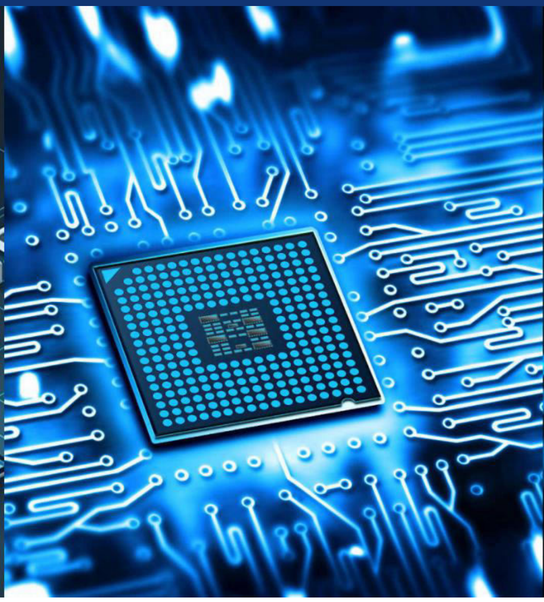
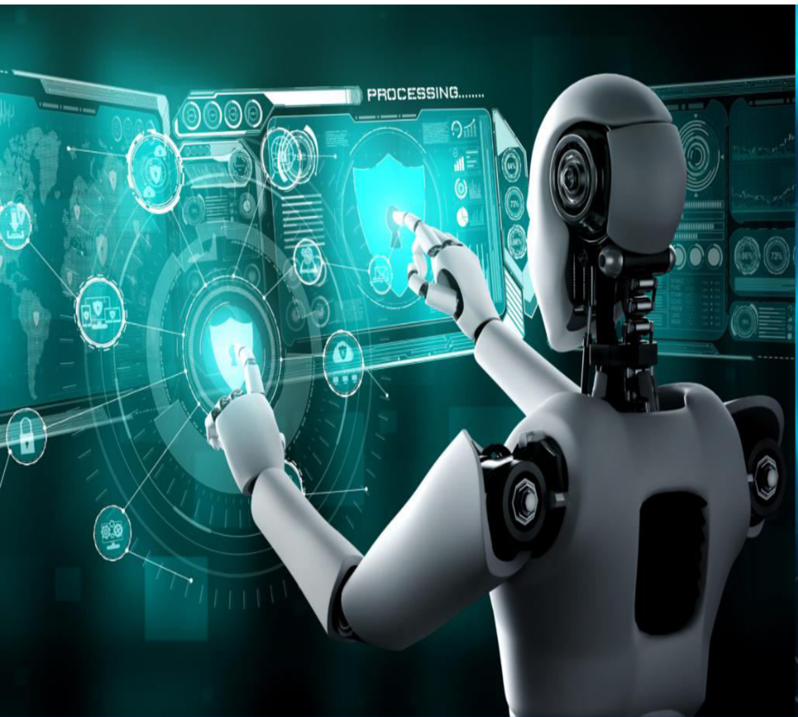
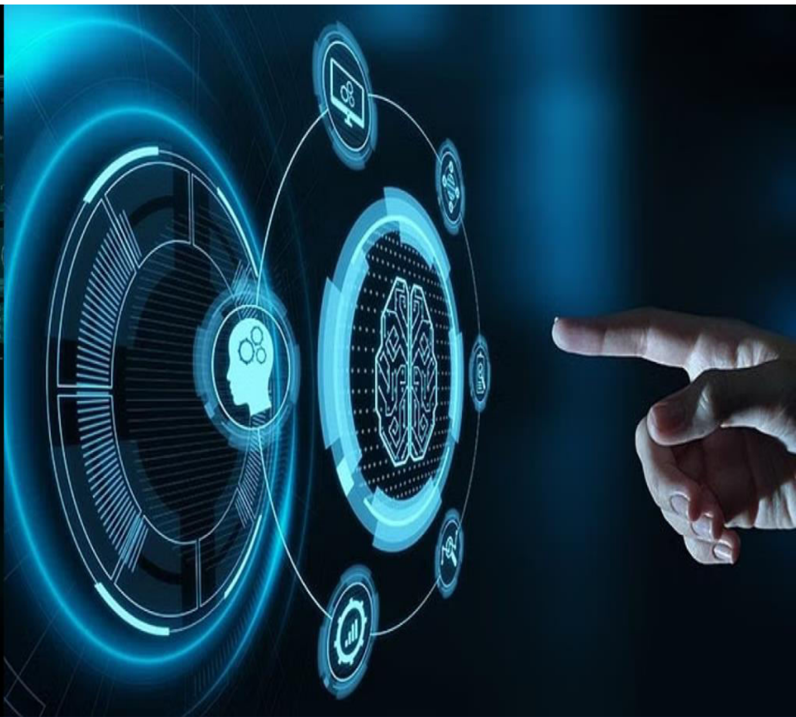
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 3, March 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303053

