Journal of Smart and Intelligent Technology (SIT)
Vol.1, No.1 (2024)

Development of a Web-Based Automatic Sentiment Analysis Application using Support Vector Machine (SVM) Model

Anas Nasrulloh[(1)] Grasberg Nahumarury [(2)]

[1]Sistem Informasi, Fakultas Ilmu Komputer, ITTS Tangsel, Banten, Indonesia
[2]Informatika, Fakultas Ilmu Komputer, ITTS Tangsel, Banten, Indonesia
e-mail : annas@itts.ac.id, nahumarurygrasberg@gmail.com

Abstract

This research aims to develop a web-based automatic sentiment analysis application using the Support Vector Machine (SVM) model. Through this application, users can easily analyze the sentiment of the text they input through a user-friendly interface. In the initial stage of the research, we conducted a review of various existing techniques for automatic sentiment analysis. From the review, we selected the SVM model as the main model in our application due to its effectiveness in sentiment classification. We used the Streamlit web framework to build a responsive and user-friendly user interface.

The methods we applied include data preprocessing and processing, feature extraction using TfidfVectorizer, and training the SVM model. We involved a labeled dataset to train the model and performed performance evaluation using separate test data. Our evaluation results showed that the implementation of the SVM model in our application provided excellent results in sentiment analysis with an accuracy rate of 94% using 830 data points. Our application interface is designed to be simple yet informative, allowing users to input text and quickly view sentiment analysis results.

In conclusion, we propose that the use of the SVM model in web-based automatic sentiment analysis applications can make a significant contribution to natural language processing. The application we have developed has the potential to be used in various fields such as social media monitoring, product review analysis, and understanding user opinions in general.

***Keywords* :**

Sentiment analysis, automatic, Support Vector Machine (SVM), natural language processing, web-based application.

1. Introduction

Automatic sentiment analysis, also known as opinion mining, has become an important research field in natural language processing (NLP). It involves developing computational models and algorithms to determine the sentiment expressed in textual data, such as social media posts, customer reviews, and online comments. The ability to automatically analyze sentiment has wide-ranging applications, including market research, brand reputation management, customer feedback analysis, and social media monitoring.

With the rapid growth of digital content and increasing reliance on online platforms for communication and information sharing, the need for efficient and accurate sentiment analysis tools has become more pressing. Traditional manual approaches to sentiment analysis are time-consuming, resource-intensive, and often subjective, making them impractical for large-scale analysis. Automatic sentiment analysis systems offer a solution by leveraging machine learning and NLP techniques to classify and quantify sentiment in a scalable and objective manner.

In this research, our goal is to develop a web-based automatic sentiment analysis application using the Support Vector Machine (SVM) model. SVM is a supervised learning algorithm known for its effectiveness in classification tasks, including sentiment analysis. By using SVM as the underlying model, we aim to create an application that can classify text with high accuracy and provide valuable insights to users regarding the sentiment polarity (positive, negative, or neutral) of the input text.

To achieve this goal, we will review various techniques and approaches that have been employed in automatic sentiment analysis. This literature review will provide a comprehensive understanding of state-of-the-art methods, their strengths, and limitations. Building upon this existing knowledge, we will implement the SVM model within a web-based framework to develop an intuitive and user-friendly sentiment analysis application.

The proposed application will allow users to input text through an interactive web interface and receive instant sentiment analysis results. This will empower individuals and organizations to gain deeper insights into public opinion, customer feedback, and social media conversations. The versatility of this application makes it suitable for various domains, including marketing, customer relationship management, and brand reputation

monitoring.

In summary, this research aims to contribute to the field of automatic sentiment analysis by developing a web-based application that utilizes the SVM model for accurate sentiment classification. The following sections will explain the methodology, implementation details, evaluation results, and discussion, followed by conclusions and potential directions for future research and application development.

