

# Analisis Sentimen Saran Pengguna Mandatory E-Learning Menggunakan Text Mining pada Learning Management System

## Sentiment Analysis of User Suggestions for Mandatory E-Learning Using Text Mining on the Learning Management System

Andi Nur Syamsudin<sup>1\*</sup> , Utomo Budiyanto<sup>2</sup> 

<sup>1,2</sup>Program Studi Magister Ilmu Komputer, Universitas Budi Luhur, Indonesia

<sup>1</sup>2111601619@student.budiluhur.ac.id, <sup>2</sup>utomo.budiyanto@budiluhur.ac.id

\*Penulis Korespondensi

### Artikel Info

#### Riwayat Artikel:

Penyerahan September 23, 2024

Revisi Januari 22, 2025

Diterima Januari 30, 2025

Diterbitkan Februari 15, 2025

#### Kata Kunci:

Text Mining

Naïve Bayes

Support Vector Machine

K-Nearest Neighbor

Random Forest

#### Keywords:

Text Mining

Naïve Bayes

Support Vector Machine

K-Nearest Neighbor

Random Forest



### ABSTRAK

**Mandatory E-learning** merupakan pembelajaran wajib pegawai Kemenkeu di LMS Kemenkeu Learning Center/KLC di mana pelaporan rekapitulasi saran peserta berupa teks tidak tersedia disebabkan besarnya jumlah data evaluasi peserta. **Analisis sentimen dengan metode text mining** diperlukan guna mengklasifikasi saran peserta ke label positif, negatif, dan netral agar rekapitulasi terotomatisasi, cepat, dan akurat. **Dengan framework** Knowledge Discovery in Databases/ KDD, dilakukan pemilihan data dan pelabelan manual, text preprocessing (data cleansing, case folding, stop word removal, stemming, tokenizing, filter token by length), transformasi data (pembobotan TF-IDF, pengukuran cosine similarity, dan resampling menggunakan random undersampling/RUS untuk mengurangi label mayoritas). Tahap pemodelan membandingkan kombinasi algoritma terbaik dari Support Vector Machine/SVM, Multinomial Naïve Bayes, K-Nearest Neighbor/KNN, dan Random Forest pada rasio data training: testing = 90:10. **Hasil penelitian** menunjukkan SVM dengan cosine similarity sebagai skenario algoritma terbaik dengan nilai akurasi, presisi, recall, dan f1-score pada label negatif berturut-turut, yaitu 97,01%, 96,22%, 95,82%, dan 96,02%, dalam waktu 48,71 detik, sehingga **dapat dimanfaatkan** untuk meningkatkan kualitas pelaporan e-learning lebih cepat, akurat, dan terotomatisasi.

Ini adalah artikel akses terbuka di bawah [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



### ABSTRACT

**Mandatory E-learning** is a required training for Ministry of Finance's employees through Kemenkeu Learning Center (KLC) as LMS, where text-based recapitulation reports for participant's feedback are not available due to large volume of participant's evaluation data. **Sentiment analysis using** text mining is necessary to classify the feedback into positive, negative, and neutral labels, enabling the recapitulation process to be automated, faster, and more accurate. **Using Knowledge Discovery in Databases (KDD) framework**, the process involves data selection and manual labeling, text preprocessing (data cleansing, case folding, stop word removal, stemming, tokenizing, filtering tokens by length), data transformation (TF-IDF weighting, cosine similarity measurement, and resampling using random undersampling/RUS to

reduce majority label). Modeling phase compares the best combination of algorithms covers Support Vector Machine (SVM), Multinomial Naïve Bayes, K-Nearest Neighbor (KNN), and Random Forest using a 90:10 training-to-testing data ratio. **This research** show that SVM with cosine similarity is the best algorithm scenario, achieving accuracy, precision, recall, and f1-score for negative label of 97.01%, 96.22%, 95.82%, and 96.02%, respectively, within 48.71 seconds, which **can be leveraged** to improve quality of e-learning's report faster, more accurate, and to be automated.

Ini adalah artikel akses terbuka di bawah [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



DOI: <https://doi.org/10.33050/tmj.v9i3.2368>

Ini adalah artikel akses terbuka di bawah CC-BY license (<https://creativecommons.org/licenses/by/4.0/>)

©Penulis memegang semua hak cipta

## 1. PENDAHULUAN

Perkembangan teknologi memungkinkan kebutuhan pembelajaran tak lagi hanya secara tatap muka klasikal, melainkan juga memanfaatkan LMS/*learning management system*. LMS merupakan paket perangkat lunak yang digunakan untuk menyampaikan materi pembelajaran dan *resources* multimedia secara *online* berbasis *web* [1]. Pembelajaran melalui LMS disebut sebagai *e-learning*, yaitu metode pembelajaran interaktif menggunakan media elektronik sebagai media pembelajaran [2]. Kombinasi penerapan *e-learning* dan LMS dimanfaatkan Kementerian Keuangan/Kemenkeu dalam menyusun *project charter* inisiatif strategis *Modern E-learning* sejak 2019 s.d. 2022 dengan platform LMS KLC (Kemenkeu Learning Center). *Output* strategik yang dihasilkan meliputi: capaian peserta yang dikelola meningkat 10 kali lipat daripada tanpa KLC (45.463 peserta pada 2016 berbanding 451.954 peserta pada 2022), waktu kelola pembelajaran berkurang dari 125 hari menjadi 3 hari (tahun 2023), serta biaya bahan dan perjalanan dinas dihemat hingga 125% (tahun 2023).

Kesuksesan penerapan KLC di Kemenkeu didukung dengan program *mandatory e-learning* yang wajib diikuti seluruh pegawai, dimulai tahun 2019 s.d. 2022 (dari mulai 2 program s.d. 7 program dalam setahun). Dengan tujuan menyosialisasikan dan memberikan pemahaman tema-tema penting yang wajib diketahui seluruh pegawai, *mandatory e-learning* dilaksanakan dengan konsep *asynchronous*, artinya seluruh peserta bisa mengakses dan berkomunikasi disesuaikan dengan waktu masing-masing [3, 4]. Hal ini berbeda dengan konsep *synchronous* di mana materi diakses bersamaan oleh semua peserta dan pengajar melalui media *web-conference* seperti Zoom atau Google Meet maupun aplikasi *chating*. Dengan *asynchronous*, jumlah peserta dapat dikelola secara massal, seperti pada *mandatory e-learning* Penguatan Kesadaran Bela Negara tahun 2022-2023, terdiri atas 15 angkatan, dengan capaian total peserta 60.556 pegawai Kemenkeu (sekitar 78% dari seluruh pegawai Kemenkeu).

Konten *Mandatory e-learning* pada KLC terdiri atas beberapa *section* (mata pelatihan) yang terbagi ke dalam unit-unit. Setiap unit mengandung konten pembelajaran spesifik dengan beragam tipe, meliputi: teks, video, PDF, penugasan, kuis, survei, bahkan media interaktif seperti SCORM (*Shareable Content Object Reference Model*). Unit survei mengakomodasi kebutuhan evaluasi level 1 Kirkpatrick yang mengukur *reaction* peserta, seberapa mereka menemukan kepuasan dalam pembelajaran, keterlibatan, serta relevansinya dengan pekerjaan [5]. Unit survei ditampilkan pada KLC akses peserta berupa formulir yang diisi peserta untuk kemudian direkapitulasi pada setiap angkatan *e-learning* menjadi laporan/informasi yang bermanfaat, meliputi: data evaluasi penyelenggaraan dan data evaluasi lainnya (kuantitatif), hasil *individual performance analysis* (IPA) model evaluasi penyelenggaraan, serta saran/masukan dari peserta (kualitatif). Laporan ini merupakan salah satu *key performance indicator* dari administrator evaluasi, *role/PIC* KLC yang bertugas mengekstrak rekapitulasi survei dari KLC akses admin, agar disusun menjadi laporan evaluasi tersebut. *Output* atas pemanfaatan fitur unit survei KLC sampai dengan menjadi laporan akan dapat lebih berkualitas jika mampu memenuhi indikator: akurat, tepat waktu, relevan, dan lengkap [5].

