Analysing Amazon Product Reviews Using Machine Learning

Kavuru.Pavani [1]
Katam.Kusuma [2]

Medicharla. Hema Venkata Lakshmi [3]
³Swarnandhra College of Engineering & Technology.

Abstract:- Reviews from customers are now an integral part of most online e-commerce (Social media , amazon ,Flipkart, meesho) Companies for their daily operations. customers reviews are becoming a major factor in determining what consumer decide to buy. For analyzing the growth of e-commerce companies based on customer reviews or feedback.

Organizations typically don't have the time or resource to scour the internet and read and analyse every piece of data relating to their products, services and brand . Sentiment analysis is an important way for organizations to understand how customers perceive and experience their products and brands. Increasingly, customer feedback is given online through a variety of unconnected platforms, such as Amazon product reviews and posts on social media platforms.

In e-commerce Companies, there are huge amount of reviews it is difficult to measure their growth based on these customer reviews .In this we are using SENTIMENT ANALYSIS (SA) is a machine learning algorithm is used to determine whether a given text contains positive(excellent, good),negative(bad, wrost) or neutral comments(average). Sentiment analysis ,also knowns as opinion mining , is the process of extracting subjective information from text and determining the sentiment expressed within it .It is a text of NPL(natural language processing) . IIn the context of product reviews data , sentiment analysis involves understanding the emotions and opinions of customers towards specific products or brands .

In the Sentiment Analysis spilt into various types like Emotion Detection(ED), Aspect Based Sentimental Analysis (ABSA), Fine Grained Sentimental Analysis(FGSA) , Multilingual Sentimental Analysis (MSA).These types are used to analyse the customer reviews is either positive or negative .Emotion detection is the process of identifying human emotion. ED is widely helpful for recognizing the emotions of other .Aspect based sentimental analysis is Breaks down text into aspects (component of products), and then allocates each one a sentimental level(positive ,negative ,neutral).Fine grained sentimental analysis is done at text and sentence level . Multingual sentimental analysis done in multiple languages and also done by the use of complex neural network architecture. The techniques in sentiment analysis are logistic regression, naivebayes, random forest classifier, SVM(support vector machine) etc.

By analyzing the sentimental expressed in these reviews, businesses can gain a comprehensive understanding of how their product are perceived by customers.

Keywords:- Sentiment Analysis; Logistic Regression ;Naive Bayes ;Random Forest; SVM.

I. INTRODUCTION

With the rise of online shopping, many consumers published their shopping experiences on the e-commerce platform, which became online product reviews. Today, it only takes one click and we are just one day away from getting the product to your doorstep. As technology changes every day and more options are available, people prefer to shop online instead of going to a store (Xiong et al., 2020). Online shopping giants like Amazon, Flipkart and Myntra make it easy by offering different products in different categories throughout the year.

In online ,the product prices are less than the traditional market. Due to the rapid development of e-commerce, the study of online product reviews is increasing. The article examines these literatures in relation to the creation of online product reviews, implications and influencing factors..

There are few generations are available in Online Product Reviews.

A. Consumer Participation Motivation:

Consumer motivation is an internal state that drives people to identify and buy products or services that fulfill conscious and unconscious needs or desires.

B. The Popularity of the Product:

From the point of view of product popularity, Chrysanthos and Guodong Gao investigated whether the online review behavior of consumers can influence the quiz between popular products and niche products.

C. User Interaction:

Result is an in-depth external analysis of your website or product focused on optimizing user experience and increasing value. From high bounce rates to accessibility features, it helps your company develop its products better.

D. Self-Selection Biases:

Self-selection bias in market research is a type of non-sampling error in which the respondents who complete a survey are demographically or behaviorally different than the intended sample.

In the third generation ,we have to called as Outcome Variable For Online Product Reviews. Intheir, they have four(4) generations are available like product sales and revenue, consumer's beliefs, competition between upstream Manufactures, consumer's preference for product attributes.

