

Smart Decisions with Opinion Mining

Dinesh M¹, Dr. R. Priya²

¹Vels Universit, India

²Professor, Department of Computer Applications, VISTAS

DOI : <https://doi.org/10.51583/IJLTEMAS.2025.140400047>

Received: 11 April 2025; Accepted: 22 April 2025; Published: 09 May 2025

Abstract: The runaway growth of web technology has resulted in an unprecedented volume of data being produced and published on the web each day. Social networking sites such as Twitter and Facebook have turned into indispensable zones for individuals to share thoughts, experiences, and opinions around the world. Sentiment analysis which involves the extraction and analysis of opinion from text, is central to gauging public feeling, monitoring trends, business strategy, and customer satisfaction with regards to unstructured and heterogeneous nature of Twitter data, most research has been conducted on how to use sentiment analysis methods to classify opinion as positive, negative, or neutral. In this paper, sentiment analysis of social media data is investigated based on a Twitter dataset, utilizing machine learning methods such as Long Short-Term Memory (LSTM) networks for precise sentiment classification.

Keywords: Web technology, social networking sites, Twitter, Facebook, Sentiment analysis, Opinion extraction, public sentiment, trend monitoring, business strategy, customer satisfaction, machine learning, long short-term memory (LSTM) networks.

I. Introduction

The internet age has revolutionized the way individuals voice opinions via blogs, forums, reviews, and social media. Millions utilize sites such as Facebook and

Twitter to voice opinions and sway others. Social media creates huge emotional data in the form of posts, comments, and reviews, offering businesses a chance to connect for decisions. Such as reading reviews prior to buying. The sheer amount of data requires automation via sentiment analysis (SA). Aids to find out whether a product is pleasing, helping companies know what users like. If targets opinions, feelings, and sentiment instead. the pure facts. With the growth of web content, SA allows the creation of applications that examine sentiment. Companies use SA to improve marketing and user interaction. Recommendation systems utilize SA to forecast user preference. Module description: Overall overview of Smart, Decisions with opinion mining. Collection – Python advanced data structure comprising counter, defaultdict, OrderedDict and namedtuple that increases performance and dependability in complex impressions. Matplotlib.pyplot – A library used for creating static, animated and interactive plots such as line chart, bar chart and histograms with customization option. nltk – An advanced NLP library that provides tools for tokenization, stemming, lemmatization, stop word elimination, and part-of-speech tagging for sentiment analysis and linguistic studies. nltk.corpus – offers access to large corpora of languages such as Brown Corpus, Gutenberg Corpus, and WorldNet that can be helpful in text categorization and syntactic parsing. nltk.stem – Contains stemmers such as Porter and Lancaster to cut down on words to their base form for search engines and text normalization purposes. nltk.tokenize – Divide text into words or sentence effectively with support for different languages and types. NumPy - A core package for numerical computing. Multi-dimensional array linear algebra and mathematical functions supported. Pandas- Data analysis and manipulation library with data frame and series structure for statistical analysis and efficient data handling. Sklearn.metrics – Offers evaluation metrics of classification, regression, and clustering models to retrieve a machine learning performance. Sklearn.model selection – Library for dataset splitting, cross-validation and hyperparameters tuning such as train set split and GridSearchCV. TensorFlow – Deep learning library with support for neural networks through high-level APIs such as keras and low-level computational operations. tensorflow.keras.preprocessing.sequence – Utilities for sequence-based data in NLP such as sequence padding and embeddings. tensorflow.keras.preprocessing.text – Functions or text preprocessing such as tokenized text to sequence conversion and one-hot encoding. Textblob - A top-level NLP library for sentiment, part of speech, tagging, and text translation that make complicated language processes easier.