## 2. PERMASALAHAN

Salah satu tantangan pengelolaan proses rekapitulasi survei pada *mandatory e-learning* adalah jumlah peserta massal, di mana satu angkatan (periode yang sama) dapat dipisahkan kedalam 5-10 kelas untuk kemudahan unduh data, sebagaimana dijelaskan pada tabel 1.

Tabel 1. Data Frame Komposisi *e-learning* Penguatan Kesadaran Bela Negara 2022-2023

No	Angkatan - Tahun	Waktu Mulai s.d. Selesai	Jumlah Kelas	Total Peserta
1	12 - 2022	23 s.d. 29 November 2022	5	2.541
2	13 - 2022	7 s.d. 9 Desember 2022	10	3.942
3	1 - 2023	25 s.d. 31 Januari 2023	5	2.902
4	2 - 2023	8 s.d. 15 Februari 2023	11	8.909
5	3 - 2023	22 s.d. 28 Januari 2023	10	8.834

Dengan jumlah massal tersebut, konten pelaporan evaluasi penyelenggaraan *mandatory e-learning* belum memenuhi aspek kelengkapan dan keakuratan dikarenakan dalam hasil tangkapan layar laporan evaluasi penyelenggaraan *e-learning* Penguatan Kesadaran Bela Negara angkatan 3 - 2023 sebagaimana tabel 1, terdapat data/informasi yang kosong. Konten saran/masukan peserta merupakan data teks yang kosong. Rekapitulasi data dapat diunduh di *role* admin evaluasi KLC, akan tetapi tidak menjadi prioritas untuk ditampilkan di pelaporan sebagaimana gambar, dikarenakan membutuhkan waktu untuk mengolahnya.

FREK.	PERSENTASE	SARAN/MASUKAN PESERTA
		PENYELENGGARAAN
		BUTIR 9 DAN 10
-	-	-
		VIDEO PEMBELAJARAN
-	-	-

Gambar 1. Cuplikan Laporan Penyelenggaraan *Mandatory E-Learning*

Sumber: Nota Dinas Pelaporan E-learning Penguatan Kesadaran Bela Negara Angkatan III

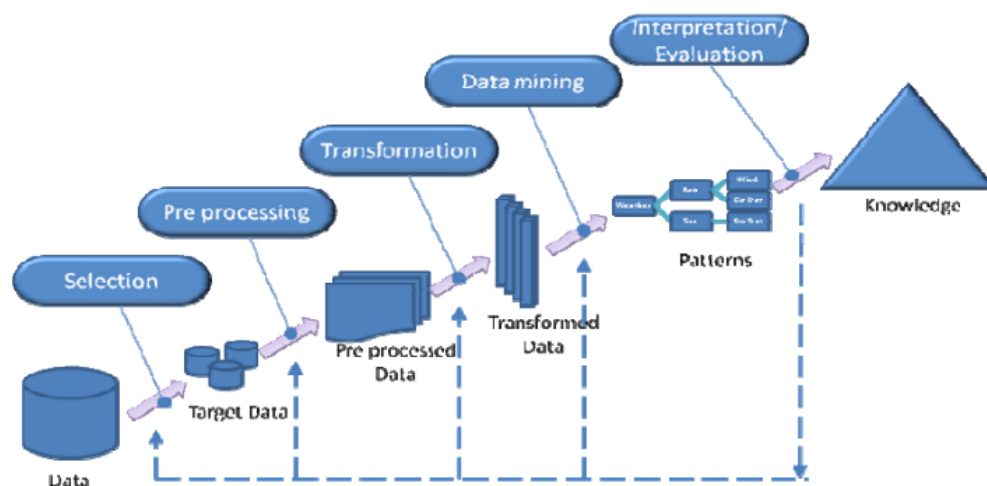
Gambar 1 menyajikan saran dan masukan dari peserta terkait penyelenggaraan *Mandatory E-Learning*. Berdasarkan tabel tersebut, terdapat beberapa aspek yang menjadi perhatian peserta, seperti penyelenggaraan secara keseluruhan, butir materi tertentu (butir 9 dan 10), serta video pembelajaran. Hal ini menunjukkan bahwa terdapat beberapa bagian dalam *e-learning* yang perlu dievaluasi lebih lanjut agar dapat meningkatkan efektivitas dan pengalaman belajar peserta.

Masukan dari peserta yang tercantum dalam gambar 1 dapat digunakan sebagai dasar untuk perbaikan pada pelaksanaan *e-learning* di masa mendatang. Evaluasi ini menjadi penting untuk meningkatkan kualitas materi, metode penyampaian, serta fitur pendukung dalam platform pembelajaran daring. Dengan adanya saran yang terstruktur, penyelenggara dapat lebih mudah mengidentifikasi aspek mana yang memerlukan perbaikan dan inovasi untuk memastikan bahwa peserta memperoleh pengalaman belajar yang optimal.

Analisis sentimen merupakan proses mempelajari data teks secara otomatis agar menghasilkan informasi yang terdapat pada sebuah kalimat [6]. Melalui analisis sentimen, saran/masukan peserta *e-learning* dipetakan ke label tertentu untuk melengkapi laporan penyelenggaraan yang belum lengkap. Analisis sentimen dapat diwujudkan melalui proses *data mining*, yaitu proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar, mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data [7]. *Text mining* menggali informasi dari data-data berupa tulisan, dokumen, atau teks untuk kebutuhan proses klasifikasi, yaitu proses menyatakan suatu objek data sebagai salah satu kategori/kelas berdasarkan riwayat data-data yang telah didefinisikan sebelumnya [8]. Dengan teknologi LMS yang diterapkan di Kemenkeu, lembaga pelatihan dapat memusatkan dan mengotomasi administrasi pembelajaran berserta hasil-hasilnya [3], sehingga analisis sentimen dan *text mining* berupaya merekapitulasi survei *mandatory e-learning* KLC agar klasifikasi label dapat lebih cepat, akurat, dan lengkap. Klasifikasi diperlukan agar laporan penyelenggaraan merepresentasikan perbaikan berkelanjutan untuk angkatan *e-learning* berikutnya.

### 3. METODOLOGI PENELITIAN

Guna mendukung analisis sentimen memanfaatkan *text mining*, digunakan kerangka kerja Knowledge Discovery in Databases (KDD). KDD versi lengkap memiliki 8 (delapan) tahapan, yaitu: identifikasi tujuan, spesifikasi masalah, seleksi data, *preprocessing* data, *data mining*, pola/model, interpretasi/evaluasi, dan pengetahuan [7]. KDD versi ringkas memiliki 5 tahapan, yaitu: pemilihan data, *pre-processing*, transformasi, pemodelan, dan evaluasi [9, 10].



