

PERBANDINGAN KINERJA METODE PEMBELAJARAN MESIN UNTUK ANALISIS SENTIMEN ULASAN FILM

Eugene Vincent Arends

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
email : eugene.535210082@stu.untar.ac.id

ABSTRACT

Analisis dari sentimen dapat membantu pembuat film dan studio membentuk strategi marketing, menilai kualitas film, analisis kompetisi, dan menganalisa tren yang lebih luas dalam industri film. Sentimen analisis ini dilakukan pada 50.000 ulasan film di IMDB dan menggunakan metode K-Nearest Neighbor, Naïve Bayes, dan Random Forest untuk mengklasifikasi positif atau negatifnya sentiment sebuah ulasan. Ulasan dalam format teks diubah menjadi representasi vector dengan menggunakan TF-IDF (term frequency-inverse document frequency) vectorizer. K-Nearest Neighbor (K-NN) menghasilkan akurasi terbaik 74%, Naïve Bayes (NB) menghasilkan akurasi 85%, dan Random Forest (RF) menghasilkan akurasi 85%. Naïve Bayes dan Random Forest mendapatkan hasil dengan akurasi terbaik. Naïve Bayes memprediksi ulasan negatif yang lebih akurat, sedangkan Random Forest memprediksi ulasan positif yang lebih akurat.

Key words

Analisis Sentimen, Machine Learning, K-Nearest Neighbor, Naïve Bayes, Random Forest

1. Pendahuluan

Pada zaman modern ini, manusia semakin terhubung satu sama lain. Setiap orang memiliki opini tentang suatu hal, dan lebih sering dari tidak, akan mengutarakannya kepada teman, keluarga, kerabat, atau bahkan orang asing. Internet telah semakin mempererat hubungan manusia di seluruh dunia [1], dan setiap opini dan ulasan sekarang dapat dilihat oleh orang-orang di seluruh penjuru dunia.

Sentiment Analysis (analisis sentimen) atau *Opinion Mining* dalam pembelajaran mesin adalah upaya oleh mesin untuk mengerti pendapat penulis mengenai sebuah objek, jasa, organisasi, individu, isu, kegiatan, topik, dan atributnya dari sebuah bagian atau tubuh teks [2]. Analisis sentimen dapat membantu marketer untuk memahami opini dan ulasan yang diberikan oleh konsumen dan memudahkan perubahan dari jasa atau produk yang disediakan [3].

Analisis sentimen dapat digunakan untuk berbagai hal, seperti analisis sentimen dalam respons pandemik COVID-19 [4] [5] [6] [7], edukasi [8] [9], politik [10], dan berbagai aplikasi lainnya. Analisis sentimen juga dapat diterapkan untuk ulasan yang diberikan oleh pembeli untuk mengukur kepuasan pembeli [11]. Dunia perfilman, sebagai bagian dari dunia kesenian, adalah sebuah industri yang sangat kenal dengan opini subjektif penonton. Opini dan ulasan yang diberikan oleh penonton dapat menentukan suksesnya sebuah film. Maka dari itu, sentimen dari sebuah film dapat membentuk strategi yang akan digunakan oleh pembuat film dan studio dalam pemasaran, promosi, dan keputusan bisnis lainnya.

Internet Movie Database (IMDB) adalah sebuah situs informasi untuk film, series televisi, podcast, *video game*, dan sebagainya. IMDB adalah salah satu situs populer untuk mencari informasi mengenai konten hiburan. Data untuk setiap konten diisi oleh para pengguna IMDB. Selain untuk informasi, IMDB juga menyediakan sistem ulasan untuk setiap konten tersebut.

Dalam penelitian ini, beberapa metode pembelajaran mesin akan diterapkan untuk analisis sentimen ulasan film. Ulasan dibagi menjadi sentimen ‘positif’ atau ‘negatif’. Penelitian ini bertujuan untuk membandingkan kinerja metode pembelajaran mesin dengan membandingkan akurasi prediksi dari setiap model.

2. Tinjauan Literatur

K. Garcia dan B. Lilian (2021) melakukan analisis sentimen dan identifikasi topik pada respons pandemi COVID-19 di media sosial Twitter untuk pengguna Amerika Serikat dan Brazil, dua negara yang terdampak besar dari pandemi. Penelitian ini menemukan bahwa sentimen yang dikemukakan sama atau seringkali jauh lebih banyak dibanding sentimen positif untuk semua topik, berbeda dengan studi-studi sebelumnya yang menemukan mayoritas sentimen yang positif. Para peneliti berhipotesa bahwa sentimen berubah seiring dengan memburuknya situasi pandemi [4].

