

Klasifikasi Teks Umpan Balik Kompetensi Kepribadian di Perguruan Tinggi Menggunakan Ekstraksi Fitur TF-IDF dan Algoritma Logistic Regression

Vina Ayumi^{1a}, Mariana Purba^{2b}, Siska Mailana^{1c}

¹Fakultas Teknik dan Informatika, Universitas Dian Nusantara, Indonesia

²Fakultas Ilmu Komputer, Universitas Sjakhyakirti, Palembang, Indonesia

^avina.ayumi@dosen.undira.ac.id, ^bmariana_purba@unisti.ac.id, ^c4112110714@mahasiswa.undira.ac.id

Article Info

Article history:

Received, 2025-06-16

Revised, 2025-06-19

Accepted, 2025-06-30

Kata Kunci:

TF-IDF,
Logistic Regression,
Klasifikasi Teks,
Sentimen Analisis,
Kompetensi Kepribadian

ABSTRAK

Penelitian ini bertujuan mengembangkan dan mengevaluasi model klasifikasi teks untuk mengidentifikasi sentimen pada umpan balik kompetensi kepribadian dosen di perguruan tinggi menggunakan metode ekstraksi fitur TF-IDF dan algoritma Logistic Regression (LR). Data yang digunakan berasal dari evaluasi mahasiswa terhadap kompetensi kepribadian dosen di Universitas Sjakhyakirti, dengan total 6.112 teks yang sudah diberi label sentimen positif sebanyak 3.700 dan negatif sebanyak 2.412. Dataset kemudian dibagi menjadi tiga bagian yakni pelatihan (70%), validasi (10%), dan pengujian (20%). Tahapan penelitian terdiri dari pra-pemrosesan teks yaitu proses pembersihan data, normalisasi huruf, dan penghapusan kata-kata umum, tahap selanjutnya pembobotan term menggunakan metode TF-IDF dan model LR untuk melakukan klasifikasi sentimen positif dan negatif. Model dievaluasi menggunakan metrik akurasi, precision, recall, serta confusion matrix. Hasil eksperimen menunjukkan bahwa pada epoch ke-50, model mencapai akurasi pelatihan sebesar 81,90% dan akurasi validasi sebesar 78,30% sedangkan pada pengujian model TF-IDF-LR mendapatkan akurasi sebesar 75,1%.

ABSTRACT

This study aimed to develop and evaluate a text classification model to identify sentiment in feedback on lecturers' personality competencies at a university using TF-IDF feature extraction and Logistic Regression (LR) algorithms. The data originated from student evaluations of lecturers' personality competencies at Universitas Sjakhyakirti, consisting of a total of 6,112 texts labeled as positive sentiment (3,700) and negative sentiment (2,412). The dataset was then divided into three parts: training (70%), validation (10%), and testing (20%). The research stages included text preprocessing, which involved data cleaning, letter normalization, and the removal of common words, followed by term weighting using the TF-IDF method and classification using the LR model to categorize texts as positive or negative sentiment. The model was evaluated using accuracy, precision, recall metrics, and a confusion matrix. Experimental results showed that at the 50th epoch, the model achieved a training accuracy of 81.90% and a validation accuracy of 78.30%, while on the testing data, the TF-IDF-LR model reached an accuracy of 75.1%.

This is an open access article under the CC BY-SA license.



Penulis Korespondensi:

Vina Ayumi,
Fakultas Teknik dan Informatika,
Universitas Dian Nusantara, Indonesia
Email: vina.ayumi@dosen.undira.ac.id

1. PENDAHULUAN

Teknologi kecerdasan buatan telah mendukung kemajuan pada bidang kesehatan dan pendidikan [1]–[3]. Evaluasi kompetensi kepribadian dosen merupakan aspek penting perlu menerapkan teknologi dalam meningkatkan kualitas pendidikan di perguruan tinggi [4]. Kepribadian dosen tidak hanya memengaruhi proses belajar mengajar, tetapi juga berdampak pada motivasi dan kepuasan mahasiswa selama mengikuti perkuliahan. Oleh karena itu, pengumpulan umpan balik dari mahasiswa mengenai kompetensi kepribadian dosen menjadi kebutuhan utama untuk mengukur dan meningkatkan mutu pengajaran di institusi pendidikan tinggi [5]–[7]. Umpan balik tersebut umumnya berupa data teks yang bersifat tidak terstruktur, sehingga analisis manual terhadap data dalam jumlah besar menjadi sangat tidak efisien dan rentan kesalahan.