Gambar 2. Lima Tahapan pada Knowledge Data in Databases/KDD

Tahap 1 - pemilihan data memastikan data tersedia, relevan, dan memenuhi ruang lingkup. Ruang lingkup yang dimaksud adalah LMS yang digunakan (KLC), *e-learning* yang dipilih (*mandatory e-learning* berjudul Penguatan Kesadaran Bela Negara tahun 2023-2024), dan aspek yang dinilai (saran peserta terhadap *e-learning*). Data ini kemudian dilakukan tahap 2 - *preprocessing* data atau *text preprocessing*, yaitu upaya menjamin integritas data agar proses pengolahan data lebih berkualitas, tidak bias, dan memperbaiki presisi [9, 11]. Tahap ini meliputi prosedur seperti: *data cleansing* (eliminasi data dari atribut yang tidak diperlukan), *data integration* (menggabungkan data yang relevan), dan *data reduction* (menyederhanakan kompleksitas data). Tahap ini meliputi prosedur-prosedur seperti yang ditunjukkan pada gambar 2, yang menggambarkan alur *preprocessing* data. Beberapa prosedur utama dalam tahap ini meliputi:

- *Stemming*: pengelompokan kata-kata dengan kata dasar dan arti serupa namun dalam bentuk berbeda/berimbuhan [12], tujuannya memperkecil jumlah indeks dari suatu data.
- *Tokenizing*: memisahkan kalimat ke dalam kata kunci/token [7].
- *Reguler Expression/Regex*: menghilangkan tanda baca yang tidak diperlukan.
- *Case Folding*: menyeragamkan case menjadi huruf kecil agar indeksasi lebih spesifik.
- *Stop Word Removal*: menghapus kata-kata yang tidak penting [13, 14], seperti 'dan', 'ke', 'di'.
- *Filtering Token By Length*: menghapus kata kunci/token dengan panjang karakter  $< 2$ .

Keluaran *preprocessing data* masuk ke tahap 3 - transformasi data, yaitu normalisasi dan agregasi data [7]. Pada tahap ini, dilakukan:

- Teknik pembobotan dengan algoritma TF-IDF (*Term Frequency Inverse Document Frequency*) mencakup 2 metode: (1) frekuensi kemunculan *term*/token kata dalam sebuah dokumen, (2) inversi frekuensi dokumen dengan *term* tersebut dalam meningkatkan proporsi jumlah dokumen yang ditemukan kembali dan dianggap relevan [13] dengan keluaran berupa matriks TF-IDF yang siap dimodelkan sebagaimana dirumuskan tabel 2.

- Teknik pengukuran kemiripan dilakukan dengan menghitung kosinus vektor *term* dua dokumen [14] yaitu hasil keluaran TF-IDF dengan dokumen tersebut menghasilkan *cosine similarity*. Hal ini bertujuan mengekstraksi informasi dalam kumpulan teks terhadap data lain. Keluaran teknik ini berupa matriks *cosine similarity* dirumuskan pada tabel 2.

Tabel 2. Rumus Transformasi Data TF-IDF dan *Cosine Similarity*

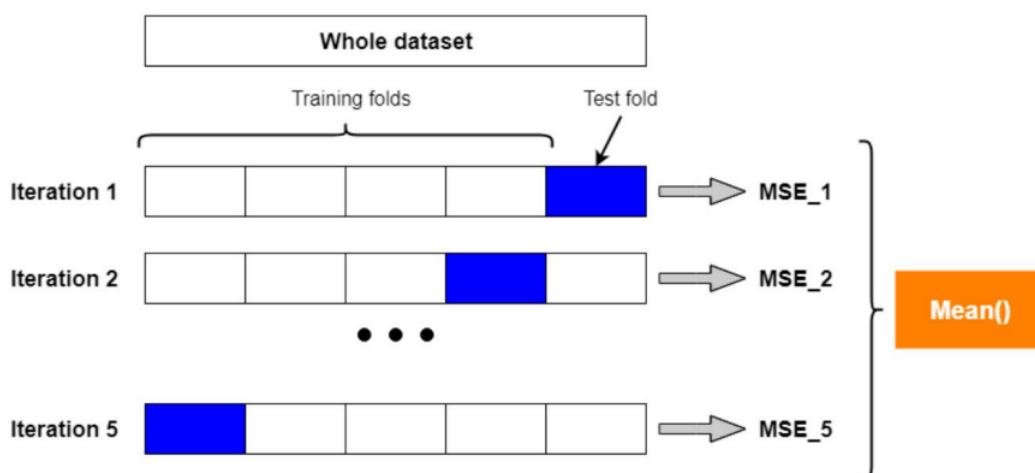
Rumus TF-IDF = TF x IDF	Rumus <i>Cosine Similarity</i>
$W_{dt} = 0.5 + 0.5 \times \frac{tf}{\max(tf)} \times \log \left( \frac{D}{df_t} \right)$	$\text{Sim}(q, d) = \frac{q \cdot d}{ q  d } = \frac{\sum_{k=1}^t w_{qk} \times w_{dk}}{\sqrt{\sum_{k=1}^t (w_{qk})^2} \times \sqrt{\sum_{k=1}^t (w_{dk})^2}}$

Pada prosedur ini, kedua matriks (TF-IDF maupun *cosine similarity*) secara bergantian dan terpisah dapat menjadi menjadi input bagi pemodelan algoritma klasifikasi *text mining*.

- Bila klasifikasi label masih *imbalance* atau tidak seimbang, seperti jumlah data positif jauh melebihi negatif dan netral, maka signifikansi perbedaan tersebut mengakibatkan model berisiko mengambil keputusan keliru. Kondisi ini membutuhkan pendekatan level data (*resampling* data) yang terbukti berkinerja lebih baik daripada level algoritmik [15]. Metode oversampling yang menduplikasi kelas minoritas lebih rentan terhadap *overfitting*, sedangkan metode *undersampling* menghilangkan sejumlah besar informasi yang mungkin berguna [16]. Berdasarkan penelitian [15], metode *undersampling* yang berkinerja paling baik adalah RUS/Random Undersampling. Metode ini menghapus data label yang terlalu banyak agar seimbang dengan data lainnya.

Hasil transformasi data masuk ke tahap 4 – *text mining* atau pemodelan menggunakan algoritma klasifikasi (dengan rasio data training dan testing yang ditentukan), yang bertujuan untuk mengidentifikasi pola serta membuat prediksi berdasarkan data yang telah diproses. Tahap ini menjadi krusial dalam memastikan bahwa model yang dihasilkan mampu memberikan hasil klasifikasi yang akurat dan dapat diandalkan. Berikut adalah beberapa algoritma yang digunakan dalam proses ini:

- Naïve Bayes: Teknik klasifikasi biner dan multikelas yang merepresentasikan tiap kelas berdasarkan kesimpulan atau rekapitulasi probabilistik berdasar atribut atau variabel-variabel yang telah diketahui nilainya [7, 17], di mana untuk kebutuhan text mining pada data berimbang/hasil penyeimbangan data digunakan *Multinomial Naïve Bayes* (MNB). Algoritma ini sering digunakan karena kesederhanaannya dan kemampuannya dalam menangani data teks secara efektif.
- K-Nearest Neighbor (KNN): Teknik klasifikasi terhadap objek berdasarkan data latih yang tingkat kemiripannya paling mendekati data ujinya [2]. KNN melihat kesamaan antar objek dengan parameter k (tingkat kemiripan) dan d (jarak objek data uji dengan semua data latihnya) sehingga dapat ditentukan letak objek *testing* dari semua tetangganya. Metode ini cocok untuk dataset berukuran kecil hingga menengah, tetapi bisa menjadi kurang efisien pada dataset yang sangat besar.
- Support Vector Machine (SVM): teknik prediksi (klasifikasi maupun regresi) bertargetkan pembelajaran atas objek-objek data *training*, dengan mencari *hyperplane*/garis batas terbaik dari fungsi-fungsi tanpa batas jumlah [18]. SVM dikenal memiliki performa yang baik dalam menangani data dengan dimensi tinggi serta mampu menghindari *overfitting* dalam proses klasifikasi.
- Random Forest (RF): merupakan ensemble/kumpulan metode pembelajaran algoritma *decision tree* sebagai pengklasifikasian [19]. Jumlah pohon keputusan dibangun dan dikombinasikan untuk memprediksi klasifikasi tertentu dengan metode *bootstrap sampling* (sampel acak). Proses ini dilakukan secara iteratif sehingga lebih stabil, informatif, kuat terhadap *outliers* dan *noise*, sederhana, dan mudah diparalelkan. Algoritma ini sering dipilih karena kemampuannya dalam menghasilkan prediksi yang lebih akurat dibandingkan metode individu lainnya.