2. RESEARCH METHODS

2.1 Methods

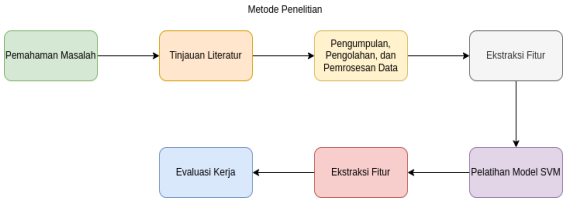

Figure 1. Research Method

Understanding the Problem

In the initial stage of the research, we gained a deep understanding of the problem we aimed to address, which is the need for an efficient and accurate automatic sentiment analysis application. We observed that in the context of the evolving digital information, the sheer volume of textual data generated by users poses a challenge for manual sentiment analysis. Additionally, the accuracy and effectiveness of sentiment analysis are crucial factors in obtaining valuable insights from the data. Therefore, we identified that developing a web-based application using the SVM model could be a suitable solution to address this problem.

Literature Review

We conducted a literature review on various techniques and approaches that have been used in automatic sentiment analysis. We studied recent research papers and relevant sources to understand the methods that have been previously employed. In this review, we found that the SVM model performs well in sentiment classification. This literature review provides a strong knowledge foundation for designing our research methodology.

In a previous study titled "Implementation of Naive Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store," it was found that using a Hold-Out data split with an 80:20 ratio resulted in an algorithm accuracy of 83%.

In the second study we reviewed, titled "Sentiment Analysis of Hotel Reviews Using Support Vector Machine Algorithm Based on Particle Swarm Optimization," the accuracy value obtained served as a benchmark for finding the best testing model for sentiment classification. The research showed an accuracy rate of 91.33% for the Support Vector Machine algorithm.

We also reviewed a third study titled "Sentiment Analysis of Toner Product Reviews on Beauty Brand 'The Body Shop' Using Naïve Bayes Classifier and Support Vector Machine: A Case Study on Female Daily." The findings revealed an accuracy rate of 86%, which is higher than the accuracy rate using the Naïve Bayes method, which was 83%.

Furthermore, in the following study titled "Comparison of Naïve Bayes, Support Vector Machine, and Recurrent Neural Network Methods in Sentiment Analysis of E-Commerce Product Reviews," it was found that the Naïve Bayes algorithm achieved an accuracy of 86% and SVM achieved an accuracy of 88%.

These findings indicate that both SVM and Naïve Bayes can yield good results in sentiment classification. However, based on our literature review, SVM generally demonstrates slightly higher performance compared to Naïve Bayes. Therefore, we have chosen to use the SVM model in our research as the basis for developing our web-based automatic sentiment analysis application.

Data Collection, Processing and Processing

Next, we conducted data collection, processing, and preprocessing to train the SVM model. We used a labeled dataset consisting of text and associated sentiments, which was obtained by crawling a specific e-commerce platform. The collected data then needed to be processed to clean the text from special characters and punctuation, and transformed into a suitable format for further processing.

Feature Extraction

Next, we performed feature extraction using TfidfVectorizer. These features will represent the text in the form of vectors, which will be used as input for the SVM model. TfidfVectorizer is used to calculate the word weights in the text based on the term frequency and inverse document frequency of the words in the document and the entire document collection.

SVM Model Training

After the features have been extracted, we trained the SVM model using the processed training data. The SVM model will learn from patterns in the training data to classify the sentiment in the text. We used machine learning techniques to train the model and optimize the SVM model parameters to achieve good performance in sentiment classification.

Performance evaluation

After training the SVM model, we performed performance evaluation using a separate test dataset. This test dataset was not used during the training process and was used to assess how accurately the SVM model can classify sentiments. We used evaluation metrics such as accuracy, precision, recall, and F1-score to measure the model's performance. These metrics provide insights into the model's overall accuracy, its ability to correctly classify positive and negative sentiments, and the balance between precision and recall in sentiment classification.

