

Analisis Sentimen Pada Agen Perjalanan *Online* Menggunakan *Naïve Bayes* dan *K-Nearest Neighbor* (*Sentiment Analysis of Online Travel Agent Using Naïve Bayes and K-Nearest Neighbor*)

Eka Wahyu Sholeha^[1], Selviana Yunita^{[2]*}, Rifqi Hammad^[3], Veny Cahya Hardita^[4], Kaharuddin^[5]

^[1]Teknologi Rekayasa Komputer Jaringan, Politeknik Tanah Laut

^[2]Sistem Informasi, Universitas Darwan Ali

^[3]Rekayasa Perangkat Lunak, Universitas Bumigora

^[4]Teknik Informatika, STMIK Palangka Raya

^[5]Teknik Perangkat Lunak, Universitas Universal

E-mail: ekwahyus@politala.ac.id, selviana.yunita@unda.ac.id, rifqi.hammad@universitasbumigora.ac.id,
venycahya@stmikplk.ac.id, kahar.olvaldo@gmail.com

KEYWORDS:

Sentiment Analysis, Online Travel Agent, Naïve Bayes, K-Nearest Neighbor

ABSTRACT

Social media has impact for decision maker to get more insights broadly. Including for online travel agent company, where customer's interest to use online travel agent for their chosen agent will grows along with the high number of customer's satisfaction. As a one of the most important point in distribution, company provides a platform that reliable and effective to purchase a trip and share information of their experience through Online travel agent. It is important to know how consumer considerate which one the online travel agent they choose. One of their method is looking at the reviews. Facebook is one of social media that provide numerous reviews through comments sections. The research purposes are twofold, algorithm comparison and reveal the effect of uppercase as well as punctuation mark. The accuracy comparison between Naïve Bayes and K-Nearest Neighbor is provided against the datasets. This research collects the data from user comments on Facebook about the biggest three online travel agents in Indonesia. We classify the comments into three categories which are positive, negative, and neutral. The result of this research is found that K-Nearest Neighbor have slightly higher accuracy than the Naïve Bayes. Moreover, lowercase text without punctuation achieves better accuracy for both of algorithm.

KATA KUNCI:

Analisis Sentimen, Agen Perjalanan Online, Naïve Bayes, K-Nearest Neighbor

ABSTRAK

Sosial media memiliki dampak yang besar bagi para pembuat keputusan untuk mendapatkan informasi secara luas. Termasuk bagi perusahaan agen perjalanan online, dimana ketertarikan pelanggan untuk pemilihan agen perjalanan akan bertumbuh sejalan dengan tingginya kepuasan pelanggan. Sebagai salah satu poin terpenting dalam distribusi, perusahaan menyediakan media yang terpercaya dan efektif untuk membeli tiket perjalanan dan berbagi informasi tentang pengalamannya dalam menggunakan agen perjalanan online. Sangat penting bagi perusahaan untuk mengetahui bagaimana pelanggan mempertimbangkan agen perjalanan online yang ingin dipilih. Salah satu metode yang digunakan adalah melihat ulasan yang diberikan. Facebook merupakan salah satu sosial media yang menyediakan ulasan yang tak terhitung pada bagian komentar. Penelitian ini bertujuan ganda, yaitu untuk membandingkan algoritma, serta menemukan efek dari huruf kapital serta tanda baca. Perbandingan akurasi antara Naïve Bayes dan K-Nearest Neighbor diberikan terhadap kumpulan data. Penelitian ini mengumpulkan data dari komentar para pengguna Facebook terhadap tiga agen perjalanan online terbesar di Indonesia. Komentar di kelompokkan menjadi tiga kategori, yaitu positif, negatif, dan netral. Hasil dari penelitian ini menemukan jika K-Nearest Neighbor memiliki akurasi yang sedikit lebih tinggi dibandingkan Naïve Bayes. Selain itu, penggunaan huruf kecil tanpa tanda baca memiliki akurasi yang lebih baik untuk kedua algoritma.