Gambar 3. Ilustrasi untuk Hasil Evaluasi Rata-rata dari 5-Fold Cross-Validation

Tahap 5 - evaluasi memanfaatkan *stratified k-fold cross validation* yaitu strategi partisi data merata melalui validasi silang dari data *training* ke data *testing* bergilir sebanyak k iterasi, bertujuan untuk mengembangkan model agar lebih umum dan performa baik pada data yang tidak terlihat [20]. Variasi evaluasi dilakukan dengan menghitung rata-rata 10 iterasi (*10-fold cross validation*), diharapkan hasilnya lebih stabil dan representatif sebagaimana gambar 3.

Selanjutnya, hasil dari validasi silang ditampilkan dalam tabel *confusion matrix* yaitu matriks dari hasil prediksi atas klasifikasi pemodelan algoritma dibandingkan klasifikasi manual.

Tabel 3. Konsep *Confusion Matrix*

		Prediksi Kelas	
		Positive	Negative
Aktual	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Dari tabel 3, dilakukan pengukuran/metrik ke dalam 4 aspek (rumus sebagaimana tabel 4):

- Akurasi: rasio jumlah klasifikasi uji yang tepat dibandingkan seluruh hasil klasifikasi.
- Presisi: rasio kepastian klasifikasi uji suatu kelas benar dibandingkan seluruh kelas.
- Recall: rasio sensitivitas klasifikasi uji suatu kelas benar dibandingkan yang salah.
- F1-score: bobot rata-rata harmoni antara presisi dan recall

Tabel 4. Rumus Pengukuran dalam *Confusion Matrix*

Akurasi	Presisi	Recall	F1-score
$\frac{TN + TP}{TN + TP + FN + FP} \times 100\%$	$\frac{TP}{TP + FP} \times 100\%$	$\frac{TP}{TP + FN} \times 100\%$	$2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\%$

Angka pada tabel *confusion matrix* diukur presentase akurasi, presisi, recall, dan f1-score-nya menghasilkan pengetahuan pada tahap akhir. Pengetahuan didapat dari perbandingan atas metrik masing-masing model *text mining* dari 4 skenario algoritma berbeda. Pengetahuan atas data dimanfaatkan untuk memecahkan masalah penelitian.



Studi terdahulu berkaitan dengan pembelajaran/*e-learning* dimulai dari penelitian [2] mengevaluasi pembelajaran *e-learning* Bahasa Inggris sehingga 50 dataset *feedback* dalam kelas positif atau negatif menghasilkan akurasi 87,71% ( $k=1$ ) dan 71,42% ( $k=2$ ). Proses utama meliputi: *text-preprocessing*, TF-IDF, *cosine similarity*, hingga pemodelan dengan algoritma KNN. Selanjutnya, penelitian [21] menganalisis sentimen dataset 251 komentar peserta terhadap performa *trainer* Educato.Id dengan proses utama: *text-preprocessing*, TF-IDF, *cosine similarity*, hingga pemodelan KNN menghasilkan akurasi 94,23%.

Penggunaan SVM dilakukan pada penelitian [22] mengklasifikasi sentimen ulasan LMS Quipper diawali tahap *case folding*, *text cleaning*, *tokenization*, *stemming*, dan *stopword removal*. Setelah pelabelan, transformasi data, dan pemodelan SVM, diperoleh akurasi 90,25% dari 1000 data ulasan. Penelitian [19] mengumpulkan 1050 data opini peserta pelatihan dengan rasio data uji:latih 20:80 memanfaatkan TF-IDF, pemodelan SVM 5 kali iterasi, sehingga menghasilkan akurasi, presisi, *recall*, dan *f1-score* 82,08%, 83,42%, 81,16%, dan 81,82% (secara berturut-turut). Kuesioner positif, negatif, dan netral penilaian dosen diteliti [23] dengan 307 data melalui proses: *tokenisasi*, *case folding*, dan *stopword removal* dengan metode evaluasi *6-fold cross validation* mengubah akurasi dari semula 49,54% menjadi 67,83%.

Penggunaan MNB pada penelitian [24, 25] menganalisis sentimen 300 data ulasan pada media sosial berfokus pada label positif, negatif, dan netral pada rasio data *training: testing* = 100:200. Perbandingan algoritma MNB, SVM, dan *Deep Learning* pada penelitian [26] mencakup proses *case folding*, *tokenization*, *filter token by length*, *filter stopwords*, *n-grams*, dan TF-IDF menghasilkan algoritma terbaik MNB dengan akurasi, presisi, dan *recall*: 82,54%, 68,08%, dan 69,81%. Selain itu, MNB, SVM, dan KNN juga digunakan pada penelitian [23, 27] terhadap sentimen LMS Ruangguru, dengan model evaluasi *10-fold cross validation* menghasilkan akurasi terbaik pada algoritma SVM (78,55%). KNN pada penelitian [28] dengan *text preprocessing* berupa *case folding*, *tokenizing*, *filtering*, dan *stemming*, mampu menghasilkan tingkat akurasi, presisi, dan *recall*: 91%, 93%, 91% atas klasifikasi materi pembelajaran menggunakan adaptive e-learning.

Penggunaan RF pada penelitian [20] menjadi algoritma terbaik dengan 97,16% dibandingkan MNB (94,16%) dan SVM (96,01%) pada ulasan aplikasi LMS Ruangguru di Playstore dengan *preprocessing* meliputi *tokenizing*, *filtering*, *stemming*, dan *case folding* terhadap 1000 data (rasio *training: testing* = 90:10). Selain itu, penelitian [29, 30] terhadap 1.040 data komentar mahasiswa terhadap kinerja dosen (rasio *training: testing* = 66:33) menghasilkan SVM sebagai algoritma terbaik pada 85,17% melalui evaluasi *10-fold cross validation*, berbeda tipis dengan RF 84,86%.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Alat Penelitian

Proses KDD dilakukan dengan bahasa pemrograman Python, *Text Editor* PyCharm, dan alat bantu *framework* Streamlit sehingga dapat menampilkan *user interface* yang memadai.

### 4.2. Pemilihan Data

Dari 22 *field* pertanyaan survey *E-learning* Penguatan Kesadaran Bela Negara tahun 2023-2024, *scope* pemilihan data berfokus pada *field*: "Secara keseluruhan, bagaimana pengalaman Bapak/Ibu selama mengikuti *E-Learning* ini? Mohon sampaikan apabila Bapak/Ibu menemui kesulitan, hambatan, atau kendala selama mengikuti *E-Learning*. Kami juga sangat mengharapkan saran/masukan dari Bapak/Ibu untuk meningkatkan kualitas *E-Learning*." Jumlah data pada awalnya 2.902 (terdiri atas 5 kelas pada angkatan 1 tahun 2023), di mana pelabelan manual dilakukan berdasarkan validasi dari admin evaluator penyelenggara e-learning sebagai pakar. Validasi tersebut memenuhi kriteria 3 label data:

- positif: teks berisi apresiasi, kalimat positif secara umum, bukan saran/masukan.
- negatif: teks sinisme, sarkasme, area perbaikan, bersifat negatif, meskipun ada apresiasi.
- netral: teks tanpa makna, misal: strip, spasi, huruf tak bermakna dan berbeda konteks.