Dengan berkembangnya teknologi *natural language processing*, analisis sentimen secara otomatis dapat menjadi metode untuk mengolah umpan balik tersebut, serta mengidentifikasi sentimen positif maupun negatif yang ada dalam teks dataset [8]–[10]. Metode ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) telah banyak digunakan dalam mengekstrak kata-kata penting yang membedakan konten dokumen. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan dan keunikannya di seluruh dataset, sehingga memudahkan proses klasifikasi teks. Metode *Logistic Regression* (LR) dipilih sebagai algoritma klasifikasi dalam penelitian ini karena dapat menyelesaikan masalah klasifikasi biner dengan baik serta interpretasi yang mudah dari hasil model. Kombinasi TF-IDF sebagai teknik ekstraksi fitur dan LR sebagai model klasifikasi dipilih berdasarkan kinerja model pada dataset teks bahasa Indonesia pada penelitian sebelumnya [11], [12].

Penelitian Antonio et al (2022) menganalisis sentimen terkait COVID-19 di Indonesia menggunakan data Twitter dengan metode ekstraksi fitur TF-IDF dan algoritma Stochastic Gradient Descent (SGD). Dataset yang digunakan berjumlah 830 tweet, terdiri dari 583 tweet positif dan 246 tweet negatif. Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian, dengan total 189 data uji. Hasil klasifikasi menunjukkan model mencapai akurasi sebesar 85,14%, menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen tweet tentang COVID-19 di Indonesia [13].

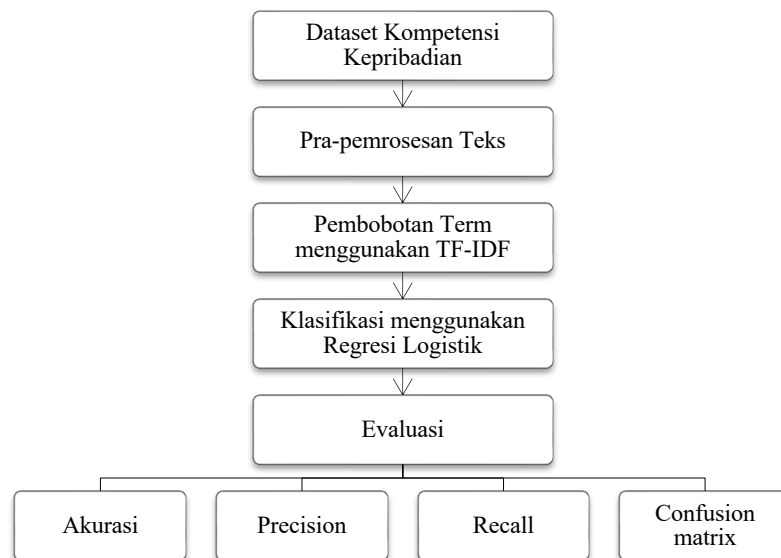
Penelitian oleh Sari et al. (2023) memprediksi jenis kasus hukum di Indonesia dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *K-Nearest Neighbor* (KNN). Metode TF-IDF digunakan untuk mengubah data kasus hukum menjadi format yang dapat diproses komputer, kemudian dilakukan klasifikasi menggunakan algoritma KNN untuk memprediksi jenis kasus hukum tersebut. Dataset yang digunakan berjumlah 90 data, dibagi secara acak menjadi 60 data pelatihan dan 30 data pengujian. Hasil pengujian menunjukkan tingkat akurasi klasifikasi sebesar 86,67%, dengan nilai precision rata-rata tertimbang 85,29% dan recall rata-rata tertimbang 86,67%, yang menunjukkan kinerja model yang baik dalam memprediksi jenis kasus hukum [14].

Penelitian oleh Syahputra et al. (2023) membandingkan algoritma Support Vector Machine (SVM) dan Random Forest untuk deteksi konten negatif pada situs web. Berdasarkan pengujian menggunakan 526 URL situs web, algoritma Random Forest mencapai akurasi 92%, precision 71%, dan recall 86%. Namun, hasil penelitian menunjukkan bahwa SVM memiliki performa yang lebih baik dibandingkan Random Forest dalam mendeteksi konten negatif pada studi [15].