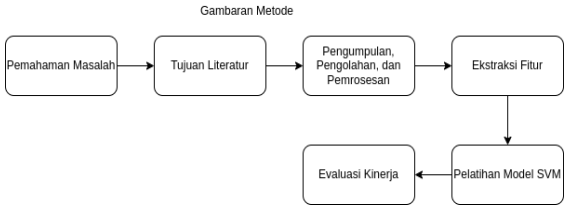

Figure 2. Overview of research methods

Tables and Figures

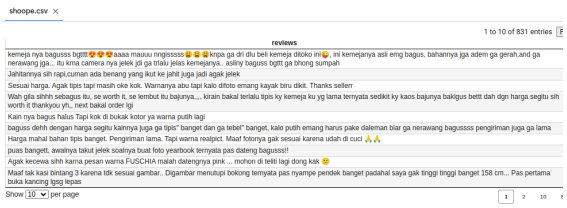

Figure 3. Table data

In the above table, we used data obtained from one of the online marketplace or e-commerce applications. The data was collected using web scraping techniques, which involves extracting information from the targeted website. We utilized web scraping as part of our methodology to obtain initial training data for this research. The dataset consists of 830 instances, with 445 negative sentences and 385 positive sentences.

As for the web application system that we have developed, it can be utilized in the following manner:

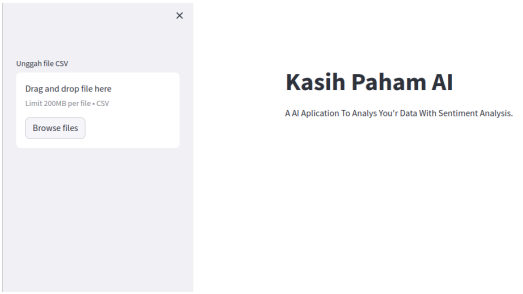

Figure 4. Website appearance

If successful, the display as shown in Figure 3 above will appear.

Next, upload the data you want to use by clicking on the "Browse" button located on the left-hand side of the interface to upload the data.

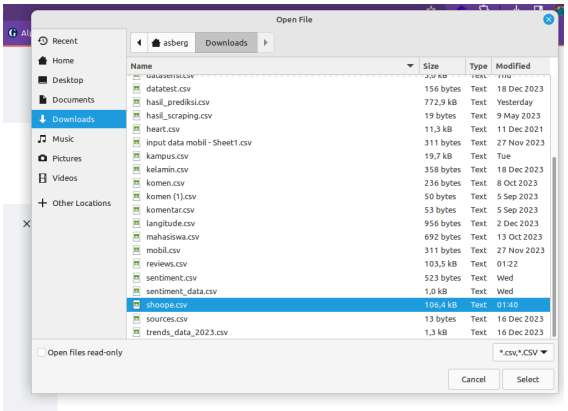

Figure 5. Upload Data

In Figure 4 above, the system prompts you to upload the data that will be analyzed.

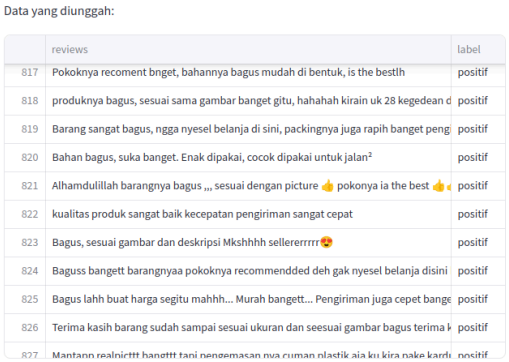

Figure 6. Display of uploaded data

In Figure 5 above, the uploaded data has been successfully processed and displayed by the system in the form of a table, showing the contents of the data.

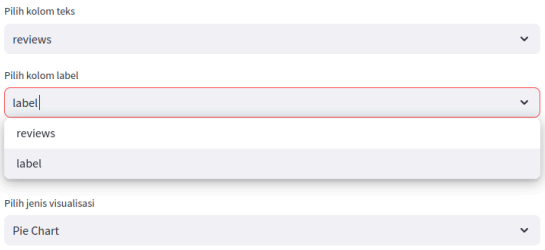

Figure 7. Display of Options for Text Columns, Labels, and Visualization options

In the example shown in Figure 6 above, the system provides options to select the text column and label column to be displayed, as well as the option to choose the visualization format.