L. Yang et al (2020) melakukan analisis sentimen menggunakan model analisis sentimen baru yang dinamakan SLCABG, sebuah gabungan dari *Sentiment*

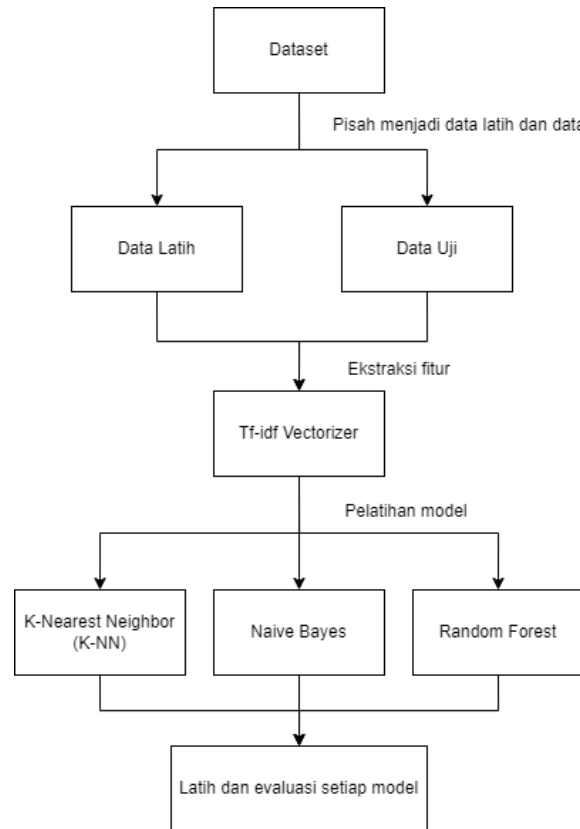
Lexicon (SL), Convolutional Neural Network (CNN), dan attention-based Bidirectional Gated Recurrent Unit (BiGRU). Model SLCABG ini menunjukkan peningkatan kinerja dibanding dengan performa model secara individu (CNN, CNN dan Attention Layer, BiGRU, BiGRU dan Attention Layer) serta model analisis sentiment analisis lainnya (Naïve Bayes dan Support Vector Machine), dengan akurasi sebesar 93.5% [11].

R. S. Jagdale, V. S. Shirsat, dan S. N. Deshmukh (2019) melakukan analisis sentimen pada ulasan produk elektronik. Produk elektronik dibagi menjadi beberapa tipe produk, seperti ‘Camera’, ‘Laptops’, ‘Mobile phones’, ‘Tablets’, ‘TVs’, dan ‘Video surveillance’. Klasifikasi dilakukan dengan model Naïve Bayes (NB) dan Support Vector Machine (SVM). Produk dengan akurasi terendah untuk NB adalah ‘TVs’ dengan akurasi 90.16% dan ‘Camera’ adalah produk tertinggi untuk model NB dengan akurasi 98.17. Untuk SVM, produk dengan akurasi klasifikasi terendah adalah ‘Video surveillance’ dengan 79.43% dan tertinggi adalah ‘Camera’ dengan akurasi 93.54%. Dapat disimpulkan bahwa metode Naïve Bayes adalah model terbaik untuk kumpulan data yang digunakan oleh peneliti [12].

H. M. Ahmed et al. (2021) melakukan analisis sentimen pada ulasan makanan. Kumpulan data sejumlah 82007 ulasan negatif dan 486447 ulasan positif diseimbangkan dengan metode Random Undersampling sehingga ulasan yang digunakan berjumlah 82007 untuk ulasan positif dan negatif. Tiga algoritma digunakan, yaitu Linear Support Vector Classifier (LinearSVC), Naïve Bayes (NB), dan Logistic Regression (LR). Penelitian ini menemukan bahwa algoritma NB adalah algoritma akurasi rendah dan kecepatan tertinggi dengan akurasi 83.43% dan lama kerja sekitar 21 detik total. Algoritma LinearSVC adalah algoritma akurasi tinggi dan kecepatan rendah dengan akurasi 88.38% dan lama kerja sekitar 347 detik. Dan terakhir adalah LR dengan akurasi dan kecepatan yang diantara keduanya dengan akurasi 87.38% dan lama kerja sekitar 41 detik total [13].

