



ANALISIS SENTIMEN PENGGUNA TWITTER MENGGUNAKAN METODE SVM – KNN

Analysis of Twiter User Sentiment Using SVM – KNN Method

Filipus Upa^{*1}, Nurhalifah²

^{1,2}STMIK Profesional Makassar

Email: upafilipus@gmail.com

Abstract

Sentiment analysis is synonymous with opinion mining, is a type of data mining that refers to the analysis of data obtained from microblogging sites, social media updates, online news reports, user reviews, etc., to study people's sentiments towards an event, organization, product, brand, people and more. In this work, sentiment classification is carried out into several classes. The proposed methodology based on KNN classification algorithm shows improvement over one of the existing methodologies based on SVM classification algorithm. The data used for analysis was taken from Twitter, which is the most popular microblogging site. Source data has been extracted from Twitter. N-Gram modeling technique has been used for feature extraction and k-nearest neighbor supervised machine learning algorithm has been used for sentiment classification. The performance of the proposed and existing techniques is compared in terms of accuracy, precision, and recall. It has been analyzed and concluded that the proposed technique performs better in terms of all standard evaluation parameters.

Keywords: KNN, SVM, Twitter

Abstrak

Analisis sentimen merupakan sinonim dengan penambangan opini, adalah jenis penambangan data yang mengacu pada analisis data yang diperoleh dari situs microblogging, pembaruan media sosial, laporan berita online, ulasan pengguna, dll., untuk mempelajari sentimen orang terhadap suatu peristiwa, organisasi, produk, merek, orang, dan lainnya. Dalam pekerjaan ini, klasifikasi sentimen dilakukan ke dalam beberapa kelas. Metodologi yang diusulkan berdasarkan algoritma klasifikasi KNN menunjukkan peningkatan dibandingkan salah satu metodologi yang ada yang berdasarkan algoritma klasifikasi SVM. Data yang digunakan untuk analisis diambil dari Twitter, yang merupakan situs microblogging paling populer. Data sumber telah diekstrak dari Twitter. Teknik pemodelan N-Gram telah digunakan untuk ekstraksi fitur dan algoritma pembelajaran mesin terawasi k-nearest neighbor telah digunakan untuk klasifikasi sentimen. Kinerja teknik yang diusulkan dan yang ada dibandingkan dalam hal akurasi, presisi, dan recall. Telah dianalisis dan disimpulkan bahwa teknik yang diusulkan berkinerja lebih baik dalam hal semua parameter evaluasi standar.

Kata Kunci: KNN, SVM, Twitter

PENDAHULUAN

Platform jejaring sosial telah mendapatkan popularitas yang besar di kalangan masyarakat dalam beberapa tahun terakhir. Situs-situs online ini banyak digunakan oleh orang-orang untuk mengekspresikan emosi, keyakinan, serta pendapat mereka terhadap berbagai entitas, mulai dari produk, orang, peristiwa, dan sebagainya. Situs jejaring ini menyediakan platform bagi pengguna untuk

memposting umpan balik dan ulasan mereka, dan data yang dihasilkan dimanfaatkan oleh perusahaan bisnis untuk mendapatkan wawasan tentang bagaimana produk dan layanan mereka berkinerja di pasar. Pengetahuan ini membantu analis bisnis dan manajer dalam pengambilan keputusan yang lebih baik. Selain itu, analisis sentimen dari komentar pengguna sangat berguna bagi pembeli juga. Misalnya, jika seseorang ingin membeli produk atau mengakses layanan tertentu, biasanya langkah awal yang akan dilakukan adalah membaca ulasan online dan menghasilkan diskusi mengenai hal tersebut di media sosial sebelum mengambil keputusan apapun. Namun, tidak mungkin bagi pengguna untuk menganalisis semua ulasan mengingat jumlah ulasan dan komentar pengguna yang sangat banyak di platform online. Oleh karena itu, beberapa teknik analisis sentimen telah diusulkan untuk mengotomatiskan proses analisis ini [1], [2]. Melalui teknik-teknik ini, pengguna dapat mengetahui pandangan positif maupun negatif yang dimiliki pengguna lain terhadap suatu produk. Dengan demikian, pengguna mendapatkan gambaran yang jelas tentang produk dan layanan, dan dapat menilai apakah sesuai dengan kebutuhan mereka [3]. Untuk mengumpulkan, memproses, dan menganalisis data faktual, teknik pengambilan informasi teks digunakan. Komentar teks terdiri dari komponen objektif dan subjektif. Analisis dan identifikasi komponen subjektif yang mengungkapkan opini, sentimen, dan sikap sangat membantu bagi produsen dan penyedia layanan dalam menyempurnakan strategi bisnis mereka. Demikian pula, hal ini membantu pelanggan dan klien untuk mengambil keputusan yang lebih baik saat mereka membeli produk atau memanfaatkan layanan^{[4],[5]}.