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Sentiment Analysis System: Transforming Product Reviews into Business Intelligence

Ilakkiya N, Lavanyagayathri V, Jisha Aravind, Ms. Sajitha L P

Department of Computer Science and Business Systems, R.M.K Engineering College, An Autonomous Institution, Thiruvallur, India

Department of Computer Science and Business Systems, R.M.K Engineering College, An Autonomous Institution, Thiruvallur, India

Department of Computer Science and Business Systems, R.M.K Engineering College, An Autonomous Institution, Thiruvallur, India

Assistant Professor, Department of Computer Science and Business Systems, R.M.K Engineering College, An Autonomous Institution, Thiruvallur, India

ABSTRACT: In the modern digital economy, customer reviews significantly influence business decisions. However, analyzing large volumes of textual feedback manually is inefficient and prone to bias. This project presents an AI-powered Sentiment Analysis System that automates sentiment classification, transforming customer reviews into actionable business intelligence. The system employs Natural Language Processing (NLP) and Machine Learning (ML) techniques to classify text into positive, negative, or neutral sentiments. Built using Flask (backend), Python (ML processing), and HTML, CSS, and JavaScript (frontend), it provides an interactive web-based interface for sentiment analysis. Users can analyze individual reviews or upload CSV files for batch processing, making it suitable for businesses handling large-scale customer feedback. The sentiment classification model utilizes TF-IDF vectorization and a Multinomial Naïve Bayes classifier, ensuring high accuracy. Additionally, the system features a dashboard for data visualization, including confusion matrices, sentiment distributions, and heatmaps, providing deeper insights into customer perceptions. By automating sentiment analysis, this project enhances decision-making, reduces manual effort, and improves response strategies for businesses. The combination of machine learning, scalability, and real-time visualization makes it a powerful tool for companies aiming to optimize their products and services based on customer feedback.

KEYWORDS: Business Analytics, Sentiment Analysis, Machine Learning, TextBlob, Python, Streamlit, Real-time Analysis, Product Reviews, .CSV Processing

I. INTRODUCTION

In the modern digital landscape, customer reviews significantly influence business strategies by offering direct insights into consumer experiences, satisfaction levels, and product performance. With the widespread availability of online platforms, customers can freely express their opinions, making it essential for businesses to analyze feedback effectively. Sentiment analysis, a technique that utilizes natural language processing (NLP) and machine learning, helps determine the overall sentiment conveyed in textual data. By leveraging sentiment analysis, businesses can monitor public perception, recognize emerging trends, and respond to feedback in a timely manner. This study focuses on developing an AI-driven sentiment analysis system designed to classify customer reviews into positive, negative, or neutral categories. Furthermore, it aims to integrate an interactive web-based platform where users can upload CSV files containing reviews for batch processing and gain valuable insights through dynamic data visualization, ultimately aiding in more informed decision-making.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE SURVEY