Literature Survey

Sentiment Analysis of Twitter Data (Asafuzzaman et al., 2020) – Examines Sentiment analysis methods on Twitter, such as lexicon based and machine learning approaches [1]. Deep Sentiment Analysis Learning (Zhang et al., 2018) – Focuses on deep learning model, (CNNs and RNNs) in sentiment analysis and importance of feature, extraction, and pre-trained embeddings such as word2Vec [2]. Comparative Study of sentiment analysis Methodologies (Hossain et al., 2019) – Compares naïve Bayes, SVM, deep learning models and concludes deep learning in superior and preprocessing increases the performances [3]. Sentiment Classification Using Machine learning (Kumar et al., 2021) – Discusses supervised and focusing hybrid methods for enhanced accuracy [4].

Exploring Sentiment Analysis for social media (Gupta et al., 2020) – Focusing on rule-based and machine methods, highlighting the effect of linguistic features such as hashtags and emoticons [5]. Novel Deep Learning Approaches for Sentiment Analysis (Ramesh et al., 2022) – Suggests an LSTM- based attention mechanism to enhance classification accuracy in twitter data [6]. Real-Time Sentiment Analysis on Twitter (Nair et al., 2021) – Presents a distributed computing system based on Apache Kafka and Spark for real time sentiment Analysis [7]. Hybrid Approaches in Sentiment Analysis (Chen et al., 2019) – Discusses hybrids of machine learning and NLP methods for better classification performance [8]. Impact of Preprocessing on Sentiment Analysis (Singh et al., 2020) – Examines the impact of preprocessing methods (stemming stop-word elimination) on model accuracy [9]. Challenges and opportunities in Social Media Sentiment Analysis (Patel et al., 2023) – Points out challenges such as slang abbreviations, and emojis necessitating more adaptive models [10].

II. Methodology

Sentiment analysis employs NLP to draw out opinions, attitudes, and emotions from text, speech, or databases. Also referred to as opinion mining, it categorizes sentiments into positive, negative, or neutral. Twitter Sentiment Analysis has been researched using binary classification via Naïve Bayes, Maximum Entropy, and SVM, with SVM usually being the best performer. Machine learning models have been experimenting with features such as unigrams, hashtags, and the bag-of-words model. Scientists have developed techniques to differentiate subjective and objective tweets as well as identifying emotions through WordNet. Spam, among other challenges, has been addressed, and techniques such as stochastic gradient descent and k-nearest neighbors have been applied with mixed success.

Algorithms

Natural Language Processing Algorithms

Natural language processing involves a range of methods for analyzing and comprehending human language. Some of the popular algorithms used are:

Tokenization

Overview

Tokenization is the act of decomposing text into individual units referred to as tokens, which may be words, phrases, or symbols.

How It Works

The algorithm reads the text and divides it based on defined delimiter (e.g., spaces, punctuation)

Advantages

Simplifies text processing by breaking it into manageable pieces.

Disadvantages

May struggle with contractions or compound words if not properly configured.

Word Embeddings

Overview

Word embeddings are dense vector representations of words that capture semantic meaning based on their context within a corpus.

How It Works

Algorithms such as Word2Vec or Glove learn associations from large text corpora and represent words as vectors in a continuous vector space.

Advantages

Captures semantic relationships between words (e.g., “king” – “man” + “woman” = “queen”)

Disadvantages

I need large datasets for effective training.

Part-of-Speech Tagging

Overview

Part-of-speech Tagging involves assigning parts of speech (e.g., noun, Verb) to each word in a sentence.

How it Works

Algorithms employ statistical models or rule-based systems to analyze sentence structure and identify word roles.

Advantages

Improves understanding of grammatical relationships in sentences.

Disadvantages

Might not perform well uncertain words without enough context.

Data Preprocessing Methods

Data preprocessing plays an important role in making raw data ready to be analyzed and modelled. The following are the primary methods that usually come into the picture:

Text Cleaning

Introduction

Text Cleaning is a method of stripping redundant characters, stop words, and other noise away from raw textual data.

How It Works.

Typical actions include lowercasing text, stripping off punctuation, and excluding stop words based on predefined lists.

Stemming and Lemmatization

Introduction

Reducing words into their root or base form are done using methods like stemming and lemmatization.

How It Works

Stemming: Trims off prefixes to reach a base form (e.g., “running” is reduced to “run”).