Belakangan ini, analisis sentimen otomatis serta penambahan opini telah menjadi topik paling populer untuk studi dan penelitian^{[6],[7]}. Proporsi signifikan dari pelanggan dan klien yang menggunakan produk atau layanan menghasilkan jumlah data yang besar dalam bentuk komentar, umpan balik, dan ulasan yang mengungkapkan opini mereka. Namun, pengembangan aplikasi untuk menganalisis data semacam itu melibatkan beberapa tantangan, mengingat ukuran data yang sangat besar dan struktur data tersebut^{[8],[9]}. Kehadiran kata-kata informal, kata slang, dan singkatan membuat sulit untuk mengidentifikasi dan mengklasifikasikan sentimen. Selain itu, untuk kata yang sama, dapat diturunkan beberapa makna dengan polaritas sentimen yang sepenuhnya berbeda; tugas ini dikenal sebagai "Pengelolaan Polisemi". Makna dan polaritas kata slang diidentifikasi dengan merujuk kamus slang, kata-kata informal digantikan dengan sinonimnya, dan singkatan digantikan dengan bentuk yang diperluas. Klasifikasi sentimen biasanya bersifat biner atau ternary dalam sebagian besar pendekatan yang diusulkan, yaitu klasifikasi teks dilakukan ke dalam "positif" dan "negatif", atau "positif", "negatif", dan "netral". Twitter adalah situs microblogging di mana pengguna dapat memposting pesan waktu nyata yang dikenal sebagai tweet^[10]. Sifat unik dari tweet menimbulkan tantangan baru dalam metode klasifikasi sentimen. Beberapa karakteristik penting dari tweet dibahas lebih lanjut. Awalnya, panjang maksimum tweet adalah 140 karakter yang kemudian digandakan pada Januari 2024, untuk semua bahasa kecuali Jepang, Korea, dan Cina. Pengguna sering menggunakan beberapa akronim, kesalahan penulisan, bahasa gaul siber, emotikon, dan karakter lain dengan makna khusus untuk membuat pesan mereka cepat dan singkat. Karakteristik menarik lainnya adalah bahwa Twitter memungkinkan tweet tentang berbagai topik alih-alih berfokus pada topik atau

tema tertentu. Tweet dapat diperbarui secara waktu nyata karena ukurannya yang terbatas juga membuatnya kurang memakan waktu^[11]. Hal ini membuat Twitter lebih baik dari blog untuk analisis sentimen karena blog bersifat lebih panjang dan lebih memakan waktu yang mengapa pembaruan blog dilakukan dalam interval yang lebih lama.

Terdapat beberapa terminologi dasar yang digunakan dalam aplikasi Twitter. Representasi ekspresi wajah yang dihasilkan menggunakan kombinasi tanda baca dan huruf dikenal sebagai emotikon. Representasi bergambar ini secara efektif menyampaikan suasana hati pengguna. Terminologi penting lainnya berkaitan dengan fasilitas untuk menyebut nama pengguna target bersama dengan simbol "@". Jika pengguna disebutkan dengan cara ini, mereka secara otomatis menerima pemberitahuan. Tagar adalah fitur penting lainnya dari tweet. Untuk menandai topik, "#" yang dikenal sebagai tagar digunakan. Melalui tagar, memungkinkan untuk jumlah audiens yang lebih besar untuk melihat tweet. Semua properti ini dari tweet diambil dalam pertimbangan bersama dengan fitur teks reguler selama pemilihan fitur^{[12-15], [16-17]}. Twitter adalah situs microblogging paling populer dan oleh karena itu menghasilkan jumlah data yang besar yang cocok untuk penambangan opini dibandingkan dengan media sosial lainnya.

