STM JOURNALS

Journal of
Software Engineering Tools & Technology Trends

ISSN: 2394-7292
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoSETTT

https://journals.stmjournals.com/josettt

Review JoSETTT

Applying Text Analysis Methods for Emotion Recognition

Anushka Yeole[1], Kalyani Alisetty[2,*]

Abstract

This article presents a comprehensive study of sentiment analysis, a vital task in the realms of natural language processing (NLP) and artificial intelligence (AI). Sentiment analysis involves the extraction and classification of subjective information from textual data, determining whether the sentiment expressed is positive or negative. This paper investigates different approaches and methodologies used in sentiment analysis, encompassing machine learning models as well. Additionally, it discusses the challenges faced in sentiment analysis, such as handling sarcasm, irony, and context-dependent sentiment. Furthermore, the paper highlights the intersection of sentiment analysis with other NLP tasks, such as text summarization using Bag of Words (BoW) and term frequency–inverse document frequency (TF-IDF) and emphasizes the importance of understanding human language nuances for accurate sentiment interpretation. This project leverages MLflow and Prefect to manage, track, and orchestrate sentiment analysis machine learning experiments, ensuring efficient workflow automation and comprehensive performance evaluation. The integration of MLflow and Prefect not only enhances the reproducibility of experiments but also facilitates the seamless deployment of sentiment analysis models into production environments. By addressing the evolving complexities in textual data interpretation, this project aims to advance the field of sentiment analysis, contributing to more nuanced and accurate sentiment detection systems. Through an examination of recent research trends and applications, this paper underscores the significance of sentiment analysis in diverse domains ranging from social media monitoring to market research.

Keywords: Text summarization, sentiment analysis, natural language processing (NLP), artificial intelligence (AI), human language, feature extraction techniques, data science life cycle

INTRODUCTION

Sentiment analysis is a process that uses artificial intelligence to identify and analyze the emotions and feelings expressed in words. With the ever-increasing volume of data generated daily, gaining insights into customer sentiments and detecting urgent issues in real-time has become crucial for businesses to maintain their competitive edge and ensure customer satisfaction. Businesses can use sentiment analysis to understand public opinion of their products, services, and reputation, as well as to build trust with customers and understand their target audience's needs [1].

Sentiment analysis is crucial for businesses, enabling them to stay updated on customer feelings, swiftly detect emerging issues, and proactively

***Author for Correspondence**
Kalyani Alisetty
E-mail: kalyani.koukuntla@sihgad.edu

[1]Student, Department of Management, Sinhgad Institute of Business Administration and Research, Pune, Maharashtra, India
[2]Assistant Professor, MCA, Sinhgad Institute of Business Administration and Research, Pune, Maharashtra, India

Received Date: June 04, 2024
Accepted Date: July 11, 2024
Published Date: July 22, 2024

Citation: Anushka Yeole, Kalyani Alisetty. Applying Text Analysis Methods for Emotion Recognition. Journal of Software Engineering Tools & Technology Trends. 2024; 11(2): 12–22p.

resolve them. By utilizing tools and techniques for sentiment analysis, businesses can boost customer satisfaction, foster growth, and uphold a competitive advantage in today's data-centric market.

Here is how sentiment analysis can help address this need.

Real-Time Monitoring

Sentiment analysis tools can continuously monitor social media platforms, review websites, customer support tickets, and other sources of customer feedback in real time. This enables businesses to promptly identify emerging issues or trends related to their products, services, or brand reputation.

Issue Detection and Prioritization

By analyzing the sentiment of customer feedback, businesses can quickly detect and prioritize urgent issues that require immediate attention. Sentiment analysis algorithms can automatically flag negative sentiments, enabling support teams to address customer concerns promptly and prevent them from escalating.

Trend Analysis

Sentiment analysis aids businesses in recognizing common themes or subjects in customer feedback. By examining sentiment patterns over time, businesses can gather valuable insights into shifting customer preferences, emerging market trends, and opportunities for enhancement.

Proactive Customer Engagement

Armed with real-time insights from sentiment analysis, businesses can proactively engage with customers to address their concerns, offer assistance, or gather additional feedback. This proactive approach not only demonstrates a commitment to customer satisfaction but also helps in building stronger customer relationships.