In the fourth generation ,is usually called as Factors That Influence The Usefulness of Online Product Reviews. Assame as ,we have fifth generation called as Research Result.

In the sixth generation of online product review is Research Prospect. They contains some generations are try to explore the deep mechanism of influence, look for internal regulators, establish theoretical models to clarify the complex effects.

In conclusion, that product sales are positively influenced by reviews with higher affective and social content and by those that use more informal language.

II. LITERATURE SURVEY

[1]. H. M. Ahmed[1] says **Nowadays, theoretical** analysis has **gained importance to he lp** researchers use methods and techniques, especially for **big data.** Amazon's food **information** is the **information** that will solve this problem. In this article, we explore various methods and techniques for sentiment analysis of Amazon Fine Food **r eview big data** using Apache Spark data **processing.**

[2]. A. Mabrouk [2] **says It has recently been found that e-commerce** (EC) websites **can** provide more valuable informat ion than human intelligence. To help consumers compare alter natives when **shopping,** the authors previously developed a **de cision** based on consumer reviews. They do not care about the **standard profile** provided by the manufacturer, but **the** descri ption contains the **best** products and the text is based on facts, unlike reviews. Therefore, this article **proposes** a method calle d SEOpinion (Opinion and Analysis), which combines **two im portant levels of** data **structures** and customer analysis to **gat her context** and **try** to **think about** these **thoughts.**

[3]. A. DadhichOver[3]
says the past few **years,** the retail industry has **made heavy us e** of online sales to **provide** advice, feedback and **ideas to cust omers.** The purpose of this article is to provide an **accurate** an alysis. It should classify **analyzes** as **good, bad** and **average** u sing five main learning **methods** such as NB, LR, SentiWord Net, RF and KNN. This paper also discusses the experimental results and **the** problems **inherent** in **the research results.**

[4]. Al Qahtani [4] says We have worked on many machine learning **algorithms such as** logistic regression, random forest, naive bias, **binary** long-term **memory,** and bert. We then evaluate the model using acc uracy, f1 score, precision, recall, and cross
entropy loss functions. We then **analyze** the **best performing** models to **understand** their **distribution. The evaluation** is d one by classifying **several groups,** then we select the best mo del and **reinform** it **into** binary classification.

III. METHODOLOGY

This Study aims to create a Sentimental Analysis using machine learning for online product reviews. Sentimental Analysis is nothing about it uses machine learning models to perform text analysis of human language.
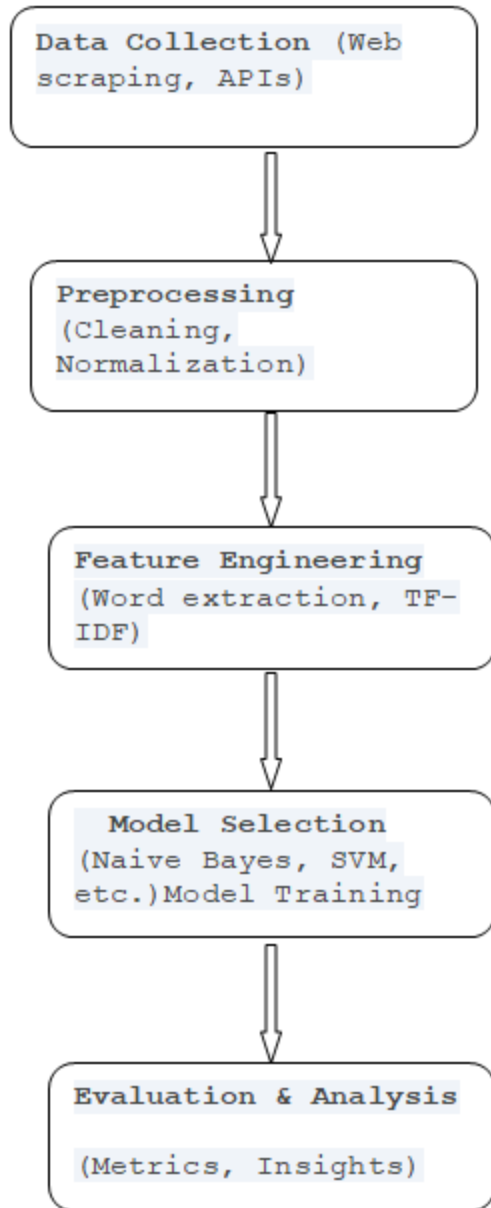
Fig 1: Process of Sentimental Analysis