I. PENDAHULUAN

Big Data yang dihasilkan melalui media sosial telah menciptakan berbagai kesempatan bagi para pembuat keputusan untuk memperoleh wawasan lebih baik [1]. Sebagai salah satu poin terpenting dalam distribusi, perusahaan menyediakan media yang terpercaya dan efektif untuk membeli tiket perjalanan dan berbagi informasi tentang pengalaman perjalanannya dalam menggunakan agen perjalanan *online* [2]. *Daily Social* telah menyelenggarakan sebuah survei tentang agen perjalanan *online* pada tahun 2018 di Indonesia, dan hasilnya menunjukkan bahwa 71,44% responden menggunakan travel agen *online* untuk reservasi tiket pesawat [3]. Berdasarkan fakta ini, penting untuk mengetahui bagaimana pelanggan mempertimbangkan pilihan agen perjalanan *online* yang ingin digunakan. Salah satu metode yang digunakan adalah dengan melihat pada ulasan-ulasan yang ada, baik itu ulasan yang baik ataupun buruk. *Facebook* adalah salah satu media sosial yang menyediakan ulasan yang tak terhitung melalui bagian komentar dan banyak pelanggan menggunakan komentar-komentar ini untuk menilai kinerja agen perjalanan *online*. Penilaian terhadap komentar oleh pelanggan terbagi menjadi penilaian positif, negatif, dan netral.

Penilaian terhadap analisis sentimen perlu dilakukan untuk mengetahui tingkat komentar konsumen di media sosial, khususnya dalam penelitian ini melalui *Facebook*. Adopsi luas dari *web* dan media sosial melibatkan pengguna dengan memberikan komentar, ulasan, dan umpan balik. Selain itu, keuntungan yang diperoleh oleh perusahaan yaitu perusahaan dapat memberikan rancangan produk yang lebih baik, pelayanan, meningkatkan pengalaman pengguna, serta mengatur kinerja perusahaan. Keuntungan-keuntungan tersebut dapat dicapai dengan melakukan analisis terhadap masukan pengguna [4]. Analisis sentimen adalah salah satu bidang ilmu yang menggunakan *Natural Language Processing* (NLP), teks analisis, dan teknik komputasional untuk mengotomisasi ekstraksi atau klasifikasi ulasan sentimen [5]. Hal ini tersebar dalam banyak bisnis dan non bisnis aspek seperti *website*, pemasaran, informasi pelanggan, aplikasi, buku-buku, dan mempengaruhi dalam proses pengambilan

keputusan. Saat ini pengambilan keputusan banyak didasari oleh analisis sentimen [6]. Teorema Bayesian adalah dasar dari algoritma *Naïve Bayes* yang bekerja dengan asumsi independen yang tinggi dan klasifikasi probabilistik sederhana [7]. Untuk membandingkan hasil akurasi, penelitian ini juga menggunakan *K-Nearest Neighbor* karena *K-Nearest Neighbor* adalah salah satu dari algoritma umum untuk menyelesaikan masalah klasifikasi [8]. *K-Nearest Neighbor* memiliki prinsip kerja dengan cara mencari jarak paling dekat diantara data yang akan dievaluasi dengan tetangga yang ada pada data pelatihan [9], algoritma ini merupakan salah satu algoritma yang sederhana untuk memecahkan klasifikasi serta mampu memberikan hasil yang signifikan dan kompetitif [10].

Penelitian ini akan membahas tentang analisis sentimen konsumen terhadap tiga agen perjalanan *online* terbesar di Indonesia, dan pengukurannya akan berdasarkan komentar konsumen di halaman *Facebook* agen perjalanan *online* tersebut. Komentar-komentar tersebut akan diberikan label positif, negatif, atau netral. Pengujian dilakukan dengan menggunakan *Waikato Environment for Knowledge Analysis* atau yang lebih dikenal dengan nama *Weka* versi 3.9. *Weka* adalah sebuah aplikasi *opensource* yang digunakan hampir pada semua algoritma *machine learning* dengan antarmuka pengguna grafis yang berisi kumpulan algoritma dan visualisasi untuk pemodelan prediktif dan analisis data [11]. *Weka* dipilih sebagai alat pengujian karena *Weka* mendukung beberapa tugas *data mining* standar khususnya pre-prosesing data, regresi, visualisasi, *clustering*, seleksi fitur, dan klasifikasi [12]. Tujuan penelitian ini adalah untuk menemukan akurasi yang lebih baik antara algoritma *Naïve Bayes* dan *K-Nearest Neighbor*, juga untuk menemukan penggunaan huruf kecil serta tanda baca dapat mempengaruhi akurasi.