Penelitian ini menggunakan dataset yang berasal dari evaluasi mahasiswa di Universitas Sjakhyakirti yang terdiri dari 6.112 teks dengan label sentimen positif sebanyak 3.700 dan negatif sebanyak 2.412. Dataset tersebut dibagi menjadi data pelatihan (70%), validasi (10%), dan pengujian (20%). Melalui serangkaian tahapan pra-pemrosesan teks, ekstraksi fitur TF-IDF, dan klasifikasi menggunakan Logistic Regression, model dievaluasi dengan metrik akurasi, precision, recall, dan confusion matrix. Hasil penelitian ini diharapkan dapat membantu perguruan tinggi dalam melakukan analisis otomatis terhadap umpan balik dosen guna meningkatkan kualitas pembelajaran secara berkelanjutan.

2. METODE PENELITIAN

Tujuan penelitian ini adalah untuk mengembangkan dan mengevaluasi model klasifikasi teks menggunakan metode TF-IDF dan Logistic Regression untuk klasifikasi teks umpan balik kompetensi kepribadian dosen berdasarkan kelas sentimen positif dan negatif. Data umpan balik kompetensi kepribadian dosen berasal dari evaluasi studi menggunakan kuesioner di Universitas Sjakhyakirti. Dataset terdiri dari 6.112 teks berlabel sentimen positif (3.700) dan negatif (2.412). Teks umpan balik tersebut dilakukan pra-pemrosesan, ekstraksi fitur menggunakan TF-IDF dan diklasifikasi Logistic Regression menggunakan bahasa pemrograman Python menggunakan komputer dengan spesifikasi Intel Core i7, RAM 16 GB SSD NVMe 512 GB Windows 10. Adapun tahap penelitian dapat dilihat pada Gambar 1.



Gambar 1 Tahap Penelitian

Data yang digunakan adalah kumpulan teks umpan balik mahasiswa mengenai kompetensi kepribadian dosen di perguruan tinggi. Dataset ini terdiri dari 6.112 teks yang telah diberi label sentimen positif dan negatif, yang kemudian dibagi menjadi data pelatihan, validasi, dan pengujian. Dataset umpan balik kompetensi kepribadian dosen terdiri dari 6.112 data teks yang terbagi menjadi 3.700 data berlabel sentimen positif dan 2.412 data berlabel sentimen negatif. Data tersebut selanjutnya dibagi menjadi tiga folder, yaitu 4.278 data untuk pelatihan (70%), 611 data untuk validasi (10%), dan 1.223 data untuk pengujian (20%). Dalam pembagian ini, data sentimen positif tersebar sebanyak 2.583 di pelatihan, 367 di validasi, dan 724 di pengujian, sedangkan data sentimen negatif terdiri dari 1.695 di pelatihan, 244 di validasi, dan 499 di pengujian dengan komposisi yang dijelaskan pada Tabel 1.

Tabel 1 Dataset teks kompetensi kepribadian

Kolom	Jumlah Data	Keterangan
Sentimen Positif	3.700	Jumlah baris data kelas positif
Sentimen Negatif	2.412	Jumlah baris data kelas negatif
Pelatihan (70%)	4.278	Jumlah data untuk pelatihan
Validasi (10%)	611	Jumlah data untuk validasi
Pengujian (20%)	1.223	Jumlah data untuk pengujian

Teks umpan balik kompetensi kepribadian dosen dibersihkan dengan menghilangkan karakter tidak relevan, mengubah huruf menjadi kecil, serta menghapus kata-kata umum yang tidak berkontribusi pada makna sebelum data dianalisis. Setelah dataset telah dilakukan pra-pemrosesan, setiap kata dalam umpan balik kompetensi kepribadian dosen diberikan bobot berdasarkan TF-IDF. Metode ini akan menganalisis kata-kata penting yang membedakan karakter dan kepribadian dosen dalam teks. Fitur TF-IDF yang mewakili karakteristik teks umpan balik kompetensi kepribadian dosen digunakan sebagai input untuk model Logistic Regression (LR) yang digunakan sebagai model untuk mengklasifikasikan sentimen positif atau negatif dari setiap teks. Model TF-IDF LR ini akan dievaluasi dengan metrik akurasi, precision, dan recall untuk melihat kinerja prediksi sentimen terhadap umpan balik kompetensi kepribadian dosen sedangkan evaluasi *confusion matrix* digunakan untuk mengidentifikasi jenis kesalahan klasifikasi yang terjadi.