Total hasil pelabelan menghasilkan *imbalance* data, sehingga dilakukan penambahan data pada tahap seleksi data dan skenario lanjutan pada tahap-tahap berikutnya sebagaimana tabel 5:

Tabel 5. Penyesuaian Data Awal Tahap Pemilihan Data KDD

No.	Label	Data awal	Penambahan Data	Penyesuaian Data
1	Positif	1.427 (49%)	1.427 (33%)	799** (31%)
2	Negatif	164 (6%)	799 (18%)	799 (31%)
3	Netral	1.311 (45%)	2.174 (49%)	1.008* (38%)
<b>Total</b>		2.902	4.400	2.606

Penyesuaian data meliputi *data cleansing* pada tahap *text preprocessing*\* dan metode *resampling data*\*\* pada tahap transformasi data.

Tabel 5 menyajikan proses penyesuaian data awal dalam tahap pemilihan data pada proses *Knowledge Discovery in Databases* (KDD). Data awal menunjukkan adanya ketidakseimbangan distribusi kelas (*imbalance data*), di mana kelas positif memiliki 1.427 data (49%), kelas negatif hanya 164 data (6%), dan kelas netral 1.311 data (45%).

Untuk mengatasi ketidakseimbangan ini, dilakukan penambahan data melalui berbagai teknik, menghasilkan tambahan 1.427 data (33%) untuk kelas positif, 799 data (18%) untuk kelas negatif, dan 2.174 data (49%) untuk kelas netral. Setelah melalui proses penyesuaian, jumlah data akhir yang digunakan dalam analisis adalah 799 data untuk kelas positif (31%), 799 data untuk kelas negatif (31%), dan 1.008 data untuk kelas netral (38%).

Dengan demikian, Tabel 5 menunjukkan bagaimana teknik *text preprocessing* dan *resampling* data digunakan untuk menyeimbangkan distribusi kelas, sehingga meningkatkan kualitas hasil klasifikasi pada tahap selanjutnya.

### 4.3. Text Preprocessing Data

#### Tabel Hasil Cleansing

♥ Jumlah data: 3234

↑	Secara keseluruhan, bagaimana pengalaman Bapak/Ibu selama mengikuti E-Learning ini? Mohon	Label3
1,461	Good	1
1,462	soalnya cukup lumayan cuma waktu yang kurang di perpanjang	2
1,463	Terima kasih	1
1,464	bagus	1
1,465	oke	1
1,466	tidak ada kesulitan	3
1,468	cukup	3
1,469	hambatan hanya ada di jaringan internet yang lambat	2
1,471	oke	1
1,473	good	1
1,474	Agar waktunya disesuaikan karena kami juga melaksanakan tugas rutin	2

#### Sebelum cleansing:

Jumlah label 1	Jumlah label 2	Jumlah label 3
1427	799	2174
↑ 0.32	↑ 0.18	↑ 0.49

#### Setelah cleansing:

Jumlah label 1	Jumlah label 2	Jumlah label 3
1427	799	1008
↑ 0.44	↑ 0.25	↑ 0.31

Gambar 4. Data Frame 11 Data (Data Nomor 1.467, 1.470, 1.472 Tidak Ada Karena Dihapus)

Tahap ini dilakukan secara simultan, dimulai dari: *data cleansing* yang meliputi: menghapus data *null*, menyeragamkan kata umum, mengubah *special* karakter menjadi spasi, menghilangkan spasi ganda, menghapus huruf berulang pada kata dengan pengecualian tertentu, dan menghapus data yang tidak bermakna. *Data cleansing* mereduksi data label netral dari semula 2.174 menjadi 1.008, sehingga total data menjadi 3.234, sebagaimana gambar 4.

*Data cleansing* dilanjutkan dengan strategi *text preprocessing* lain sebagaimana gambar 5:



♥ Jumlah data: 3234      ♥ Jumlah seluruh token: 13738

Unduh CSV

+	Secara keseluruhan, bagaimana pengalaman Bapak/Ibu selama mengikuti E-Learning	Label3	preprocessed_t	token	cosine_similarity
1,461	Good	1	good	good	0 0 0 0 0 0 0
1,462	soalnya cukup lumayan cuma waktu yang kurang di perpanjang	2	soal cukup lum	soal cukup lumayan cuma waktu kurang panjang	0 0 0 0 0 0 0
1,463	Terima kasih	1	terima kasih	terima kasih	0 0 0 0 0 0 0
1,464	bagus	1	bagus	bagus	0.49201735361140475 0 1
1,465	oke	1	oke	oke	0 0 0 0 0 0 0
1,466	tidak ada kesulitan	3	tidak ada sulit	tidak ada sulit	0 0 0 0 0 0.0584899981
1,468	cukup	3	cukup	cukup	0 0 0 0 0 0 0
1,469	hambatan hanya ada di jaringan internet yang lambat	2	hambat ada jaring	hambat ada jaring internet lambat	0 0 0 0 0 0 0
1,471	oke	1	oke	oke	0 0 0 0 0 0 0
1,473	good	1	good	good	0 0 0 0 0 0 0
1,474	Agar waktunya disesuaikan karena kami juga melaksanakan tugas rutin	2	waktu sesuai la	waktu sesuai laksana tugas rutin	0 0 0 0 0 0 0
1,476	mantap	1	mantap	mantap	0 0 0 0 0 0 0
1,477	Tidak Ada	3	tidak ada	tidak ada	0 0 0 0 0 0.0920931268
1,481	tidak ada kesulitan	3	tidak ada sulit	tidak ada sulit	0 0 0 0 0 0.0584899981
1,482	Penjelasan pelajaran dan video cukup bagus	1	jelas ajar video	jelas ajar video cukup bagus	0.17151010961547777 0 0
1,484	tidak ada hambatan kesulitan ataupun kendala yang berarti	3	tidak ada hamb	tidak ada hambatan sulit atau kendala arti	0 0 0 0 0 0.0238378187
1,495	pengalaman selama mengikuti elearning ini semakin memahami upaya bela negara	1	alam lama ikut	alam lama ikut elearning makin paham upaya bela negara	0 0 0 0 0 0 0.33

Gambar 5. Data Frame 11 menampilkan *Text Preprocessing*, *Tokenizing*, dan *Cosine Similarity*

Hasil *cleansing data* ditunjukkan pada kolom paling kiri gambar 5 (kolom "secara keseluruhan dan seterusnya"). Pelabelan manual (kolom "label3") merepresentasikan label 1 positif, 2 negatif, dan 3 netral. Adapun kolom *preprocessed text* dan token dijelaskan sebagai berikut:

- Good => *case folding* dan pengecualian penghilangan huruf ganda => menjadi good;
- soalnya, perpanjang, kesulitan => setelah proses *stemming* => soal, panjang, sulit;
- yang, di, dan, agar => *stop word removal* => dihilangkan;
- tidak, ada => dikecualikan dari *stop word removal* => tidak, ada;

Adapun *regular expression/regex* dan *filtering by length* telah dilakukan pada *data cleansing*.

#### 4.4. Transformasi Data

Kolom "token" pada gambar 5 menunjukkan pemisahan token kata hasil tokenizing yang kemudian di vektorkan menjadi TF-IDF matriks. Proses ini bertujuan untuk merepresentasikan teks dalam bentuk numerik sehingga dapat digunakan sebagai input pada model klasifikasi. TF-IDF membantu menyoroti kata-kata yang lebih relevan dalam dokumen dengan memberikan bobot lebih tinggi pada kata yang jarang muncul di seluruh dataset, namun sering muncul dalam dokumen tertentu. Pada titik ini, terdapat dua cabang:

- Matriks TF-IDF langsung menjadi input pada tahap pemodelan, atau;
- Matriks TF-IDF diolah untuk dihitung *cosine similarity*-nya berdasarkan input tiap pasangan matriks TF-IDF, menghasilkan *matrix cosine similarity*.