METODE

Zhao Jianqiang dan koleganya^[18] telah mengusulkan peningkatan terhadap pendekatan analisis sentimen konvensional yang fokus pada analisis leksikal setiap unit dalam tweet seperti kata-kata, tanda seru, emotikon, dll. Pendekatan yang diusulkan berbasis pada metode pembelajaran tidak terawasi yang membentuk embedding kata dengan merujuk pada korpora Twitter yang besar dengan mempertimbangkan hubungan semantik laten kata-kata berdasarkan makna kontekstual, serta properti statistik yang muncul bersamaan antar kata dalam tweet. Untuk menghasilkan set fitur dari tweet, embedding kata diintegrasikan dengan fitur n-gram dan fitur skor polaritas sentimen kata. Untuk melatih dan memberi label pada klasifikasi sentimen, set fitur diintegrasikan ke dalam jaringan saraf konvolusi dalam. Skor kinerja dari model yang diusulkan dibandingkan dengan model dasar pada lima dataset Twitter, di mana model yang diusulkan menunjukkan kinerja yang lebih baik dalam hal akurasi dan ukuran F1. K Lavanya dan koleganya^[19] telah mengusulkan mekanisme klasifikasi berdasarkan metode pelatihan yang dapat disesuaikan dengan topik tweet. Mekanisme ini diusulkan untuk mengatasi masalah keragaman topik yang sangat luas di Twitter. Hal ini membuat algoritma klasifikasi bersifat dinamis. Dalam pendekatan ini, fitur non-teksual dari tweet juga digunakan untuk pelatihan algoritma klasifikasi. Metodologi yang diusulkan dapat diterapkan pada data statis yang berasal dari topik yang sepenuhnya berbeda serta pada data dinamis untuk garis waktu tertentu selagi data tersebut masuk secara streaming. Algoritma klasifikasi mengkategorikan tweet ke dalam tiga kelas utama yaitu positif, netral, dan negatif. Ketiga label kelas ini dapat diperluas lebih lanjut untuk membentuk lima label kelas yaitu netral, positif, sangat positif, negatif, dan sangat negatif. Pendekatan yang diusulkan menunjukkan peningkatan kinerja dengan mengacu pada recall, presisi, dan skor F1-Score.

Chintan Dedhia dan koleganya^[20] mengusulkan metode tambahan untuk meningkatkan kekuatan klasifikasi SVM untuk analisis sentimen, yang merupakan

salah satu algoritma pembelajaran mesin yang sering digunakan untuk analisis dan klasifikasi sentimen. Model ansambel ini dirancang dengan mengintegrasikan algoritma SVM sebagai klasifikasi dasar yang dipasangkan dengan algoritma Adaboost untuk penguatan ansambel. Algoritma yang diusulkan memanfaatkan detail terstruktur yang terkait dengan tweet seperti retweet, pengikut tweet, tag dalam tweet bersama dengan semua karakteristik penting lainnya dari tweet untuk menemukan informasi hubungan antar tweet. Ini membantu dalam mempelajari sifat-sifat interaksi sosial. Model yang diusulkan digunakan untuk mengklasifikasikan data Twitter menjadi label positif dan negatif. Pendekatan yang diusulkan telah dibandingkan dengan algoritma SVM dasar dan menunjukkan kinerja yang lebih baik dengan mengacu pada presisi, recall, dan skor F-Score. Yeqing Yan dan koleganya^[21] memperkenalkan dua klasifikasi ansambel yang sederhana namun kuat untuk melakukan klasifikasi sentimen data Twitter. Dua model ansambel dikembangkan menggunakan klasifikasi Naive Bayes dengan MaxEnt Mallet, dan SentiStrength dengan Pattern of Textblob. Model ini mengatasi masalah yang muncul dalam melatih klasifikasi dengan tepat ketika tidak cukup data pelatihan; yaitu, untuk mengategorikan tweet yang terkait dengan satu produk, tweet yang terkait dengan produk serupa lainnya dapat digunakan untuk belajar klasifikasi. Kedua model tersebut telah terbukti efisien dengan menunjukkan akurasi tinggi dalam klasifikasi sentimen dari dua belas dataset yang berbeda. Kedua klasifikasi ansambel tersebut dapat dijalankan secara paralel dan mampu menangani set data Twitter yang besar.