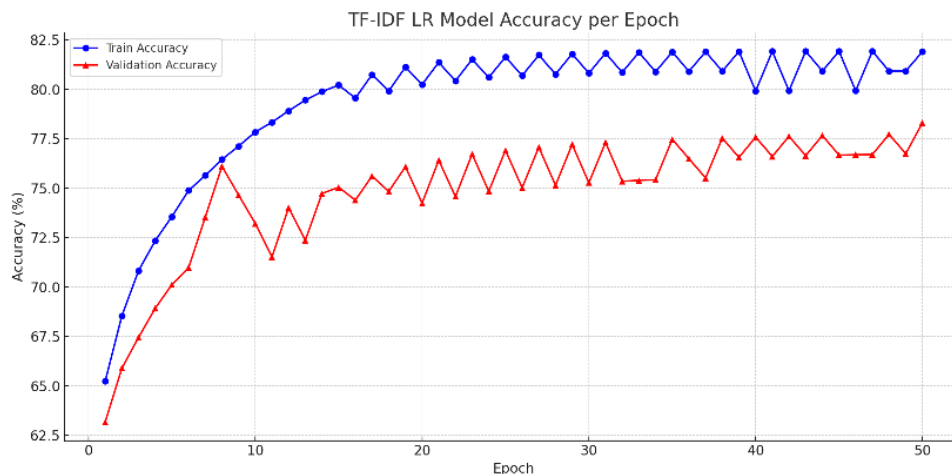
3. HASIL DAN ANALISIS

Eksperimen dalam penelitian ini dilakukan dalam tiga tahap utama yaitu pelatihan, validasi, dan pengujian menggunakan dataset umpan balik kompetensi kepribadian dosen yang terdiri dari 6.112 data teks. Data tersebut dibagi ke dalam set pelatihan sebanyak 4.278 data, validasi 611 data, dan pengujian 1.223 data. Pada setiap tahap, distribusi sentimen positif sebanyak 2.583 data di pelatihan, 367 di validasi, dan 724 di pengujian, serta sentimen negatif sebanyak 1.695 di pelatihan, 244 di validasi, dan 499 di pengujian dengan jumlah data seperti yang terlihat pada Tabel 1.

Tabel 2 Komposisi dataset eksperimen

Kategori	Pelatihan	Validasi	Pengujian	Jumlah
Positif	2.583	367	724	3.674
Negatif	1.695	244	499	2.438
Jumlah Total	4.278	611	1.223	6.112

Tujuan eksperimen ini adalah untuk mengevaluasi nilai akurasi model klasifikasi teks berbasis TF-IDF dan LR dalam mengklasifikasikan sentimen umpan balik kompetensi kepribadian dosen selama 50 epoch pelatihan. Dengan menganalisis akurasi pada data pelatihan dan validasi maka dapat diketahui kinerja model dalam mempelajari dan menggeneralisasi dataset eksperimen. Grafik akurasi model TF-IDF dan LR pada tahap pelatihan dan validasi dapat dilihat pada Gambar 2.



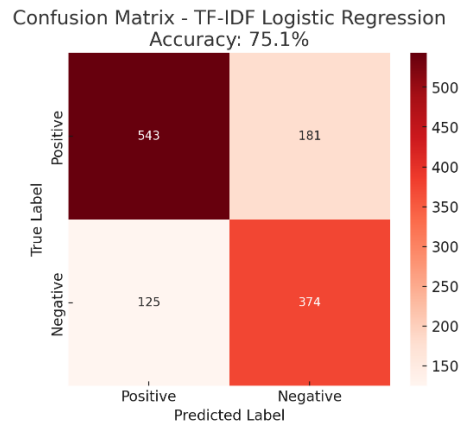
Gambar 2 Akurasi TF-IDF LR per epoch

Hasil eksperimen menunjukkan peningkatan akurasi yang signifikan sejak epoch awal hingga akhir. Akurasi pada data pelatihan meningkat secara bertahap dari sekitar 65,23% pada epoch pertama hingga mencapai 81,90% pada epoch ke-50, menandakan model berhasil mempelajari pola-pola dalam data dengan baik. Sementara itu, akurasi validasi yang mencerminkan kemampuan generalisasi model juga mengalami peningkatan dari 63,17% menjadi 78,30%, meskipun fluktuasi kecil terlihat di beberapa epoch. Perbedaan nilai akurasi antara pelatihan dan validasi yang relatif stabil menunjukkan model tidak mengalami overfitting yang signifikan, sehingga hasil klasifikasi sentimen kompetensi kepribadian dosen dapat dipercaya dan akurat. Evaluasi akurasi model TF-IDF LR per epoch pada tahap pelatihan dan validasi dapat dilihat pada Tabel 2.