Dua jenis matriks, baik TF-IDF maupun cosine similarity sebelum dimodelkan, dilakukan terlebih dahulu teknik *resampling data* memanfaatkan metode *Random Undersampling* (RUS) untuk mereduksi 44% data label positif dari 1.427 menjadi 799 sehingga mengurangi *data training* pada tahap pemodelan (komposisi data dijelaskan sebagaimana tabel 5).

#### 4.5. Pemodelan/Text Mining dan Evaluasi Data

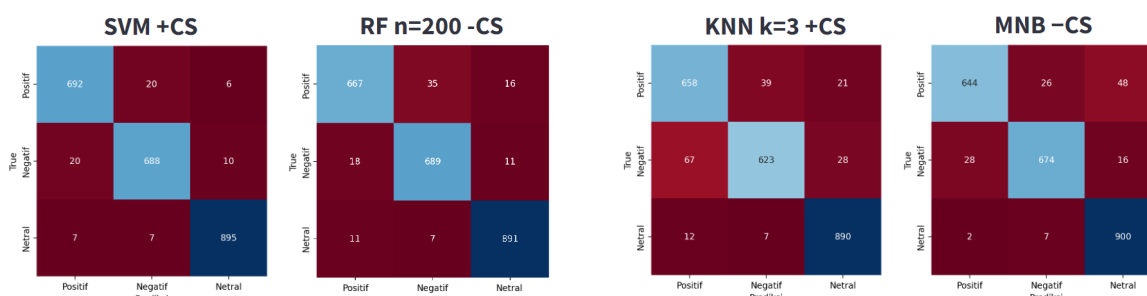
Hasil transformasi data dimodelkan ke masing-masing algoritma klasifikasi text mining. Model kemudian dilakukan stratified cross-validation dengan nilai average k=10, menghasilkan komparasi tabel akurasi sebagaimana Tabel 6, yang divisualisasikan dalam bentuk heatmap untuk mempermudah interpretasi nilai akurasi tiap algoritma. Warna pada heatmap menunjukkan perbedaan tingkat akurasi, di mana semakin gelap atau terang warnanya mencerminkan perbedaan performa antar model dalam melakukan klasifikasi.

Dari hasil *modeling*, dilakukan proses di mana:

Tabel 6. *Heatmap* Akurasi (dalam %) 4 Algoritma dengan 3 Jenis Rasio *Training-Testing*

Rasio Data <i>Training: Testing</i>	Support Vector Machine (SVM)		Multinomial Naïve Bayes (MNB)		3-Nearest Neighbor (KNN k=3)		Random Forest (RF n=200)	
	+CS	-CS	+CS	-CS	+CS	-CS	+CS	-CS
90:10	97.01	96.37	82.39	94.58	92.58	88.02	95.44	95.82
80:20	96.64	96.26	82.44	94.15	92.37	87.95	95.35	95.78
70:30	96.55	96.16	82.18	93.81	91.83	87.39	94.85	95.34

1. Algoritma KNN menggunakan k=3 sebagai hasil *f1-score* terbaik di antara k=1 s.d. k=5.
2. Algoritma RF dengan n atau jumlah pohon keputusan dengan hasil akurasi terbaik antara n=10, 50, 100, dan 200, maka didapat hasil terbaik adalah 200. Dengan banyaknya jumlah pohon keputusan, proses komputasi relatif lebih lama daripada algoritma lain.
3. Nilai akurasi merupakan total *true* positif, negatif, dan netral dibagi dengan total data.
4. +CS menunjukkan penggunaan matriks *cosine similarity* pada *modeling*. -CS berarti model menggunakan matriks TF-IDF saja.
5. Rasio data 90:10 menghasilkan nilai akurasi relatif lebih baik daripada rasio data lainnya.
6. Dari hasil tersebut, model dengan algoritma SVM dan KNN memiliki nilai akurasi yang tinggi jika memanfaatkan matriks *cosine similarity* di dalam modelnya. Sebaliknya, algoritma MNB dan RF menghasilkan nilai akurasi yang tinggi jika tidak menggunakan matriks *cosine similarity* (hanya matriks TF-IDF saja).
7. Perbandingan *confusion matrix* 4 model terbaik tiap algoritma sebagaimana gambar 6:

Gambar 6. Perbandingan *Confusion Matrix* 4 Algoritma Terbaik

Selanjutnya, diperhitungkan durasi proses terhadap ke-4 algoritma terbaik pada gambar 6:

Gambar 6 menunjukkan *confusion matrix* dari empat algoritma klasifikasi terbaik yang digunakan dalam penelitian ini, yaitu *Support Vector Machine* (SVM), *Random Forest* (RF), *K-Nearest Neighbor* (KNN), dan *Multinomial Naïve Bayes* (MNB). *Confusion matrix* ini menggambarkan performa masing-masing model dalam mengklasifikasikan data, termasuk jumlah prediksi yang benar dan salah dalam kategori positif, negatif, dan netral.

Tabel 7 menyajikan perbandingan akurasi dan durasi *training* serta evaluasi model dari keempat algoritma tersebut. Berdasarkan tabel, SVM memiliki akurasi tertinggi sebesar 97,01%, diikuti oleh RF (95,82%), KNN (92,58%), dan MNB (94,58%). Namun, dari segi efisiensi waktu, MNB memiliki durasi *training* tercepat, yaitu hanya 0,11 detik, sedangkan RF membutuhkan waktu paling lama, yaitu 84,47 detik. Dengan demikian, pemilihan algoritma yang optimal bergantung pada kebutuhan antara akurasi yang lebih tinggi atau waktu komputasi yang lebih singkat.

Tabel 7. Perbandingan Akurasi dan Durasi Training Model

Indikator	SVM	MNB	KNN	RF
	+CS	-CS	+CS	-CS
Akurasi	97,01%	94,58%	92,58%	95,82%
Durasi Training dan Evaluasi Model	48,71 detik	0,11 detik	11,57 detik	84,47 detik

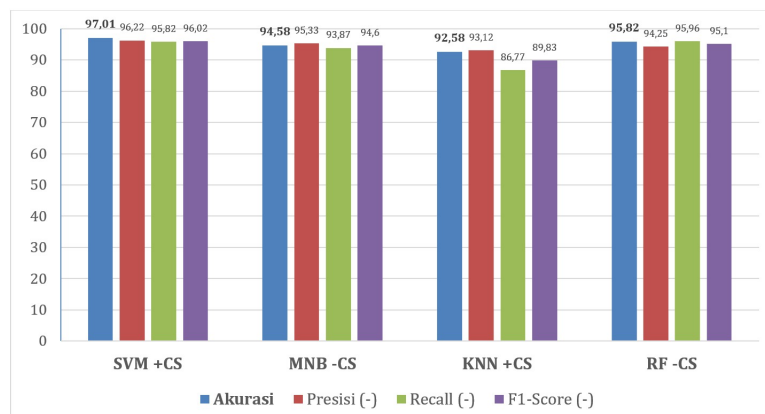
Selain mempertimbangkan akurasi dan durasi *training*, pemilihan algoritma juga dapat dipengaruhi oleh kompleksitas data serta kebutuhan aplikasi. SVM, dengan akurasi tertinggi, cocok digunakan jika prioritas utama adalah hasil prediksi yang lebih presisi, meskipun membutuhkan waktu pemrosesan yang lebih lama. Sebaliknya, MNB yang memiliki durasi *training* tercepat dapat menjadi pilihan jika sistem memerlukan respons *real-time* atau bekerja dengan data dalam jumlah besar. RF dan KNN berada di antara kedua ekstrem tersebut, menawarkan keseimbangan antara akurasi dan efisiensi komputasi. Oleh karena itu, pemilihan model harus disesuaikan dengan tujuan dan batasan sistem yang digunakan.

