Project Summary: Stock Movement Analysis Using Reddit API

Objective:

The goal of this project is to analyze and predict stock price movements by leveraging discussions and sentiments expressed in various subreddits, particularly those related to stock trading (r/wallstreetbets). By analyzing the data from Reddit, this project aims to uncover correlations between social media sentiment and stock price fluctuations.

Project Components:

1. Reddit API Integration:

- Use Reddit's API (through `PRAW`) to gather historical posts, comments, and other interactions from specific subreddits.

- Collect data such as post titles, comments, upvotes, and timestamps from subreddits that are highly active in financial discussions.

- Apply filters to focus on stock-related keywords or specific stocks, such as those tagged with $ symbols or stock tickers.

2. Data Preprocessing:

- Clean and preprocess Reddit data by removing stopwords, special characters, and irrelevant posts.

- Extract sentiment from the text (positive, negative, neutral) using Natural Language Processing (NLP) techniques, such as VADER

- Aggregate posts to analyze sentiment over time, tracking how sentiment evolves as market events unfold.

3. Stock Price Data Acquisition:

- Use stock market APIs (such as Yahoo Finance) to collect historical price data of stocks mentioned in Reddit posts.

- Include data on stock open/close prices, volume, high/low prices, and daily price changes.

4. Sentiment Analysis:

- Perform sentiment analysis on the Reddit data, classifying posts and comments into positive, neutral, or negative sentiment categories.

- Use machine learning or rule-based approaches (like a Naive Bayes classifier) to enhance accuracy.

- Analyze how sentiment correlates with price movements over a given time period.

5. Correlation and Prediction Models:

- Investigate if there is a correlation between the volume of Reddit posts on a particular stock and its price movements.

- Use statistical methods (like Pearson correlation or regression analysis) to test the relationship between Reddit sentiment and stock prices.

6. Visualization:

- Visualize stock price movements and Reddit sentiment using libraries such as `matplotlib`, `seaborn`.

- Create graphs that show how Reddit activity (post volume, sentiment) impacts stock prices over time.

- Implement dashboards for real-time stock sentiment analysis and predictions.

7. Conclusion & Insights:

- Identify patterns, such as whether strong positive or negative sentiment in Reddit correlates with significant stock price movement.

- Determine if Reddit can be used as an early indicator of stock volatility.

- Evaluate the potential use of social media sentiment in trading strategies.

Potential Challenges:

- Data noise: Reddit discussions can be noisy and may include unrelated or spammy content that could skew sentiment analysis.

- Market volatility: External market factors not covered in Reddit discussions may influence stock price movements.

- Sentiment analysis accuracy: Accurately determining sentiment in informal and slang-heavy language may be challenging.

Technologies & Tools:

- APIs: Reddit API (PRAW), Stock market APIs (Yahoo Finance)

- Libraries: Pandas, NLTK, VADER, Matplotlib, Seaborn

- Visualization: Matplotlib, Seaborn