Various researchers have researched and developed this project. However,they serve different purposes and have different technologies being implemented. Some of those papers are mentioned below stating their technology and application. Several studies utilize machine learning and lexicon-based methods for sentiment analysis. D'Andrea et al. [1] explore various implementations,highlighting the advantages of accuracy, versatility, and interpretability, while acknowledging data dependency and contextual challenges. Alsaeedi and Khan [5] and Chakraborty and Bhattacharyya [6] focus on Twitter and social media data, respectively, emphasizing real-time analysis and large data volumes, but note issues with noisy data and privacy. Abirami and Gayathri [7] also present a survey on sentiment analysis methods, reiterating the same pros and cons. Mudinas et al. [9] combine lexicon and learning-based approaches, noting benefits like simplicity and domain adaptation, but also limitations in handling context. Yi and Liu [14] demonstrate the use of machine learning for customer sentiment analysis in recommendation systems, showcasing personalized recommendations but raising concerns about data privacy. Kumar and Sebastian [15] provide a perspective on the past, present, and future of sentiment analysis, discussing the evolution of techniques and highlighting both the advancements and persistent challenges. Rashid Rana et al. [2] propose a hybrid computational framework for aspect-based sentiment analysis in social multimedia, enabling fine-grained and multimodal analysis, though it introduces complexity. Pozzi et al. [8] provide an overview of the challenges of sentiment analysis in social networks, discussing aspect-based sentiment analysis and emotion detection, and highlighting issues such as language bias and contextual ambiguity. Kaminska et al. [11] utilize Fuzzy Rough Nearest Neighbour methods for aspect-based sentiment analysis, showcasing granular analysis and uncertainty handling, but noting computational complexity. Derbentsev et al. [3] conduct a comparative study of deep learning models for sentiment analysis of social media texts, emphasizing high accuracy and scalability, but noting the need for large datasets and interpretability challenges. Dragoni et al. [4] and Srishti and Vashishtha [10] explore fuzzy logic for concept-level sentiment analysis, offering granularity and adaptability, but also introducing complexity and resource intensity. Wanniarachchi et al. [12] combine sentiment analysis with topic modeling and discourse analysis to study hate speech patterns, providing a holistic understanding but facing complexity and subjectivity. Bautin et al. [13] focus on international sentiment analysis for news and blogs, highlighting global perspectives but acknowledging language challenges and bias.

III. SYSTEM ARCHITECTURE

Being run on artificial intelligence, the Sentiment Analysis System is meant to allow automatic classification of customer reviews using methods from Natural Language Processing (NLP) and Machine Learning (ML). Through uploads in CSV form, the system lets users either study specific reviews or handle a lot of data. With CSS animations used to augment the user experience overall, the frontend is created using HTML, CSS, JavaScript, and Flask templates (Jinja2). Flask runs the backend, which integrates machine learning models, handles requests, and offers user sessions. For feature extraction and sentiment analysis with reliable predictions, the ML model uses a TFIDF vectorizer together with a Multinomial Naïve Bayes classifier. Temporary storage using dictionaries helps with user session handling. The Seaborn and Matplotlib data visualization libraries produce informative visualizations including confusion matrices, sentiment distribution, and heatmaps. The system offers effective realtime sentiment analysis, therefore making it interactive, scalable, and usable for businesses handling great numbers of consumer comments.

