SENTIMENT ANALYSIS OF AMAZON REVIEWS USING MACHINE LEARNING MODEL

OMAR MOHAMMED ALI ALBAAGARI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Your Degree

Master in Data Science
Faculty of Computing
Universiti Teknologi Malaysia

OCTOBER 2023

TABLE OF CONTENTS

	TITLE	PAGE
TABLE OF CONTENTS		ii
CHAPTER 1	**INTRODUCTION**	**1**
1.1	Problem Background	1
1.2	Problem Statement	2
1.3	Research Objectives	2
1.4	Scope of the research	3
REFERENCES		**5**

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Emotions are present in every single situation in which people engage with one another [1] . In a variety of contexts, they have the ability to mould an individual's view of an experience, a subject, an issue, and so on. Through a number of channels, Message boards, comments, and reviews are all examples of places where we could get feedback. and opinions about a wide range of items, both online and offline. These feedback and ideas may be presented in the form of text, video, image, and other forms. Each and every kind of feedback has a certain degree of emotion, such as whether the experience as a whole was positive, negative, or neutral.

In the twenty-first century, the internet has grown into a technology that is indispensable to our everyday life [2]. People are purchasing things from a large number of e-commerce websites in the current period, and it is more likely that they would first assess the products before purchasing them [3]. In this modern era, all commercial enterprises have turned their focus to the internet technology. These days, the majority of websites that are dedicated to e-commerce now have a specific part where customers may submit evaluations of products or services.

Sentimental analysis is one of the machine learning processing techniques that helps detect feelings [4]. This approach enables business owners to collect information about the perspectives of their customers via various online media, such as social media, questionnaires, and analyses of websites that allow for online shopping. A better understanding of the factors that contribute to the deterioration of the commodity will be possible as a result of this information. In this analysis, two things are taken into consideration: the line "Apple Iphone 15 battery life is good and speakers' quality is not good" is an example of sentiment analysis. Sentiment analysis represents the

behavior of the consumer with regard to the product, as well as the reputation of the company. When it comes to the quality of the speakers, there is a negative opinion, but the battery life is great [5].

1.2 Problem Statement

Customer ratings and reviews reveal the buyer's judgment on the product. It might be positive, negative, or neutral. When it comes to a product, customers could give it four or five stars, indicating contentment, while others may give it one or two stars, indicating discontent. When it comes to sentiment analysis, this does not provide any kinds of challenges. Other individuals, on the other hand, have given it three stars, despite the fact that they have definitely expressed their overall delight with it. As a result, this causes confusion among other consumers as well as among businesses that are interested in learning their genuine opinion. As a consequence of this, both consumers and businesses have challenges when it comes to assessing evaluations and comprehending the level of happiness experienced by customers. Therefore, the three-star rating does not genuinely indicate a neutral mood. This is due to the fact that in actuality, individuals who give a product or service a rating of three stars do not always imply that they are completely balanced in their judgment between both positive and negative opinion.

Based on the issue that mentioned above, the purpose of this research is to do a sentiment analysis applied to the Amazon smart device dataset reviews in order to forecast the opinions of customers. This research will use real dataset from Amazon using the counter vectorizer (CV), term frequency inverse document frequency (TF-IDF), logistic regressor, and Naïve Bayes as the principal statistical tools.

1.3 Research Objectives

(a) To extract the feelings that are expressed in customer evaluations and to conduct an analysis of these expressed feelings.

(b) To train a machine learning model that is capable of sorting client evaluations into two unique sentiment categories, namely positive, neutral, and negative categories.

(c) To develop a dashboard that summarize the analysis

1.4 Scope of the research

Objective 1:

For the purpose of pre-processing the data, a counter vectorizer, term frequency, and inverse document frequency are used. Additionally, the logistic regressor, which is utilized for the building of the model via the utilization of the scikit-learn package in Pythons.

Objective 2:

Logistic Regression, Naïve Bayes are used in this research for the sentiment analysis and selecting the best model performance

Objective 3:

Power Bi is used to build an interactive dashboard that describe the analysis

REFERENCES

1. De Saa, E. and Ranathunga, L. Self-reflective and introspective feature model for hate content detection in sinhala youtube videos. *2020 From Innovation to Impact (FITI)*. IEEE. 2020, vol. 1. 1–6.

2. Ripa, S. P., Islam, F. and Arifuzzaman, M. The emergence threat of phishing attack and the detection techniques using machine learning models. *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*. IEEE. 2021. 1–6.

3. Rathor, A. S., Agarwal, A. and Dimri, P. Comparative study of machine learning approaches for Amazon reviews. *Procedia computer science*, 2018. 132: 1552–1561.

4. Rajat, R., Jaroli, P., Kumar, N. and Kaushal, R. K. A sentiment analysis of amazon review data using machine learning model. *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*. IEEE. 2021. 1–6.

5. Gope, J. C., Tabassum, T., Mabrur, M. M., Yu, K. and Arifuzzaman, M. Sentiment analysis of Amazon product reviews using machine learning and deep learning models. *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*. IEEE. 2022. 1–6.