Brand Reputation Management

Sentiment analysis enables businesses to track online discussions concerning their brand and products, enabling early detection of potential public relations (PR) crises or adverse publicity. By promptly addressing negative sentiments and managing brand reputation effectively, businesses can mitigate potential damage to their brand image.

Product Development and Innovation

Through the analysis of customer feedback and sentiment, businesses can acquire valuable insights into customer preferences, challenges, and unfulfilled requirements. These insights can guide product development efforts, enabling businesses to craft products and services that more closely match customer expectations and preferences.

PURPOSE

The Art of Understanding Human Emotions Through Sentiment Analysis

Sentiment analysis is not just a scientific tool but also an art form. Sentiment analysis entails examining textual data to ascertain the sentiment or emotional expression it conveys.

Sentiment Analysis: A Key Tool for Predicting Consumer Behavior

It highlights the practical application of sentiment analysis in the realm of market research and consumer behavior prediction. Through the analysis of sentiment conveyed in consumer-generated content such as online reviews, social media posts, and forum discussions, businesses can glean valuable insights into consumer perceptions of products, brands, or services. These insights can inform strategic decisions related to marketing, product development, and customer engagement, making sentiment analysis a valuable tool for businesses seeking to understand and anticipate consumer behavior.

How Sentiment Analysis Can Help Companies Build Stronger Customer Relationships

Through vigilant monitoring and analysis of customer sentiment across diverse channels like social media, customer service interactions, and online reviews, organizations can pinpoint areas for enhancement, promptly resolve customer concerns, and tailor their interactions to align with customer needs and preferences. Utilizing sentiment analysis in this way can contribute to building trust, loyalty, and satisfaction among customers, ultimately leading to stronger and more enduring relationships.

The Future of Market Research: Predicting Trends with Sentiment Analysis

Traditional market research methods often rely on surveys, focus groups, and demographic data to understand consumer preferences and predict market trends. Nevertheless, sentiment analysis presents a dynamic and real-time method that involves analyzing extensive unstructured data sourced from online platforms. It enables the detection of emerging trends, recognition of shifts in consumer sentiment, and discovery of concealed insights. By leveraging these capabilities, businesses can gain a competitive advantage in foreseeing market fluctuations and adjusting their strategies proactively.

The Role of Sentiment Analysis in Enhancing Mental Health Research

Researchers can leverage sentiment analysis to study patterns of emotional expression, detect indicators of mental health issues, and monitor changes in mood or sentiment over time. By integrating sentiment analysis with other psychological assessments and clinical data, researchers can gain valuable insights into mental health trends, treatment outcomes, and the effectiveness of interventions, ultimately enhancing our understanding and management of mental health disorders.

OBJECTIVE

We are conducting customer sentiment analysis to categorize reviews into positive or negative sentiments. Our goal is to identify product improvement opportunities by addressing concerns highlighted in negative feedback, ultimately aiming to boost overall customer satisfaction [2].

We performed sentiment analysis on real-time tweets to identify market status and develop trading strategy by implementing a data science life cycle as shown in Figure 1.

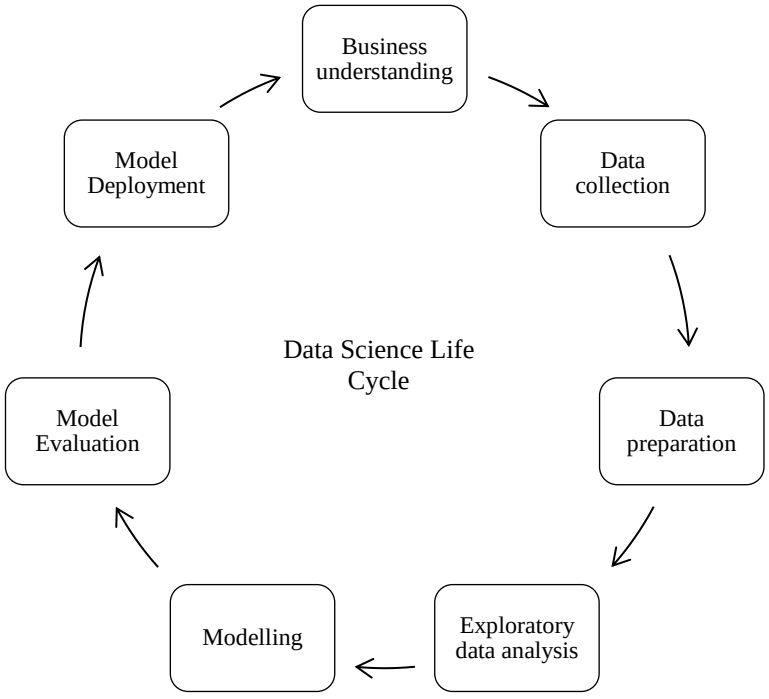


Figure 1. Data science life cycle.

APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis has widespread applications across different fields and sectors, providing valuable insights into customer sentiments, preferences, and emotional responses [3]. Some notable uses include the following:

- *Social Media Monitoring:* Examining sentiments conveyed across social media platforms like X (formerly Twitter), Facebook, and Instagram aids in comprehending public opinion, monitoring brand perception, and recognizing evolving trends.
- *Customer Feedback Analysis:* By examining customer reviews, surveys, and feedback, we assess customer satisfaction, pinpoint areas for enhancement, and customize products or services to align with customer expectations.
- *Market Research:* Analyzing market trends, consumer preferences, and competitor strategies to guide marketing strategies, product development, and business decisions through sentiment analysis.
- *Brand Monitoring:* Monitoring mentions and sentiment surrounding a brand or product in online discussions, news articles, and reviews to assess brand perception, track sentiment trends, and manage brand reputation.
- *Customer Service Optimization:* Analyzing customer interactions, such as emails, chat logs, and support tickets, to identify sentiment patterns, detect customer issues, and improve the quality of customer service.
- *Political Analysis:* Examining public sentiment regarding political candidates, parties, and policies is crucial for comprehending voter preferences, forecasting election results, and guiding political campaigns.
- *Financial Markets:* Analyzing sentiment in financial news, social media discussions, and analyst reports to gauge market sentiment, predict stock price movements, and identify investment opportunities.
- *Healthcare:* Analyzing patient reviews, social media discussions, and medical records to assess patient satisfaction, detect adverse events, and monitor public health trends.
- *Product Development:* Incorporating sentiment analysis into the product development process to gather feedback from early adopters, identify feature requests, and prioritize product improvements.
- *Human Resources:* Analyzing employee feedback, sentiment in internal communications, and reviews on employer review sites to assess employee morale, identify areas for improvement, and enhance employee engagement.

In general, sentiment analysis presents diverse applications across various sectors, delivering valuable insights that empower informed decision-making, elevate customer experiences, and optimize business results.

Data Science Life Cycle

The data science life cycle encompasses various stages, including data collection, data cleaning, data exploration, model building, and model deployment. Each phase plays a critical role in transforming raw data into actionable insights and data-driven decisions [4].

Business Understanding

In the dynamic world of business, paying close attention to the requirements and feelings of your customers is crucial. In today's interconnected world, customers are exceptionally vocal and connected, utilizing numerous online platforms to express their opinions, feedback, and emotions. The abundance of digital expression offers valuable insights to businesses [5]. Through the utilization of customer sentiment analysis, brands can develop strategies that not only meet but surpass customer expectations, thus improving their business performance.

Businesses can leverage customer sentiments for development in several ways:

1. Product development.
2. Marketing strategy.
3. Customer experience enhancement.
4. Reputation management.
5. Competitive analysis.
6. Predictive analytics.

Data Collection

In collecting data from Flipkart reviews through web scraping, it's crucial to ensure ethical and legal compliance with Flipkart's terms of service and data usage policies. The process involves extracting relevant information such as customer ratings, comments, and product details from the reviews section. It's essential to implement robust data scraping techniques to handle dynamic web pages and potential anti-scraping measures [6]. Additionally, considering the volume and diversity of reviews, effective data filtering and cleaning methods are necessary to ensure the quality and reliability of the collected data. Furthermore, it is crucial to ensure user privacy and confidentiality are upheld throughout the entirety of the data collection process.