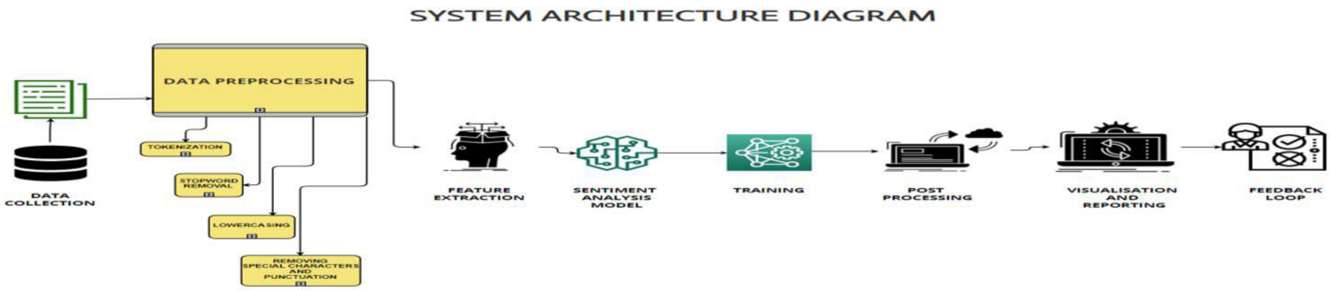


Fig. 1. System Architecture

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303053

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

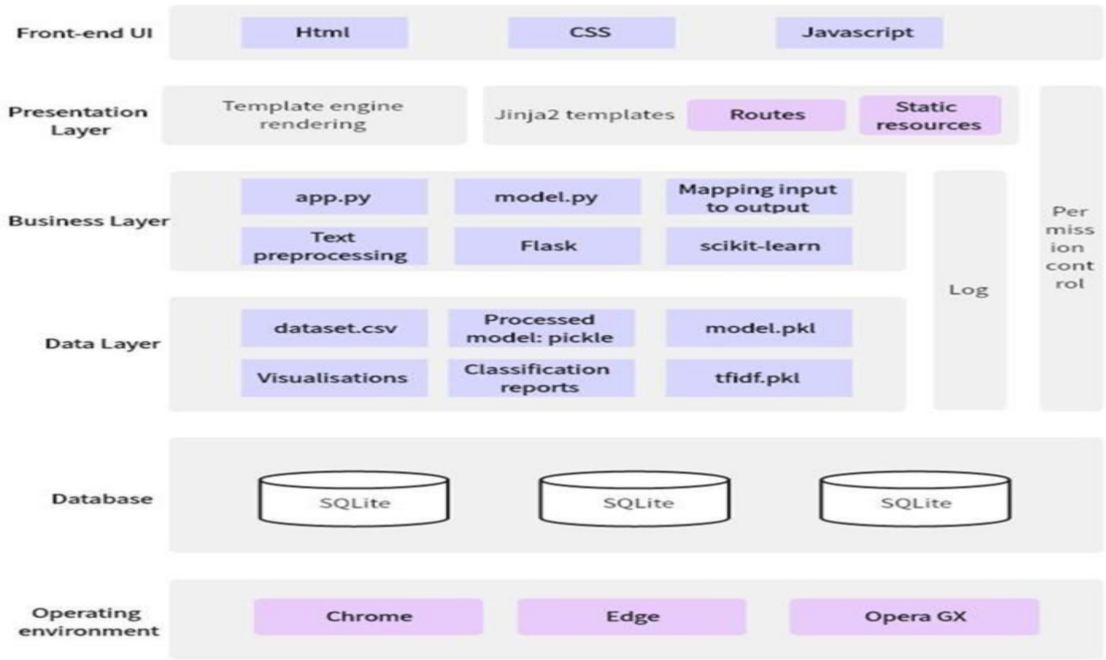


Fig. 2. Layered Architecture

Sentiment analysis relies on several layers of system architecture, each of which serves a different purpose. HTML, CSS, and JavaScript are used at the top to construct the Frontend UI, therefore guaranteeing a visually pleasing and interactive user interface. Using Jinja2 templates, the Presentation Layer renders dynamic content, controls paths, and deals static assets. Flask based application (app.py), machine learning model (model.py), and text preprocessing tools are among the core features of the Business Layer. It also uses scikit-learn to train models and make predictions, therefore mapping input data to output classifications. Supporting data visualization and analysis, the Data Layer keeps critical documents including datasets (dataset.csv), preprocessed model files (pickle, model.pkl, tfidf.pkl), and classification reports. At the Database Level, SQLite databases are employed to securely house organized data. Including Chrome, Edge, and Opera GX, the whole system runs in many browsers to guarantee cross-platform compatibility. Logging and permission control, among other functionalities, improve system security and maintenance.