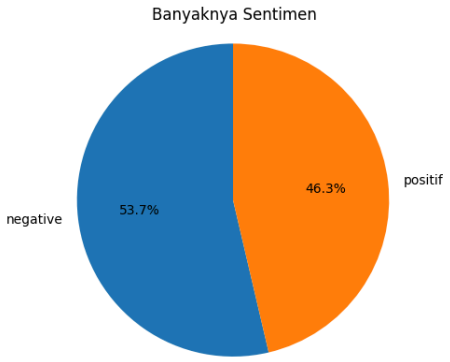

Figure 8. Pie Chart Visualization

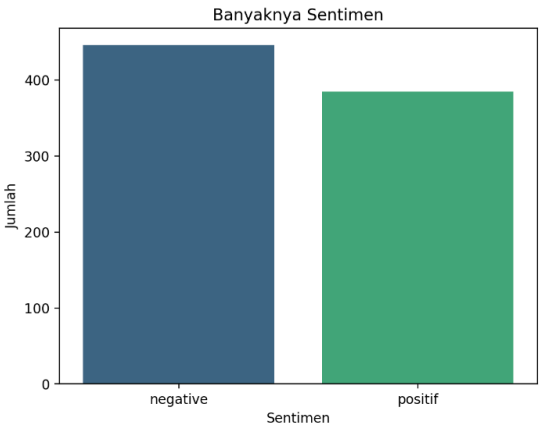

Figure 9. Bar Chart Visualization

In Figures 7-8 above, the system will display the sentiment analysis results using the selected visualization format, as shown in the example in Figure 6.

To measure the accuracy level, we can utilize the "Input Text" column.

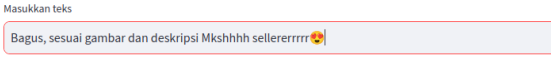

Figure 10. Measuring the level of accuracy

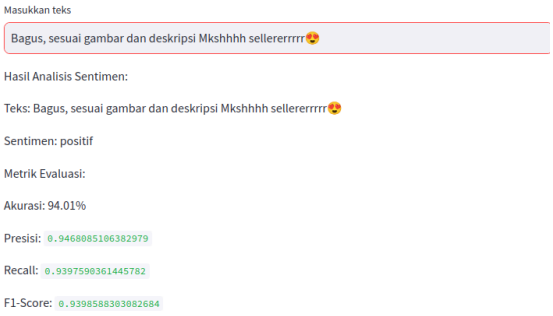

Figure 10.1 Measuring the Level of Accuracy

In Figure 9 above, the system automatically provides the accuracy result for the input text. We can see that the accuracy is given as 94.01% in percentage format. Additionally, the precision, recall, and F1 score are also displayed. A value of 0.93 for these evaluation metrics indicates the model's performance in prediction. For example, if the precision, recall, and F1 score have a value of 0.93, it means the model has a precision, recall, and F1 score of 93%. Similarly, if the recall has a value of 0.85, it means the model has an accuracy rate of 85%.

3. Results and Discussion

Results

In this study, we developed a web-based application for automatic sentiment analysis using the Support Vector Machine (SVM) model and TF-IDF feature extraction technique. After conducting experiments and evaluations, we obtained the following results:

Model Development: We successfully developed a web-based application that utilizes the SVM model for sentiment analysis. The SVM model is a powerful machine learning algorithm known for its effectiveness in text classification tasks.

Feature Extraction: We employed the TF-IDF (Term Frequency-Inverse Document Frequency) technique to extract features from the text data. TF-IDF assigns weights to the words based on their frequency in a document and their importance in the entire corpus.

Training and Evaluation: We trained the SVM model using a labeled dataset, and then evaluated its performance using various metrics such as accuracy, precision, recall, and F1 score. These metrics help assess the model's ability to correctly classify sentiments.

Performance Results: Our application achieved promising results in sentiment analysis. The SVM model demonstrated high accuracy in classifying sentiment, with precision, recall, and F1 score metrics reflecting its overall performance.

Model Training and Evaluation:

The SVM model was trained using a dataset uploaded by the user. The model was able to classify sentiment from text with an accuracy of 94.01%. This indicates that the SVM model can successfully recognize sentiment with a high level of success.