Dalam penelitian ini, tiga agen perjalanan *online* terbesar digunakan untuk menilai sentimen terhadap komentar pelanggan menggunakan metode klasifikasi *Naïve Bayes*. *Naïve Bayes* adalah metode klasifikasi mesin pembelajaran yang digunakan untuk konsep supervisi dan pembelajaran statistik.

II. METODOLOGI

A. Studi Literatur

A.1. Analisis Sentimen

Didunia, informasi tekstual secara garis besar dikategorikan menjadi dua kategori utama, yaitu fakta dan opini. Fakta adalah ekspresi objektif terhadap peristiwa, entitas, dan propertinya. Pendapat biasanya berdasarkan atas ekspresi subjek dan mendeskripsikan sentimen subjek, persetujuan, atau perasaan terhadap entitas, peristiwa dan propertinya. Analisis sentimen adalah area penelitian komputasional yang mengekstraksi pendapat antar tingkatan (positif, negatif, atau netral) dari teks atau dokumen [13][14]. Analisis sentimen juga dikenal sebagai *opinion mining*, yaitu penentuan polaritas kontekstual dalam sebuah dokumen [15]. Analisis sentimen dikategorikan kedalam dua pendekatan utama, berbasis *lexicon* dan *machine learning* [16].

Pendekatan berbasis *Lexicon* bergantung pada kamus *lexicon*. Kamus *lexicon* menampung daftar kata *lexical* dengan nilai positif atau negatif yang ditentukan [17][18]. Pendekatan *Machine learning* dikategorikan menjadi dua, *supervised learning* dan *unsupervised learning*. *Machine Learning* mengandalkan pola dalam teks dan memprediksi lebih baik untuk melatih model data [19][20]. Pada klasifikasi sentimen, *machine learning* memiliki akurasi yang lebih baik bahkan lebih cepat dibandingkan metode berbasis *Lexicon* [21][22][23].

Penelitian sebelumnya melakukan klasifikasi informasi dan mengekstraksi aspek menggunakan ontologi, kemudian mengkategorikan dalam positif, negatif, dan netral menggunakan *supervised learning*. Kemudian menggunakan Weka untuk menggunakan algoritma *machine learning*, dan penilaian efisiensi dilakukan dengan menggunakan strategi pencarian temu kembali informasi [23].

A.2. Naïve Bayes Classifier

Naïve Bayes Classifier adalah sebuah metode algoritma klasifikasi yang berdasarkan pada konsep probabilitas. Teorema Bayes adalah dasar konsep untuk metode *Naïve Bayes*. *Naïve Bayes* merupakan salah satu alat klasifikasi yang kompeten [24]. Pekerjaan terkait *Naïve Bayes* dilakukan dengan membandingkan tiga metode *machine learning* seperti *Naïve Bayes*, *Support Vector Machines* (SVM), dan *Maximum Entropy*. Penelitian ini mengkaji efisiensi pembelajaran mesin menggunakan

ulasan yang ada pada Amazon API, dan untuk melatih klasifikasi *machine learning*, digunakan unigram. Hasil penelitian menunjukkan bahwa *machine learning* bekerja dengan baik dalam unigram berbobot, dan juga SVM memiliki akurasi terbaik [25]. Wongkar melakukan analisis sentimen melalui twitter pada proses pemilihan presiden di Indonesia yang bertujuan untuk menemukan yang terbaik antara *lexicon* sentimen seperti W-WSD[26], *SentiWordNet* [27], *TextBlob* kemudian diuji dengan *supervised machine learning* seperti *Naïve Bayes* dan SVM. Hasilnya menunjukkan bahwa *TextBlob* berkinerja relatif lebih baik [28].