Tabel 3 Nilai akurasi TF-IDF LR per epoch

Epoch	Akurasi Train (%)	Akurasi Validation (%)
1	65.23	63.17
2	68.54	65.89
3	70.82	67.45
4	72.34	68.92
5	73.56	70.12
6	74.88	70.97
7	75.64	73.53
8	76.45	76.10
9	77.12	74.65
10	77.84	73.21
...	...	...
41	81.92	76.60
42	79.92	77.62
43	81.92	76.64
44	80.92	77.66
45	81.92	76.67
46	79.92	76.69
47	81.92	76.70
48	80.92	77.72
49	80.92	76.74
50	81.90	78.30

Pada tahap pengujian, evaluasi menggunakan confusion matrix untuk menggambarkan performa model klasifikasi TF-IDF LR dalam mengklasifikasikan sentimen umpan balik kompetensi kepribadian dosen pada dataset pengujian. Pada matriks ini, baris mewakili kelas sebenarnya (*true label*) dan kolom menunjukkan hasil prediksi model (*predicted label*). Model berhasil mengklasifikasikan dengan benar 543 data sebagai sentimen positif (*true positive*) dan 374 data sebagai sentimen negatif (*true negative*). Namun, terdapat 181 data positif yang salah diklasifikasikan sebagai negatif (*false negative*) serta 125 data negatif yang salah diklasifikasikan sebagai positif (*false positive*). Model confusion matrix dari model TF-IDF LR pada tahap pengujian dapat dilihat pada Gambar 3.



Gambar 3 Confusion matrix model TF-IDF LR

Dari hasil confusion matrix, model TF-IDF-LR berhasil mendapatkan akurasi pengujian sebesar 75.1%, dengan beberapa kesalahan klasifikasi yang perlu diperhatikan, terutama pada *false negative* dan *false positive*. Kondisi *false negative* berarti model TF-IDF-LR gagal mengenali beberapa umpan balik positif dan mengkategorikannya sebagai negatif, sedangkan *false positive* menunjukkan model TF-IDF-LR salah mengklasifikasikan data negatif sebagai positif. Berdasarkan confusion matrix yang diperoleh dari model TF-IDF-LR, performa klasifikasi sentimen terhadap umpan balik kompetensi kepribadian dosen dapat diukur menggunakan metrik precision, recall, dan F1-score yang dapat dilihat pada Tabel 3.

Tabel 4 Evaluasi kinerja TF-IDF-LR

Kelas	Precision	Recall	F1-score
Positif	0.81	0.75	0.78
Negatif	0.67	0.75	0.71

Untuk kelas positif, model TF-IDF-LR mendapatkan nilai precision sebesar 81%, yang berarti bahwa 81% dari data yang diprediksi sebagai positif memang benar-benar positif. Nilai *recall* untuk kelas positif adalah 75%, yang menunjukkan bahwa model TF-IDF-LR dapat berhasil memprediksi 75% dari seluruh data positif yang sebenarnya ada dalam dataset. Nilai F1-score dari model adalah TF-IDF-LR 78%, menunjukkan keseimbangan yang baik dalam mendeteksi sentimen positif.

Untuk kelas negatif, nilai precision model TF-IDF-LR lebih rendah yaitu sebesar 67%, yang berarti prediksi positif terhadap data negatif lebih tinggi dibanding kelas positif. Namun, nilai *recall* untuk kelas negatif cukup baik dengan nilai 75%, menunjukkan kemampuan model TF-IDF-LR dalam memprediksi data negatif yang sebenarnya. Nilai *f1-score* dari TF-IDF-LR untuk kelas negatif adalah 71%, yang menunjukkan performa model yang masih bisa ditingkatkan terutama dalam mengurangi kesalahan prediksi positif pada data negatif. Hasil evaluasi menunjukkan model TF-IDF-LR harus dilakukan perbaikan terutama meningkatkan ketepatan klasifikasi kelas negatif dengan cara melakukan evaluasi kembali dataset kelas negatif yang digunakan.

