International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

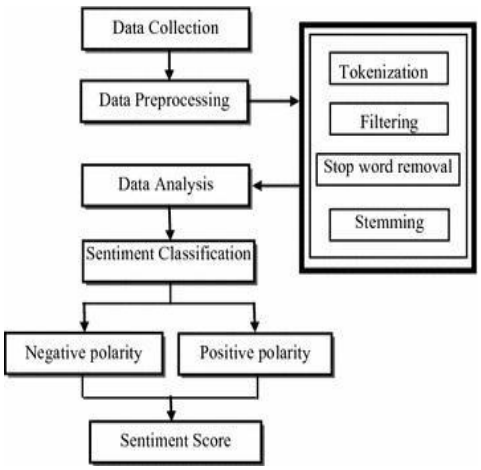
A Comprehensive Review of Sentiment Analysis Software for Airline Reviews

Goutam Mahajan, Grashal Vishwa, Hemant Gehlod, Goshen Masih

Abstract
Sentiment analysis falls within the realm of Natural Language Processing, focusing on the examination of subjective viewpoints or emotions gathered from diverse sources regarding a specific topic. This field provides a collection of techniques for recognizing and extracting opinions, which can be leveraged to enhance business operations.

Introduction
Sentiment analysis involves the process of categorizing a block of text as positive, negative, or neutral. It entails context-based examination of language to gauge the social perception of a brand, assisting businesses in forecasting the market demand for their products. The ultimate objective of sentiment analysis is to distill people's opinions in a manner that facilitates business expansion. This analysis doesn't merely focus on polarity (positive, negative, or neutral) but also delves into emotional aspects (such as happiness, sadness, or anger).
The significance of sentiment analysis stems from the vast amount of unstructured data generated daily by humans, which can be challenging and time-consuming to interpret. This is where sentiment analysis shines, as it simplifies the task of unraveling unstructured data using automated techniques and algorithms.
Various approaches are employed in conducting sentiment analysis:

1. Lexicon based approach: Count number of positive and negative words in each text and the larger count will be the sentiment of the text.



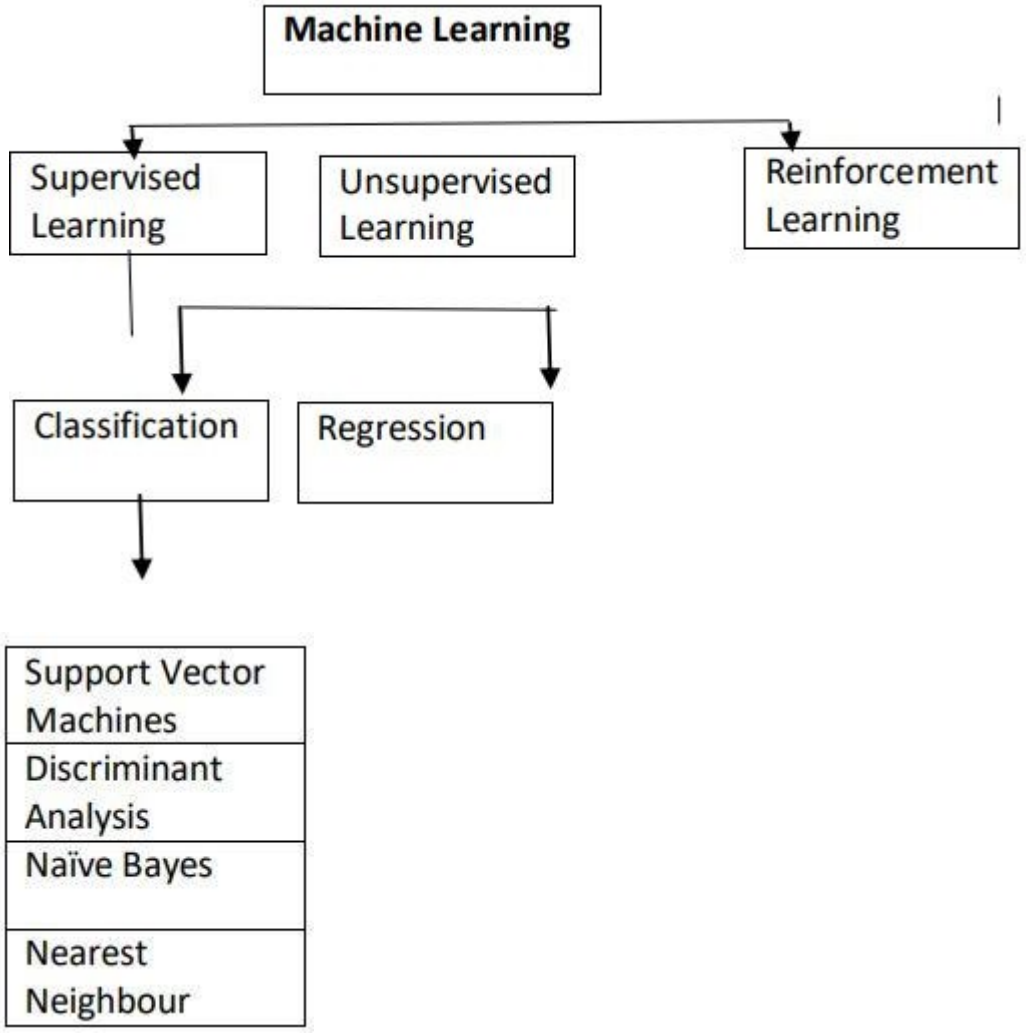
-Initial Processing: The text undergoes a series of initial processing steps, including POS tagging, stemming, stop-word removal, and tokenization into N- grams. The result of this pre-processing is a collection of tokens or a bag-of- words.

