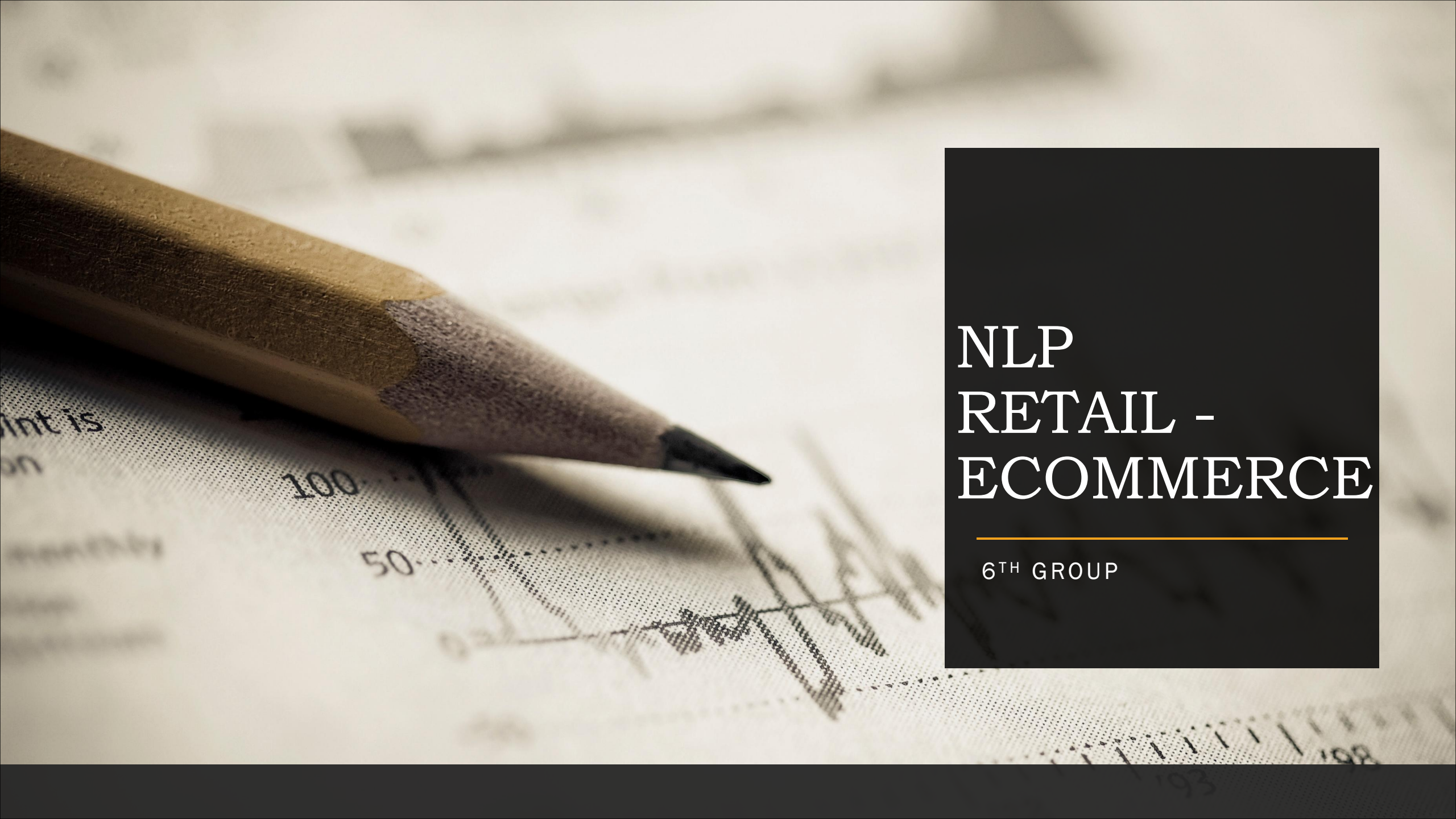


NLP RETAIL - ECOMMERCE

6TH GROUP

Problem Statement:

Extracting customer reviews on a product, Doing Sentiment Analysis and then Model Building a Classification model to Predict the Sentiment.

NLP:

*The main motive of Natural Language Processing (NLP) is to enable computers to understand, interpret, and generate human language in a way that is both useful and meaningful.

