|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nos. | Paper name | Methods and algo used | Dataset used | result | Advantages of methodologies used | disadvantages of methodologies used | summary |
| 1 | [Sentiment Analysis and Emojification of Tweets](https://drive.google.com/file/d/1YkwU15exvX0OA1JF62MGV7lYMQvg1JB6/view?usp=drive_link) | 1.Data Acquisition  2.Data Pre-processing  3.Tweet Classification: The TextBlob Python library was used to classify each tweet as positive, negative, or neutral based on its polarity value ranging from -1 to +1.  3.Machine Learning Models: Five machine learning models were used to test the classification performance of the dataset:  Gradient Boosting, Logistic Regression, Naive Bayes, Random Forest, Support Vector Machines  4.Model Evaluation  5.Logistic Regression  6.API Development  7.Web Interface Development: | 168,274 English tweets were collected using the Twitter API | The Logistic Regression model achieved the highest accuracy of 94.05% and was selected as the most successful method | Efficient Data Acquisition, Comprehensive Data Cleaning, Accurate Tweet Classification, Comparative Model Evaluation, Efficient Model Deployment | Limitations of TextBlob, Emojification Limitation, Data Quality and Noise(spam) | The researchers collected 168,274 English tweets using the Twitter API.  The tweets underwent a data cleaning process to remove URLs, hashtags, mentions, emojis, punctuation, and numerical values.  The cleaned tweets were classified as positive, negative, or neutral using the TextBlob Python library based on polarity values ranging from -1 to +1.  Five machine learning models were tested on the classified dataset: Gradient Boosting, Logistic Regression, Naive Bayes, Random Forest, and Support Vector Machines.  The Logistic Regression model achieved the highest accuracy of 94.05% and was selected as the most successful method for sentiment analysis.  The trained Logistic Regression model was saved and integrated into a Google Cloud Functions API.  A Flask-based web application was developed that retrieves the last 50 tweets of a user and appends relevant emojis based on the sentiment analysis results |
| 2 | [Sentimental Analysis of Twitter Data with respect to General elections in india](https://drive.google.com/file/d/12KWiP7CFaOda1tJGU3hvqMXNuAoRlcba/view?usp=drive_link) | Data CollectionPre-processing, Sentiment Analysis Approaches-Lexicon-based Approach, Unsupervised Techniques, Evaluation Metrics  Sentiment polarity (positive, negative, neutral)  Sentiment strength  Sentiment subjectivity | tweets collected from January 2019 to March 2019 related to the general elections in India. Two sample candidates were used for comparison | Candidate-1 was more liked and popular compared to Candidate-2 based on sentiment analysis of Twitter data collected from January to March 2019. | Efficient Sentiment Analysis, Utilization of Twitter APIs, Insightful Data Processing, Open Access and Reproducibility | Limited Context, Noise in Twitter Data, Lack of Ground Truth | The paper used Twitter as a source of opinionated data to perform sentiment analysis on tweets related to two election candidates, referred to as Candidate-1 and Candidate-2, during the 2019 Indian general elections.  Twitter APIs were used to collect tweets from January to March 2019, which were then pre-processed to handle noisy, unstructured data.  Sentiment analysis was performed using both lexicon-based and unsupervised techniques. The "SentimentAnalysis" and "sentimentr" packages in R, as well as Rapid Miner's ALYLIEN extension, were utilized for sentiment analysis at the sentence level.  The sentiment analysis evaluated sentiment polarity (positive, negative, neutral), sentiment strength, and sentiment subjectivity for the two candidates.  The results concluded that Candidate-1 was more liked and popular compared to Candidate-2, which aligned with the actual election results obtained in May 2019.  The paper highlighted the advantages of using social media data and sentiment analysis techniques to gain insights into public opinions and emotions during election campaigns.  The limitations of the study included the noisy nature of Twitter data, dependence on sentiment lexicons, potential biases, and the time-dependent nature of the data collected. |