Lemmatization: Applies vocabulary and morphological analysis to return the base form (e.g., “better” is reduced to “good”).

Feature Extraction

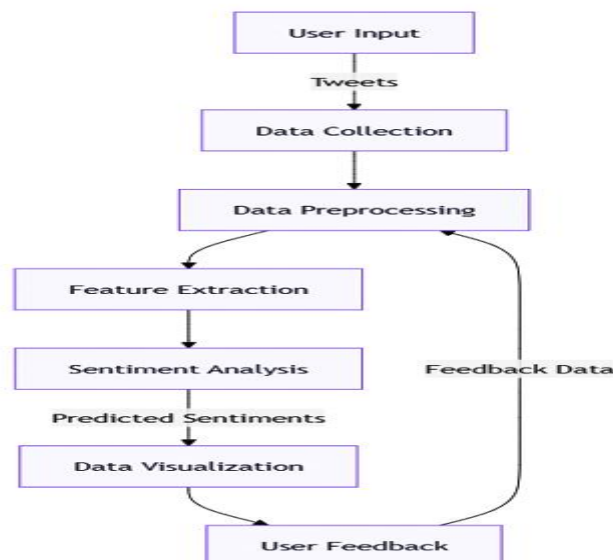
Overview

Feature Extraction converts raw data into numerical features that are acceptable for machine learning models.

How It Works

Methods such as Term Frequency – Inverse Document Frequency (TF-IDF) transform text into numerical vectors by words frequency in documents.

Data Flow Diagram



This process of sentiment analysis begins with user input, where tweets are gathered and preprocessing. Feature extraction, after cleaning, identifies important patterns, which are examined in sentiment analysis to make predictions about sentiments. The outcome is then displayed through data visualization. Users give feedback, which is fed back into preprocessing to increase accuracy and enhance future predictions.

