Decoding Data Science(DDS) Academy

AI Application Building Challenge

For students (instructor review required)

Shivraj Karewar

4th Jan 2025

Day 1 Idea Submission Template

Project Title:

NewsBot: LLM API for News Summary (Sentiment Analysis and Market Directions)

Concept Summary:

Creating an LLM-based API for news summarization and sentiment analysis involves leveraging NLP capabilities to process and analyze content from various platforms, providing concise summaries and sentiment insights for users.

The proposed project involves building an LLM-powered API for analyzing news sentiment to aid stock market prediction. The purpose of this project is to aggregate real-time financial news data, extract sentiment insights, and align them with stock market trends to provide actionable predictions for traders and investors. This solution addresses the challenge of processing vast amounts of unstructured financial news, helping users make informed investment decisions based on sentiment-driven market analysis.

Target Audience:

This project will primarily benefit retail investors, financial analysts, and stock market traders. These stakeholders need timely and accurate insights from financial news to make data-driven investment decisions. By providing sentiment analysis aligned with stock trends, the project addresses their need for quick, reliable, and actionable intelligence to navigate volatile markets and optimize investment strategies.

Key Features:

The key feature of the project is news summarization, which involves aggregating and processing content from various web sources (e.g., X platform, financial news websites) and presenting it in a concise, easy-to-read format for users. This ensures that users can quickly

grasp essential insights without sifting through extensive articles or posts, saving time and enhancing decision-making efficiency.

1. **Real-Time News Aggregation:** Collect and process financial news articles and social media data using APIs or web scraping, ensuring up-to-date information.
2. **Sentiment Analysis:** Utilize Natural Language Processing (NLP) models to analyze the sentiment (positive, negative, or neutral) of news articles and tweets related to specific stocks or industries.
3. **Customizable Insights:** Provide users with customizable dashboards to filter news sentiment and stock data based on industries, keywords, or time frames.
4. **API for Integration:** Offer an API for developers and organizations to integrate sentiment-driven stock predictions into their own systems.

Future directions:

Depending on the available time for the project implementation, I will try to implement a couple of additional things mentioned below:

1. **Predictive Analytics:** Use machine learning models to forecast stock price movements based on sentiment patterns and historical trends.
2. **Stock Data Integration:** Align sentiment scores with real-time or historical stock market data to uncover correlations and trends.

Technical Approach:

The project will be implemented using NLP techniques. Key technologies include:

1. Large Language Models (LLMs): Models like OpenAI's GPT or similar for news summarization and sentiment analysis.
2. **Web Scraping and APIs:** Tools like BeautifulSoup, Scrapy, or financial news APIs (e.g., NewsAPI, Bing News) to collect data from platforms and news sources.
3. **Data Preprocessing:** Python libraries such as pandas and NLTK for cleaning and preparing text data for analysis.
4. **Deployment:** Flask or FastAPI to create a RESTful API for integrating the summarization and sentiment analysis service with user applications.

The workflow will involve data collection, NLP-driven text summarization, sentiment scoring, and presenting results in an easy-to-read format.

Expected Challenges and potential solutions:

1. **Data Accessibility and Quality:**

Obstacle: Limited access to premium financial news APIs or incomplete datasets from web scraping.

Solution: Consider using a mix of free and paid APIs, implement robust web scraping techniques with error handling, and ensure compliance with content licensing regulations.

2. **Language and Context Challenges:**

Obstacle: Difficulty in accurately analyzing nuanced language in financial news, such as sarcasm or complex jargon.

Solution: Fine-tune pre-trained LLMs on financial datasets to improve domain-specific accuracy and incorporate context-aware sentiment scoring models.

3. **User Trust and Interpretation:**

Obstacle: Users might misinterpret sentiment analysis results or rely too heavily on them for decision-making.

Solution: Provide clear explanations of the sentiment scoring process and include disclaimers emphasizing that the tool is for informational purposes only.

Submission Format:

The submission will include a video demonstration showcasing the functionality of the news sentiment analysis tool, along with a link for users to interact with and experience the system firsthand.

Expected Outcome:

The project will deliver an LLM-based API for summarizing financial news and analyzing sentiment, providing concise, sentiment-driven insights for users. This will help financial analysts, traders, and investors make more informed decisions by saving time on manual analysis. Additionally, the project will deepen my understanding of NLP, sentiment analysis, API development, and financial market prediction, enhancing my skills in integrating these technologies for real-world applications.