| 3 | [Twitter Sentiment Analysis on Airline Tweets in India Using R Language](https://drive.google.com/file/d/1MckMbw4TA-ccJk0DUa-18c3u7J84Y8i1/view?usp=drive_link) | 1.Data Collection 2.Accessing Twitter Data Sets (#indigo, #airindia, #spicejet)  4.Text Pre-processing 5.Sentiment Analysis 6.sentiment Visualization | The data is collected from Twitter using hashtags related to the airlines, such as #spicejet, #indigo, and #airindia. 1,000 tweets are randomly selected. These tweets cover various emotions and sentiments | Air India received a high number of positive responses and was highly trusted among customers.  indigo-Most responses were positive, indicating good customer experiences. Sentiments like anticipation, joy, and trust were also high.  SpiceJet had a higher number of negative sentiments compared to positive ones. Issues like fear and sadness were also significant. | Efficient Data Extraction and Processing, Comprehensive Sentiment Analysis, Data Visualization Capabilities(R's strong data visualization capabilities), Customer Understanding | Sarcasm Detection, Dependency on Hashtags, Sentiment Analysis Limitations(degrees of sentiments) | sentiment analysis using Twitter data, providing insights into customer satisfaction and areas needing improvement for Indian airlines. |
| 4 | [Twitter Sentiment Analysis](https://drive.google.com/file/d/16qQH2eH_8xX_N3DO4rna3Un_yW1ijZTC/view?usp=drive_link) | 1.Text Preprocessing  2.Machine Learning Tools and Libraries (Tweepy, Sklearn (Scikit-learn))  3.Machine Learning Algorithms(Support Vector Machine (SVM), Naive Bayes, Decision Trees and Random Forests)  4.Sentiment Analysis(textblob, natural language toolkit)  5.Sentiment Classification | The tweets were collected using the Twitter API, leveraging the Tweepy library based on certain keywords and hashtags | Support Vector Machine (SVM) achieved an accuracy of 86.35% Naïve Bayes classifiers showed varying accuracy, with one study achieving 86.7% accuracy and another achieving 81.5%.  Random Forest methods had an accuracy range, with one study achieving up to 86.1%  CNNs achieving up to 88.7% | Text Preprocessing tools, Real-time Sentiment Analysis, data refining. | Lang limitation, Negation and Sarcasm Detection, Handling Ambiguity and Noise | The paper references various studies and approaches used in sentiment analysis, highlighting their methodologies and accuracies. It includes comparisons between different algorithms such as SVM, Naïve Bayes, and Random Forest, demonstrating their effectiveness in sentiment classification. |
| 5 | [Twitter Sentiment Analysis using Supervised Machine Learning](https://drive.google.com/file/d/1QjyFrIQmV_gJi2Id7Z2wfIH8g5L7K268/view?usp=drive_link) | following methods and algorithms in this paper(Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Linear SVC)  Data cleaning  Data pre-processing  Feature extraction  Feature selection | p. The data is extracted using SNS services which are done using twitter’s streaming API. The tweets are loaded into Hadoop and are pre-processed using map-reduce functions | The results show that logistic regression achieves the highest accuracy of 82.47%, followed by linear SVC with 83.71% and multinomial Naive Bayes with 80.61%.  The paper concludes that logistic regression is a suitable choice for sentiment analysis | the paper highlights the strengths of logistic regression in handling complex datasets with correlated features and achieving a high accuracy with reliable results | may not capture the full range of sentiments and topics expressed on Twitter,  The process of labeling data can be time-consuming and expensive,  issue of sarcasm or other complex emotions. | the paper presents a supervised machine learning approach for Twitter sentiment analysis, specifically for product reviews, and compares the performance of different classifiers. The study finds that logistic regression performs the best, and highlights the importance of data cleaning and pre-processing. The authors also acknowledge the limitations of the study and suggest future directions for research. |
| 6 | [Twitter Sentiment Analysis Approaches: A Survey](https://drive.google.com/file/d/19HI4PYOdm-NCNewt_QAlMNhy3A1P9dXR/view?usp=drive_link) | * Machine learning approaches using feature extraction, feature selection, and classification algorithms like Naive Bayes, SVM, and Logistic Regression * Lexicon-based approaches leveraging sentiment lexicons to determine tweet sentiment polarity * Hybrid approaches combining machine learning and lexicon-based methods * Graph-based approaches representing tweets as graphs to analyze sentiment using social network analysis techniques * Other approaches utilizing interactive visualization, cognitive science theories, big data platforms, and semantic analysis | 1. Twitter data:    * Tweets Sentiment Scores, Stock Market Values, and Hybrid Data used for forecasting monthly total vehicle sales in the USA    * Real-time Twitter Data used for processing real-time data to view people's reactions to events    * Twitter Streaming API data used for collecting tweets for training and testing sentiment analysis models 2. Other datasets:    * Movie Reviews used for classifying reviews into positive, negative, and neutral polarity    * Stock Prices used for forecasting movements of individual stock prices    * Event Mentions in Microblogs used for quantifying user interests using similarity-based region networks | 1. Machine learning approaches achieved accuracy ranging from 85% to 91%, with SVM and Logistic Regression being the best performing classifiers. 