Paramita Ray dan rekan-rekannya^[22] mengusulkan kerangka kerja berbasis R yang didasarkan pada pendekatan leksikon untuk analisis sentimen dan kategorisasi ulasan produk, yang dapat membantu dalam pengambilan keputusan yang lebih baik terkait dengan produk dan layanan. Pra-pemrosesan tweet masukan terutama mencakup penggantian akronim dengan bentuk yang diperluas, penggantian emotikon dengan kata-kata yang menggambarkan sentimen terkait, penghapusan kata stop, dan penanganan negasi. Metodologi yang diusulkan menganalisis teks baik pada tingkat dokumen maupun aspek. Penulis bertujuan mengembangkan model kerja hibrida di masa depan yang berbasis pada teknik pembelajaran mesin untuk analisis dan klasifikasi sentimen. Ranjan Satapathy dan rekan-rekannya^[23] dalam makalahnya menyebutkan bahwa konsep mikroteks yang menjadi populer karena penggunaan teknologi Web 2.0 menimbulkan tantangan bagi alat pemrosesan bahasa alami standar yang dirancang untuk menangani teks yang terbentuk dengan baik. Normalisasi mikroteks membantu mengatasi tantangan-tantangan ini. Oleh karena itu, penulis mengusulkan pendekatan berbasis fonetika untuk mengubah mikroteks menjadi teks biasa dalam bahasa Inggris. Hasil demonstrasi menunjukkan bahwa indeks kesamaan antara tweet yang dinormalisasi dengan teknik yang diusulkan dan tweet yang dinormalisasi oleh anotator manusia setara atau lebih dari 0,8 untuk 85,31% tweet. Juga ditemukan bahwa normalisasi tweet meningkatkan akurasi dalam deteksi polaritas lebih dari 4%.

Klasifikasi SVM (*Support Vector Machine*)

Support Vector Machine (SVM) adalah algoritma yang dapat digunakan untuk klasifikasi linear dan non-linear, serta regresi linear dan non-linear. Bentuk awal dari klasifikasi SVM adalah klasifikasi biner, di mana fungsi yang dipelajari mengklasifikasikan data menjadi kelas positif dan negatif^[24]. Klasifikasi

multikelas dapat dimodelkan dengan mengintegrasikan beberapa klasifikasi biner menggunakan metode penggabungan pasangan (pair-wise coupling). Metode ini memungkinkan SVM untuk menangani lebih dari dua kelas dengan membandingkan setiap pasangan kelas dalam data, sehingga menghasilkan model yang mampu melakukan prediksi kelas pada data yang lebih kompleks. Margin maksimum yang memisahkan secara linear titik data n -dimensi $X = \{x_1, x_2, x_3 \dots x_n\}$ dengan bobot atribut $W = \{w_1, w_2, w_3 \dots w_n\}$ menjadi dua kelas ditulis dalam formula:

$Y_i = \{+1, -1\}$ is defined by the margins

$$H1: w_0 + \sum_{k=1}^n w_k x_k \geq +1 \text{ for } y_i = +1 \quad (1)$$

$$H2: w_0 + \sum_{k=1}^n w_k x_k \leq -1 \text{ for } y_i = -1 \quad (2)$$

dimana w_0 adalah sebuah skala pengetahuan sebagaimana dikombinasikan antara formula (1) dan (2). Ini adalah masalah optimasi kuadratik konveks yang diselesaikan untuk memperoleh vektor pendukung dan bidang margin maksimum, sebagaimana ditulis dalam formulasi:

$$D(X^T) = \sum_{i=1}^k y_i \alpha_i X X^T + b \quad (4)$$

dimana, y_i adalah label kelas dari vektor pendukung X_i , k adalah jumlah vektor pendukung, X^T adalah type uji, b adalah parameter numerik yang ditentukan secara otomatis oleh optimasi algoritma SVM. Untuk mendukung klasifikasi non-linear, pemetaan non-linear dilakukan untuk mengubah data asli di ruang masukan ke ruang berdimensi lebih tinggi. Dengan menghindari formulasi eksak dari fungsi pemetaan, trik kernel digunakan yang melalui itu kutukan dimensionalitas dihasilkan. Dengan demikian, klasifikasi non-linear dalam ruang asli disamakan dengan klasifikasi linear di ruang baru. Dalam dimensi baru, klasifikasi SVM mencari bidang margin maksimum yang memisahkan titik data satu kelas dari kelas lainnya.

