

Project Title:

Identifying the Best Intern Using Naive Bayes on WhatsApp Chat Data

Project Description:

- **Scenario:** As the head of 20 interns, the objective is to identify the most active, vibrant, energetic, target-achieving, and responsive intern of the month using WhatsApp chat data.
- **Approach:** The project analyzes WhatsApp chat data (.txt file) to evaluate each intern's activity. Machine learning (Naive Bayes algorithm) is applied to classify interns based on their performance and behavior. Results are visualized using Tableau to provide actionable insights.

Skills Demonstrated:

1. **Data Collection and Preprocessing:**
 - Extracted and cleaned WhatsApp chat data from a .txt file.
 - Preprocessed the text data by removing noise (e.g., timestamps, metadata, and irrelevant chats) to focus on meaningful interactions.
 - Tokenized the chat data for sentiment and frequency analysis.
2. **Machine Learning (Naive Bayes Algorithm):**
 - Built a classification model using Naive Bayes to identify the most active and responsive intern based on chat engagement metrics.
 - Applied text-based features like message count, sentiment analysis, and activity frequency as model inputs.
3. **Data Visualization:**
 - Visualized intern performance metrics using Tableau to create interactive dashboards that compare engagement levels, responsiveness, and overall contributions.
 - Designed clear visual representations (bar charts, heatmaps, etc.) to highlight the most responsive and energetic interns.
4. **Problem Solving:**
 - Developed a custom evaluation system to quantify intern performance based on multiple factors such as message frequency, content positivity, and responsiveness.
 - Created a scalable solution that allows for monthly evaluations using data-driven techniques.
5. **Predictive Analytics:**
 - Used the Naive Bayes model to predict future performance trends based on current and historical chat activity.
 - Offered insights for improving intern engagement and productivity through data analysis.

Tools and Technologies Used:

- **Programming:** Python (pandas, scikit-learn, NLTK for NLP)
- **Machine Learning:** Naive Bayes classification
- **Data Visualization:** Tableau

- **Text Preprocessing:** Regular expressions, tokenization, and sentiment analysis

This project showcases your **data analysis, machine learning, problem-solving,** and **data visualization** skills, which are highly relevant for a data science internship.