2. Lexicon-based approaches showed varying accuracy, ranging from 52% to 84%, depending on the sentiment lexicons used. 3. Hybrid approaches combining machine learning and lexicon-based methods demonstrated improved performance over individual approaches. 4. Graph-based and other approaches like using cognitive science theories, big data platforms, and semantic analysis showed promising results for Twitter sentiment analysis. | * Machine learning approaches like SVM and Logistic Regression achieved high accuracy (85-91%) * Hybrid approaches combining machine learning and lexicon-based methods demonstrated improved performance * Graph-based approaches using social network analysis can identify influential users and optimize machine learning classifiers | * Lexicon-based approaches showed varying accuracy (52-84%) depending on the sentiment lexicons used * Challenges remain in handling context-dependent sentiment, sarcasm, and domain-specific language * Lack of standardized datasets and evaluation metrics makes it difficult to compare approaches | he research paper "Twitter Sentiment Analysis Approaches: A Survey" provides an overview of various methodologies for analyzing sentiment in tweets, including machine learning, lexicon-based, hybrid, and graph-based approaches. Machine learning methods like SVM and Logistic Regression achieved high accuracy (85-91%), while lexicon-based approaches showed varied results (52-84%). Hybrid approaches combining machine learning and lexicon-based methods demonstrated improved performance. Graph-based approaches using Social Network Analysis identified influential users. The paper also discusses future directions incorporating cognitive science theories, big data platforms, and semantic analysis to enhance Twitter sentiment analysis. |
| 7 | [Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers](https://drive.google.com/file/d/1GDF__8N_8mGBa2lA1Y8w7pt3XhaJtZHe/view?usp=drive_link) | * Data collected from Twitter for sentiment analysis of mobile providers in Indonesia. * Preprocessing involved removing duplicates, converting to lowercase, and eliminating URLs and stopwords. * Classification using Naïve Bayes, SVM, and Decision Tree. * Evaluation based on accuracy compared to manual classification. * Brand reputation measured using Net Brand Reputation index | The data set used in this paper consists of tweets collected from the official Twitter accounts of three major mobile providers in Indonesia: XL Axiata, Telkomsel, and Indosat | * NBR scores for the mobile providers: XL Axiata (32.3%), Telkomsel (19.0%), Indosat (10.9%). * SVM classifier had the best performance with an AUC score of 0.854, outperforming Naive Bayes and Decision Tree. * SMS services had the highest NBR scores, while 4G services had the lowest, especially for Indosat at -34.01% | * Utilizes Twitter data to analyze brand reputation of mobile providers, which is a relevant and timely data source. * Employs a comprehensive data preprocessing approach, including removing duplicates, URLs, stopwords, and converting to lowercase. * Applies Part-of-Speech (POS) tagging to identify important features like adjectives, nouns, verbs, and adverbs for sentiment analysis. * Compares the performance of three popular classification algorithms - Naive Bayes, SVM, and Decision Tree - to identify the best model. * Evaluates the classifiers using metrics like precision, recall, F-measure, and ROC curve to ensure robust performance. * Introduces a novel "Net Brand Reputation" (NBR) metric to quantify brand reputation, similar to Net Promoter Score (NPS). * Visualizes the results using Tableau dashboards for real-time monitoring and analysis | * The dataset is relatively small, with only 10,000 tweets collected over a 3-month period. A larger dataset may provide more comprehensive insights. * The manual labeling of tweets into positive, negative, and unknown categories by 5 respondents could introduce subjectivity and bias. * The study focuses only on the top 3 mobile providers in Indonesia, limiting the generalizability of the findings. * The paper does not provide details on the performance of the POS tagger developed for Bahasa Indonesia, which is a critical component of the methodology. * The comparison between NBR and NPS is limited, and the advantages of NBR over NPS are not thoroughly explored | This paper presents a methodology to analyze brand reputation of mobile providers in Indonesia using Twitter sentiment analysis. The authors collected 10,000 tweets about three major providers - XL Axiata, Telkomsel, and Indosat. After preprocessing the data, they used Naive Bayes, SVM, and Decision Tree classifiers to categorize tweets as positive, negative, or unknown. The SVM model performed best, with an AUC of 0.854. The authors then calculated a "Net Brand Reputation" (NBR) score for each provider, finding XL Axiata had the highest at 32.3%, followed by Telkomsel at 19.0% and Indosat at 10.9%. The results show the value of social media sentiment analysis for measuring brand reputation. |