IV. METHODOLOGY

A. NLP

A technique used in natural language processing (NLP). NLP technology allows the system to effectively process thestructuring, optimization, and readiness of the large amount of textual data and important properties needed for effective sentiment analysis. The main NLP technologies used in this system are tokenization, stop word removal, trunk, and lemmatization. It is an important process for subsequent word processing, as opposed to sentences and whole paragraphs, allowing the system to process manageable text units. Splitting the text into tokens enables sample analysis to be performed and enables the effective use of algorithms for machine learning. Some words in sentences are less wise and have redundant information values for sentiment analysis purposes. Deleting such ubiquitous words minimizes noise and calculation optimization in the data record. Deleting stop words ensures that only words that help classify sentiment are left for further analysis. Stemming performs word reduction by removing suffixes and prefixes to achieve rough but effective transformations (e.g. "execute") and contribute to creating grammatically correct words. Lemmatization is more advanced and converts words into a dictionary with the correct form, but maintains semantic meaning (for example, "better" is converted to "good"). Together they normalize textual data, minimizing redundancy and improving model accuracy by treating words of different forms as a single unit.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. ML (Multinomial Naïve Bayes*)*

Using them, the textual data is improved, formatted and enables efficient sentiment analysis, and provides clear and uniform data for mutual circulation. Sentiment classification is performed in the Multinomial Naïve Bayes classifier, a probabilistic classification used in text classification problems, particularly in the Multinomial Naïve Bayes classifier. This special classifier shows knowledge about dealing with discrete properties like the frequency of words and terminology, and is thus, a very suitable candidate for sentiment analysis. Each data record check is linked to a specified sentiment label (positive, negative, or neutral). This monitored learning technology allows models to identify patterns and associations of word and sentiment designations and thus make intelligent predictions about new, invisible texts. This process changes the oral information to a numerical formula that reflects the value of each word in the body.

C. Text Illustration with TF-IDF Vectorization

Using Term Frequency Inverse Document Frequency (TFIDF) vectorization, this research changes text data into well organized numerical features appropriate for machine learning. Text preprocessing, which guarantees consistency by removing superfluous components like mentions, URLs, special characters, and uppercase letters, starts off the process. After preprocessing, the TFIDF vectorizer assigns numerical weights twofold to terms based on two factors: term frequency (TF), which counts word occurrences inside a document, and inverse document frequency (IDF), which gauges word rarity across the entire dataset. These figures' product gives a TFIDF rating that underlines word relevance compared to the corpus. Machine learning models can now be effectively used for sentiment classification. Frequent words receive lower weights. Words that are often displayed in certain documents but rarely appear throughout the body are administered a higher weight. Vectorization of TF-IDF helps the model better understand how important a particular term is in the sentiment of a decision and improves the accuracy of classification. Based on the obtained patterns and probability calculations, the classifier examines the text input and checks the sentiment label. High accuracy of sentiment classification is guaranteed by a mix of a powerful NLP preprocessing pipeline and a complete Naïve Bayesian model of wells. Therefore, this system is suitable for practical employment.

D. Data Preprocessing

Valuable data points, these ratings are the opinions, comments and feelings of users being promoted. One of the three sentiment categories is either manually assigned to each rating or neutral is positive or neutral. Training monitored machine models that can be effectively generalized to new, unobserved reviews of this labelled dataset. Text cleaning that does not require components such as users (such as "http://example.com"), URLs (such as "http://example.com"), special signs, and excessive distances for preprocessing. This step ensures that your data remains in important textual material and removes unnecessary noise. Normalizing text with small letters avoids duplication and ensures uniformity of the data record. This is because we consider machine learning models to be entities with different characters, large and small. This step improves model functionality extraction and model learning efficiency. After text cleaning and normalization. Deleting stop words eliminates unnecessary words. Tribal and lemmatization reduce words to root format for consistency. Tokenization divides text into individual words. These methods change the text to a structured format suitable for machine learning. This change allows the model to turn text into quantities, improving its ability to improve patterns related to emotion.

