Sentiment Analysis on Product Reviews

1. Introduction

Sentiment analysis is a Natural Language Processing (NLP) technique used to determine whether textual data
is positive, negative, or neutral. In this project, we implement a simple sentiment analysis model on
product reviews using Python. The project uses text preprocessing, TF-IDF feature extraction, and a
Naïve Bayes classifier to classify reviews based on their sentiment.

2. Objectives

1. To preprocess text data for effective sentiment classification.
2. To apply TF-IDF vectorization to convert text into numerical features.
3. To train a Naïve Bayes model on labeled product reviews.
4. To evaluate model performance using accuracy, confusion matrix, and classification report.
5. To predict sentiment for new unseen product reviews.

3. Installation and Library Import

The following libraries are required for the project: nltk, scikit-learn, pandas, and matplotlib.
They can be installed using pip and imported as follows:

```
!pip install nltk scikit-learn pandas matplotlib

import pandas as pd
import numpy as np
import nltk
import re
import string
import matplotlib.pyplot as plt
```

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

4. Dataset

A small dataset of 10 product reviews is used, labeled as 'positive' or 'negative'. The data is stored in
a pandas DataFrame for easy processing.

Example dataset:
- "I love this phone, the battery life is amazing!" → positive
- "Worst purchase ever. It stopped working in 2 days." → negative
- "Excellent quality and fast delivery. Worth every penny!" → positive

5. Text Preprocessing

Text preprocessing is an essential step in sentiment analysis. It involves cleaning and preparing the data
for feature extraction. The following steps are performed:
1. Convert text to lowercase.
2. Remove URLs, punctuation, and numbers.
3. Tokenize text using NLTK.
4. Remove stopwords.

6. Feature Extraction using TF-IDF

TF-IDF (Term Frequency–Inverse Document Frequency) converts textual data into numerical feature vectors
based on the importance of words. In this project, TfidfVectorizer from scikit-learn is used with a
maximum of 1000 features.

7. Model Training

The dataset is split into training and testing sets (70% train, 30% test). A Multinomial Naïve Bayes model
is trained on the TF-IDF vectors to learn sentiment patterns.

8. Model Evaluation

After training, the model is evaluated on the test set using accuracy score, classification report, and
confusion matrix. The results for the small dataset are as follows:

Accuracy: 33%
Confusion Matrix:
[[0 2]
 [0 1]]

9. Discussion of Results

The model achieves an accuracy of 33%, which is relatively low due to the limited size of the dataset.
However, it demonstrates the working of a basic sentiment analysis pipeline. With a larger and more balanced
dataset, the accuracy would improve significantly.

10. Predicting Sentiment for New Reviews

The trained model is used to predict sentiment for new product reviews. Examples include:
- "This product is fantastic and works perfectly!" → Predicted: Positive
- "Completely useless, I want a refund." → Predicted: Positive (misclassified due to small dataset)
- "Average quality but acceptable for the price." → Predicted: Positive

11. Conclusion

This project successfully implements a basic sentiment analysis model using machine learning. It showcases
the importance of preprocessing, feature extraction, and model evaluation in NLP applications. Although the
dataset is small, the approach can be easily scaled for larger datasets and improved using advanced models
like Logistic Regression, SVM, or Deep Learning.

12. Future Scope

1. Increase dataset size for better accuracy.
2. Implement deep learning models such as LSTM or BERT for improved performance.
3. Use data augmentation techniques for class balancing.
4. Deploy the model as a web application using Flask or Streamlit.

13. References

1. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
2. scikit-learn documentation: https://scikit-learn.org/
3. NLTK documentation: https://www.nltk.org/
4. TF-IDF Vectorizer Guide: https://scikit-learn.org/stable/modules/feature_extraction.html#tfidf

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,

market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems

to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.

In natural language processing, sentiment analysis plays a key role in understanding customer opinions,
market trends, and social media behavior. It is used extensively in business intelligence to assess
public response to products, services, and brands. Companies like Amazon and Flipkart use similar systems
to filter and recommend products based on customer reviews.