| 8. | [Sentiment analysis on large scale Amazon product reviews](https://drive.google.com/file/d/1mab-3VbE8jQjnmHBkIXjuZV26wEcqcRl/view?usp=drive_link). | The methodology involves acquiring Amazon review data in JSON format and labeling it using both manual and active learning approaches. The dataset includes electronics, cell phone accessories, and musical instrument reviews. Preprocessing is performed to clean the data, and features are extracted using bag-of-words and TF-IDF methods. Various classifiers, including Naive Bayes and Support Vector Machines, are applied to classify the sentiment of the reviews, aiming for high accuracy in sentiment analysis. | The dataset used in the paper consists of Amazon product reviews. Specifically, it includes reviews from three categories: Electronics, Cell Phones and Accessories, and Musical Instruments. The dataset comprises approximately 48,500 product reviews, with 21,600 reviews from mobile phones, 24,352 from electronics, and 2,548 from musical instruments. The data was provided in JSON format, including fields such as reviewer ID, product ID, review text, overall rating, summary, and review time. | The paper achieved high accuracy in sentiment analysis of Amazon product reviews by combining manual and active learning for dataset labeling, and employing feature extraction techniques like bag-of-words and TF-IDF. Various classifiers, including Naive Bayes and Support Vector Machines, were used to classify sentiments. The results demonstrated that their approach, particularly the use of active learning and diverse feature extraction methods, yielded more accurate sentiment predictions compared to existing methods. |  **High Accuracy**: Achieved high accuracy in sentiment analysis through the combination of active learning and feature extraction techniques.   **Large Dataset**: Utilized a large and diverse dataset, enhancing the robustness of the results.   **Multiple Classifiers**: Employed various classifiers, including Naive Bayes and SVM, to ensure comprehensive analysis.   **Innovative Labeling**: Used active learning for efficient dataset labeling, saving time and effort. |  **Complexity**: The methodology is complex and may require significant computational resources.   **Generalization**: Results may be specific to the Amazon review dataset and might not generalize well to other datasets.   **Manual Effort**: Despite using active learning, some manual labeling was still necessary. | "Sentiment analysis on large scale Amazon product reviews" presents a methodology combining manual and active learning to label a large Amazon review dataset. Using feature extraction techniques like bag-of-words and TF-IDF, various classifiers, including Naive Bayes and SVM, were applied to achieve high accuracy in sentiment classification. The study demonstrates the effectiveness of their approach in accurately polarizing reviews, contributing to more reliable sentiment analysis in e-commerce. |
| 9. | [Twitter Sentiment Analysis for Product Review Using Lexicon Method](https://drive.google.com/file/d/1V75M5V44AsukF7NZU65n2S8FAiqA97-r/view?usp=drive_link). | The methodology includes creating a Twitter application to collect tweets via Twitter API, followed by preprocessing to clean and structure the data. The sentiment analysis uses a lexicon-based approach, matching words in the tweets with a predefined dictionary of positive and negative terms to assign sentiment scores. Steps involve removing punctuation, URLs, and stop-words, handling emoticons and acronyms, and classifying the tweets into positive, negative, or neutral categories based on the sentiment scores | The dataset used in the paper "Twitter Sentiment Analysis for Product Review Using Lexicon Method" involved collecting tweets from Twitter using the Twitter API with different keywords related to a specific topic within a particular time period and date. The tweets were then pre-processed to eliminate unwanted information such as punctuation, digits, URLs, and emoticons. The dataset comprised a large number of tweets from Twitter, focusing on sentiment analysis of user opinions regarding particular products or services. The sentiment analysis was conducted at both the document level and aspect level, analyzing aspects like voice quality, battery life, service, size, picture quality, and price of the product. The sentiment of the tweets was categorized as positive, negative, or neutral to estimate the overall sentiment of customers or users about the products or services | The paper presents a framework for sentiment analysis of Twitter data using a lexicon-based approach. It collects tweets related to the iPhone, pre-processes the data, and performs