V. CONCLUSION

This study successfully develops an AI-driven sentiment analysis system designed to classify customer reviews into positive, negative, or neutral sentiments, providing businesses with actionable insights for better decision-making. The system incorporates an interactive web-based interface that enables batch processing and data visualization, ensuring ease of use and accessibility. The findings highlight the effectiveness of the Multinomial Naïve Bayes classifier and emphasize the growing importance of sentiment analysis in today's digital economy. By leveraging this system, businesses can refine their products, enhance customer satisfaction, and stay competitive by addressing consumer needs based on real-time feedback. Future improvements will focus on integrating advanced machine learning models, expanding the dataset to include diverse sources, and incorporating real-time sentiment analysis capabilities. Additionally, implementing deep learning techniques and transfer learning will further enhance accuracy, while refining the user interface will provide more intuitive visualizations to support data-driven decision-making.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. RESULTS AND FUTURE ENHANCEMENTS

The sentiment analysis model demonstrates strong performance in classifying customer reviews, with accuracy, precision, recall, and F1-score validating its effectiveness. The system efficiently processes large datasets while maintaining reliable sentiment classification, though areas such as handling ambiguous sentiments and mixed reviews offer scope for improvement. Real-world applications across industries like e-commerce, hospitality, and entertainment showcase its ability to monitor customer feedback, identify trends, and refine business strategies. By analyzing sentiment trends, companies gain valuable insights into customer preferences, allowing them to improve products, enhance satisfaction, and make data-driven decisions. Case studies highlight its impact, such as e-commerce platforms optimizing product recommendations, hotels tracking guest satisfaction, and entertainment services gauging audience sentiment. While the system effectively transforms textual data into actionable intelligence, challenges remain in nuanced sentiment interpretation, sarcasm detection, and multilingual analysis. Future enhancements, including deep learning integration, aspect-based sentiment analysis, and real-time sentiment tracking, will further improve its accuracy and adaptability, ensuring its continued relevance in the evolving digital landscape.

REFERENCES

[1] D'Andrea et al., "Approaches, Tools, and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications, 2015.
[2] Rashid Rana et al., "Aspect-Based Sentiment Analysis for Social Multimedia: A Hybrid Computational Framework", Tech Science Press.
[3] Derbentsev et al., "A Comparative study of deep Learning models for sentiment analysis of social media texts", CEUR-WS.
[4] Dragoni et al., "A Fuzzy System for Concept-Level Sentiment Analysis", Springer.
[5] Alsaeedi and Khan, "A Study on Sentiment Analysis Techniques of Twitter Data", IJACSA.
[6] Chakraborty and Bhattacharyya, "A Survey of Sentiment Analysis from Social Media Data", IEEE.
[7] Abirami and Gayathri, "A Survey On Sentiment Analysis Methods And Approach ", IEEE.
[8] Pozzi et al., "Challenges Of Sentiment Analysis In Social Networks: An Overview", Morgan Kaufmann.
[9] Mudinas et al., "Combining Lexicon and Learning-based Approaches for Concept-Level Sentiment Analysis", ACM.
[10] Srishti and Vashishtha, "Design And Development Of A Fuzzy Logic-Based Sentiment Analysis System For Online Reviews & Social Media Posts", Dspace JSPUI.
[11] Kaminska et al., "Fuzzy Rough Nearest Neighbour Methods for Aspect-Based Sentiment Analysis", MDPI.
[12] Wanniarachchi et al., "Hate Speech Patterns in Social Media: A Methodological Framework and Fat Stigma Investigation Incorporating Sentiment Analysis, Topic Modelling and Discourse Analysis", PKP.
[13] Bautin et al., "International Sentiment Analysis for News and Blogs", PKP.
[14] Yi and Liu, "Machine learning based customer sentiment analysis for recommending shoppers, and shops based on customers' review", Springer.
[15] Kumar and Sebastian, "Sentiment Analysis: A Perspective on its Past, Present and Future", MECS.