Sentiment Analysis Architecture

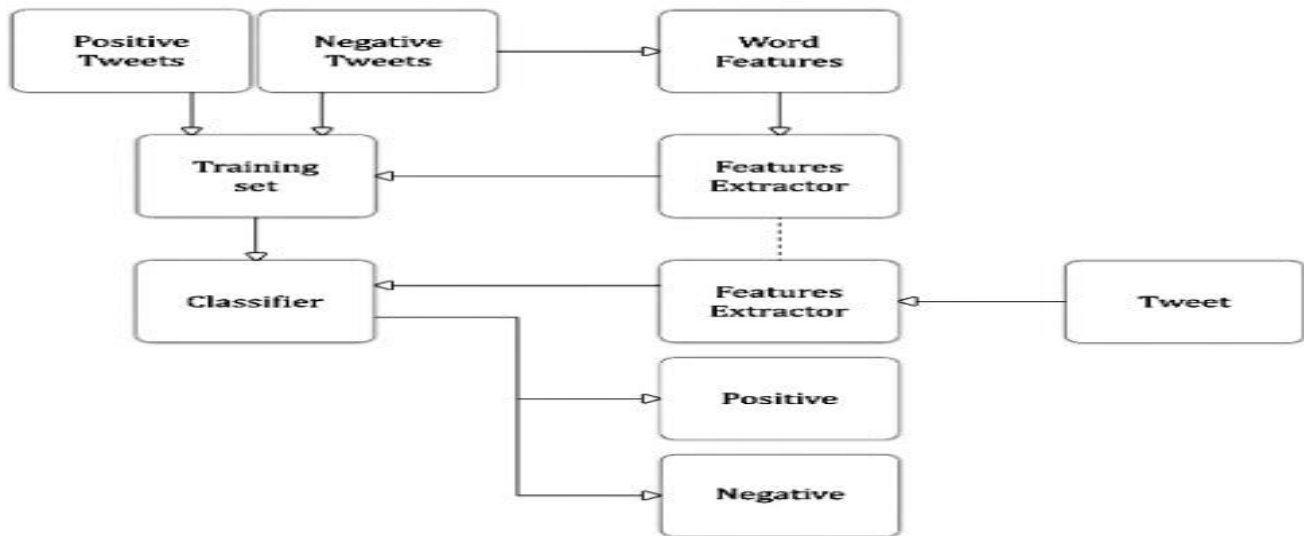


Figure – 1: Sentiment Analysis Architecture

The architecture of sentiment analysis operates by identifying tweets as either positive or negative through machine learning methods. It starts with a set of positive and negative tweets, which are utilized to train a classifier. Significant words, referred to as word features, are extracted from the tweets via a feature extractor, transforming textual data into numerical features that are comprehensible by the classifier. The classifier is trained on these extracted features and labelled tweets so that it can be taught how to differentiate between positive and negative sentiments. When a new tweet is presented, it is subjected to the same features extraction process prior to being inspected by the trained classifier. From the extracted features, the classifier makes a prediction of whether the sentiment of the tweets is positive or negative. This machine learning based sentiment analysis is commonly applied for social media tracking, customer complaint analysis, and opinion extraction for insights into people's opinions.

Sentiment Classification System Based on Emoticons

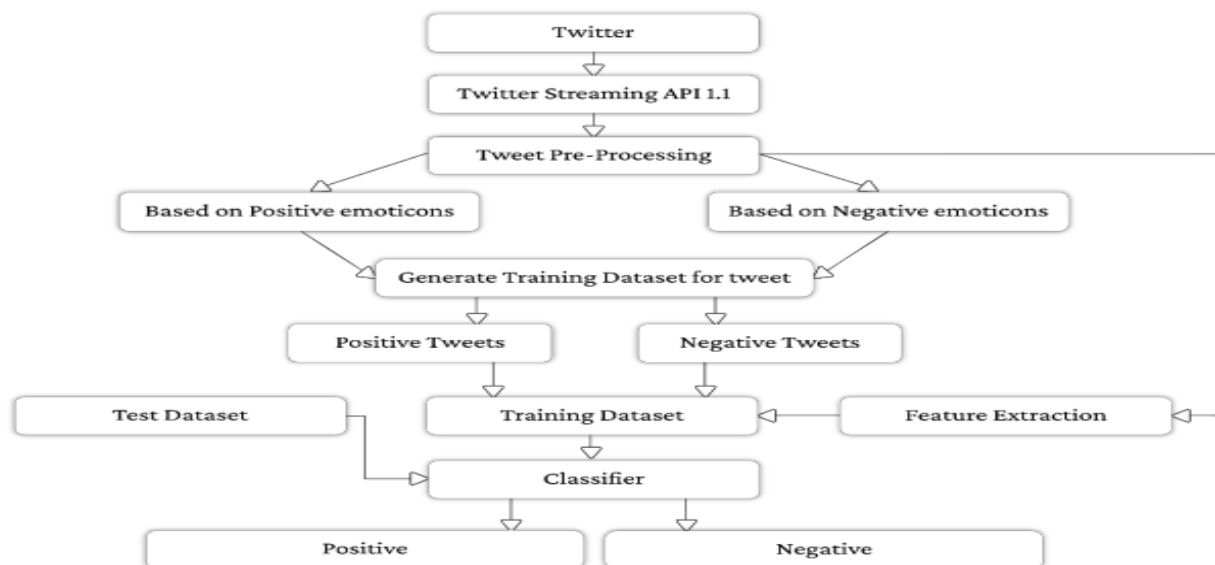


Figure – 2: Sentiment Classification Based On Emoticons

The figure depicts the sentiment classification process Twitter emoticons. It starts with gathering tweets via the Twitter Streaming API 1.1, followed by a pre-processing stage in which tweets are classified according to positive and negative emotions. The classified tweets are then utilized to create a training tweet. The training dataset is then subjected to feature extraction to make it ready for classification. A classifier is then used to separate positive and negative sentiments. Also, a test dataset is employed separately to test the performance of the model. This method uses emoticons as sentiment markers to classify tweets automatically as positive or negative.