- ➢ *Steps to be followed*
- Importing Libraries and Datasets
- Preprocessing and cleaning the reviews
- Analysis of the Dataset
- Converting text into Vectors
- Model training, Evaluation, and Prediction

Once the dataset is ready, we will clean the review column by removing the stopwords. Once we have done with the preprocess, observe rows to see the improved dataset. Once analysis and vectorization is done. We can now explore any machine learning model to train the data. But before that perform the train-test split. Now we can train any model. we can use sentiment analysis techniques like naïve bayes, logistic regression, random forest classifier etc.

Random Forest is a popular machine learning algorithm used for classification and regression tasks due to its high accuracy, robustness, feature importance, versatility, and scalability. Random Forest reduces overfitting by averaging multiple decision trees and is less sensitive to noise and outliers in the data.
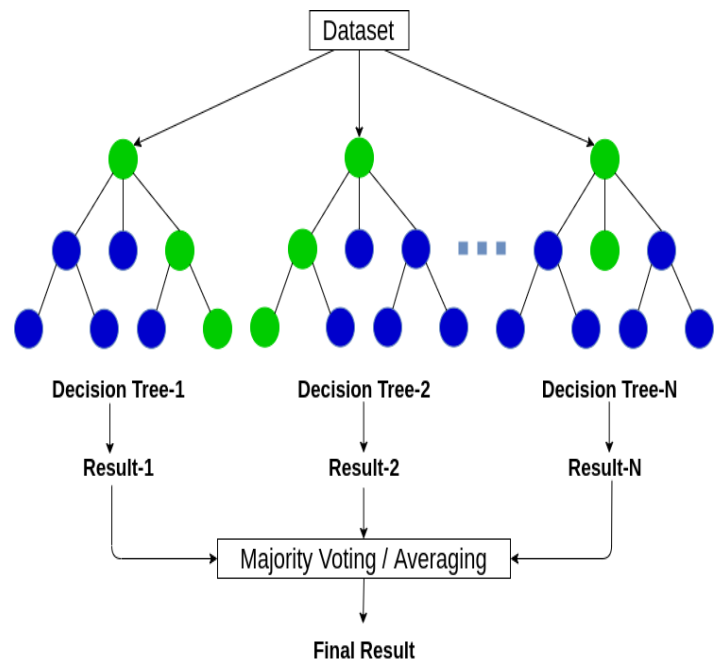

Fig 2: Random Forest

Logistic regression is a data analysis technique that uses mathematics to find the relationships between two data factors. It then uses this relationship to predict the value of one of those factors based on the other. The prediction usually has a finite number of outcomes, like yes or no.

Here, we implemented logistic regression and random forest classifier for the data set mentioned below. These are used to predict sentiment for new reviews.

|	marketplace	customer_id	review_id	product_id	product_parent	product_title	product_category	star_rating	helpful_votes	total_votes	vine	verified_purchase	review_headline	re
0	US	11555559	R1QXC7AHHJBQ3O	B00IKPX4GY	2693241	Fire HD 7, 7" HD Display, Wi-Fi, 8 GB	PC	5	0	0	N	Y	Five Stars	(
1	US	31469372	R175VSRV6ZETOP	B00IKPYKWG	2693241	Fire HD 7, 7" HD Display, Wi-Fi, 8 GB	PC	3	0	0	N	N	Lots of ads Slow processing speed Occasionally...	Lots of ads processing speed
2	US	26843895	R2HRFF78MWGY19	B00IKPW0UA	2693241	Fire HD 7, 7" HD Display, Wi-Fi, 8 GB	PC	5	0	0	N	Y	Well thought out device	Excellent unit. Th of
3	US	19844868	R8Q39WPKYVSTX	B00LCHSHMS	2693241	Fire HD 7, 7" HD Display, Wi-Fi, 8 GB	PC	4	0	0	N	N	Not all apps/games we were looking forward to ...	I bought this Prime so I en
4	US	1189852	R3RL4C8YP2ZCJL	B00IKPZ5V6	2693241	Fire HD 7, 7" HD Display, Wi-Fi, 8 GB	PC	5	0	0	N	Y	Five Stars	All Amazon produc to meet