3. Metode Penelitian

Penelitian ini dilaksanakan sesuai dengan alur yang digambarkan pada diagram Gambar 1 berikut.



Gambar 1 Alur eksperimen

3.1 Dataset

Sentimen yang akan digunakan sebagai data uji dan data latih dalam penelitian ini adalah ulasan film dari situs Internet Movie Database (IMDB). Kumpulan data yang digunakan memiliki 50.000 ulasan film yang telah dibersihkan dan diberikan label sentimen positif atau negatif. Dataset ini dapat disebut seimbang karena kedua kelas tersebar merata, seperti ditunjukkan pada Tabel 1.

Tabel 1 Pembagian kelas dataset

Kelas	Jumlah
Positif	25000
Negatif	25000

3.2 Data splitting

Data dibagi menjadi data yang akan digunakan untuk melatih model (data latih) dan data yang akan digunakan untuk mengetes akurasi model (data uji). Data dipisah 80% data latih dan 20% data uji.

3.3 TF-IDF

TF-IDF (singkatan dari term frequency-inverse document frequency) adalah sebuah teknik pembobotan yang umum digunakan untuk mengevaluasi seberapa

pentingnya sebuah kata dalam sebuah teks [14]. TF-IDF merupakan gabungan dari dua konsep: *term frequency* dan *inverse document frequency*.

Term Frequency (TF) adalah sebuah metrik yang mengukur seringnya (frekuensi) sebuah kata (*term*) muncul dalam sebuah teks atau dokumen. Rumus TF dapat dijabarkan [14] dalam rumus (1).

$$TF = \frac{n_{t,d}}{\sum_k n_{k,d}} \tag{1}$$

TF adalah *term frequency*, $n_{t,d}$ adalah jumlah munculnya istilah t dalam dokumen d , dan $\sum_k n_{k,d}$ adalah total jumlah semua istilah k yang digunakan dalam dokumen d .

Inverse Document Frequency (IDF) adalah sebuah metrik yang mengukur keunikan sebuah kata dalam semua teks/dokumen. Rumus IDF dapat dijabarkan [15] [16] [17] dalam rumus (2).

$$IDF = \log\left(\frac{D}{df_t}\right) \tag{2}$$

IDF adalah *inverse document frequency*, D adalah jumlah dokumen atau teks, dan df_t adalah jumlah dokumen yang mengandung istilah t .

Term frequency-inverse document frequency diperoleh dengan mengalikan nilai TF dan IDF [15] [16] [17] dan dapat dijabarkan dalam rumus (3).

$$W_{t,d} = TF_t \times IDF_t \tag{3}$$

$W_{t,d}$ adalah TF-IDF atau bobot istilah t dalam dokumen d , TF_t adalah jumlah kemunculan istilah t , dan IDF_t adalah *inverse document frequency* yang mengandung istilah t . $W_{t,d}$ dapat dijabarkan lagi menjadi rumus (4).

$$W_{t,d} = \frac{n_{t,d}}{\sum_k n_{k,d}} \times \log\left(\frac{D}{df_t}\right) \tag{4}$$

$W_{t,d}$ adalah TF-IDF atau bobot istilah t dalam dokumen d , $n_{t,d}$ adalah jumlah munculnya istilah t dalam dokumen d , $\sum_k n_{k,d}$ adalah total jumlah semua istilah k yang digunakan dalam dokumen d , D adalah jumlah dokumen atau teks, dan df_t adalah jumlah dokumen yang mengandung istilah t .

TF-IDF menaruh bobot besar kepada istilah-istilah penting yang sering digunakan dalam teks, sedangkan istilah-istilah umum yang ditemukan di semua teks diberi bobot kecil. Hal ini berarti *stop words*, kata-kata yang memiliki fungsi tata bahasa kalimat tetapi tidak memiliki makna, secara matematis dihiraukan oleh TF-IDF [17].