-Evaluating Token Polarity: Each token in the bag-of-words is assessed for its polarity using a lexicon. Every word from the bag of words is cross-referenced with the lexicon. If a word is present in the lexicon, its polarity (Wi) is incorporated into the text's sentiment score. If the word is not found in the lexicon, its polarity is treated as zero.

-Calculating the sentiment score of the text. After assigning polarity scores toall words comprising the text, the final sentiment score of the text is calculated by dividing

$$\text{Score}_{AVG} = \frac{1}{m}\sum_{i=1}^{m} W_i$$

2. Machine learning-based approach: Develop a classification model, which is trained using the pre-labelled dataset of positive, negative.



Fundamentals of Supervised Sentiment Analysis

Sentiment analysis is a valuable tool for discerning the emotional tone (positive, negative, or neutral) within data. In the context of businesses, it provides a

straightforward means of gauging customer reactions to their products or services, enabling swift detection of any shifts in sentiment that may necessitate immediate attention. A fundamental strategy to tackle this challenge involves utilizing supervised learning, where human annotators assess and label the sentiment of the data, effectively treating it as a text classification problem.

Supervised learning is a machine learning approach characterized by its reliance on labeled datasets. These datasets are meticulously crafted to train algorithms, guiding them in accurately categorizing data or making precise predictions. By comparing labeled inputs with expected outputs, the model can continually refine its performance over time.

Supervised learning can be categorized into two primary data mining problems: classification and regression. In classification problems, algorithms are employed to accurately assign test data into distinct categories, analogous to separating apples from oranges or filtering spam emails into a separate folder from your inbox. Common classification algorithms include linear classifiers, support vector machines, decision trees, and random forests.

Among these classification methods, the Naive Bayes classifier stands out as a frequently used approach in Sentiment Analysis. This supervised machine learning classifier, named after Thomas Bayes, applies Bayesian theorem to calculate conditional probabilities. For instance, it estimates the probability of event p1 occurring when event p2 has already taken place, using the following mathematical formula:

$$P(p1|p2) = P(p2|p1)P(p1)/P(p2)$$

Where A = Sentiment, B=Sentence

And the conditional probability of a word is given by

$$P(word|A) = C+1/(D+E)$$

C=no. of word occurrence in class D= no of words belonging to a class E= total no. of words

Sentiment analysis serves a broad spectrum of purposes, encompassing the evaluation of emotional expressions in text as positive, negative, or neutral. Within the realm of business, it provides a straightforward method to gauge customer sentiments regarding products or services, enabling the prompt identification of shifts in emotions that demand immediate attention.