4. KESIMPULAN

Penelitian ini mengembangkan dan mengevaluasi model klasifikasi teks untuk mengidentifikasi sentimen pada umpan balik kompetensi kepribadian dosen menggunakan TF-IDF dan Logistic Regression (LR). Dataset berisi 6.112 teks dengan 3.700 positif dan 2.412 negatif, dibagi menjadi pelatihan (70%), validasi (10%), dan pengujian (20%). Proses meliputi pra-pemrosesan teks, ekstraksi fitur TF-IDF, dan klasifikasi dengan LR. Model dievaluasi menggunakan akurasi, precision, recall, dan confusion matrix. Pada epoch ke-50, akurasi pelatihan mencapai 81,90% dan validasi 78,30%, sedangkan pengujian mencapai 75,1%. Untuk kelas positif, precision 81%, recall 75%, dan F1-score 78%, menunjukkan kinerja yang baik. Untuk kelas negatif, precision

67%, recall 75%, dan F1-score 71%, menandakan kebutuhan peningkatan, terutama dalam mengurangi kesalahan prediksi positif. Pada penelitian selanjutnya adalah eksperimen perbaikan pada klasifikasi kelas negatif dengan meninjau ulang dataset negatif.

UCAPAN TERIMA KASIH

Terima kasih kepada Lembaga Riset dan Pengabdian kepada Masyarakat (LRPM) Universitas Dian Nusantara (UNDIRA) yang telah mendanai penelitian ini melalui skema penelitian internal.

REFERENSI

- [1] H. Noprisson, E. Ermatita, A. Abdiansah, V. Ayumi, M. Purba, and H. Setiawan, "Fine-Tuning Transfer Learning Model in Woven Fabric Pattern Classification," *Int. J. Innov. Comput. Inf. Control*, vol. 18, no. 06, p. 1885, 2022.
- [2] V. Ayumi, "Performance Evaluation of Support Vector Machine Algorithm for Human Gesture Recognition," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 7, no. 6, pp. 204–210, 2020.
- [3] A. Ratnasari, Y. Jumaryadi, and G. Gata, "Sistem Pakar Deteksi Penyakit Ginekologi Menggunakan Metode Forward Chaining," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 3, no. 5, pp. 321–327, 2023.
- [4] M. Purba, S. D. Asri, V. Ayumi, U. Salamah, and L. Iryani, "Klasifikasi Dataset Teks Pengaduan Masyarakat Terhadap Pemerintah di Sosial Media Menggunakan Logistic Regression," *JSAI (Journal Sci. Appl. Informatics)*, vol. 7, no. 1, pp. 78–83, 2024.
- [5] A. K. I. Syahwani and A. Soeyono, "The effect of lecturer competence in online learning methods on student college satisfaction during the pandemic," *Procedia Soc. Sci. Humanit.*, vol. 1, pp. 239–250, 2021.
- [6] A. Amoozegar, M. Abdelmagid, and T. Anjum, "Course satisfaction and perceived learning among distance learners in Malaysian research universities: The impact of motivation, self-efficacy, self-regulated learning, and instructor immediacy behaviour," *Open Learn. J. Open, Distance e-Learning*, vol. 39, no. 4, pp. 387–413, 2024.
- [7] T. Salsabila and O. Usman, "Influence of Student Motivation, Student Learning Facilities and Lecturer Teaching Skills Towards Learning Interest," *Student Learn. Facil. Lect. Teach. Ski. Towar. Learn. Interes. (January 15, 2021)*, 2021.
- [8] T. Shaik *et al.*, "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," *Ieee Access*, vol. 10, pp. 56720–56739, 2022.
- [9] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study," *Appl. Sci.*, vol. 11, no. 9, p. 3986, 2021.
- [10] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, "Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review," *BMJ Heal. Care Informatics*, vol. 28, no. 1, p. e100262, 2021.
- [11] Y. A. Permana, I. Setiawan, and F. Diani, "Enhancing Sundanese News Articles Classification: A Comparative Study of Models and Feature Extraction Techniques," *Tech-E*, vol. 8, no. 2, pp. 109–124, 2025.
- [12] D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," *Bull. Electr. Eng. Informatics*, vol. 10, no. 5, pp. 2780–2788, 2021.
- [13] V. D. Antonio, S. Efendi, and H. Mawengkang, "Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent," *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, pp. 1367–1373, 2022.
- [14] I. Sari and R. Kosasih, "Prediction types of legal cases in Indonesia using the TF-IDF method and the KNN algorithm," in *AIP Conference Proceedings*, 2023, vol. 2689, no. 1.
- [15] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 165–173, 2023.