DS Capstone Three: Project Proposal

Analyzing Social Media Sentiments and Trends Using Natural Language Processing

Problem Identification: This project aims to leverage Natural Language Processing (NLP) techniques to analyze sentiments and trends in the Social Media Sentiments Analysis Dataset. By applying NLP methods, we seek to extract meaningful insights from user-generated content across various social media platforms.

Context: In the modern world, social media is an integral tool for anyone seeking to understand their potential customer base and market to them appropriately. Understanding the sentiments and trends prevalent in social media content can provide valuable insights for businesses seeking to capitalize on this easy access to consumers. By analyzing this dataset, we can gain deeper insights into user behavior, preferences, and societal trends.

Criteria for Success: The success of this project will be measured by the accuracy and depth of the sentiment analysis, the identification of significant trends over time and across different platforms, and the extraction of actionable insights for stakeholders.

Scope of Solution Space: The solution space encompasses applications of NLP techniques such as sentiment analysis, topic modeling, and trend analysis on the Social Media Sentiments Analysis Dataset. The project will focus on extracting insights related to emotions, trends, user engagement, and geographical variations present in the dataset.

Constraints: Constraints may include data privacy concerns, limitations in computational resources for processing large volumes of text data, and challenges in accurately categorizing nuanced sentiments expressed in user-generated content.

Stakeholders: Stakeholders for this project include marketers, social media analysts, businesses, researchers, and decision-makers who seek to understand and leverage the sentiments and trends prevalent in social media content.

Data Sources: The primary data source for this project is the Social Media Sentiments Analysis Dataset from Kaggle, which contains user-generated content, sentiment labels, timestamps, user identifiers, platform information, hashtags, engagement metrics, and geographical details.

Deliverables: The deliverables will include a comprehensive Jupyter Notebook containing all code, from data processing to final analysis, with detailed comments explaining each step, a final report summarizing findings, methodologies, and implications, formatted as a slide deck for presentation to stakeholders, and supplemental materials including data visualizations and a summary of actionable insights for enhancing social media engagement based on sentiment analysis.