A.3. K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu algoritma artificial learning yang berbasis pada kesamaan yang bekerja berdasarkan tetangga terdekat karena. *K-Nearest Neighbor* merupakan algoritma yang membutuhkan pelatihan data dan nilai K yang telah ditentukan sebelumnya untuk mencari K terdekat berdasarkan jarak komputasi [29].

Penelitian terdahulu bertujuan untuk mempelajari kinerja jaringan klasifikasi K-nearest ketika menggunakan pengukuran jarak yang berbeda. Termsauk juga sebelas matrik jarak seperti *Spearman*, korelasi, *Cosine*, *Minkowski*, *Euclidean*, *Standardized euclidean*, *City block*, *Hamming*, *Jaccard*, *Chebychev*, *Mahalanobis*, dan Korelasi. Deret eksperimen dilakukan pada delapan deret syntetics *dataset* dan hasilnya menunjukkan bahwa *City Block*, *Chebychev*, *Euclidean*, *Mahalanobis*, *Minkowski*, dan *Standarized Euclidean* memberikan akurasi tertinggi dalam perhitungan jarak [30].

Tripathi menggabungkan algoritma N-gram dan jaringan K-terdekat untuk menilai analisis sentimen di *Twitter*. Algoritma N-gram berlaku untuk ekstraksi ciri, dan KNN untuk mengklasifikasikan data menjadi positif, negatif, atau netral. Hasilnya menunjukkan bahwa sistem yang diusulkan berkinerja baik dibandingkan dengan sistem yang ada dengan berbasis SVM [31].

Penelitian sebelumnya yang lain menggabungkan Algoritma *Crow search* dengan algoritma KNN. Algoritma *Crow Search* berlaku untuk pemilihan fitur dan untuk mengklasifikasikan, menggunakan teknik *K-nearest* tetangga dan TF ditambahkan untuk menghitung dan menghitung frekuensi kata. Hasilnya menunjukkan bahwa model yang diusulkan

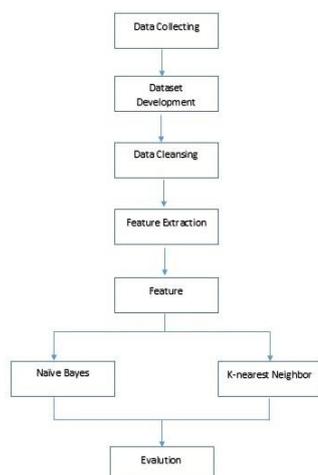
memiliki akurasi yang lebih baik daripada model KNN, dan memiliki *F-measure* yang lebih besar dibandingkan dengan KNN dan C4.5, dan dengan pemilihan fitur, meningkatkan akurasi hingga 27%, daripada KNN [32].

B. Metode Penelitian

Proses pada metodologi penelitian ini dimulai dengan mengumpulkan data yang berasal dari *Facebook*. Data dikumpulkan dari 3 fanpage agen perjalanan *online* yang ada di *Facebook*. Jumlah data yang dikumpulkan adalah 1500 komentar. Aplikasi yang digunakan untuk mengumpulkan data adalah aplikasi *facepager* [33]. Jangka waktu pengumpulan data adalah tiga bulan, yaitu selama periode september, oktober, dan november 2020.

Langkah selanjutnya adalah melakukan *dataset development* dengan memberikan label secara manual dengan label positif, negatif, dan netral. Setelah proses pemberian label selesai, proses selanjutnya adalah proses pembersihan data. Pembersihan data dilakukan dengan menghilangkan huruf kapital dan tanda-tanda baca, dan kemudian ekstraksi fitur. Fitur-fitur tersebut akan diproses dengan menggunakan *machine learning* yaitu *Naïve Bayes* dan *K-Nearest Neighbor*. Dalam penelitian ini, proses *machine learning* menggunakan Weka versi 3.9.

