Sentiment Analysis using Machine Learning Algorithms

Dhrumil Mistry, Kalp Ranpura & Rashika Jain

BS Honours Computer Science, Ahmedabad University

Ahmedabad, India

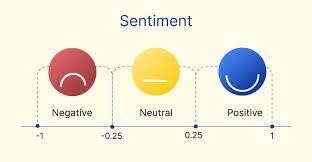
**Abstract- Sentiment analysis or opinion mining is the computational study of people’s opinions, sentiments, attitudes, and emotions expressed in written language. It has been one of the most active analysis areas in natural language processing and text mining in recent years. There are various existing models and algorithms for analyzing the user sentiments, and they can be classified into knowledge-based, statistical, or hybrid.**

**Keywords-Sentiment Analysis, Natural Language Processing, Prediction, Artificial Intelligence, Machine Learning, Deep Learning, Python.**

1. Introduction

Sentiment Analysis is a powerful approach that enables companies to understand the user emotions in their marketing campaigns. It plays a vital role in consumer loyalty satisfaction, advertisements, and promotional success, monitoring brand reputation on social media, gaining insights from customer feedback, and most importantly, product acceptance. Understanding consumer psychology allows them to alter their product roadmap with more precision and increase product and brand recall.

Sentiment analysis is a tool used to analyze texts for polarity, i.e., positive to negative. Training the machines automatically learns to detect sentiment without human input. The term emotion-based marketing encompasses emotional consumer responses, such as "positive," "negative," "neutral," "uptight," "disgust," "frustration," and others. These models can detect beyond mere definitions like sarcasm, context, or misapplied words. We can save hundreds of hours by using sentiment analysis as it can process the data and understand it as a human would.



The figure above depicts how the variation in the consumer reviews will help the model learn and predict a likely and accurate output.

1. Literature Survey

Much work has been done on Sentiment analysis, where individuals have utilized different methodologies and technologies. The most utilized approaches are the Bag of Words (BoW), AI algorithms, and deep learning procedures. The most normal methodologies used by individuals from all over the side world incorporate characterization models like Naive Bayes, Support Vector Machine, Random Forest, Decision Tree, KNN Classifier, etc. Albeit every one of the calculations functioned well when the data was preprocessed accordingly. This substantial number of models fundamentally centers around one of these two things, for example, subjectivity/objectivity and feature/aspect. We can't completely automate the sentiment analysis as, eventually, the outcome is to be validated by a human analyst.

1. Implementation

Our entire project was written and compiled in python. We will provide our implementation in a stepwise process:

1. Fetching Data

For our data collection, we looked up numerous datasets online. We needed data that had personal opinions in textual form as reviews. So, we chose a data set of Amazon baby products for training our model. It also had rating information which further helped in model training.

1. Pre-processing the Data

The data we retrieved from Amazon first had to be cleaned for training the model. We used Natural Language Processing Techniques for processing our data. First, we removed the neutral ratings(ratings=3) from our data. Then, we converted all the collected data into lower case. Next. Then, we removed all the duplicate data from the datasets. After that, we manipulated the data and converted the review ratings to 1(positive) and -1(negative). For the next step, punctuations had to be removed. We performed the process of lemmatization instead of stemming because it provided us with actual words compared to stemmed terms. The last step was to remove the stop words from the data to focus more on the essential information in the dataset.

1. Building Our Sentiment Analysis Model

The basic idea of the Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes. We will be using the Multinomial Naive Bayes classifier we imported from sklearn that is suitable for classification with discrete features, including text classification.

1. Initial Results

We achieved 84.8 percent accuracy on the testing data when applying the Naive Bayes model. The other model we used to our data is the Decision Tree classifier which gave us an accuracy of 75.4% on the testing data.

1. Conclusions And Future Work

The results are quite preliminary. Even though the accuracy of our model is quite high on the testing data, we are yet to assess real-world data on our model. Further, we have to optimize the Decision Tree model for future work. Also, we are planning to try other classification models like Random Forest, KNN classifier, and SVM on the dataset.

ACKNOWLEDGEMENT

We want to thank Prof. Mehul Raval and Ahmedabad University for allowing us to conduct this study as part of our mid-semester evaluation project for CSE 523 Machine Learning. Further, we extend our gratitude to the TAs for guiding us through this study.

REFERENCES

1. Sentiment Analysis & Machine Learning. MonkeyLearn Blog. (2020, April 20). Retrieved March 20, 2022, from [https://monkeylearn.com/blog/sentiment](https://monkeylearn.com/blog/sentiment-analysis-machine-learning/%23:~:text=Sentiment%20analysis%20is%20a%20machine,detect%20sentiment%20without%20human%20input)analysis-machinelearning/#:~:text=Sentiment%20analysis% 20is%20a%20machine,detect%20sentimen t%20without%20human%20input
2. Performing sentiment analysis with naive Bayes classifier! Analytics Vidhya. (2021, July 13). Retrieved March 20, 2022, from [https://www.analyticsvidhya.com/blog/202](https://www.analyticsvidhya.com/blog/2021/07/performing-sentiment-analysis-with-naive-bayes-classifier/%20) 1/07/performing-sentiment-analysis-withnaive-bayes-classifier/
3. Foy, P. (2021, July 26). Naive Bayes for sentiment analysis &amp; Natural Language Processing (NLP). MLQ.ai. Retrieved March 20, 2022, from

[https://www.mlq.ai/sentiment-analysis](https://www.mlq.ai/sentiment-analysis-with-naive-bayes/)with-naive-bayes/

1. Kuzminykh, N. (2020, October 23). Sentiment Analysis in python with textblob. Stack Abuse. Retrieved March 20,

2022, from

[https://stackabuse.com/sentiment-analysis](https://stackabuse.com/sentiment-analysis-in-python-with-textblob/)in-python-with-textblob/

1. K. Nigam, J. Lafferty, and A.

McCallum, “Using maximum entropy for text classification,” in IJCAI-99 workshop on machine learning for information filtering, 1999, vol. 1, pp. 61–67.

1. H. Wang, D. Can, A. Kazemzadeh, F.

Bar, and S. Narayanan, “A System for

Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,”

Proc. 50th Annu. Meet. Assoc. Comput. Linguist., no. July, pp. 115–120, 2012.

1. P. K. Singh and M. Shahid Husain,

“Methodological Study Of Opinion

Mining And Sentiment Analysis Techniques,” Int. J. Soft Comput., vol. 5, no. 1, pp. 11–21, 2014.

[8]Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. *Int. J. Multidiscip. Sci. Eng*, *8*(3), 27.