Fig 3: sample data for processing.
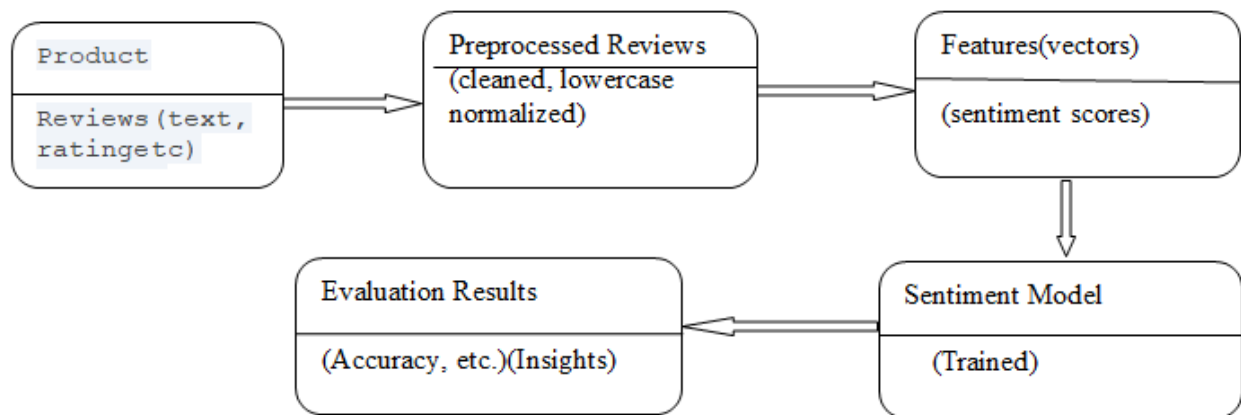
Fig 4: Process Flow of Logistic Regression

By comparing both logistic regression and random forest classifier, logistic regression is a best fit with 91% accuracy.

IV. APPLYING LOGISTIC REGRESSION

For this dataset we got the accuracy upto 91% by using logistic regression algorithm

```
lr = LogisticRegression()
lr.fit(X_train_tvec, Y_train)
```

```
▼ LogisticRegression
LogisticRegression()
```

```
lr_predictions = lr.predict(tvec.transform(X_test_cleaned))
report = classification_report(Y_test,lr_predictions, output_dict=True)
data_report = pd.DataFrame(report).transpose().round(2)
cm = sns.light_palette("purple", as_cmap=True)
data_report.style.background_gradient(cmap=cm)
```