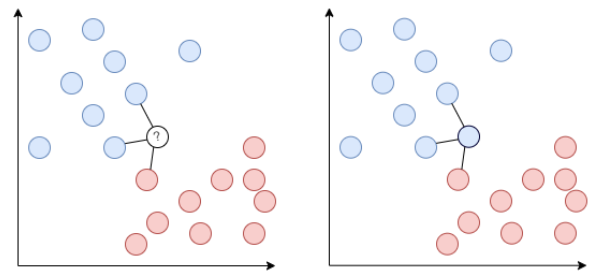
3.4 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor adalah sebuah metode pembelajaran mesin yang digunakan untuk

mengklasifikasi sebuah objek berdasarkan kedekatan objek dengan sejumlah k tetangganya [18]. Dalam kata lain, kelas objek yang ingin diprediksi ditentukan oleh mayoritas kelas tetangganya [19]. Kedekatan ini dihitung dengan rumus *Euclidean distance*, yang dijabarkan dalam rumus (5).

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

Dengan n adalah jumlah dimensi, dan x_i serta y_i sebagai dua titik koordinat yang ingin dicari jarak *Euclidean*-nya. k objek terdekat akan menentukan kelas dari objek yang dicari, seperti ditunjukkan pada Gambar 2.



Gambar 2 Diagram K-Nearest Neighbor dengan $k = 3$

3.5 Naive Bayes (NB)

Algoritma *Naive Bayes* adalah metode pembelajaran yang memprediksi kelas objek dengan memodelkan distribusi, mengikuti asumsi bahwa data mengikuti sebuah distribusi probabilitas tertentu [20].

Diambil dari nama ahli matematika abad ke-18, Thomas Bayes, algoritma *Naive Bayes* mengikuti teorema Bayes [21] [22], yang dijabarkan dalam rumus (6).

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \tag{6}$$

Dimana $P(A|B)$ adalah peluang terjadinya kejadian A pada saat kejadian B terjadi, $P(A)$ adalah peluang terjadinya A , $P(B|A)$ adalah peluang terjadinya kejadian B pada saat kejadian terjadi, dan $P(B)$ adalah peluang terjadinya B .

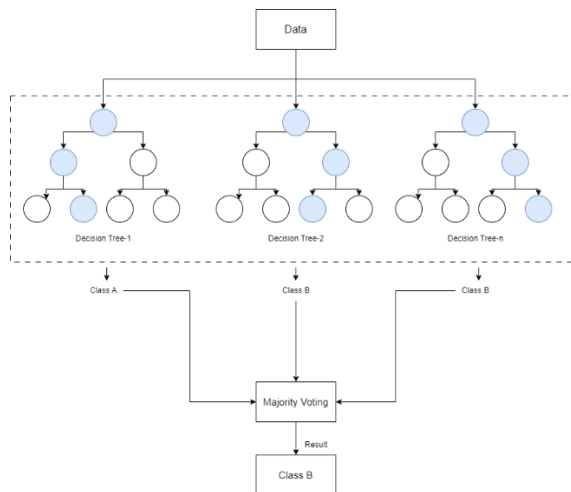
Dalam eksperimen yang dilakukan dalam penelitian ini, algoritma *Naive Bayes* yang digunakan adalah varian *Bernoulli Naive Bayes*, yang sesuai untuk klasifikasi biner [22], dan dapat dijabarkan dalam rumus (7).

$$p(x|C_k) = \prod_{i=1}^n p_{ki}^{x_i} (1 - p_{ki})^{(1-x_i)} \tag{7}$$

Dengan p_{ki} adalah probabilitas dari kelas C_k membuat istilah x_i .

3.6 Random Forest (RF)

Random Forest adalah metode pembelajaran mesin yang menggabungkan prediksi dari beberapa model untuk melakukan klasifikasi. Metode pembelajaran tipe ini disebut juga dengan *ensemble learning*. Random Forest menggunakan beberapa *Decision Tree* dengan data acak, lalu mengambil keputusan menurut keputusan mayoritas *Decision Tree* tersebut [18]. Algoritma ini dapat divisualisasi seperti pada Gambar 3.



