# A Project Based Learning Report on:



Amazon Sentiment Analysis

*In partial fulfilment of the requirement For Project Based Learning of* **BACHELOR OF TECHNOLOGY**

*In*

**COMPUTER ENGINEERING**

*For*

# Big Data Analytics

****

### BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY) COLLEGE OF ENGINEERING, PUNE-43

**DEPARTMENT OF COMPUTER ENGINEERING**

2024-25



SUBMITTED TO:

Prof Sunita Dhotre (Coordinator)



SUBMITTED BY:

PRN

Abhishek Nimbalkar (127) 2314110743

Shreyas Pande (120) 2314110728

|  |  |
| --- | --- |
| Jayesh Nimbalkar (125) | 2314110741 |
| Shivam Debadwar (126) | 2314110742 |
| Sumit Chavan(124) | 2314110740 |

# BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY) COLLEGE OF ENGINEERING, PUNE-43

**DEPARTMENT OF COMPUTER ENGINEERING**

****

## CERTIFICATE

This is to certify that the Project Based Learning report title Amazon Sentiment Analysis



, submitted by **Abhishek Nimbalkar (127), Shreyas Pande (120), Sumit Chavan (124), Jayesh Nimbalkar (125)**,**Shivam Debadwar(126)** to the Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune-43 for the award of the degree of **BACHELOR OF TECHNOLOGY** in Computer Engineering is a bonafide record of the PBL work done by them under my supervision.

Place: Pune Prof. Sunita Dhotre

(Course Coordinator)

Date:

## Acknowledgement

To begin with, we would like to thank our Course Coordinator for providing us this golden opportunity. We are very grateful to our Course Coordinator, Prof. Sunita Dhotre ma’am for her valuable support and guidance.

We are also very grateful to all those who helped us in our entire project. We would like to thank Prof. Veena J. Jadhav ma’am, our Course Coordinator and our HOD Dr. S. Vanjale sir for their support and guidance in completing our project based learning on the topic **Amazon Sentiment Analysis** Through Message Passing. It was a great learning experience.

I would like to take this opportunity to express my gratitude to all my group members Abhishek Nimbalkar, Shreyas Pande, Shivam Debadwar, Sumit Chavan, Jayesh Nimbalkar. The project based learning would not have been successful without their cooperation and inputs. At last, but not the least, we are thankful to our parents and families who encourage us to study and who support us financially and who gave us the opportunity to spend our golden time in learning.

Thanks again to all who helped us.

## Index

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Title** | **Page No.** |
| 1. | Introduction And Overview | 5 |
| 2. | Data Sources | 6 |
| 3. | Methodologies | 7 |
| 4. | Code and Output | 8-12 |
| 5. | Challanges | 13 |
| 6. | Conclusion | 14 |

### INTRODUCTION AND OVERVIEW

This presentation examines the significance of sentiment analysis for Amazon, detailing methodologies, data sources, challenges, and the insights gleaned from consumer sentiment.

### Definition of Sentiment Analysis:

Sentiment analysis is a subset of Natural Language Processing (NLP) that focuses on extracting subjective information from textual data. It aims to determine the emotional tone behind a series of words, providing insights into consumer attitudes and opinions.

### Importance in Business:

Sentiment analysis is crucial for businesses as it enables the understanding of customer opinions, improves product development, enhances customer service, and provides competitive insights. This analytical tool helps organizations adapt strategies based on consumer sentiment.

### Application in Amazon:

Amazon utilizes sentiment analysis to monitor customer feedback on products, identify trends, optimize the shopping experience, and refine recommendations. By analyzing user sentiments, Amazon enhances customer satisfaction and drives sales.

### Data Sources:

**Customer Reviews:**

Customer reviews on Amazon offer critical insights into user experiences and satisfaction levels. By analyzing these reviews, businesses can spot strengths and weaknesses in products and services**.**

### Product Ratings:

Product ratings provide a quantitative measure of customer satisfaction. Analyzing the correlation between ratings and sentiment allows companies to assess overall product acceptance.

### Social Media Feedback:

Social media platforms present a rich source of user sentiment. Monitoring mentions and discussions about Amazon products helps gauge public perception and adapt marketing strategies accordingly.

### Methodologies

**Natural Language Processing:**

Natural Language Processing (NLP) is a crucial methodology for sentiment analysis, enabling the extraction of meaningful insights from unstructured text. Through techniques such as tokenization, stemming, and lemmatization, NLP transforms raw data into a format suitable for analysis. Algorithms analyze language patterns and sentiments, ensuring accurate interpretation of customer opinions.