*Ecommerce businesses use sentiment analysis powered by NLP to analyze customer feedback, reviews, and social media posts. This helps in understanding customer sentiments towards products and services, identifying trends, and taking proactive measures to address concerns or capitalize on positive feedback.

*Overall, the main motive of NLP is to bridge the gap between human language and computer understanding, making it possible for machines to interact with and process human-generated content effectively across a wide range of applications and domains.

Product we choose: Amazon Echo Dot 4th Gen with clock

---

url=https://www.amazon.in/dp/B085M5R82K?ref=cm_sw_r_apan_dp_8T0C5C34D20K4DQEFE5S&ref_=cm_sw_r_apan_dp_8T0C5C34D20K4DQEFE5S&social_share=cm_sw_r_apan_dp_8T0C5C34D20K4DQEF

*. It can serve as a hub for controlling compatible smart home devices, such as lights, thermostats, and locks, adding convenience and automation to your home.

* The integrated LED display shows the time, making it convenient to glance at the time from across the room without needing to ask Alexa.

* The Echo Dot with clock is designed to fit into different environments, whether it's in the bedroom, kitchen, office, or living room, providing utility and connectivity wherever it's placed.

* It is one of the advanced Technology, Which can make work fast and save time for Humans.


echo

First week: EDA

Extracting reviews:

Scratching or scraping reviews from Amazon (or any website) using WebDriver (such as Selenium WebDriver in Python) involves automating the process of visiting the webpage, extracting relevant HTML elements, and then parsing the data you need. Here's a step-by-step guide on how you can do this:

Step 1: Setup
1.Install Python and Selenium:

- Make sure Python is installed on your system.
- Install Selenium WebDriver using pip:

pip install selenium

Step 2: Download web Driver

Downloading required libraries

```
import numpy as np

import pandas as pd

import time

import warnings

from selenium import webdriver

from selenium.webdriver.common.by import By

from selenium.webdriver.support.ui import WebDriverWait

from selenium.webdriver.support import expected_conditions as EC

from bs4 import BeautifulSoup

import nltk

from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

EDA:

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often using visual methods. It helps understand the data and its underlying structure before applying more advanced statistical techniques.

Steps in EDA:

1. **Data Collection:** Obtain the dataset from appropriate sources, ensuring data integrity and completeness.
2. **Data Cleaning:** Address missing values, handle outliers, and standardize formats for consistency.
3. **Initial Exploration:** Calculate summary statistics, generate simple visualizations to understand the overall structure and distribution of the data.
4. **Detailed Exploration:** Dive deeper into relationships between variables, use more advanced visualization techniques, and consider transformations or data normalization.
5. **Identify Patterns and Insights:** Draw conclusions based on patterns observed, form hypotheses for further analysis, and prepare data for modeling.

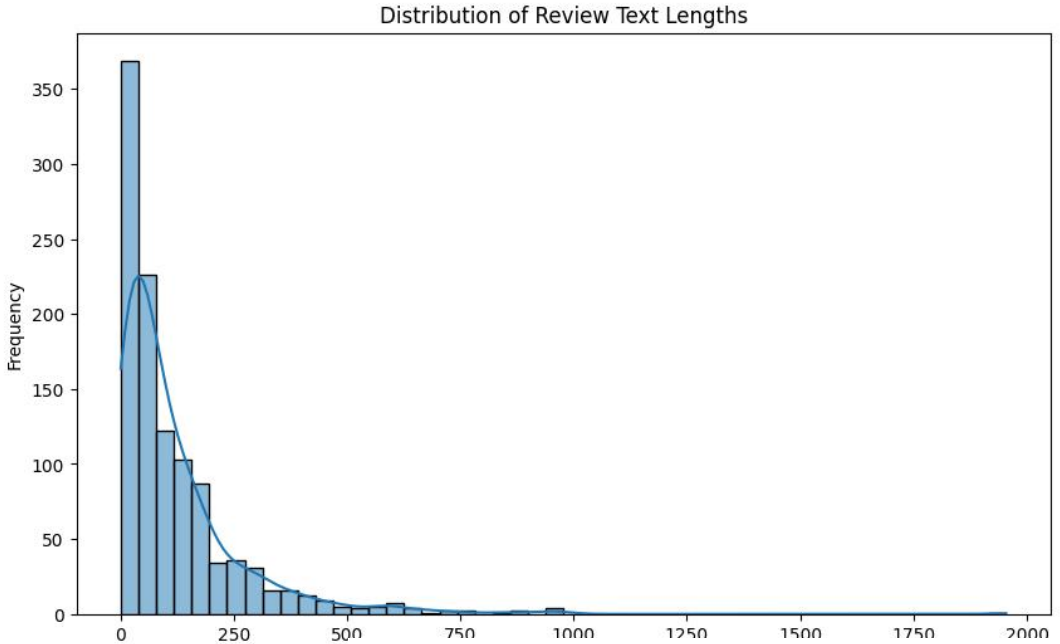
*Removed Non-English Reviews and Removed Emojis and special Symbols.

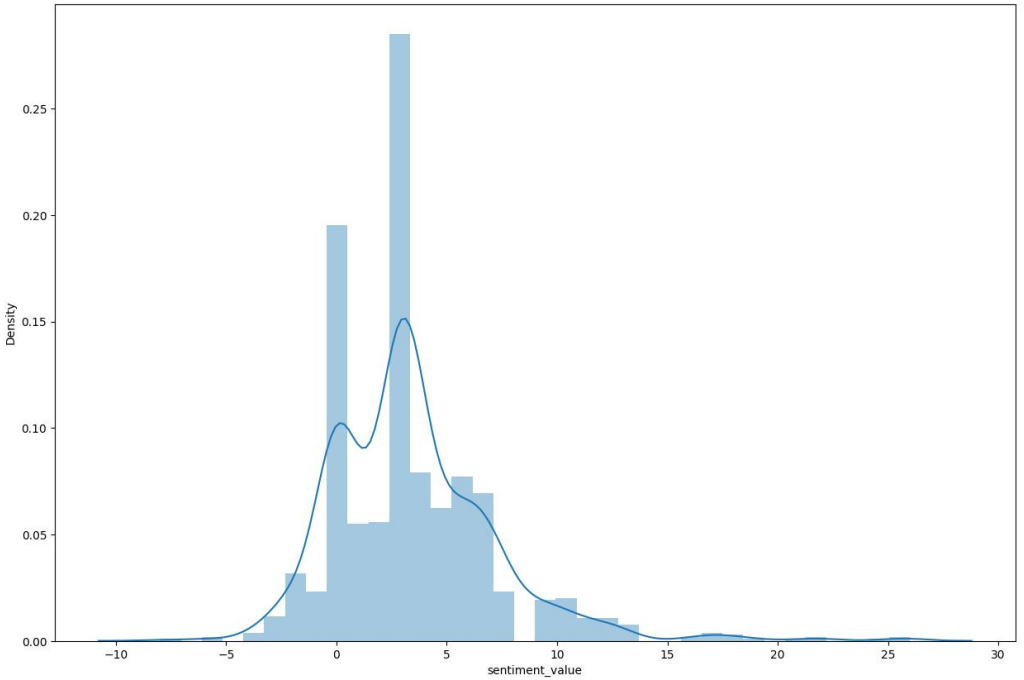
*Removed Stopwords , Numerics, punctuations. Performed Tokenization,Lemmatization and visualized the reviews using wordcloud.

* **Histograms:** Show the distribution of numerical data.
•**Box plots:** Display summary statistics such as median, quartiles, and outliers.
•**Scatter plots:** Explore relationships between two numerical variables.
•**Heatmaps:** Visualize correlations between variables.
•**Bar plots and pie charts:** Display categorical data distributions

* Exploratory Data Analysis is a critical step in any data science project, providing a comprehensive understanding of the dataset and guiding subsequent analysis and modeling efforts. It combines statistical techniques with visualizations to uncover patterns, relationships, and anomalies that drive insights and decision-making.