- Data Scrapped From: Flipkart Website
- Customer Reviews for Product: Sports – YONEX MAVIS 350 Nylon Shuttle
- Website Link: https://www.flipkart.com/yonex-mavis-350-nylon-shuttle-yellow/p/itmfcjdyhnghfyey?pid=STLEFJ7UFQGRUUR3

Data Preparation

To prepare the data and to train it, we go through several steps of data preparation which are listed below:

- *Data Cleaning:* Addressing missing values, duplicate entries, and inconsistencies in the dataset to ensure data quality and accuracy.
- *Emoji Conversion:* Utilizing the demojize function to convert emojis into text format, enabling their inclusion in text analysis and modeling tasks.
- *Outlier Detection:* Detecting and addressing outliers in the data is crucial to mitigate their potential impact on model performance.
- *Text Preprocessing:* Applying techniques such as stemming and lemmatization to standardize and normalize text data, reducing variation and improving model generalization.
- *Feature Engineering:* Developing new features or modifying existing ones to extract valuable insights and improve the predictive accuracy of models.
- Bag of Words and term frequency–inverse document frequency: Text data can be transformed into numerical format using methods such as Bag of Words (BoW) or term frequency–inverse document frequency (TF-IDF). BoW converts text into a matrix based on word frequencies, while TF-IDF evaluates word significance within a document compared to a corpus.
- *Data Preparation for Model Training:* Converting the processed text data into a format suitable for training machine learning models involves organizing it into input features (X) and corresponding target labels (Y).

Cleaned/Preprocessed Data

The raw data undergoes various preprocessing steps such as tokenization, stop word removal, and normalization to enhance the quality and consistency of the dataset [7]. This cleaned data is essential for improving the accuracy and performance of sentiment analysis models as shown in Figure 2.

Exploratory Data Analysis

By conducting exploratory data analysis (EDA), you gain insights into the underlying structure and characteristics of the dataset, which informs subsequent steps in the data preparation and modeling

	Review text	Review Title	clean_text_stem	text_length_stem	clean_text_lemma	text_length_lemma
6289	nice.. shuttle	Good choice	nice shuttlgood choic	4	nice shuttlegood choice	4
549	product are good only, but we received the sto...	Nice	product good receiv stock month befor manufact...	8	product good received stock month befor manufa...	8
4707	excellent service, got in one day even at remo...	Waste of money!	excel servic got one day even remot locat like...	12	excellent service got one day even remote loca...	12
764	good but high price...	Good quality product	good high pricegood qualiti product	6	good high pricegood quality product	6
6861	there are 2 damaged shuttles out of 6...not sat...	Awesome	damag shuttl not satisfiawesom	5	damaged shuttle not satisfiedawesome	5

Figure 2. Cleaned/preprocessed data.

pipeline [8]. It helps in making informed decisions about feature selection, preprocessing techniques, and modeling approaches, ultimately leading to more robust and accurate machine learning models.

Word Cloud

A word cloud is a visual representation of text data, where the size of each word indicates its frequency or importance within the dataset [9]. It is commonly used to quickly identify prominent terms and themes in textual information as shown in Figure 3.

- The word cloud provides insights into common themes and sentiments expressed by reviewers.
- Positive terms like "good," "nice," and "quality" are prevalent, suggesting overall satisfaction.
- Negative terms like "bad," "poor quality," and "worst" also appear, indicating areas for improvement.

Heatmap for Rating Distribution

A heatmap is a powerful visualization tool that represents data in a matrix format where individual values are indicated by varying colors [10]. When applied to rating distributions, a heatmap can provide insightful visual summaries of how ratings are distributed across different categories or periods as shown in Figure 4.

- The heatmap provides insights into how these three variables relate to each other. For example, the positive correlation between Up Votes and Down Votes suggests that when one increases, the other tends to increase as well.
- Figure 4 is a heatmap which is divided into three rows and columns, representing three variables: "Up Votes," "Down Votes," and "Ratings."
- Each cell in the matrix displays a correlation coefficient ranging from -0.01 to 1.00.
- Cells are color-coded according to their correlation coefficients:
  - Red indicates positive correlation.
  - Blue indicates negative correlation.
  - The intensity of the color represents the strength of the correlation.
- Up Votes, Down Votes, and Ratings with itself has a coefficient of 1.00 (red).
- Up Votes and Down Votes have a positive correlation of 0.63 (light red).
- Both combinations involving Ratings have almost no significant correlations (-0.01), colored in blue.

Interactive Histogram Plot Showing the Ratings of the Product

Figure 5 shows the bar plot distribution of Ratings which fall under:

- The majority of the ratings fall into either the lowest (1) or highest (5) categories.
- Very few ratings fall in the middle range (2, 3, and 4).