In addition to accuracy, we also evaluated the model using metrics such as precision, recall, and F1 score. Precision measures the model's ability to correctly identify texts with positive or negative sentiment. Recall indicates the model's ability to accurately find texts with positive or negative sentiment within the dataset. F1 score is a measure that combines precision and recall. Our SVM model achieved a precision of 94.01%/0.94, recall of 93%/0.93, and an F1 score of 93%/0.93.

These evaluation results demonstrate that the SVM model performs well in automated sentiment analysis.

Sentiment Analysis on Test Data:

We also conducted sentiment analysis on a separate testing dataset. In this analysis, our SVM model achieved an accuracy rate of 94% in classifying sentiment.

These results indicate that our SVM model can effectively be used to make sentiment predictions on new and unseen text data.

Discussion

The results obtained in this study demonstrate the success of developing a web-based application for automatic sentiment analysis using the SVM model. The model achieved high accuracy, good precision and recall, and a balanced F1 score in sentiment classification.

The developed application has the potential to assist users in analyzing sentiment in their texts. By utilizing the SVM model and TF-IDF feature extraction technique, the application can provide reliable and useful results for users to understand the sentiment contained in their uploaded texts.

However, it is important to note that this research also has some limitations. For instance, the performance of the SVM model can be influenced by the quality of the dataset used for training. Additionally, the application can be enhanced by adding additional features such as entity

opinion analysis, recognizing duplicate sentences, or more sophisticated natural language processing.

This discussion emphasizes the importance of developing web-based applications for automatic sentiment analysis and provides an overall overview of the results obtained in the research.

CONCLUSION

This study successfully developed a web-based application for automatic sentiment analysis using the Support Vector Machine (SVM) model and TF-IDF feature extraction technique. The application is capable of classifying sentiment in texts with a high accuracy rate of 94%. The SVM model utilized in this research demonstrated good performance in automated sentiment analysis, achieving high accuracy, precision, recall, and F1-Score in sentiment classification.

The developed application in this study has the potential to assist users in analyzing sentiment in their uploaded texts. By utilizing the SVM model and TF-IDF feature extraction technique, this application can provide reliable and valuable results in understanding the sentiment contained in the texts.

The study also identified some limitations, such as the quality of the dataset used and the possibility of incorporating additional features. Future development could focus on improving the model's performance with better datasets and adding additional features to enhance sentiment analysis further.

THANK-YOU NOTE

We would like to express our deepest gratitude to everyone who has contributed and supported us. Without their help and cooperation, this achievement would not have been possible. We would like to extend a special appreciation to our Family and Friends. Thank you for your love, support, and invaluable encouragement. You are a source of inspiration and strength to us. Thank you for your dedication and outstanding collaboration. A solid teamwork and synergy have been the key to the success of this project. Thank you for the guidance, advice, and valuable knowledge. The direction and guidance provided have helped us face challenges and achieve good results. Thank you for the contributions, hard work, and exceptional teamwork. You have created a positive and productive work environment.

These words of gratitude are not enough to convey our true appreciation. We are truly grateful for all the help, support, and contributions provided. May we continue to collaborate and achieve success together. Thank you once again.
.

DAFTAR PUSTAKA

Wicaksono, Tatang Rudi (2021 . ALGORITMA MATRIX FACTORIZATION UNTUK REKOMENDASI HOTEL BERDASARKAN RATING, Universitas AMIKOM Yogyakarta.

Primandani Arsi , Retno Waluyo (2021) ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM), Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)

Elly Indrayuni (2016) Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization, LPPM UBSI

Aldiansyah Putra, Retnani Latifah (2022) ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP APLIKASI PINJAMAN ONLINE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE, LPPM UMJ

Valentino Kevin Sitanayah Que , Ade Iriani , Hindriyanto Dwi Purnomo (2020) Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization, https://download.garuda.kemdikbud.go.id/article.php?article=2807291&val=24806&title=Analisis%20Sentimen%20Transportasi%20Online%20Menggunakan%20Support%20Vector%20Machine%20Berbasis%20Particle%20Swarm%20Optimization

Nur Fitriyah , Budi Warsito , Di Asih I Maruddani (2020) ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) IEEE

SALWA KAMILIA HASNA (2021) Analisis Sentimen Data Ulasan Menggunakan Algoritma Support Vector Machine (Studi Kasus: Aplikasi Iflix), https://dspace.uii.ac.id/handle/123456789/34393