### Machine Learning Techniques:

Machine learning techniques are fundamental in enhancing the accuracy of sentiment analysis. Supervised learning models, such as Support Vector Machines and Random Forests, can be trained on labeled datasets to classify sentiments effectively. Unsupervised learning approaches, such as clustering and topic modeling, can identify hidden sentiments within vast amounts of unlabelled customer feedback.

### Text Classification Approaches:

Text classification approaches play a pivotal role in sentiment analysis by categorizing text data based on the sentiment expressed. Common methods include rule-based systems, where predefined rules identify sentiments, and deep learning approaches like recurrent neural networks (RNNs) that learn from data to enhance classification. Each method offers unique advantages and trade- offs.

### Code And Output:

**Code:**

install.packages(c("tidytext", "dplyr", "ggplot2", "readr", "lubridate", "stringr", "tidyr")) library(tidytext)

library(dplyr) library(ggplot2) library(readr) library(lubridate) library(stringr) library(tidyr)

reviews <- read\_csv("customer\_reviews.csv")

reviews$date <- ymd(reviews$date) # convert to Date format glimpse(reviews)

data("stop\_words")

tidy\_reviews <- reviews %>% unnest\_tokens(word, review\_text) %>% anti\_join(stop\_words)

sentiment\_words <- tidy\_reviews %>% inner\_join(get\_sentiments("bing"), by = "word")

product\_sentiment <- sentiment\_words %>% count(product, sentiment) %>%

pivot\_wider(names\_from = sentiment, values\_from = n, values\_fill = 0) %>% mutate(net\_sentiment = positive - negative)

sentiment\_words <- tidy\_reviews %>% inner\_join(get\_sentiments("bing"))

ggplot(product\_sentiment, aes(x = reorder(product, net\_sentiment), y = net\_sentiment, fill = net\_sentiment > 0)) +

geom\_col(show.legend = FALSE) + coord\_flip() +

labs(title = "Net Sentiment by Product", x = "Product", y = "Net Sentiment Score") + scale\_fill\_manual(values = c("FALSE" = "red", "TRUE" = "green"))

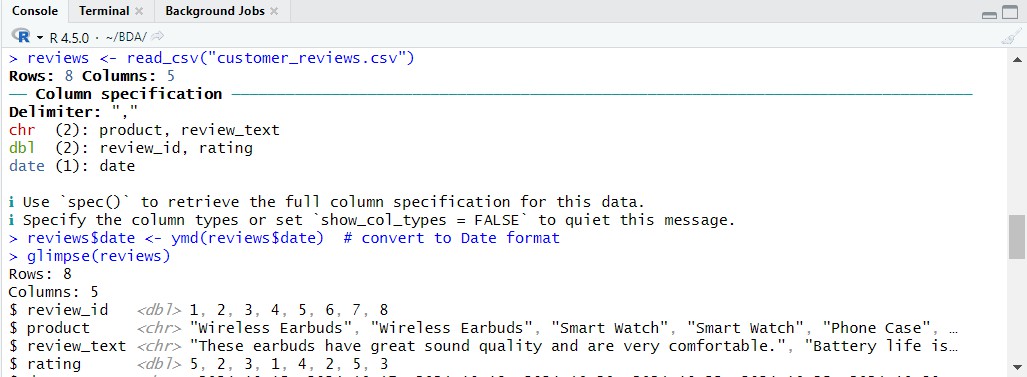
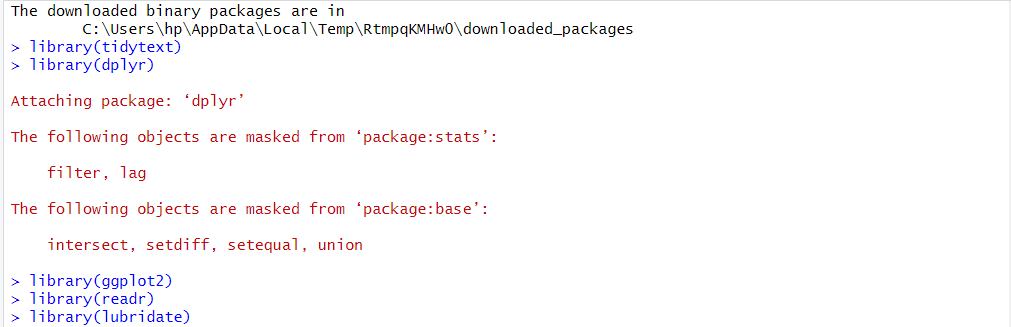
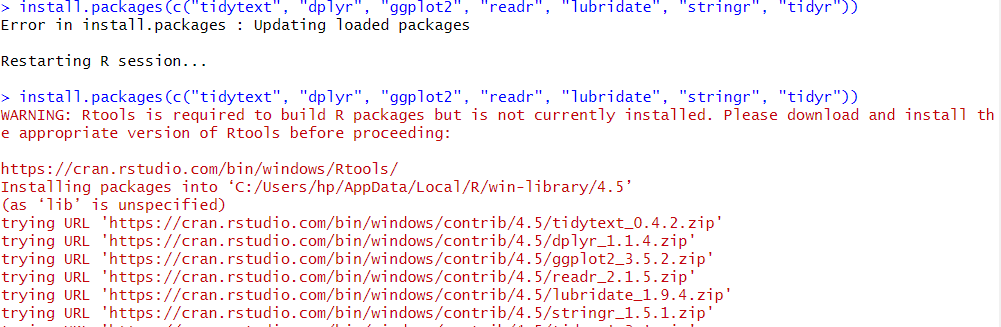
daily\_sentiment <- sentiment\_words %>% count(date, sentiment) %>%