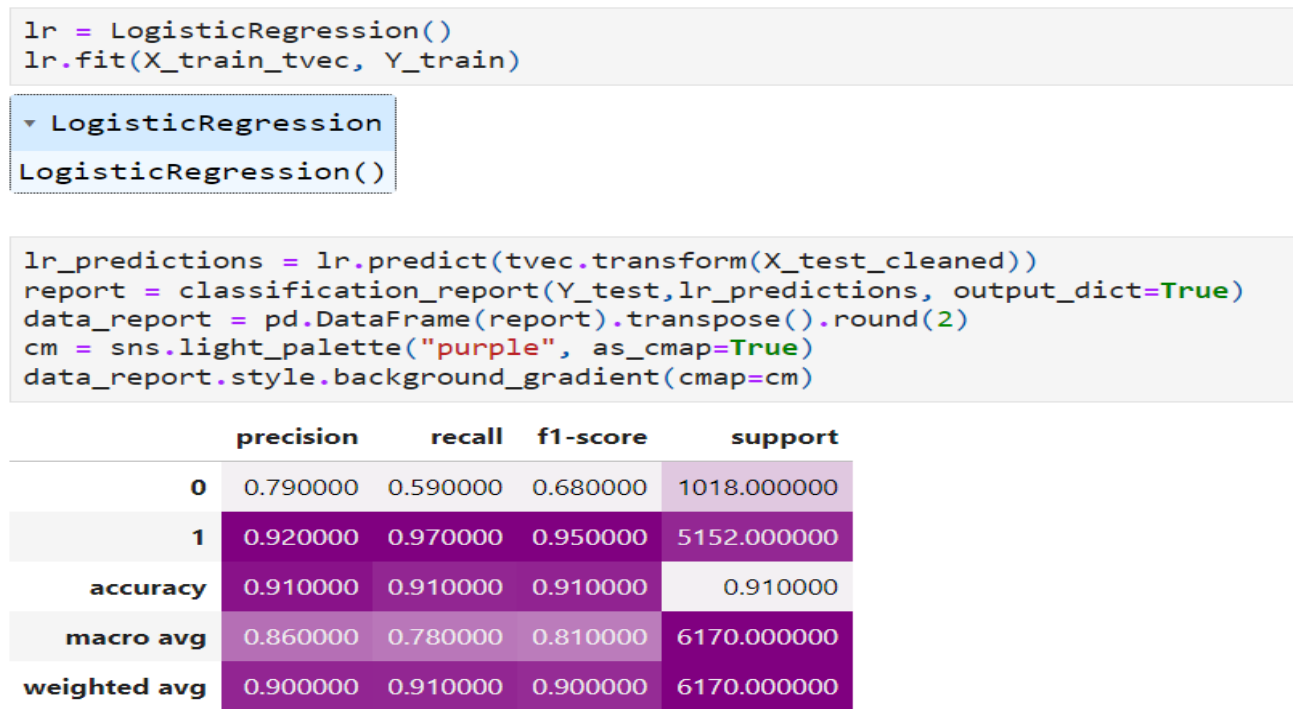|	precision	recall	f1-score	support
0	0.790000	0.590000	0.680000	1018.000000
1	0.920000	0.970000	0.950000	5152.000000
accuracy	0.910000	0.910000	0.910000	0.910000
macro avg	0.860000	0.780000	0.810000	6170.000000
weighted avg	0.900000	0.910000	0.900000	6170.000000

Fig 5: code for implementation of logistic regression.

V. CONCLUSION

In this analysis, we have taken the dataset from github based on Sentimental analysis of amazon customer reviews. In -conclusion, online product reviews serve as valuable insights for potential buyers, offering a diverse range of opinions and experiences. By analyzing the sentiment of product reviews, e-commerce platforms can better understand their customers' needs, improve the quality of their products and services, and ultimately increase customer satisfaction and loyalty.

REFERENCES

[1]. H. M. Ahmed, M. J. Awan, N. S. Khan, A. Yasin and H. M. F. Shehzad, "Sentiment analysis of online food reviews using big data analytics", *Elementary Educ. Online*, vol. 20, no. 2, pp. 827-836, 2021.

[2]. A. Mabrouk, R. P. D. Redondo and M. Kayed, "SEOpinion: Summarization and exploration of opinion from E-commerce websites", *Sensors*, vol. 21, no. 2, pp. 636, Jan. 2021.

[3]. A. Dadhich, B. Thankachan Sentiment analysis of amazon product reviews using hybrid rule-based approach Smart Systems: Innovations in Computing, Springer, Singapore (2022), pp. 173-193.

[4]. Al Qahtani, Arwa S. M., "Product Sentiment Analysis for Amazon Reviews (2021)". International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 3, June 2021.