Project - 2

Data Analyst Project: Twitter Sentiment Analysis

Project Title: Twitter Sentiment Analysis

Dataset: (https://www.kaggle.com/datasets/kazanova/sentiment140)

Project Overview:

Twitter Sentiment Analysis is a data analytics project that involves analyzing a dataset of tweets to determine the sentiment expressed in each tweet—whether it is positive, negative, or neutral. The project aims to gain insights into public opinions, trends, and sentiments shared on Twitter, utilizing data analytics techniques.

Project Objectives:

1. Data Exploration:
   - Explore the Sentiment dataset to understand its structure, features, and size.
   - Identify key variables such as tweet content, timestamp, and sentiment labels.

2. Data Cleaning:
   - Perform data cleaning tasks to handle missing values, duplicate entries, and irrelevant information.
   - Ensure data quality by addressing any anomalies or inconsistencies in the dataset.

3. Exploratory Data Analysis (EDA):
   - Conduct exploratory data analysis to gain initial insights into tweet patterns, sentiment distributions, and temporal trends.
   - Utilize visualizations (e.g., histograms, word clouds) to represent key aspects of the dataset.

4. Sentiment Distribution:
   - Visualize the distribution of sentiment labels (positive, negative, neutral) in the dataset.
   - Analyze the balance of sentiment classes to understand potential biases.

5. Word Frequency Analysis:
   - Analyze the frequency of words in tweets to identify common terms and themes.
   - Create word clouds or bar charts to visualize the most frequent words in positive and negative sentiments.

6. Temporal Analysis:
   - Explore how sentiment varies over time by analyzing tweet timestamps.
   - Identify patterns, peaks, or trends in sentiment within specific time frames.

7. Text Preprocessing:
   - Preprocess tweet text by removing stop words, special characters, and URLs.
   - Tokenize and lemmatize words to prepare the text for sentiment analysis.

8. Sentiment Prediction Model:
   - Implement a sentiment prediction model using machine learning or natural language processing techniques.
   - Train the model on a subset of the dataset and evaluate its performance using metrics like accuracy and F1 score.

9. Feature Importance:
   - Identify the most important features (words or phrases) contributing to sentiment predictions.
   - Visualize feature importance using techniques such as bar charts or word clouds.

10. User Interface (Optional):
    - Develop a simple user interface allowing users to input custom text for sentiment analysis.
    - Showcase the sentiment prediction results in a user-friendly manner.

11. Documentation:
    - Create comprehensive documentation covering data preprocessing steps, model implementation, and analysis findings.
    - Include code snippets, visualizations, and explanations to aid understanding.

12. Insights and Recommendations:
    - Summarize key insights gained from the analysis.
    - Provide recommendations or suggestions based on the sentiment trends observed.

Submission Instructions:

1. Follow all the instructions mentioned in the attached Instruction PDF.
2. Zip the entire project file, including all necessary documents, assets, and source code.
3. Submit the zipped file through Google Classroom.
4. Include the GitHub repository link (if applicable) and the live website link in the submission form.

Important Dates:

- Last Date for Submission: February 05, 2024

We encourage you to approach this project with creativity and professionalism. Success in this project will further enhance your internship profile. Ensure timely project submission.