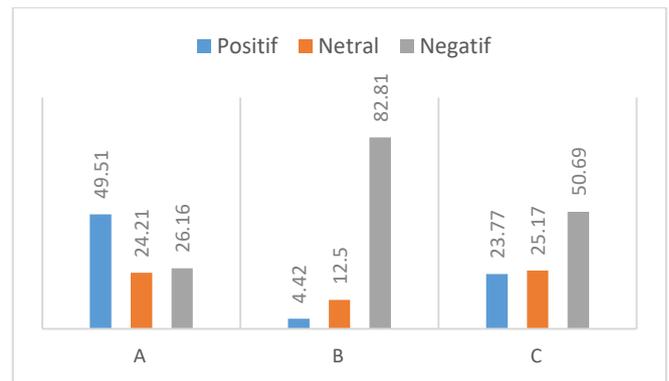
Setelah proses dengan *machine learning* selesai, langkah selanjutnya adalah evaluasi. Evaluasi dilakukan dengan dua metode, yaitu *10-Fold Cross Validation* [34] dan *Confusion Matrix* [35]. Adapun alur penelitian pada penelitian ini dapat dilihat pada Gbr 1.



Gbr 1. Alur Penelitian

III. HASIL DAN PEMBAHASAN

Setelah data dikumpulkan, langkah selanjutnya adalah pengembangan *dataset* dengan memberikan label secara manual dengan positif, negatif, atau netral. Gambar 2 menunjukkan bahwa agen perjalanan *online* A memiliki jumlah terbanyak pada komentar positif yaitu 29,51%, kedua yaitu agen perjalanan *online* C dengan 23,77% komentar positif, dan yang ketiga adalah agen perjalanan *online* B dengan 4,42% komentar positif. Nilai dari data komentar tersebut diperoleh dengan memberikan label secara manual.



Gbr 2. Diagram Jumlah Dataset

Ketika proses pengembangan *dataset* selesai, langkah selanjutnya adalah pembersihan data atau *stemming*. Proses *Stemming* adalah proses normalisasi komentar untuk mengurangi jumlah variasi data dengan arti yang mirip. Proses *stemming* diharapkan dapat meningkatkan akurasi data dan proses analisis lebih mudah. Proses *stemming* dilakukan dengan dua langkah berikut: 1) Mengeliminasi huruf kapital, sehingga diharapkan dapat mempersempit variasi kata dengan hasil yang sama namun terdeteksi berbeda karena adanya perbedaan dalam penggunaan huruf besar dan kecil. 2) Menghilangkan tanda baca sehingga membuat proses analisis lebih fokus terhadap analisis kata. Hasil dari proses *stemming* dapat dilihat pada Tabel 1.

TABEL 1
DATASET SETELAH PROSES STEMMING

Komentar	Analisis
izin share yaa	Positif
keren saya suka	Positif
buruk sekali penanganan masalahnya	Negatif
cara pesennya gimana ya	Netral
seharusnya ga usah di bikin ribet	Negatif

Komentar	Analisis
gag konsisten nih	Negatif

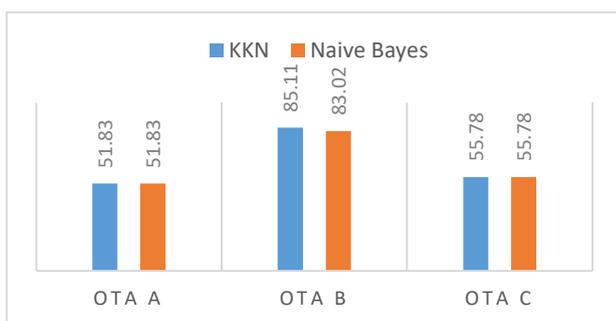
Setelah proses *stemming*, proses pengujian dilakukan dengan *Naïve Bayes* dan *K-Nearest Neighbor*. Data kemudian diuji dengan menggunakan beberapa variasi, antara lain data tanpa tanda baca, dengan huruf kecil, dan data dengan tanda baca, gabungan huruf kapital dan huruf kecil (campuran). Pengujian dilakukan dengan menggunakan Weka versi 9.2. Hasil pengujian dapat dilihat pada Tabel 2.

