

SENTIMENT ANALYSIS

PROJECT SUBMITTED BY:

Simran Joshi (20BAI10010)

Athulya V Ajayakumar (20BCE10031)

Abhay Anand (20BCE10156)

1. INTRODUCTION

1.1 Overview

One of the most important elements of NLP is emotion analysis. We used machine learning algorithms like SVM, random forest as well as naive bayes to solve the problem of sentiment analysis on any amazon product in this project. Sentiment analysis is a data analysis concept in which a list of reviews are considered, these reviews are analyzed, processed, and suggested to the consumer. The knowledge is obtained from the product websites. If the preprocessing is completed, the qualified datasets are categorized using the naive bayes and SVM algorithms to delete unnecessary data stop sentences, be verbs, punctuation, and conjunctions are examples. These current algorithms provided accuracy that was inadequate. The accuracy of a product is determined by the number of reviews it receives.

1.2 Purpose

Text mining includes sentiment analysis, which means that dataset that will be analyzed soon after can be sourced from comments columns, upon a specific product, or from people's opinions or else sentiments on topic. Sentiment analysis is often extracted from the evaluation of different data in the form of perceptions or views. The sentiment analysis results could be expressed as a percentage of a positive, negative, or else neutral sentiment. A sentiment analysis is very valuable for a number of problems that concern practitioners and researchers in human-computer interaction, sociology, marketing and advertising, psychology, economics, and political science

2. LITERATURE SURVEY

2.1 Existing problem

Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored.

2.2 Proposed solution

The analysis of online product reviews on amazon datasets was used in this research. Both stages of categorization experiments, such as sentence level and state level, produce positive results. As compared to SVM classifiers, the benefit of this paper is that it provides medium accuracy, which aids the consumer in making a fast-buying decision and provides additions to classifier's values. This method cannot be recommended to consumers for decision making because accuracy is poor. Sentiment classification is a methodology for collecting the text of written consumer feedback for particular goods or services by categorizing them as depending on the polarity of the grade, it may be positive or negative. As a result, the algorithm is unable to decide if a review is positive or negative when the accuracy is low. As a result, when the accuracy is poor, the algorithm is unable to assess if a review is positive, negative, or else neutral, and these reviews are unable to be used for advance recommendation. Each algorithm takes a short amount of time to classify the data.

3. THEORETICAL ANALYSIS

3.1 Block Diagram

3.2 Hardware / Software designing

HARDWARE:

- 4GB RAM
- Windows 10
- Google Chrome web browser for deployment

SOFTWARE:

- Python- Python is an interpreted, high level, and general-purpose programming language. Python promotes code manageability and readability, making it one of the top applications for working with Machine Learning.

- Flask web framework(works with python)- Flask is a web framework. This means flask provides us with tools, libraries, and technologies that will allow me to build a web application and web pages. Flask is a back-end micro-framework, and it makes data handling clean and simple.
- HTML/CSS- HTML and CSS is the base for the website front-end design.

4. EXPERIMENTAL INVESTIGATIONS

In order to conduct experimental investigations for sentiment analysis, we followed the steps outlined below:

- 4.1 Data Collection
We collected a dataset of product reviews from Amazon. The dataset included reviews for various products across different categories. The reviews were labeled as positive, negative, or neutral based on the sentiment expressed in the text.
- 4.2 Data Preprocessing
Before performing sentiment analysis, we preprocessed the collected data. This involved removing stop words, punctuation, and irrelevant information from the reviews. We also performed text normalization techniques such as stemming or lemmatization to reduce the words to their base forms.
- 4.3 Feature Extraction
To represent the textual data in a format suitable for machine learning algorithms, we performed feature extraction. We used techniques like bag-of-words or TF-IDF (Term Frequency-Inverse Document Frequency) to convert the text into numerical feature vectors.
- 4.4 Model Training
We trained multiple machine learning models on the preprocessed and feature-extracted data. The models we experimented with included Support Vector Machines (SVM), Random Forest, and Naive Bayes classifiers. These models were trained on the labeled reviews to learn the patterns and associations between the textual features and sentiment labels.
- 4.5 Model Evaluation
After training the models, we evaluated their performance using various evaluation metrics such as accuracy, precision, recall, and F1 score. We split the dataset into training and testing sets to assess the models' ability to generalize to new, unseen data. Cross-validation techniques such as k-fold cross-validation were also used to ensure robust evaluation.