Figure 3. Word cloud of review text.

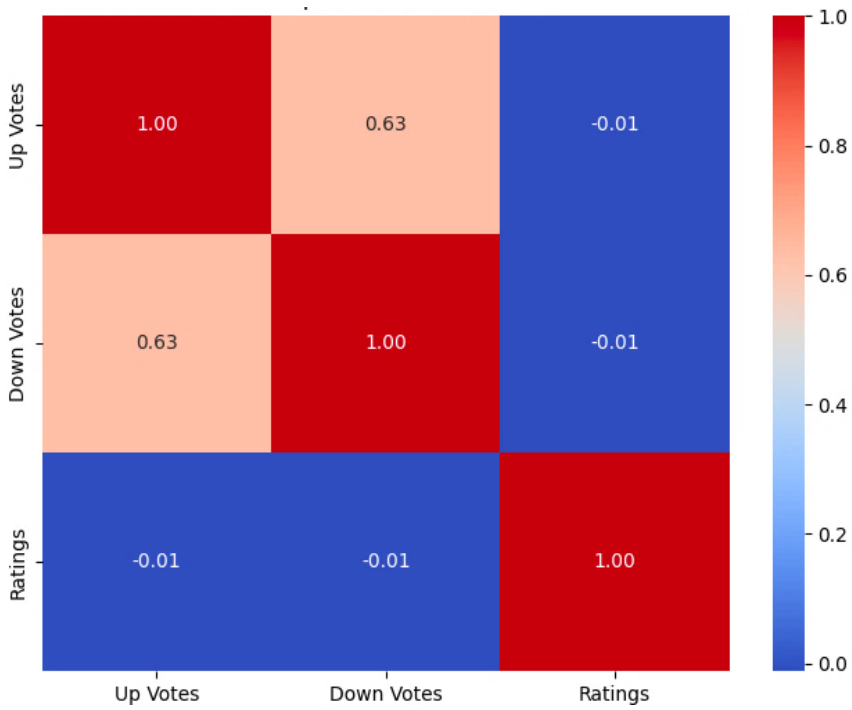


Figure 4. Heatmap of correlation matrix.

Overall, this histogram suggests that the product or service being rated tends to receive extreme ratings (either very positive or very negative), with fewer moderate ratings.

Modeling

We do modeling to extract insights, make predictions, or classify data based on patterns and relationships discovered within the dataset, helping to solve problems, make informed decisions, and optimize processes. Throughout the modeling phase, it is essential to iterate on the process, experimenting with different algorithms, features, and parameters to improve model performance continually. Collaboration between data scientists, domain experts, and stakeholders is key to developing models that address the business problem effectively and deliver actionable insights.

The Trained Models Accuracy Score

The trained model's accuracy score reflects its performance in correctly predicting sentiment. A high accuracy score indicates the model's effectiveness in understanding and classifying sentiments accurately as shown in Table 1.

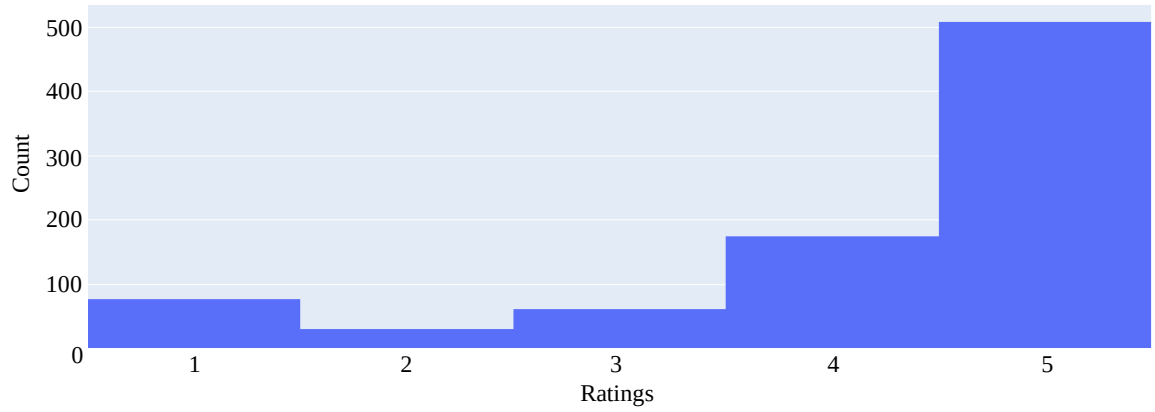
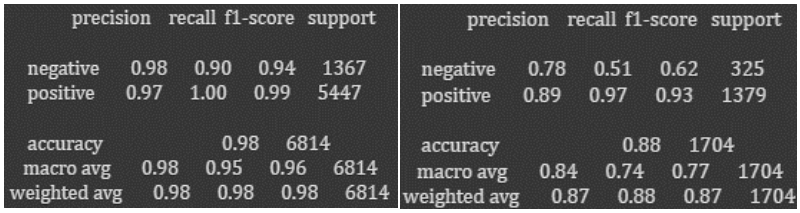