Visualization:





Week-2:Sentiment Analysis

Sentiment analysis is a natural language processing (NLP) technique used to determine the sentiment (positive, negative, or neutral) expressed in a piece of text. It aims to automatically extract subjective information from text, helping to understand the attitudes, opinions, and emotions of individuals or groups towards certain topics or entities.

Step 1: Load Affin Dataset which contains sentiment values for the word.

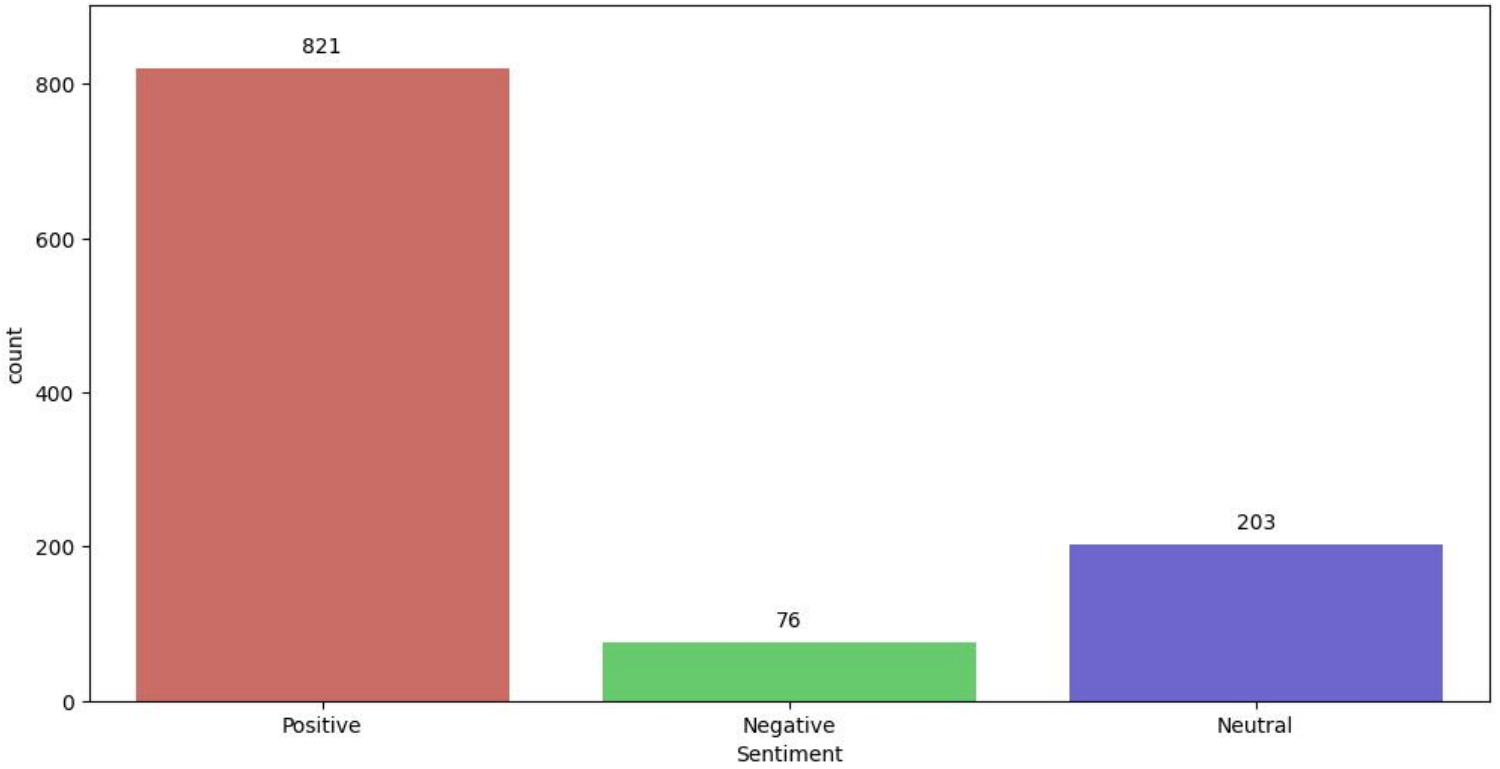
Step 2: Apply sentiment value to the Review.

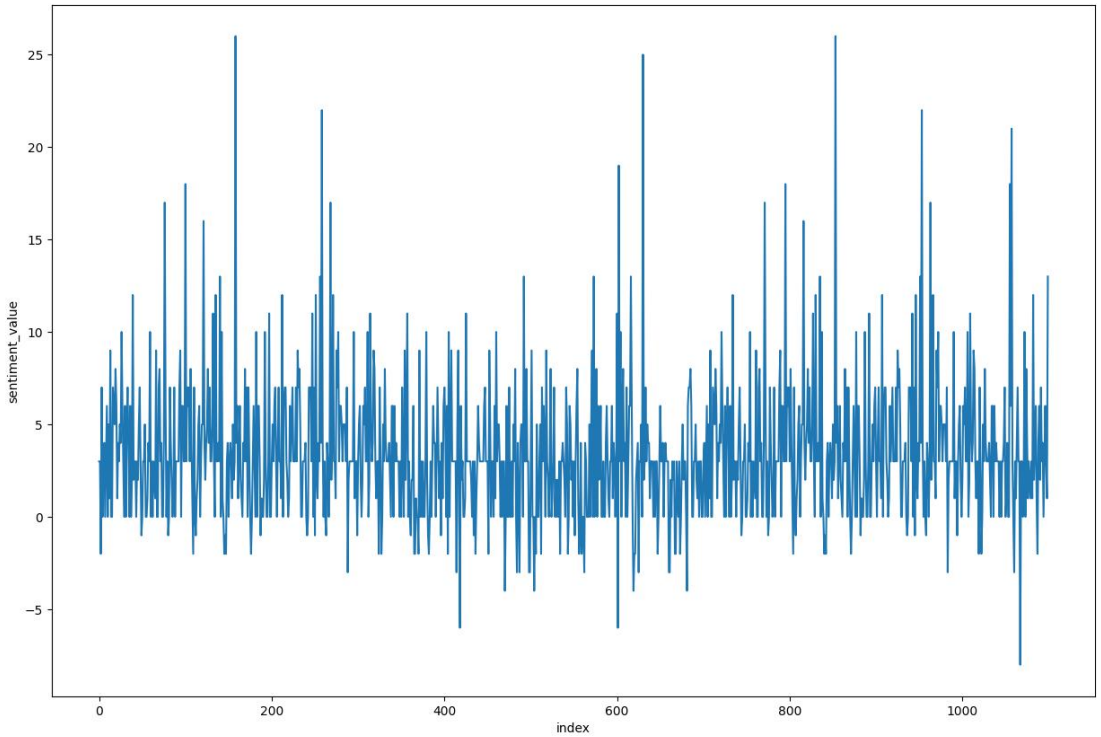
Step 3:Visualize sentiment value with Review using seaborn like dist plot, scatter plot and Trend.

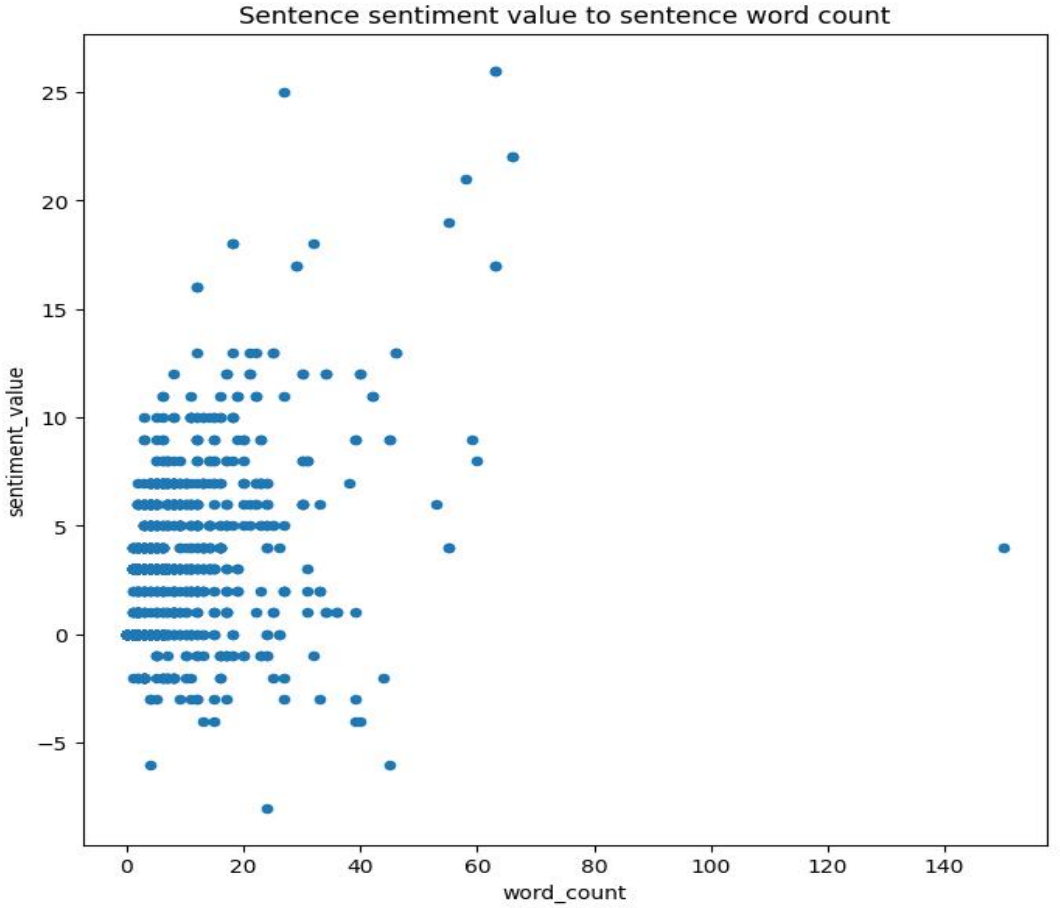
Step 4:Then perform Sentiment Analysis if Sentiment value is >0 it is a Positive Review, <0 it is a Negative Review

And if it is 0 it is Neutral Review.

Sentiment Analysis in Visualization:




sentiment_value
index
25
20
15
10
5
0
−5
0
200
400
600
800
1000


Sentence sentiment value to sentence word count