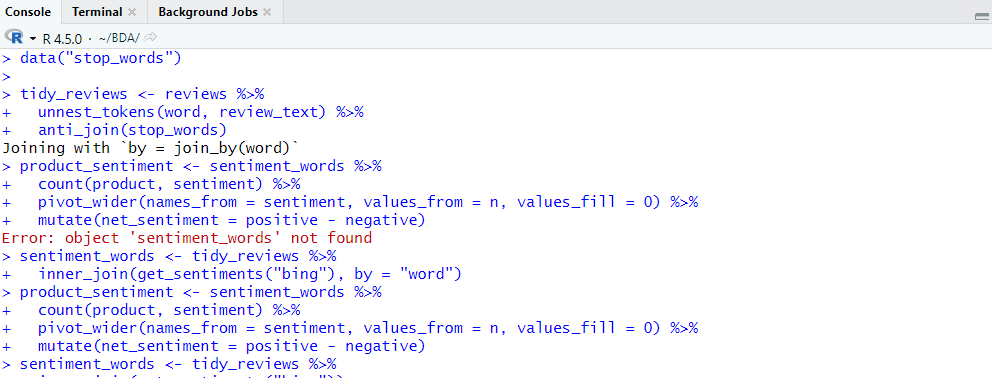
pivot\_wider(names\_from = sentiment, values\_from = n, values\_fill = 0) %>% mutate(net\_sentiment = positive - negative)

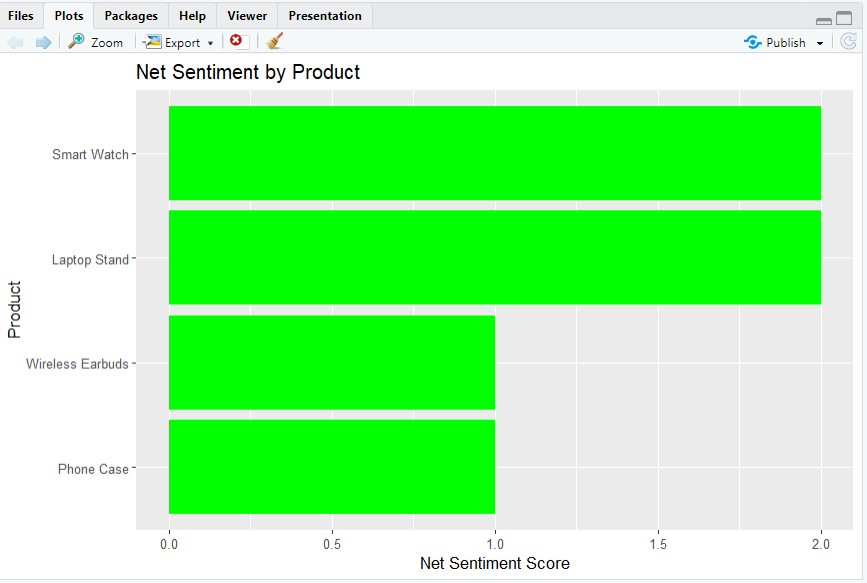
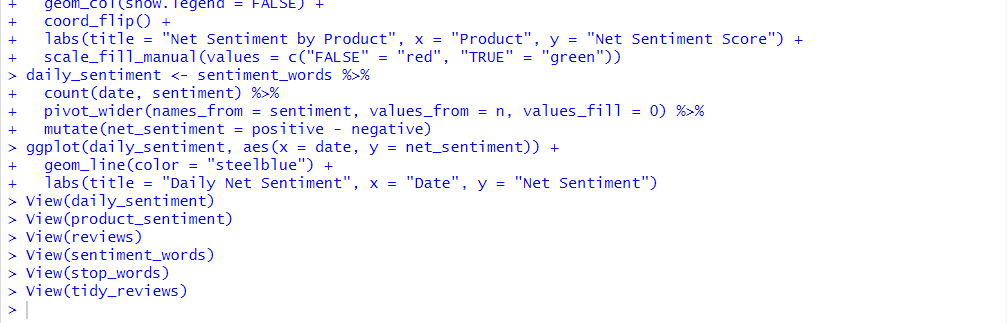
ggplot(daily\_sentiment, aes(x = date, y = net\_sentiment)) + geom\_line(color = "steelblue") +