Lexicon-Based Model

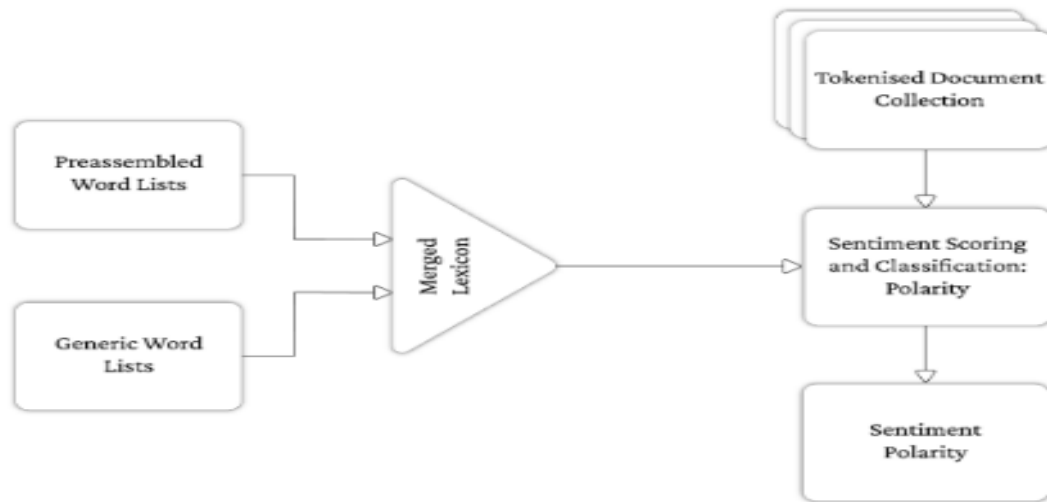


Figure – 3: Lexicon-Based Model

The Lexicon-Based Model processes sentiment according to predefined lists of word. Preassembled and general word lists are combined into a lexicon that is then used on tokenized document. Sentiment is scored from the words, and the polarity (positive, negative, or neutral) is decided. The approach is commonly applied in text analysis, like reviews and social media sentiment analysis.

Sentiment Analysis Tasks

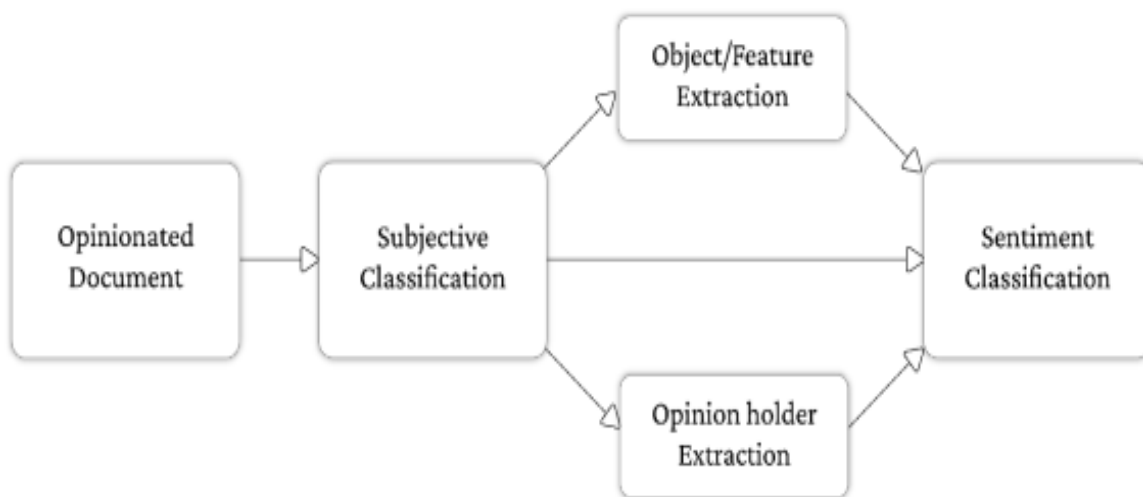


Figure – 4: Sentiment Analysis Tasks

The image illustrates the most important tasks of sentiment analysis. It begins with an opinionated document that is subjected to subjective classification to ascertain whether or not it holds opinions. If subjective, the document is processed for object/feature extraction to detect certain aspects, opinions holder extraction to see who held the sentiment, and sentiment classification to classify the opinions as being positive, negative, or neutral. These are the processes that analyse sentiment in texts for use in customer feedback and social media monitoring.