sentiment_value
word_count
25
20
15
10
5
0
−5
0
20
40
60
80
100
120
140

Challenges in Sentiment Analysis:

- **Contextual Understanding:** Text can contain sarcasm, irony, or ambiguity that may affect sentiment classification accuracy.
- **Language Ambiguity:** Different languages or dialects may express sentiment differently, requiring language-specific models or lexicons.
- **Domain-Specific Language:** Sentiment analysis models trained on general text may not perform well on domain-specific texts (e.g., medical or legal documents).

Sentiment analysis plays a crucial role in understanding and summarizing large amounts of textual data, enabling businesses, researchers, and organizations to make data-driven decisions based on public opinion and sentiment trends. Advances in NLP techniques continue to improve the accuracy and applicability of sentiment analysis across various domains and languages.

Week 3:Model Building

*Download and import required libraries.

* Split the data into training and testing .

*Using Different Machine Learning algorithms like Naïve Bayes, Random Forest, SVM, Logistic Regression, Gradient Boosting and Neural Networks.

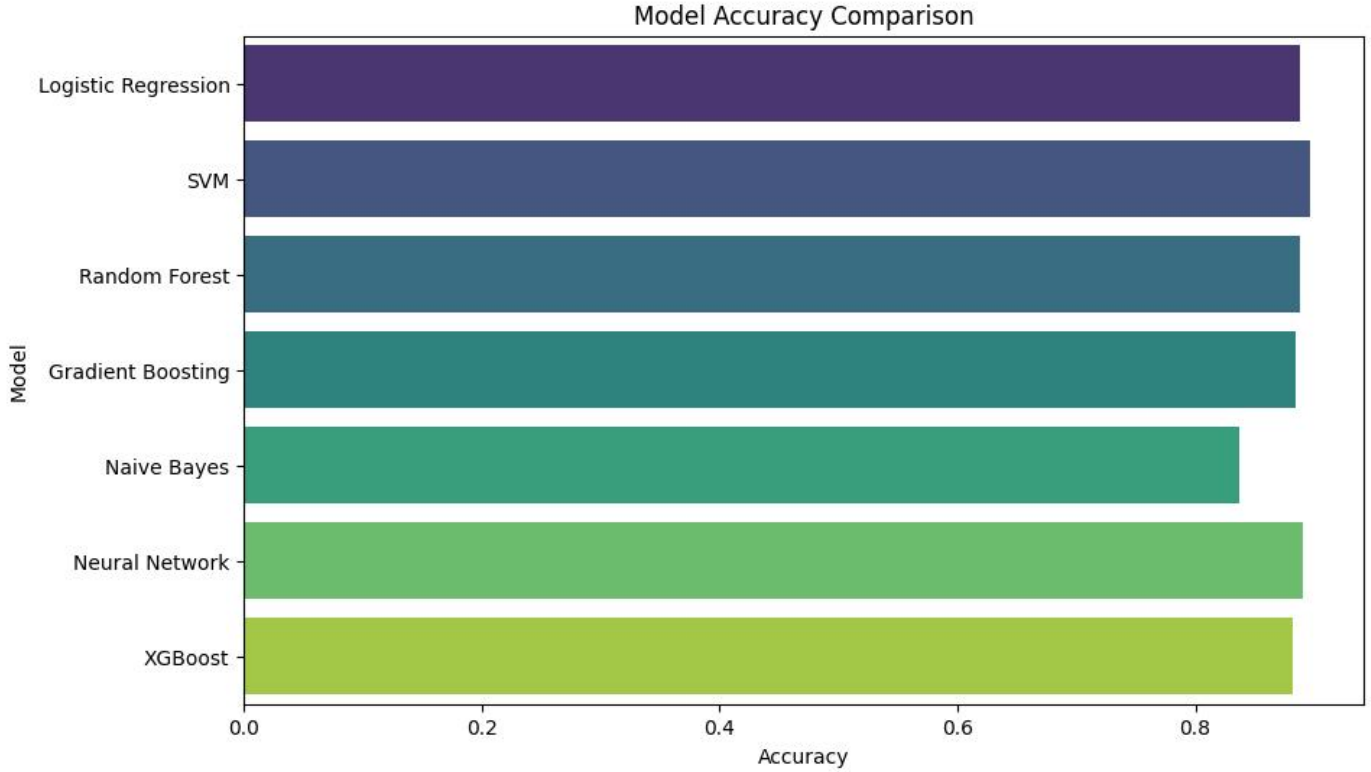
* Performing Parameter Tuning with various parameters and using GridSearchCV to find the best parameters.

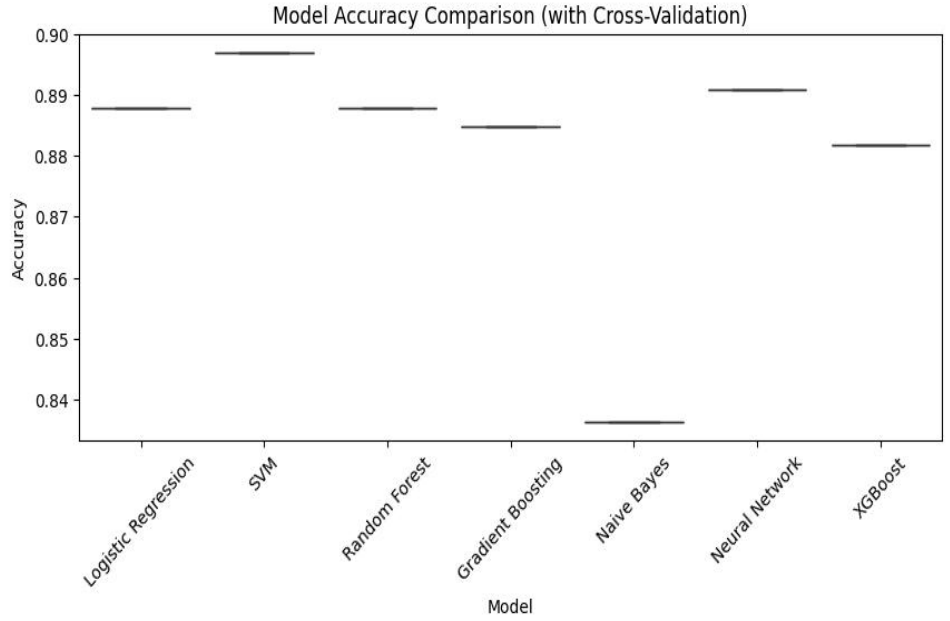
*And then find it’s Performance metrics such as F1Score, recall, Precision and Accuracy.

* Visualize performance using Bar plot , Box plot etc.

* Best model: SVM with F1 score 0.8910 Best Params: {'clf__C': 10, 'clf__gamma': 'scale', 'clf__kernel': 'rbf'}

Visualizing Accuracy Of a Models:





Deployment:

*create the data into a pickle file and store it into a new variable.

*Import streamlit to create the webpage and deploy the model.

*Make the requirements according the need and then download the pickle file and the code file with(.py) extension.

*copy both the files into the Anaconda Prompt location.

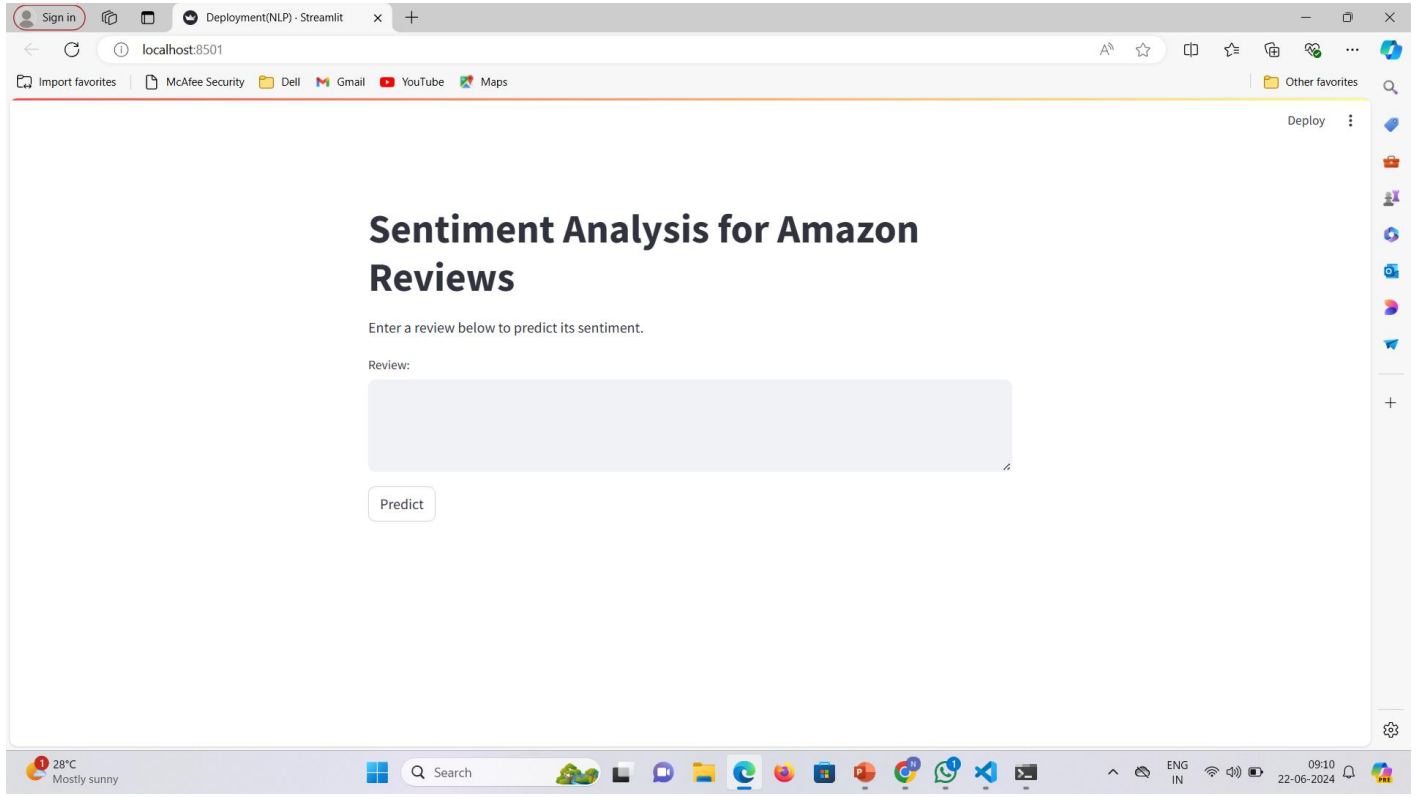
*Then go to Anaconda Prompt Type as “streamlit run (file name).py” and then click enter button.

*It automatically takes you to the webpage and then you can enter the review in the review box and then click predict button.

*Then it will predict whether the review is Positive or Negative or Neutral.

*This is the deployment of the model. Successfully Project completed.

Web Page:


Sign in
Deployment(NLP) - Streamlit
localhost:8501
Import favorites
McAfee Security
Dell
Gmail
YouTube
Maps
Other favorites
Deploy
Sentiment Analysis for Amazon Reviews
Enter a review below to predict its sentiment.
Review:
Predict
28°C
Mostly sunny
Search
ENG
IN
09:10
22-06-2024

Sentiment Analysis App.

Please Enter The Review

Horrible product, broke after one use

Analyze the Sentiment

The entered text has negative sentiments associated with it: -1.0

Sentiment Analysis Result: Sentiment(polarity=-1.0, subjectivity=1.0)

Sentiment Analysis App.

Please Enter The Review

I am extremely satisfied with this product.

Analyze the Sentiment

The entered text has positive sentiments associated with it: 0.5

Sentiment Analysis Result: Sentiment(polarity=0.5, subjectivity=1.0)

Sentiment Analysis App.

Please Enter The Review

okk

Analyze the Sentiment

The entered text seems to be neutral.

Sentiment Analysis Result: Sentiment(polarity=0.0, subjectivity=0.0)

Thank You