TABEL 2
PERSENTASE HASIL PENGUJIAN DATASET

Algoritma	Campuran	Tanpa Tanda Baca	Huruf Kecil
<i>K-Nearest Neighbor</i>	48.87%	49.34%	52.35%
<i>Naïve Bayes</i>	48.99%	49.34%	52.35%

Dengan algoritma *K-Nearest Neighbor* dan *Naïve Bayes*, terlihat bahwa akurasi tertinggi dicapai saat data seluruhnya menggunakan huruf kecil.

Proses akurasi telah selesai dilakukan pada seluruh *dataset*, langkah selanjutnya adalah menguji setiap agen perjalanan *online* menggunakan metode 10-Fold Cross Validation. Gambar 3 menunjukkan hasil dari pengujian menggunakan metode 10-Fold Cross Validation. A, B, dan C merupakan nama yang digunakan untuk merepresentasikan agen perjalanan *online* di Indonesia.



Gbr. 3 Hasil Pengujian Agen Perjalanan Online

IV. KESIMPULAN

Berdasarkan penelitian dengan 1500 data komentar dari 3 *fanpage Facebook* agen perjalanan *online*, ditemukan bahwa algoritma *K-Nearest Neighbor* memiliki akurasi yang lebih baik pada rata-rata dibandingkan *Naïve Bayes*. Penelitian juga

menunjukkan bahwa akurasi tertinggi ketika seluruh data menggunakan huruf kecil untuk kedua algoritma, dengan akurasi 52,35%. Hasil akurasi ini diperoleh dengan pengujian menggunakan Weka 3.9 menggunakan dataset dengan yang dilabeli manual sebagai *data training*.

UCAPAN TERIMA KASIH

Penelitian ini dapat dilaksanakan dengan baik berkat dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada pimpinan dan kolega kami di Politeknik Tanah Laut, Universitas Darwan Ali, Universitas Bumigora, STMIK Palangka Raya, Universitas Universal.

REFERENSI

- [1] S. . Miaha, H. . Vub, J. Gammackc, and M. McGrath, "A Big Data Analytics Method for Tourist Behaviour Analysis," *Inf. Manag.*, vol. 54, pp. 771–785, 2016.
- [2] M. . Navarro, C. . Manero, M. . Campillo, and M. P. Iglesias, *Strenght of Online Travel Agencies from the Perspective of Digital Tourist*. USA: IGI Global, 2019.
- [3] F. Zebua, "Laporan DailySocial: Survey Online Travel Agencies (OTA) 2018," *Daily Social*, 2018. <https://dailysocial.id/post/laporan-dailysocial-survey-online-travel-agencies-ota-2018> (accessed Jun. 01, 2021).
- [4] M. Ghiassi and S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach," *Expert Syst. witf Appl.*, vol. 106, pp. 197–216, 2018.
- [5] B. Agarwal, N. Mittal, P. Bansal, and S. Garg2, "Sentiment Analysis Using Common-Sense and Context Information," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–9, 2015.
- [6] J. . Mathew, G. Spencer, and Z. Andrea, "Potential applications of sentiment analysis in educational research and practice – Is SITE the friendliest conference? In: Slykhuis, D., Marks, G. (Eds.)," 2015.
- [7] R. S. Perdana and A. Pinandito, "Combining Likes-Retweet Analysis and Naïve Bayes Classifier within Twitter for Sentiment Analysis," Univesitas Brawijaya, 2018.
- [8] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD, CRISP- DM and SEMMA)," *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014.
- [9] A. G. Novianti and D. Prasetyo, "Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Prediksi Waktu Kelulusan Mahasiswa," in *Seminar Nasional APTIKOM (SEMNASTIKOM)*, 2017, pp. 108–113.
- [10] D.A.Adeniyi, Z.Wei, and Y.Yongquan, "Automated web usage data mining and recommendation system