Gambar 3 Diagram Random Forest

3.7 Evaluasi

Metode evaluasi yang digunakan dalam penelitian ini adalah akurasi. Akurasi adalah ketepatan prediksi kelas oleh sebuah model. Evaluasi lain yang dilihat adalah *True Positive* dan *True Negative*, hasil prediksi yang sesuai dengan data asli, serta *False Positive* dan *False Negative*, yaitu hasil prediksi yang tidak sesuai dengan data asli. *Positive* dan *Negative* merujuk pada prediksi mesin. *True Positive* menunjukkan bahwa sentimen prediksi dan sentimen sesungguhnya dari sebuah ulasan adalah positif, dan *False Negative* menunjukkan bahwa sentimen prediksi adalah negatif, tetapi sentimen sesungguhnya adalah positif.

4. Hasil dan Pembahasan

Setelah data latih dan data uji divektorisasi dengan TF-IDF, data latih digunakan untuk melatih ketiga metode pembelajaran: *K-Nearest Neighbor*, *Bernoulli Naïve Bayes*, dan *Random Forest*. Ketiga model lalu melakukan prediksi sentimen pada data uji.

4.1 K-Nearest Neighbor (K-NN)

Dalam eksperimen ini, evaluasi kinerja model K-NN dilakukan dengan merancang sepuluh model K-NN yang berbeda, masing-masing dikonfigurasi dengan nilai k yang beragam. Rentang nilai k yang dipilih melibatkan

penggunaan nilai k dari 1 hingga 10. Melalui proses ini, kesepuluh model tersebut dibangun dan dijalankan pada *dataset*.

Hasil dari evaluasi masing-masing model tercatat dalam Tabel 2.

Tabel 2 Hasil Prediksi Model *K-Nearest Neighbor*

Hasil	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
<i>True Positive</i>	3504	3965	3606	3965	3670
<i>True Negative</i>	3800	3338	3736	3408	3744
<i>False Positive</i>	1161	1623	1225	1553	1217
<i>False Negative</i>	1535	1074	1433	1074	1369
Akurasi	0.73	0.73	0.73	0.74	0.74

Berdasarkan hasil prediksi yang ditunjukkan pada Tabel 2, nilai k genap menghasilkan prediksi yang cenderung negatif, dan nilai k ganjil menghasilkan prediksi yang cenderung positif. Selain itu, semakin besar nilai k , semakin seimbang dan akurat hasil prediksi model K-NN, yang ditunjukkan oleh nilai akurasi yang semakin meningkat.

4.2. Naïve Bayes (NB)

Dalam eksperimen ini, sebuah model *Naïve Bayes* dilatih. Algoritma yang digunakan *Bernoulli Naïve Bayes*, sebuah varian dari *Naïve Bayes* yang baik digunakan untuk klasifikasi kelas biner [22]. Hasil prediksi model *Bernoulli Naïve Bayes* ditunjukkan pada Tabel 3.

Tabel 3 Hasil Prediksi Model *Bernoulli Naïve Bayes*

Hasil	Jumlah
<i>True Positive</i>	4304
<i>True Negative</i>	4184
<i>False Positive</i>	777
<i>False Negative</i>	735
Akurasi	0.85

Berdasarkan hasil prediksi yang ditunjukkan pada Tabel 3, model *Bernoulli Naïve Bayes* menghasilkan prediksi yang lebih akurat dibanding model *K-Nearest Neighbor*. Dapat dilihat dari angka *True Positive* dan *False Positive* yang lebih tinggi dari *True Negative* dan *False Negative*, model ini menghasilkan prediksi yang lebih sedikit cenderung ke prediksi positif.

4.3 Random Forest (RF)

Dalam eksperimen ini, sebuah model *Random Forest* dilatih. Hasil prediksi model *Random Forest* ditunjukkan pada Tabel 4.

Tabel 4 Hasil Prediksi Model *Random Forest*

Hasil	Nilai
<i>True Positive</i>	4250
<i>True Negative</i>	4254
<i>False Positive</i>	707
<i>False Negative</i>	789
Akurasi	0.85

Berdasarkan hasil yang ditunjukkan oleh Tabel 4, model *Random Forest* menghasilkan prediksi yang setara dengan model *Naïve Bayes*. Prediksi memiliki sedikit kecenderungan ke prediksi negatif.

