Final Project Proposal

Full Name: Abu Bakar

Email Address: abubakarp789@gmail.com

1. Project Title

Viral or Not? – Predicting Virality of Tweets

2. Overview

This project aims to analyze a collection of tweets and build a model that can predict whether a tweet is likely to go viral or not. The dataset will be cleaned and processed using Pandas and NumPy. Exploratory Data Analysis will be performed to uncover engagement patterns, and both basic and advanced feature engineering techniques will be applied to extract meaningful information from the tweet content and metadata.

3. Purpose

In the digital age, the ability to predict the virality of content holds immense value for marketers, influencers, and organizations. This project will help identify key attributes that contribute to a tweet's virality, such as timing, content, hashtags, sentiment, and engagement levels. The final model will classify tweets as "viral" or "not viral" based on these features.

4. Tools and Technologies

- **Languages:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, TextBlob, re, datetime
- **Dataset:** Twitter US Airline Sentiment Dataset (Kaggle)
- **Optional:** Streamlit (for UI demo)

5. Expected Output

- A cleaned and processed dataset with engineered features
- Visual analysis showing what influences virality
- A binary classification model to predict tweet virality
- A Jupyter Notebook with all code, explanations, and results
- A short summary report of insights and findings
- (Optional) A basic web app to test tweet virality using Streamlit