Project Brief: Sentiment Analysis on E-commerce Dataset

Project Overview:

To determine the sentiments conveyed in customer reviews, we are performing a sentiment analysis project utilizing an e-commerce dataset. The goal is to create a model that can correctly categorize these evaluations' sentiments into positive, negative, or neutral. Data preprocessing, exploratory data analysis (EDA), model training, and the development of a web application for real-time sentiment analysis are all aspects of the project.

Dataset Details:

- **Source**: The dataset is sourced from Kaggle and is available at the following link: [E-commerce Dataset](#).
- **Content**: The dataset comprises various attributes such as name, brand, categories, primaryCategories, reviews.date, reviews.text, and reviews.title
- **Objective**: Our primary goal is to analyze customer sentiments expressed in the **reviews.text** field of the dataset. This will help us understand the overall sentiment associated with different products and brands in the e-commerce domain

Proposed Actions:

- **Data Preprocessing:** Handled missing values and duplicated data points to ensure data quality.

- **Exploratory Data Analysis (EDA):** Explored the dataset to gain insights into the distribution of data, identify trends, and understand the composition of reviews across different attributes.

- **Class Imbalance Handling:** Addressed class imbalance in sentiment labels by applying appropriate techniques to balance the dataset, ensuring that the model is not biased towards the majority class.

- **Model Training:**
    - Trained three machine learning models: Support Vector Classifier (SVC), Decision Tree, and Naive Bayes.
    - Evaluated model performances using accuracy metrics.
    - Chose the Support Vector Classifier (SVC) algorithm based on its higher accuracy compared to other models.

- **Text Vectorization:** Employed text vectorization techniques (e.g., TF-IDF) to convert the text-based reviews into numerical features that the machine learning models can understand.

- **Web Application Development:**
    - Utilized Flask, a web framework, to create a user-friendly web application.
    - Integrated the trained SVC model into the web application to enable real-time sentiment analysis of customer reviews.

Project Outcomes:

- **Model Performance:** The Support Vector Classifier (SVC) demonstrated the highest accuracy of 64%, indicating its effectiveness in sentiment classification.

Decision Tree Classifier

Accuracy: 0.5056179775280899

Confusion matrix :
```
[[10  7  3]
 [ 5 16 11]
 [ 4 14 19]]
```

Classification Report:

	precision	recall	f1-score	support
Negative	0.53	0.50	0.51	20
Neutral	0.43	0.50	0.46	32
Positive	0.58	0.51	0.54	37
accuracy			0.51	89
macro avg	0.51	0.50	0.51	89
weighted avg	0.51	0.51	0.51	89

Naive Bayes

Accuracy: 0.6179775280898876

Confusion matrix :
```
[[ 7  8  5]
 [ 1 21 10]
 [ 0 10 27]]
```

Classification Report:

	precision	recall	f1-score	support
Negative	0.88	0.35	0.50	20
Neutral	0.54	0.66	0.59	32
Positive	0.64	0.73	0.68	37
accuracy			0.62	89
macro avg	0.69	0.58	0.59	89
weighted avg	0.66	0.62	0.61	89

SVC

Accuracy: 0.6404494382022472

Confusion matrix :
```
[[10  6  4]
 [ 5 18  9]
 [ 2  6 29]]
```

Classification Report:

	precision	recall	f1-score	support
Negative	0.59	0.50	0.54	20
Neutral	0.60	0.56	0.58	32
Positive	0.69	0.78	0.73	37
accuracy			0.64	89
macro avg	0.63	0.62	0.62	89
weighted avg	0.63	0.64	0.64	89

- **Web Application:** The developed Flask-based web application allows users to input customer reviews and receive instant sentiment analysis results.

Test with your review:

Enter review

Check Sentiment

Test with your review:

Enter review

Check Sentiment

Results

Sentiment: Positive

Review Text : I've been using the SuperGlide Electric Toothbrush for a few weeks now, and I'm absolutely thrilled with the results.

Supervisor Communication:

You can inform your supervisor that the project involves sentiment analysis on an e-commerce dataset sourced from Kaggle. The dataset contains attributes related to product information and customer reviews. The primary goal is to build a sentiment classification model that accurately categorizes customer sentiments as positive, negative, or neutral. The project includes data preprocessing, exploratory data analysis, class imbalance handling, model training, text vectorization, and the creation of a user-friendly web application for real-time sentiment analysis. The chosen model, SVC, achieved an accuracy of 64%, and the web application offers a practical interface for analyzing customer sentiments.