Figure 5. Interactive histogram of ratings.

Table 1. Training accuracy score.

Algorithm	Training accuracy score	Test accuracy score
Decision tree	0.97	0.86
Logistic regression	0.92	0.87
K-nearest neighbor (KNN)	0.88	0.85
Naïve Bayes	0.91	0.88
Random forest	0.97	0.87



	precision	recall	f1-score	support
negative	0.98	0.90	0.94	1367
positive	0.97	1.00	0.99	5447
accuracy			0.98	6814
macro avg	0.98	0.95	0.96	6814
weighted avg	0.98	0.98	0.98	6814

	precision	recall	f1-score	support
negative	0.78	0.51	0.62	325
positive	0.89	0.97	0.93	1379
accuracy			0.88	1704
macro avg	0.84	0.74	0.77	1704
weighted avg	0.87	0.88	0.87	1704

Figure 6. Classification report.

The best model from all the algorithms (Table 1) is a random forest model, which gives training accuracy score of 0.97 and test accuracy score of 0.87 whose classification report is also giving great results as shown in Figure 6. Data in the left part of Figure 6 is the classification report of random forest model for training data and the data in the right part is the classification report of random forest model for the test data.

Model Evaluation

In model evaluation, we assess the performance of machine learning models using appropriate metrics and techniques, such as train-test splitting, cross-validation, confusion matrices, receiver operating characteristic (ROC) curves, and precision-recall curves. We evaluate models, assess overfitting or underfitting, and analyze results to guide decisions on model selection, optimization, and deployment.

- Evaluating the trade-off between precision and recall in classification models becomes crucial, particularly in scenarios involving imbalanced datasets where one class significantly outweighs the other.
- Comparing the performance of multiple models using the selected evaluation metrics to identify the best-performing model for the specific problem domain.
- Checking for signs of overfitting (high training accuracy but low-test accuracy) or underfitting (low training and test accuracy) to ensure that the model generalizes well to unseen data.
- Interpreting evaluation metrics in the context of the problem domain and business objectives, considering factors such as class imbalance, cost considerations, and the desired trade-offs between different metrics.

Model Deployment

The model is deployed on an EC2 (Amazon Elastic Compute Cloud) instance by configuring the instance type and setting up the appropriate security group. The instance type is selected based on the computational requirements of the model and the expected workload. The security group defines the inbound and outbound traffic rules, specifying which IP addresses or ranges are allowed to access the instance and which ports are open for communication. This ensures that the deployed model is accessible securely and can handle incoming requests effectively.

Similarly, deploying the model on a local server involves configuring the server environment and hosting the model within it. Both deployments enable the model to be accessed and utilized for making predictions or serving requests.

Flow of How Sentiment Analysis Work in Industry to Analyze the Overall Customer Satisfaction

Sentiment analysis in industry involves collecting customer feedback from various sources, such as social media, reviews, and surveys. This data is then processed using natural language processing (NLP)