- using K-Nearest Neighbor (KNN) classification method,” *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 90–108, 2016.
- [11] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, “Educational Data Mining and Analysis of Students’ Academic Performace Using WEKA,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, pp. 447–459, 2018.
- [12] Kou, “WEKA Packages,” *Weka*, 2016. <https://weka.sourceforge.io/packageMetaData/> (accessed Jun. 01, 2021).
- [13] M. Azam, T. Ahmed, F. Sabah, and M. . Hussain, “Feature Extraction based Text Classification using K-Nearest Neighbor Algorithm,” *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 12, 2018.
- [14] A. A, Y. A, and V. DK, “Multimodal sentiment analysis via RNN variants,” in *In IEEE international conference on big data, cloud computing, data science and engineering (BCD)*, 2019, pp. 19–23.
- [15] M. S. Bhatia, K. Sharma, and K. Bhatia, “Strategies for mining opinions: A survey,” in *Proceedings of the IEEE 2nd International Conference on Computing for Sustainable Global Development*, 2015, pp. 262–266.
- [16] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [17] S. Ahire, “A Survey of Sentiment Lexicons,” 2015.
- [18] S. Akter and M. T. Aziz, “Sentiment analysis on facebook group using lexicon-based approach,” 2016.
- [19] D. Anand and D. Naorem, “Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering,” *Procedia Comput. Sci.*, vol. 84, pp. 86–93, 2016.
- [20] M. Meire, M. Ballings, and D. Van den Poel, “The added value of auxiliary data in sentiment analysis of Facebook posts,” *Decis. Support Syst.*, vol. 89, pp. 93–112, 2016.
- [21] V. A. Rohani and S. Shayaa, “Utilizing machine learning in Sentiment Analysis: SentiRobo approach,” 2015.
- [22] B. Samal, A. K. Behera, and M. Panda, “Performance analysis of supervised machine learning techniques for sentiment analysis,” 2017.
- [23] M. S. Bhatia, K. K. B. Sharma, and P. Das, “Opinion Target Extraction with Sentiment Analysis,” *Int. J. Comput.*, vol. 17, no. 3, pp. 136–142, 2018.
- [24] L. Zhang, L. Jiang, C. Li, and G. Kon, “Two feature weighting approaches for naive Bayes text classifiers,” *Knowledge-Based Syst.*, vol. 100, pp. 137–144, 2016.
- [25] A. S. Rathor, A. Agarwal, and P. Dimri, “Comparative Study of Machine Learning Approaches for Amazon Reviews,” *Procedia Comput. Sci.*, vol. 132, pp. 1552–1561, 2018.
- [26] M. Wongkar and A. Angdresey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” 2019.
- [27] M. Husnain, M. M. S. Missen, N. Akhtar, M. Coustaty, S. Mumtaz, and V. B. S. Prasath, “A systematic study on the role of SentiWordNet in opinion mining,” *Front. Comput. Sci.*, vol. 15, 2021.
- [28] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, “Machine Learning-Based Sentiment Analysis for Twitter Accounts,” *Math. Comput. Appl.*, vol. 23, no. 1, 2018.
- [29] J. Diz, G. Marreiros, and A. Freitas, “Applying Data Mining Techniques to Improve Breast Cancer Diagnosis,” *J. Med. Syst.*, vol. 40, no. 9, 2016.
- [30] K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, and N. Kerdprasop, “An Empirical Study of Distance Metrics for K-Nearest Neighbor Algorithm,” 2015.
- [31] G. Tripathi and G. Singh, “Sentiment Analysis Approach based N-gram and KNN Classifier,” *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 3, 2018.
- [32] A. Allahverdipoor and F. . Gharehchopogh, “An Improved K-nearest Neighbor with Crow Search Algorithm for Feature Selection in Text Document Classification,” *J. Adv. Comput. Res. Q. Sari Branch, Islam. Azad Univ.*, vol. 9, no. 2, pp. 37–48, 2018.
- [33] S. . Solloum, C. Mhamdi, M. Al-Emran, and K. Shaalan, “Analysis and Classification of Arabic Newspaper’s Facebook Page using Text Mining Techniques,” *Int. J. Inf. Technol. Language Stud.*, vol. 1, no. 2, pp. 8–17, 2017.
- [34] T.-T. Wong and P.-Y. Yeh, “Reliable Accuracy Estimates from k-Fold Cross Validation,” *EEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, 2019.
- [35] E. Morvant, S. Koco, and L. Ralaivola, “PAC-Bayesian Generalization Bound on Confusion Matrix for Multi-class Classification,” in *International Conference on Machine Learning*, 2018, pp. 815–822.