Klasifikasi KNN

K-Nearest Neighbors (KNN) adalah algoritma klasifikasi yang diawasi non-parametrik, yang sederhana namun efektif dalam banyak kasus. Klasifikasi KNN dianggap sebagai klasifikasi paling populer untuk pengenalan pola karena kinerjanya yang efektif dengan hasil yang efisien dan kesederhanaannya. Algoritma ini banyak digunakan dalam bidang pengenalan pola, pembelajaran mesin, kategorisasi teks, penambangan data, pengenalan objek, dan banyak lagi^[25]. Algoritma KNN mengklasifikasikan dengan analogi, yaitu dengan membandingkan titik data yang tidak diketahui dengan titik data pelatihan yang serupa dengannya. Kesamaan diukur dengan jarak Euclidean. Nilai atribut dinormalisasi untuk mencegah atribut dengan rentang yang lebih besar dari atribut dengan rentang yang lebih kecil. Dalam klasifikasi KNN, pola yang tidak diketahui diberikan kelas yang paling dominan di antara kelas dari tetangga terdekatnya. Dalam kasus ada seri antara dua kelas untuk pola tersebut, kelas yang memiliki jarak rata-rata minimum ke pola yang tidak diketahui diberikan. Melalui kombinasi sejumlah fungsi jarak lokal berdasarkan atribut individu, fungsi jarak

global bisa dihitung. Seperti yang diberikan dalam persamaan (5), cara termudah adalah dengan menjumlahkan nilai-nilainya:

$$(X^T, X) = \sum^n \text{distA}(X^T * A, X * A) \tag{5}$$

Di mana X^T adalah tupel uji, X adalah tetangga terdekat, dan A_i mewakili atribut dari titik data. Jumlah terbobot dari jarak lokal dikenal sebagai jarak global. Atribut A_i dapat diberi bobot w_i untuk menggambarkan tingkat kepentingannya dalam menentukan kelas yang tepat untuk sampel. Bobot biasanya berkisar antara 0-1. Atribut yang tidak relevan diberi bobot 0.

HASIL DAN PEMBAHASAN

Kinerja sistem klasifikasi, seperti yang Anda deskripsikan, umumnya dievaluasi menggunakan tiga metrik kunci: presisi, recall, dan akurasi. Metrik-metrik ini memberikan pandangan komprehensif tentang seberapa baik sistem tersebut berkinerja, terutama dalam mengklasifikasikan data ke dalam kategori, seperti tujuh kelas data Twitter yang Anda sebutkan. Berikut adalah penjelasan singkat tentang setiap metrik dan rumusnya masing-masing:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{8}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{9}$$

$$\text{Accuracy} = \frac{\text{tweets correctly classified}}{\text{Total no. of tweets}} \tag{10}$$

Presisi sistem yang diusulkan sekitar 82 persen, sedangkan presisi sistem yang sudah ada sekitar 79 persen. Recall sistem yang diusulkan sebesar 81,5 persen sementara nilai recall oleh sistem yang sudah ada adalah hingga 86 persen, sedangkan akurasi sistem yang sudah ada sekitar 81 persen. Hasil Evaluasi kinerja sistem ditunjukkan dalam Tabel 1.

Tabel 1. Hasil perbandingan klasifikasi

Matriks	SVM	SVM + KNN
Precision	79%	82 %
Recall	78%	81.5 %
Accuracy	81%	86 %

KESIMPULAN

Analisis sentimen merupakan sebuah teknik untuk menganalisis dan mengklasifikasikan sentimen yang disampaikan oleh pernyataan pengguna, dimana dapat dilakukan menggunakan beberapa metodologi. Salah satu metodologi yang sudah ada didasarkan pada penggunaan algoritma SVM untuk melatih pengklasifikasi yang membedakan data Twitter ke dalam beberapa kelas. Dalam penelitian ini, metodologi yang disebutkan di atas diperbaiki lebih lanjut dengan menggunakan algoritma KNN untuk melatih pengklasifikasi. Kinerja algoritma KNN meningkat secara signifikan dengan pemilihan parameter k yang tepat, penggunaan metrik jarak yang sesuai, penggabungan pembobotan atribut,

dan pemangkasan titik data yang berisik. Teknik-teknik yang sudah ada dan diusulkan diimplementasikan dalam Python dan hasil simulasi menunjukkan bahwa pendekatan yang diusulkan unggul dibandingkan dengan pendekatan yang sudah ada dalam hal presisi, recall, dan akurasi. Pendekatan yang diusulkan dapat diperbaiki lebih lanjut dengan menggunakan algoritma KNN berbobot jarak yang melibatkan penugasan bobot dengan tetangga terdekat berdasarkan kedekatan mereka dengan titik data; semakin dekat tetangganya, semakin besar bobotnya. Hal ini akan memungkinkan penyempurnaan proses klasifikasi.