Levels of Sentiment Analysis

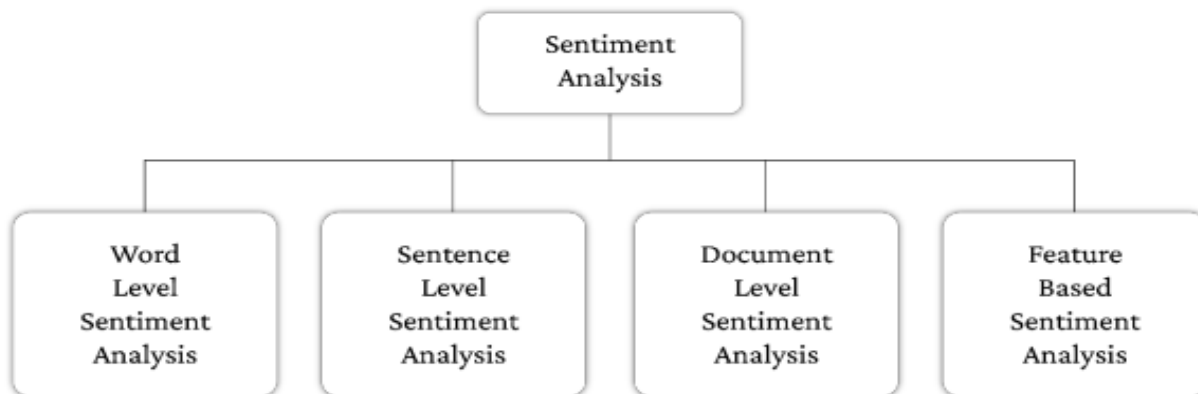


Figure – 5: Levels of Sentiment Analysis

The figure shows the various levels of sentiment analysis, which is the process of classifying the sentiment or emotion conveyed in a text. Sentiment analysis may be carried out at more than one level, such as word-level-sentiment analysis, where the sentiment of words is analysed; sentence-level sentiment analysis, where the overall sentiment of a sentence is analysed; document-level sentiment analysis, where the sentiment of an entire document is assessed; and feature-based sentiment analysis, where an emphasis is put on sentiment extraction based on aspects or features within the text. These various level assist in the deeper comprehension of expressed emotions and opinions contained in text data, which makes sentiment analysis an essential resource in marketing, customer opinion analysis, and social media tracking.

III. Conclusion

We effectively processed text data for sentiment – positive, negative, or neutral – via NLP, data preprocessing, and machine learning. Our findings identify the importance of sentiment analysis in business, customer reviews, and social media. Accuracy in the model relies on data quality, feature engineering, and algorithm choice. Future enhancement involves deep learning (e.g., BERT), bigger datasets, hyperparameter optimization, and real-time processing. This project shows the potential of sentiment analysis in analysing emotions for data-driven decisions.

References

1. Asafuzzaman, et al.(2020). Sentiment Analysis of Twitter Data.
2. Zhang, et al. (2018). Deep Sentiment Analysis Learning.
3. Hossain, et al. (2019). Comparative study of Sentiment Analysis Methodologies.
4. Kumar, et al. (2021). Sentiment Classification Using Machine Learning.
5. Gupta, et al. (2020). Exploring Sentiment Analysis for Social Media.
6. Ramesh, et al. (2022). Novel Deep Learning Approaches for Sentiment Analysis.
7. Nair, et al. (2021). Real-Time Sentiment Analysis on Twitter.
8. Chen, et al. (2019). Hybrid Approaches in Sentiment Analysis.
9. Singh, et al. (2020). Impact of Preprocessing on Sentiment Analysis.
10. Patel, et al. (2023). Challenges and Opportunities in Social Media Sentiment Analysis.