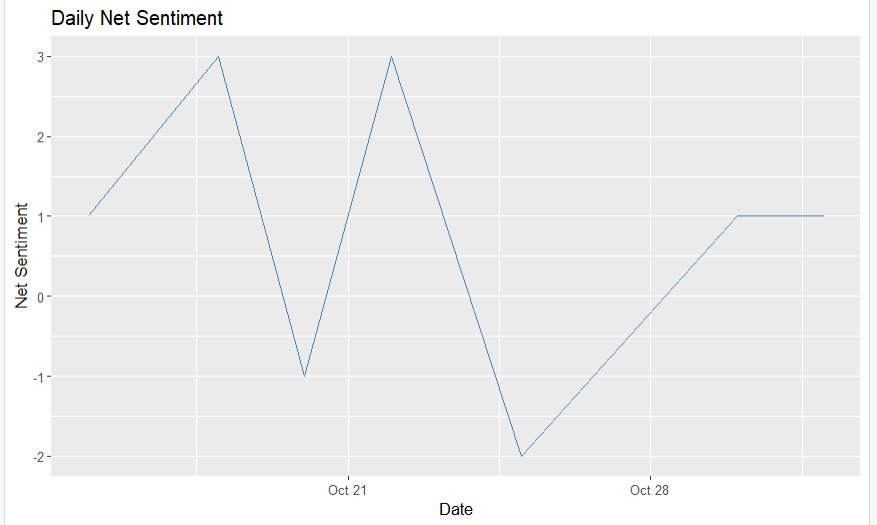
labs(title = "Daily Net Sentiment", x = "Date", y = "Net Sentiment")

**Output:**

****



****



1. **Challenges:**

**Data Quality Issues:**

Data quality issues pose significant challenges in sentiment analysis. Incomplete, biased, or noisy data can lead to erroneous interpretations and unreliable outcomes. Ensuring the integrity of datasets is critical for accurate sentiment assessment and deriving actionable insights.

### Sentiment Ambiguity:

Sentiment ambiguity arises when words or phrases have multiple meanings or emotional connotations, complicating analysis. For instance, sarcasm and irony can distort genuine sentiment. Developing models capable of understanding context is essential to minimizing these ambiguities.

### Scalability Concerns:

Scalability concerns are crucial for organizations analyzing vast amounts of data. As data grows, maintaining processing speed and accuracy becomes challenging. Implementing scalable solutions such as cloud computing and distributing workloads can help organizations manage and analyze big data effectively.

### CONCLUSION:

In conclusion, sentiment analysis plays a pivotal role in understanding consumer opinions and improving business strategies. While it presents challenges, the insights gained from effective methodologies empower organizations to make informed decisions that enhance customer engagement and satisfaction.

Youtube Link:- <https://youtu.be/AgxcIf0fRm8>

LinkedIn :- <https://www.linkedin.com/posts/abhishek-nimbalkar-063997361_bda-project-activity-7320369736486137856-GiIy?utm_source=share&utm_medium=member_desktop&rcm=ACoAAFoGRIsBKwMIBVpy5Gs4d3ju6neN36G-cCY>