Imbalanced Learning, Foundations, Algorithms, and Applications*. Hoboken: John Wiley & Sons, 2013.