UCAPAN TERIMAKASIH

Ucapan terima kasih penulis haturkan kepada STMIK Profesional Makassar yang telah memberi dukungan selama penelitian.

DAFTAR PUSTAKA

- [1] Meesala Shobha Rani, Sumathy S. (2017). Perspectives of the performance metrics in Lexicon and Hybrid based approaches: a review. *IJET*, 6 (4).
- [2] Cambria, B. Schuller, Y. Xia, and C. Havasi. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28 (2): 15-21. <https://doi.org/10.1109/MIS.2013.30>.
- [3] Sasikumar.A. N. (2017). Sentimental Analysis of Social Networking Sites for Categorization of Product Reviews. *Internation Journal of Pure and Applied Mathematics*, 117: 87-92.
- [4] J. Mannar Mannan, J, Jayavel. (2018). *An adaptive sentimental analysis using ontology for retail market*. *IJET*, 7 (1.2).
- [5] V. Uma. Ramya, K. Thirupathi Rao. (2018). Sentiment Analysis of movie review using Machine Learning techniques. *IJET*, 7 (2.7).
- [6] Thelwall, M., Buckley, K., & Paltoglou, G. (2016). Sentiment strength detection for the social web. *J. American Society for Information Science and Technology*, 63 (1): 163–173. <https://doi.org/10.1002/asi.21662>.
- [7] Paltoglou, G, & Thelwall, M. (2012). Twitter, MySpace, Digg Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems and Technology*, 3 (4): 1-19. <https://doi.org/10.1145/2337542.2337551>.
- [8] Wafa Zubair Al-Dyani, Adnan Hussein Yahya, Farzana Kabir Ahmad. (2018). Challenges of Event Detection from social media streams. *IJET*, 7 (2.15).
- [9] Socher, R., Perelygin, A., Y.Wu, J., Chuang, J., Manning, C. D., Ng, Y., & Potts, C., (2017). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Pro-Recommendation System for off-the shelf medicines. *IJET*, 6 (2.24).
- [10] Tharindu Weerasooriya, Nandula Perera, S.R. Liyanage. (2016). A method to extract essential keywords from tweet using NLP. *16th International Conference on Advances in ICT for Emerging Regions (IC- Ter)*.



- [11] Prasanna Moorthi N, Mathivanan V. (2018). An improved Wrapper based feature selection for feature mining. *IJET*, 7 (1.3).
- [12] Zhao Jianqiang, Gui Xiaolin, *Deep Convolution Neural Networks for Twitter Sentiment Analysis*, IEEE, 2017.
- [13] K Lavanya, C Deisy. (2017). Twitter Sentiment Analysis Using Multi- Class SVM. *International Conference on Intelligent Computing and Control (I2C2'17)*.
- [14] Chintan Dedhia, Mrs Jyoti Ramteke. (2017). Ensemble model for Twitter Sentiment Analysis. *International Conference on Inventive Systems and Control (ICISC)*. <https://doi.org/10.1109/ICISC.2017.8068711>.
- [15] Yeqing Yan, Hui Yang, Hui-ming Wang. (2017). Two Simple and Effective Ensemble Classifiers for Twitter Sentiment Analysis. *Computing Conference*.
- [16] Paramita Ray and Amlan Chakrabarti. (2017). Twitter Sentiment Analysis for Product Review Using Lexicon Method. *International Conference on Data Management, Analytics and Innovation (ICDMAI)*. <https://doi.org/10.1109/ICDMAI.2017.8073512>.
- [17] Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, Erik Cambria. (2017). Phonetic-Based Microtext Normalization for Twitter Sentiment Analysis. *IEEE International Conference on Data Mining Workshops*.
- [18] Shweta Rana, Archana Singh. (2016). Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. *2nd International Conference on Next Generation Computing Technologies (NGCT)*.
- [19] Pedro, J., Garcia-Laencina, Jose-Luis Sancho-Gomez, Anibal, R., Figueiras-Vidal, and Michel Verleysen. (2009). *KNN with mutual information for simultaneous classification and missing data imputation*: 1483–1493