5. Kesimpulan

Penelitian ini menerapkan beberapa metode pembelajaran mesin untuk mengklasifikasi ulasan film sebagai ulasan yang bersentimen positif atau negatif. 3 algoritma yang dipilih dalam penelitian ini adalah *K-Nearest Neighbor*, *Bernoulli Naïve Bayes*, dan *Random Forest*. Algoritma yang memiliki akurasi terbaik adalah *Bernoulli Naïve Bayes* dan *Random Forest* dengan nilai 0.85, lalu diikuti oleh model *K-Nearest Neighbor*. Akurasi *K-Nearest Neighbor* (K-NN) tertinggi adalah model K-NN dengan nilai k 9 dan 10, dengan akurasi 0.74, dan model K-NN dengan akurasi terendah adalah model K-NN dengan nilai k 6 dan 7 dengan akurasi 0.73. Model K-NN dengan nilai k ganjil lebih cenderung memprediksi sentimen dari sebuah ulasan sebagai positif, dan sentiment negatif untuk nilai k genap. Model *Bernoulli Naïve Bayes* sedikit lebih cenderung memprediksi sentimen ulasan sebagai positif dan model *Random Forest* sedikit lebih cenderung memprediksi sentiment ulasan sebagai negatif.

REFERENSI

- [1] M. S. Basarslan and F. Kayaalp, "Sentiment Analysis with Machine Learning Methods on Social Media," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, no. 3, pp. 5-15, 2020.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*, Springer Nature, 2022.
- [3] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, 2021.
- [4] K. Garcia and B. Lilian, "Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the USA," *Applied Soft Computing*, vol. 101, p. 107057, 2021.
- [5] M. Singh, A. K. Jakhar and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 33, 2021.
- [6] G. Barkur, V. and G. B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India," *Asian Journal of Psychiatry*, vol. 51, p. 102089, 2020.
- [7] K. H. Manguri, R. N. Ramadhan and P. R. M. Amin, "Twitter sentiment analysis on worldwide COVID-19 outbreaks," *Kurdistan Journal of Applied Research*, pp. 54-65, 2020.
- [8] T. T. Sasidhar, B. Premjit and S. K. Padannayil, "Emotion detection in hinglish (hindi+ english) code-mixed social media text.," *Procedia Computer Science*, vol. 171, pp. 1346-1352, 2020.
- [9] J. Zhou and J.-m. Ye, "Sentiment analysis in education research: a review of journal publications," *Interactive learning environments*, vol. 31, no. 3, pp. 1252-1264, 2023.
- [10] L. Yue, W. Chen, X. Li, W. Zuo and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, p. 617-663, 2019.
- [11] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 9, pp. 23522-23530, 2020.
- [12] R. S. Jagdale, V. S. Shirsat and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in *Cognitive Informatics and Soft Computing: Proceeding of CISC 2017*, Springer, 2019, pp. 639-647.
- [13] H. M. Ahmed, M. J. Awan, N. S. Khan, A. Yasin and H. M. Faisal Shehzad, "Sentiment analysis of online food reviews using big data analytics," *Elementary Education Online*, vol. 20, no. 2, pp. 827-836, 2021.
- [14] H. Yuan, Y. Tang, W. Sun and L. Liu, "A detection method for android application security based on TF-IDF and machine learning," *PLoS ONE*, vol. 15, no. 9, p. e0238694, 2020.
- [15] D. E. Cahyani and I. Patasik, "Performance comparison of TF-IDF and Word2Vec models for emotion text classification," *Performance comparison of TF-IDF and Word2Vec models for*, vol. 10, no. 5, pp. 2790-2788, 2021.
- [16] M. Das, P. Alphonse and S. Kamalanathan, "A Comparative Study on TF-IDF feature Weighting Method and its Analysis," *arXiv preprint arXiv*, p. 2308.04037, 2023.
- [17] A. Mee, E. Homapour, F. Chiclana and O. Engel, "Sentiment Analysis using TF-IDF Weighting of UK MPs' Tweets on Brexit," *Knowledge-Based System*, vol. 228, p. 107238, 2021.
- [18] E. Y. Boateng, J. Otoo and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," *Journal of Data*

Analysis and Information Processing, vol. 8, no. 4, pp. 341-357, 2020.

- [19] M. K. Mahinda, L. Pasi and C. Mikael, "A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean," *Pattern Recognition Letters*, vol. 140, pp. 172-178, 2020.
- [20] I. Wikramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft Computing*, vol. 25, no. 3, pp. 2277-2293, 2021.
- [21] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 7, no. 2, pp. 88-91, 2019.
- [22] M. Artur, "Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features," *Procedia Computer Science*, vol. 190, pp. 564-570, 2021.