sentiment analysis at both the document and aspect levels. The results show positive, negative, and neutral sentiments for various aspects like battery life, voice quality, service, size, picture quality, and price. The sentiment analysis helps in decision-making and can be applied to other domains such as customer reviews of airlines services or political reviews | * Presents a framework for sentiment analysis of Twitter data using a lexicon-based approach * Collects tweets related to the iPhone, pre-processes the data, and performs sentiment analysis at both the document and aspect levels * Analyzes various aspects like battery life, voice quality, service, size, picture quality, and price * Applicable to other domains such as customer reviews of airlines services or political reviews | * Limited to a specific product (iPhone) and a particular time period * Relies on a lexicon-based approach, which may not capture context-dependent sentiment * Sentiment analysis is performed at a high level, without delving into the nuances of user opinions * Lacks a comparative analysis with other sentiment analysis techniques or datasets | This paper proposes a framework for sentiment analysis of Twitter data using a lexicon-based approach. The methodology involves collecting tweets using the Twitter API, pre-processing the data to eliminate unwanted information, and then analyzing the sentiment using a dictionary-based approach. The sentiment analysis is performed at both the document and aspect levels, focusing on various aspects like battery life, voice quality, service, size, picture quality, and price of the iPhone. The results show positive, negative, and neutral sentiments for these aspects, which can be used to improve product quality and services. The approach is applicable to other domains such as customer reviews of airlines services or political reviews |
| 10. | [Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser](https://drive.google.com/file/d/1cNFyCiUbhUNiOMfb6GMUeT7Gxd4xSQT7/view?usp=drive_link) | The methodology used in this paper involves collecting 1,048,575 tweets from Twitter using the hashtag "COVID-19." Tweets were pre-processed using a Twitter tokenizer, and sentiment analysis was conducted using VADER and TextBlob analyzers. Topic modeling was done using Latent Dirichlet Allocation and visualized with Multidimensional scaling. The study aimed to analyze emotional responses to COVID-19 in Nigeria and provide insights for organizations, governments, and nations to make informed decisions | The dataset used in this paper is a collection of COVID-19 tweets from Twitter. The tweets were collected using the hashtags #StayHomeNigeria, #covid19nigeria, #coronavirusnigeria, #COVID-19Nigeria, #NCDC, and #FMOH. The data was collected between November 2019 and May 2021, resulting in a total of 1,036,716 tweets | The research paper on sentiment analysis of COVID-19 tweets in Nigeria using VADER and TextBlob analyzers found that VADER sentiment analysis returned 39.8% positive, 31.3% neutral, and 28.9% negative sentiment, while TextBlob analysis showed 46.0% neutral, 36.7% positive, and 17.3% negative sentiment. The study highlighted the importance of using social media data to aid organizations and governments in making informed decisions regarding COVID-19 and addressing challenges like misinformation | * Provides insights into the sentiment of COVID-19 tweets in Nigeria using two sentiment analysis tools, VADER and TextBlob * Identifies the most common topics discussed in the tweets using Latent Dirichlet Allocation (LDA) topic modeling * Offers a methodology for collecting and analyzing large volumes of Twitter data related to COVID-19 * Visualizes the results using word clouds and intertopic distance maps for better understanding | * The dataset is limited to tweets collected using specific hashtags, which may not represent the entire population's sentiment * The study does not compare the performance of VADER and TextBlob or provide a justification for using both tools * The paper lacks a discussion on the limitations of sentiment analysis and topic modeling techniques in the context of social media data * The results are not validated against other data sources or expert opinions to ensure their reliability and generalizability | This research paper presents a sentiment analysis of COVID-19 tweets from Nigeria using VADER and TextBlob analyzers. The study collected 1,036,716 tweets from Twitter using specific hashtags and analyzed them for sentiment using both tools. The results show that VADER returned 39.8% positive, 31.3% neutral, and 28.9% negative sentiment, while TextBlob returned 46.0% neutral, 36.7% positive, and 17.3% negative sentiment. The study highlights the importance of using social media data to aid organizations and governments in making informed decisions regarding COVID-19 and addressing challenges like misinformation |