One of the primary approaches to tackle this challenge is through supervised learning, where human annotators assess and label data sentiments, effectively framing it as a text classification task.

Supervised learning, characterized by the use of labeled datasets, offers two primary avenues: classification and regression. In classification, algorithms accurately categorize test data into distinct groups, such as distinguishing between positive and negative sentiments or classifying spam emails.
Among the classification methods, the Naive Bayes classifier stands as a commonly used technique for Sentiment Analysis. It leverages Bayesian theorem to estimate conditional probabilities and evaluate the relationship between events.

Regression, another supervised learning method, aims to discern the connection between dependent and independent variables. Regression models are particularly useful for predicting numerical values based on various data points, such as forecasting sales revenue for a business. Notable regression algorithms include linear regression, logistic regression, and polynomial regression.

Sentiment Analysis has diverse applications:

Social Media: Platforms like Instagram are used to analyze and categorize user reviews as positive, negative, or neutral.

Customer Service: Play Store reviews are often subjected to sentiment analysis to assign sentiment scores to user feedback.

Industry Sector: Evaluating whether a particular product is well-received or not within an industry.

Reviewers: Reviewers often use sentiment analysis to assess comments and provide an overall product review.

Challenges in sentiment analysis include interpreting tone, emoji-based data, and handling ironic or sarcastic comments. Additionally, distinguishing neutral statements can be a significant task.

Motivated by the potential of sentiment analysis, extensive work has been done, particularly in analyzing tweets using machine learning techniques. Twitter's massive data repository allows for the assessment of public sentiment, which is relevant for business decisions and even political assessments.
Machine learning tools are continuously evolving to handle more extensive and intricate datasets, providing faster and more accurate results. This empowers organizations to swiftly identify opportunities and potential risks.

Types of Sentiment Analysis:

Sentiment analysis encompasses various types, including:

Fine-grained sentiment analysis: Classifying sentiments into categories like very positive, positive, neutral, negative, or very negative, often rated on a scale from 1 to 5.

Emotion detection: Identifying emotions such as happiness, sadness, anger, or pleasantness, often employing lexicon-based methods.

Aspect-based sentiment analysis: Focusing on specific aspects of a product or service, such as battery life, screen quality, or camera performance.

Multilingual sentiment analysis: Assessing sentiments in different languages, classifying them as positive, negative, or neutral, which can be a complex task.

The future scope of sentiment analysis extends to a wide array of topics, aiding in understanding public sentiments on various matters. This analysis finds applications in diverse areas, from movie reviews and political assessments to corporate decision-making and customer satisfaction assessment.
Efficiently collecting relevant data and analyzing it for sentiment prediction are key components of sentiment analysis projects. Expanding the scope to include data from social media platforms like Facebook, personal blogs, and e-commerce websites enhances the accuracy and scope of results.

The goal is to create a user-friendly sentiment analysis web application, allowing individuals without technical backgrounds to make sense of the results easily. Such a tool can be instrumental for comparing various products, helping users in decision-making processes, such as choosing between different car models based on user sentiments.

Future Scope

Sentiment analysis can be used on a wide variety of motifs to figure out the sentiments of the general public about any matter. It has a wide variety of operations from getting reviews on a new movie, or a series to more serious operations like chancing out the disposition of the public on any new bill or overall satisfaction with the government. Business Intelligence is a field which regularly uses sentiment analysis for figuring out people's feelings about the products to help commercial bodies take opinions to earn better gains and make their consumers happier. All of it can be achieved by precisely opting the keywords and rooting the affiliated tweets. Majorly the design can be divided into two corridor
• Efficiently collecting as important applicable data as possible.
• Analysing that data to prognosticate the sentiments.

Right now, we're just collecting the data from twitter itself. We aren't using any data from Facebook. We can include that. Along with the fresh data from Facebook we can also explore colorful particular blogs and scrape applicable information from them. The better our data birth ways are, the better results we will be getting.

Bibliography

[1] Amitava Das, Sivaji Bandopadaya, SentiWordnet for Bangla,
KnowledgeSharing Event -4: Task, Volume 2,2010
[2] A.Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-132