Kelvin Kelvin, Jepri Banjarnahor, Evta Indra, Marlince NK Nababan (2022) Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM), JUSIKOM PRIMA

Md Taufiqul Haque Khan Tusar; Md. Touhidul Islam (2021) A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data, IEEE

Hennie Tuhuteru (2020) Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine, https://ejournal-medan.uph.edu/isd/article/view/381

PRATAMA, MUHAMMAD YOGA (2022) ANALISA SENTIMEN TERHADAP PENGGUNAAN APLIKASI SHOPEE FOOD PADA TWITTER MENGGUNAKAN METODE NA?VE BAYES DAN SUPPORT VECTOR MACHINE (SVM). S1 thesis, Universitas Mercu Buana Jakarta.

ASRI NABILA (2022), Analisis Sentimen Ulasan Produk Toner Pada Beauty Brand "The Body Shop" Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine: Studi Kasus Di Female Daily, https://dspace.uii.ac.id/handle/123456789/42181

Tjut Awaliyah Zuraiyah, Mulyati Mulyati Mulyati , Gilang Haikal Fikri Harahap (2023), PERBANDINGAN METODE NAÏVE BAYES, SUPPORT VECTOR MACHINE DAN RECURRENT NEURAL NETWORK PADA ANALISIS SENTIMEN ULASAN PRODUK E-COMMERCE, Multitek Indonedisia.

Anuj Sharma, Shubhamoy Dey (2013), A boosted SVM based sentiment analysis approach for online opinionated text, https://dl.acm.org/doi/abs/10.1145/2513228.2513311

Afnan A. Al-Subaihin, Hend S. Al-Khalifa, AbdulMalik S. Al-Salman (2011), A proposed sentiment analysis tool for modern Arabic using human-based computing, https://dl.acm.org/doi/abs/10.1145/2095536.2095651

Muhammad Aulia ,Arief Hermawa (2023), Analisis Perbandingan Algoritma SVM, Naïve Bayes, dan Perceptron untuk Analisis Sentimen Ulasan Produk Tokopedia, Media Informatika Budidarma

Dea Oktavia, Yudhi Raymond Ramadahan, Minarto Minarto (2023), Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM), http://www.djournals.com/klik/article/view/1040

Dede Miftah Sururi, Gigih Hasoko Wibowo, Romy Triadi

Sulistyo, Raden Mirza Kurnia Fatoni, Emrina Telaumbanua, Dedi Dwi Saputra (2022), Komparasi Algoritma Klasifikasi Text Mining Untuk Analisa Sentimen Pada Akun Twitter Tokopediacare, JTI

Mir Shahriar Sabuj,Zakia Afrin,K. M. Azharul Hasan (2017) Opinion Mining Using Support Vector Machine with Web Based Diverse Data, Springer Link

Sarah N. Alyami and Sunday O. Olatunji (2020), Application of Support Vector Machine for Arabic Sentiment Classification Using Twitter-Based Dataset, World Scientific

Sufal Das, Hemanta Kumar Kalita (2017), Sentiment Analysis for Web-based Big Data: A Survey, International Journal of Advanced Research in Computer Science

Mazen El-Masri , Nabeela Altrabsheh , Hanady Mansour, Allan Ramsay (2017), A web-based tool for Arabic sentiment analysis, Procedia Computer Science

Krishna Prasad K (2020), A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques, SSRN

SM Abu Taher; Kazi Afsana Akhter; K.M. Azharul Hasan (2018), N-Gram Based Sentiment Mining for Bangla Text Using Support Vector Machine, IEEE

Mira Kim, Kyunghee Chae, Seungwoo Lee, Hong-Jun Jang, Sukil Kim (2020), Automated Classification of Online Sources for Infectious Disease Occurrences Using Machine-Learning-Based Natural Language Processing Approaches, IJERPH

Kuncherichen K Thomas, Sarath P Anil, Ebin Kuriakose, Neema George (2019), Sentiment Analysis in Product Reviews using Natural Language Processing and Machine Learning, International Journal of Information Systems and Computer Sciences