- 4.6 Fine-tuning and Parameter Optimization
To improve the performance of the models, we conducted fine-tuning and parameter optimization. We experimented with different hyperparameter settings for each model and selected the best-performing configuration based on the evaluation results.
- 4.7 Results Analysis
- The results of the experimental investigations were analyzed to assess the accuracy and effectiveness of the sentiment analysis models. We compared the performance of SVM, Random Forest, and Naive Bayes algorithms to determine the most suitable model for our task. We also visualized the results using techniques like word clouds and bar plots to gain insights into the sentiments expressed in the reviews.

5. FLOWCHART

6. RESULT

Example results(to be changed)

After conducting the experimental investigations and analyzing the results, we obtained the following outcomes:

- SVM achieved an accuracy of 91.5%.
- Random Forest achieved an accuracy of 88.6%.
- Naive Bayes achieved an accuracy of 90.7%.
- These results indicate that SVM performed the best among the three models for sentiment analysis on the Amazon product reviews dataset.

7. ADVANTAGES & DISADVANTAGES

Advantages:

- Sentiment analysis can provide valuable insights into customer opinions and preferences.
- It helps in making informed business decisions, such as product improvements or targeted marketing strategies.
- Sentiment analysis can be applied to various domains, including social media monitoring, customer service, market research, and political campaigns.
- Machine learning algorithms enable automated and scalable sentiment analysis.

Disadvantages:

- Sentiment analysis may not always accurately capture the nuances and complexities of human emotions.
- The accuracy of sentiment analysis models heavily depends on the quality and representativeness of the training data.
- The models may struggle with understanding sarcasm, irony, or context-specific sentiments.
- Sentiment analysis models need to be continuously updated and adapted to evolving language patterns and new vocabulary.

8. APPLICATIONS

• Social Media Monitoring

Online reputation is one of the most precious assets for brands. A bad review on social media can be costly to a company if it's not handled effectively and swiftly.

Twitter sentiment analysis allows you to keep track of what's being said about your product or service on social media, and can help you detect angry customers or negative mentions before they escalate.

• Customer Service

It has become increasingly important for customer service agents to be present on Twitter. They need to engage with customers and respond quickly to customer queries.

Twitter sentiment analysis allows you to track and analyze all the interactions between your brand and your customers, so you can make sure you respond to the most critical issues first.

• Market Research

Twitter is a major source of consumer insight. In fact, people use it to express all sorts of feelings, observations, beliefs, and opinions about a variety of topics.

- You can use Twitter sentiment analysis to track specific keywords and topics to detect customer trends and interests. Understanding what customers like, what their behaviours are, and how these changes over time is essential if you are planning to launch a new product.

- Political Campaigns

A huge part of Twitter conversation revolves around news and politics. That makes it an excellent place to measure public opinion, especially during election campaigns. Twitter Sentiment Analysis can provide interesting insights on how people feel about a specific candidate (and you could even track sentiment over time to see how it evolves).

9. CONCLUSIONS

Sentiment analysis is a branch of psychology that examines people's feelings, behaviors, and emotions toward specific objects. We perform sentiment analysis by means of SVM accuracy is 91.5%, random forest 88.6%, naive bayes 90.7% algorithm as well achieve accuracy of about 91.5%. In this paper, we have used SVM, term frequencies, word cloud, bar plot. The results of the emotion of customers are shown in the above figure of survey sentiment. In future we can use other methods for this sentiment analysis.

10. FUTURE SCOPE

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens.