

**E-COMMERCE REVIEW  
ANALYSIS**

**Internship Report**

# 

## **BY:**

MOHAMMED KHIZAR BAIG - AP19110010167

Boyapati Sai Venkat – AP19110010174

Ishita Agrawal – AP19110010511

G.B.N.V.S. Abhinav – AP19110010166

N. Sree Aravind Bhaskar – AP19110010196

# 

# 

# **Abstract**

It is a project which we intend to make using Sentimental Analysis, it starts with the notion of an assessment of sentiments that have emerged as a method for understanding clients' emotions and an efficient knowledge representation for recognizing opinions on the products of E-Commerce companies and segregated into categories.

****

# **Introduction**

Due to advancements in the field of deep learning and improvements in the computational power of hardware systems, natural language processing, a sub-field of machine learning, has gained enormous popularity in both research and industrial applications in the last 5 years in both research and industrial applications. It is a method for computers to comprehend how human languages work by combining computational linguistics and machine learning.

Sentiment analysis is one of the primary areas where NLP has been widely applied. Understanding consumer behavior and needs in relation to a company's products and services is crucial for businesses. In general, a customer's feedback on a product can be classified as Positive, Negative, or Neutral. Using product reviews to interpret client feedback allows businesses to assess how well they're doing. For Sentimental Analysis (SA), there are two approaches which are semantic orientation and statistical Machine Learning (ML) approach. The first method determines the document's sentiment based on the sentimental words and phrases extracted. Furthermore, the latter focuses on the document's sentiment, which is based on the sentimental elements extracted as well as the ML.

In terms of SA correctness, the statistical scheme triumphs over the semantic system. Although semantic word spaces are tremendously useful, they are unable to communicate the meaning of long sentences morally. Machines must be dependable and efficient to comprehend as well as understand human emotions and sensations. Hence, we will be explaining a sentimental analysis of an amazon product by considering the product review dataset.

# 

# **Literature review**

E-commerce websites allow customers to review every product available on their site leading to one of the reasons for their increasing popularity. Purchasers can leave unbiased feedback on different products, generating millions of reviews every day. This makes it quite difficult for the product manufacturers to categorize the customer opinion of their products. So, it’s important to classify such large and complex amounts of data to derive the required information. Classification algorithms are the best way to tackle such a problem.

### SENTIMENT ANALYSIS:

It is a natural language Processing problem that identifies and extracts information from the text resources. Here the use of sentiment analysis will be to analyze the reviews written by the people and categorize them into positive or negative. So, the particular system does not have to go through the process of understanding the semantics of each phrase.

In this, the words are labeled as positive or negative, but there is a chance the errors might occur in recognizing words. For example, the word “amazing” has a prior positive polarity, but if it comes with a negation word like “not”, the context can completely change (Singla et. al 2013). Classifying text with prior negative or positive polarity will not always work.

### SENTIMENT ANALYSIS USING MACHINE LEARNING:

A lot of research has been done in the field of machine learning algorithms. One of the best approaches for sentiment analysis is machine learning algorithms.

One of the first definitions of machine learning that has been provided by Tom Mitchell (1997) in his book Machine Learning is as follow:

”A computer program is said to learn from experience E concerning some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.”

Machine learning has lots of developed algorithms to reform the performance of systems by using the data given. To solve sentiment analysis using a machine learning model, there is a two-step procedure. In the first step, we train a model that is learning from the given training data, and in the second step, we classify the unseen data with the help of the training model (Khairnar & Kinikar 2013).

The machine learning algorithm is classified into two different categories:

* Supervised learning
* Unsupervised learning

### SUPERVISED LEARNING:

In this, the algorithm learns from the training dataset by making repetitive predictions and adjusting the answer that is done to the given dataset so that accurate results can be found.

The model is taught in such a way that while giving the output, it can come up with different conclusions. Here the training data is already labeled and provided with both input and output data. The training model has to learn how to process itself to reach the output for any given input.

The more labels are provided to the classifier the more precisive is the output. The major goal is to train the model so that it can provide the required output for any new input. Working with supervised learning can be quite challenging, like if the dataset is not labeled then the machine may face issues while generating output.

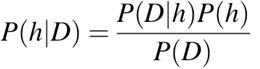
### UNSUPERVISED LEARNING:

In this, the training dataset is unlabeled with no corresponding output. The model should itself work and discover the underlying structure of the dataset which means it has to find out various similar patterns in the given data to determine the correct output without having any right answers.

Clustering is one of the most important techniques of unsupervised learning. Clustering will help to group the data into groups where every individual has the same characteristics. Clustering helps in solving sentiment analysis by grouping large and complex datasets.

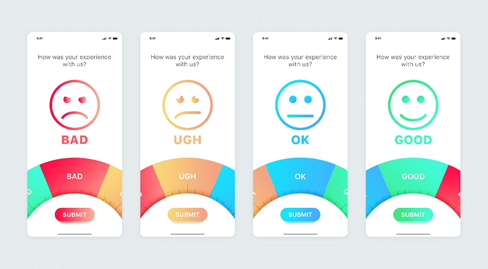
### NAIVE BAYES:

The Naive Bayes method is an easy and fast machine learning algorithm. It is based on the Bayes theorem and is suitable for classifying massive amounts of data. The assumption that is taken into account is that any individual feature in the class is independent of the others. These characteristics may be interrelated, but they are normally considered separate. This assumption is called conditional independence.



SVM:

Support vector machine is a supervised learning algorithm used for sentiment analysis problems. This method uses a decision plane to place labeled training data, and then an algorithm to generate an optimum hyperplane that divides the data into separate groups or classes.



# 

# **Proposed Methodology**

A huge amount of data is being generated every second in today’s digital world. Such large data becomes very difficult to process using manual checking and traditional methods to sort/understand. So, to handle this, we are developing a Sentimental Analysis. Understanding the sentiment of textual data is a very crucial component in data science.

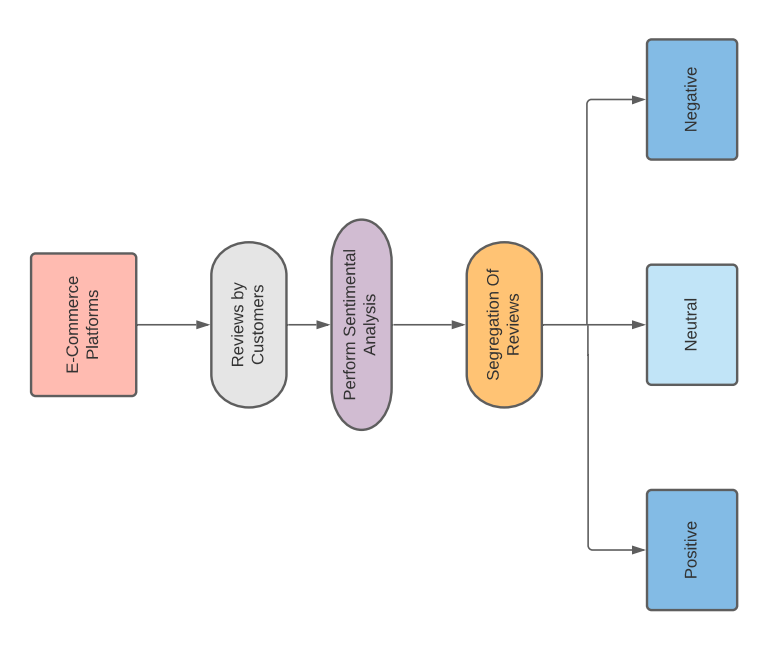
As feedback plays a crucial role when it is related to e-commerce, products, digital marketing, and so on. Every host wants to improve their product/work, it’ll more effective when he/she goes through what their customer wants, they receive the feedback through public and to go through what they want, we have a solution “Sentimental Analysis”, that sorts the feedback through the sections (Positive, Neutral, Negative) and helps them to improve.

Sentiment Analysis is also known as Opinion Mining refers to the use of natural language processing, text analysis to systematically identify, extract, quantify, and study affective states and subjective information.

Sentiment analysis studies people’s sentiments towards certain entities. It is widely applied to reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. We use a database of sentiment-based keywords along with positivity, negativity, and neutral weight in the database, and then based on these sentiment keywords mined in user review is ranked.

# 

# **Flow Chart**



# 

# **Results**

We have taken the project topic by referring to some of the existing topics such as E-Commerce review analysis and by observing the different options we generally see in the different platforms while giving a review for a product such as Online Shopping websites, Food Delivery Apps, Medical suggestions websites about medicines, etc.

So, with the help of this project by making it thought of segregating the reviews data into three categories are positive, negative, and neutral. This will determine the polarity of reviews so it helps for the product improvisation because we can’t acknowledge the huge data we can drop it down into these categories so that by the count of these categories we can analyze the reviews. This helps in enhancing the business operation.

# 

# **Conclusion**

t

# 

# 

# **References**

* Singla, Z., Randhawa, S., & Jain, S. (2017). Sentiment analysis of customer product reviews using Machine Learning. 2017 International Conference on Intelligent Computing and Control (I2C2). <https://doi.org/10.1109/i2c2.2017.8321910>
* Sklearn naive Bayes classifier python: Gaussian naive Bayes Scikit-Learn tutorial. DataCamp Community. (n.d.). Retrieved September 20, 2021, from <https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn>.
* 1.4. support vector machines. scikit. (n.d.). Retrieved September 20, 2021, from <https://scikit-learn.org/stable/modules/svm.html>.
* Michael J. Greis CFA, Michael, J., Greis, & amp; Cfa. (2018, October 25). Mainstreaming Sustainable Investing. CFA Institute. Retrieved September 20, 2021, from <https://www.cfainstitute.org/en/research/foundation/2018/mainstreaming-sustainable-investing?s_cid=ppc_RF_Google_Search_MainstreamingSustainableInvesting>.