#### 4.6. Pengetahuan

Berdasarkan gambar 6, pemodelan text mining menghasilkan pengetahuan bahwa keempat model menghasilkan nilai *true* positif, negatif, dan netral (warna biru diagonal) secara mayoritas, artinya *output* prediksi model memiliki tingkat akurasi yang tinggi dibandingkan dengan pelabelan data sebelumnya. Tingkat akurasi (efektivitas) dan durasi waktu (efisiensi):

- 97,01% untuk algoritma SVM (dengan *cosine similarity*),
- 94,58% untuk algoritma MNB (tanpa *cosine similarity*),
- 92,58% untuk algoritma KNN k=3 (dengan *cosine similarity*), dan
- 95,82% untuk algoritma RF n=200 (dengan *cosine similarity*).

Dari hasil tersebut, algoritma SVM menjadi model dengan akurasi terbaik dan efisiensi waktu yang relatif baik. Adapun tingkat presisi, *recall*, dan *f1-score* perlu dianalisis lebih lanjut pada label negatif, artinya data saran negatif perlu diperhitungkan dan ditindaklanjuti evaluator LMS sebagai laporan nantinya. Hasil perbandingan ketiga metrik tersebut sebagaimana gambar 7 memiliki kemiripan pada algoritma SVM +CS dengan RF -CS.

Gambar 7. Perbandingan Nilai Akurasi dan Presisi, *Recall*, F1-Score Terhadap Label Negatif

Keseluruhan hasil presisi, *recall*, dan *f1-score* pada gambar 7 dapat dimaknai:

1. Presisi label negatif tinggi menunjukkan semakin sedikitnya label *false* negatif yang salah prediksi ke *true* negatif, sehingga meminimalisasi label negatif yang salah prediksi ke label positif/netral. Presisi tertinggi terjadi pada algoritma SVM +CS sebesar 96,22%.
2. *Recall* label negatif tinggi menunjukkan semakin sedikitnya label *true* negatif yang salah prediksi ke *false* negatif, sehingga meminimalisasi label positif/netral yang salah prediksi ke label negatif. *Recall* tertinggi terjadi pada algoritma RF-CS sebesar 95,96%.
3. *F1-score* label negatif tinggi menunjukkan optimalisasi kestabilan/kesetimbangan *true* negatif yang diprediksi sebagai *true* negatif dan *false* negatif yang juga diprediksi sebagai *false* negatif, sehingga menunjukkan performa kombinasi yang baik antara presisi dan *recall*. *F1-score* tertinggi terjadi pada algoritma SVM +CS sebesar 96,02%.

Meskipun pemodelan RF -CS menghasilkan nilai *recall* lebih tinggi daripada SVM +CS, secara kualitas lebih diunggulkan SVM +CS. Hal ini dikarenakan jumlah pohon keputusan  $n=200$  pada RF -CS relatif banyak, sehingga waktu komputasi menjadi lebih lama, yaitu 84,47 detik atau 150% lebih lama daripada waktu komputasi algoritma SVM CS+.

## 5. KESIMPULAN

Penelitian ini menganalisis keseluruhan saran peserta secara tekstual atas evaluasi level 1 (*reaction*) pada *mandatory e-learning* sejumlah 4.400 data dilakukan proses *Knowledge Discovery in Databases* (KDD) dimulai dari tahap pemilihan dan pelabelan data manual, *text preprocessing*, transformasi data, termasuk resampling data menggunakan *Random Undersampling/RUS* sehingga menyeimbangkan sejumlah 2.602 data untuk dimodelkan. Hasil *text mining* tersebut menghasilkan 4 model terbaik pada rasio *data training* dan data testing 90:10, yaitu: *Support Vector Machine* dengan *cosine similarity* (48,71 detik), Multinomial Naive Bayes tanpa *cosine similarity* (0,11 detik), KNN  $k=3$  dengan *cosine similarity* (11,57 detik), dan Random Forest dengan  $n=200$  tanpa *cosine similarity* (84,47 detik).

Model beralgoritma *Support Vector Machine* dan *cosine similarity* menghasilkan metrik evaluasi data terbaik dengan tingkat akurasi 97,01%, sedangkan presisi, *recall*, dan *f1-score* terhadap label negatif secara berturut-turut: 96,22%, 95,82%, dan 96,02%. Meskipun *recall* pada model beralgoritma *Random Forest* tanpa *cosine similarity* lebih baik, waktu komputasi yang lebih efisien (150% lebih baik) tetap mengunggulkan algoritma SVM tersebut. Berdasarkan penelitian ini, model yang dihasilkan dari algoritma SVM dengan *cosine similarity* dapat memprediksi klasifikasi pada data evaluasi *reaction* peserta berbasis teks pada penyelenggaraan *mandatory e-learning* lain pada LMS Kemenkeu Learning Center (KLC) secara akurat, cepat, dan terotomasi. Hal ini dapat mempermudah PIC admin evaluasi memperoleh data label negatif, yaitu saran peserta yang perlu ditindaklanjuti dan dilaporkan secara berkala sebagai wujud *continuous improvement* pada penyelenggaraan pembelajaran.

## 6. SARAN

Saran dan rekomendasi atas penelitian mendatang meliputi:

1. Klasifikasi target perlu diperluas menjadi beberapa kelas (*multiclass*) yang merujuk pada kategori saran peserta yang perlu ditindaklanjuti, misalnya berkaitan dengan video, materi, fitur aplikasi LMS, jaringan/*network*, dan lain-lain sesuai topik yang relevan.
2. Untuk melakukan evaluasi dengan metode yang lebih bervariasi, jenis algoritma *deep learning* perlu dipertimbangkan untuk digunakan.
3. Untuk menguji objek penelitian dengan sample lebih menyeluruh, perlu dikolaborasikan dataset dengan jenis pelatihan lain yang diselenggarakan melalui LMS KLC, seperti *microlearning*, *e-learning open access*, dan pelatihan jarak jauh yang bersifat *synchronous*.
4. Teknik analisis sentimen yang memanfaatkan *text mining* dengan strategi pelabelan perlu ditinjau efektivitas hasilnya dibandingkan dengan teknik *summarization* pada masalah yang sama yaitu pengintisarian evaluasi level 1 (*reaction*) untuk *mandatory e-learning*.

## 7. DEKLARASI

### 7.1. Tentang Penulis

Andi Nur Syamsudin (AN)  <https://orcid.org/0009-0008-6397-7165>

Utomo Budiyanto (UB)  <https://orcid.org/0009-0009-0888-051X>

### 7.2. Kontribusi Penulis

Konseptualisasi: AN; Metodologi: UB ; Perangkat Lunak: AN; Validasi: UB dan AN; Analisis Formal: UB dan ; Investigasi: AN ; Sumber Daya: UB; Kurasi Data: AN; Penulisan Draf Asli Persiapan: UB dan AN ; Penulisan Tinjauan dan Penyuntingan: UB dan AN ; Visualisasi: UB; Semua penulis, AN dan UB, telah membaca dan menyetujui versi naskah yang diterbitkan.

### 7.3. Pernyataan Ketersediaan Data

Data yang disajikan dalam studi ini tersedia atas permintaan dari penulis terkait.

### 7.4. Pendanaan

Penulis tidak menerima dukungan finansial untuk penelitian, kepenulisan, dan/atau penerbitan artikel ini.

### 7.5. Deklarasi Konflik Kepentingan

Penulis menyatakan bahwa mereka tidak memiliki konflik kepentingan, konflik kepentingan finansial yang diketahui, atau hubungan pribadi yang dapat memengaruhi pekerjaan yang dilaporkan dalam makalah ini.

## DAFTAR PUSTAKA

- [1] F. Alifa *et al.*, “Evaluasi learning management system (lms) berbasis web menggunakan moodle pada fakultas tarbiyah uin ar-raniry,” *Evaluasi Learning Management System (LMS) Berbasis Web Menggunakan Moodle pada Fakultas tarbiyah UIN Ar-Raniry*, vol. 4, no. 4, pp. 1–6, 2024.
- [2] H. P. Maulidina and F. A. Bachtiar, “Klasifikasi komentar pada pembelajaran e-learning menggunakan analisis sentimen dengan metode k-nearest neighbor,” *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 6, no. 5, pp. 2301–2307, 2022.
- [3] A. Yauma, I. Fitri, and S. Ningsih, “Learning management system (lms) pada e-learning menggunakan metode agile dan waterfall berbasis website,” *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 5, no. 3, pp. 323–328, 2021.
- [4] S. A. Anjani and I. A. Mutiara, “Mengintegrasikan teknologi blockchain dalam pendidikan tinggi: Meningkatkan transparansi dan keamanan dalam kredensial akademik,” *ADI Pengabdian Kepada Masyarakat*, vol. 5, no. 1, pp. 65–71, 2024.
- [5] B. Setiyadi, “Pemanfaatan dan pengelolaan teknologi informasi dan komunikasi dalam menunjang proses pembelajaran,” *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 150–161, 2023.
- [6] A. Y. Setiawan, I. G. M. Darmawiguna, and G. A. Pradnyana, “Sentiment summarization evaluasi pembelajaran menggunakan algoritma lstm (long short term memory),” *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, vol. 11, no. 2, pp. 183–191, 2022.
- [7] I. R. Munthe, A. P. Juledi *et al.*, “Implementasi data mining algoritma apriori untuk meningkatkan penjualan,” *Jurnal Teknik Informatika UNIKA Santo Thomas*, vol. 6, no. 1, pp. 188–197, 2021.
- [8] M. Sholeh, S. Suraya, and D. Andayati, “Penerapan data mining pada model clustering data kuesioner mahasiswa terhadap kinerja dosen,” *Jurnal Eksplora Informatika*, vol. 13, no. 2, pp. 208–217, 2024.
- [9] W. Yustanti and N. Rochmawati, “Analisis algoritma klasifikasi untuk memprediksi karakteristik mahasiswa pada pembelajaran daring,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 57–61, 2022.
- [10] I. N. Ramadhan and K. ini berlisensi di bawah Creative, “Penerapan database redis sebagai optimalisasi pemrosesan kueri data pengguna aplikasi siresma berbasis laravel: Implementation of the redis database as optimization of user.”
- [11] I. A. Mutiara, A. Syamsuddin, M. Maharida, F. Napasti, and L. Hasnawati, “Instilling nationalism and sociopreneurship in young indonesian immigrants,” *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 1, pp. 37–47, 2025.

- [12] N. Permatasari and C. Karima, "Penerapan algoritma idris pada dokumen dengan menggunakan teks bahasa indonesia," *Jurnal Humaniora Teknologi*, vol. 10, no. 2, pp. 80–88, 2024.
- [13] M. I. A. G. Wibowo and I. Pratama, "Analisis sentimen ulasan aplikasi identitas kependudukan digital menggunakan metode support vector machine," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 4, pp. 715–722, 2024.
- [14] R. Aprianto, C. Lukita, A. Sutarman, R. A. Sunarjo, R. N. Muti, and E. Dolan, "Facing global dynamics with effective strategy: A tasted organizational change management approach," *International Journal of Cyber and IT Service Management*, vol. 5, no. 1, pp. 1–11, 2025.
- [15] Y. A. Sir, A. H. Soepranoto *et al.*, "Pendekatan resampling data untuk menangani masalah ketidakseimbangan kelas," *J-ICON: Jurnal Komputer dan Informatika*, vol. 10, no. 1, pp. 31–38, 2022.
- [16] A. Y. Taha, S. Tiun, A. H. Abd Rahman, and A. Sabah, "Multilabel over-sampling and under-sampling with class alignment for imbalanced multilabel text classification," *Journal of Information and Communication Technology*, vol. 20, no. 3, pp. 423–456, 2021.
- [17] M. Toha, E. Anoh *et al.*, "Strategi public relations dalam menangani pemberitaan negatif di media massa atas pengoperasian teknologi pionir indonesia."
- [18] A. A. Viriya, I. S. E. Maghfiroh, and N. Y. Setiawan, "Analisis sentimen ulasan pengguna aplikasi mobile gapura ub pada google play store menggunakan algoritma support vector machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, 2024.
- [19] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis sentimen terhadap aplikasi peluang menggunakan algoritma naive bayes dan support vector machine (svm): Sentiment analysis of peluang applications with naive bayes and support vector machine (svm) algorithm," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, 2024.
- [20] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation," *Electronics*, vol. 10, no. 16, p. 1973, 2021.
- [21] F. Riza, "Penggunaan text mining untuk klasifikasi tanggapan peserta pelatihan terhadap performa trainer menggunakan metode algoritma k-nearest neighbor," *ISMETEK*, vol. 13, no. 2, 2022.
- [22] A. A. Annur, A. A. Murtopo, and N. Fadilah, "Analisis sentimen aplikasi e-learning quipper selama pandemi covid-19 dengan menggunakan metode support vector machine (svm)," *Indonesian Journal of Informatics and Research*, vol. 3, no. 2, pp. 9–17, 2022.
- [23] D. S. Utami and A. Erfina, "Analisis sentimen pinjaman online di twitter menggunakan algoritma support vector machine (svm)," in *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, vol. 1, 2021, pp. 299–305.
- [24] R. N. Mauliza and Y. R. Sipayung, "Penerapan text mining dalam menganalisis pendapat masyarakat terhadap pemilu 2024 pada media sosial x menggunakan metode naive bayes," *Technomedia Journal*, vol. 9, pp. 1–16, 2024.
- [25] S. S. Wulandari, M. L. B. M. Diah, and A. Asari, "Digital proficiency and entrepreneurial mindset for sme success through market savvy and tech literacy," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 1, pp. 26–36, 2025.
- [26] N. Firdausy, I. Yuadi, and I. Puspitasari, "Analisis sentimen evaluasi reaksi e-learning menggunakan algoritma naive bayes support vector machine dan deep learning," *Techno. Com*, vol. 22, no. 3, pp. 677–689, 2023.
- [27] A. Ekawaty, E. A. Nabila, S. A. Anjani, U. Rahardja, and S. Zebua, "Utilizing sentiment analysis to enhance customer feedback systems in banking," in *2024 12th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2024, pp. 1–6.
- [28] A. A. Nugraha and U. Budiyanto, "Adaptive e-learning system berbasis vark learning style dengan klasifikasi materi pembelajaran menggunakan k-nn (k-nearest neighbor)," *Technomedia Journal*, vol. 7, no. 2, pp. 248–261, 2022.
- [29] K. Okoye, A. Arrona-Palacios, C. Camacho-Zuñiga, J. A. G. Achem, J. Escamilla, and S. Hosseini, "Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification," *Education and Information Technologies*, pp. 1–43, 2022.
- [30] R. D. Hadiwidjaja, A. I. Suroso, H. Siregar, and I. Sailah, "Performance paradigm: Entrepreneurial good university governance mediating leadership style in state universities," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 6, no. 3, pp. 492–508, 2024.
-