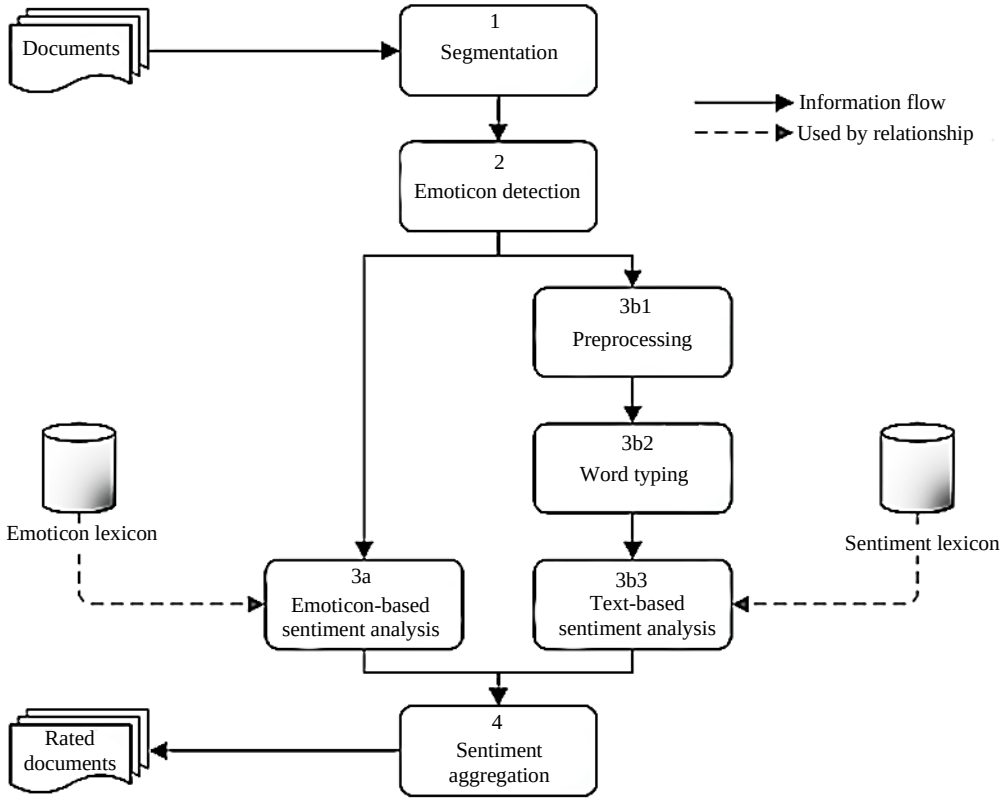


Figure 7. Sentiment analysis work in industry.

algorithms to determine overall customer satisfaction by categorizing sentiments as positive, negative, or neutral as shown in Figure 7.

CONCLUSION

In conclusion, our research paper addresses the comprehensive process of developing and deploying a sentiment analysis model. We have highlighted the importance of data preprocessing, feature extraction, and model selection in preparing the data and building effective machine learning models. Through the evaluation of various algorithms and the deployment of the model on both an EC2 instance and a local server, we have demonstrated the practical application and scalability of our approach. Overall, this research contributes to advancing the understanding and implementation of sentiment analysis in real-world scenarios, paving the way for more informed decision-making and enhanced user experiences in diverse domains.

REFERENCES

1. Hogenboom A, Bal D, Frasincar F, Bal M, De Jong F, Kaymak U. Exploiting emoticons in sentiment analysis. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, Coimbra, Portugal, March 18–22, 2013. pp. 703–710.
2. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. September 1, 2014. Available at https://arxiv.org/abs/1409.0473
3. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. [Online]. 2018. Available at https://www.mikecaptain.com/resources/pdf/GPT-1.pdf
4. Grljević O, Bošnjak Z. Sentiment analysis of customer data. Strategic Manage. 2018; 23 (3): 38–49.
5. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. San Diego, CA, USA: NeurIPS; 2017. pp. 5998–6008.

6. Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev. 2022; 55 (7): 5731–5780.
7. Rızvı M. Decoding emotions: harnessing the power of Python for sentiment analysis in social media. In: The Eurasia Proceedings of Educational and Social Sciences, October 27–30, 2023. Vol. 31, pp. 189–195.
8. Prathi JK, Raparthi PK, Gopalachari MV. Real-time aspect-based sentiment analysis on consumer reviews. In: Raju KS, Senkerik R, Lanka SP, Rajagopal V, editors. Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19. Singapore: Springer; 2020. pp. 801–810.
9. Shinde PP, Shah S. A review of machine learning and deep learning applications. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, August 16–18, 2018. pp. 1–6.
10. Gupta S. Sentiment Analysis: Concept, Analysis and Applications. Medium. Towards Data Science. [Online]. 2018. Available at https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

