# Stock Movement Analysis Based on Social Media Sentiment

**Objective**: To develop a machine learning model that predicts stock movements by scraping data from social media platforms like Twitter, Reddit, or Telegram. The model should extract insights from user-generated content, such as stock discussions, predictions, or sentiment analysis, and accurately forecast stock price trends.

## Description of the Features Extracted and Their Relevance to Stock Movement Predictions

The following features were extracted from the scraped Reddit posts to aid in stock movement predictions:

### Text Data (Preprocessed):

The raw text from Reddit posts was cleaned and preprocessed by removing stopwords, special characters, and lowercasing all text. This preprocessed text was then used for sentiment analysis and feature extraction.

### Sentiment Polarity:

Using **VADER Sentiment Analysis** (Valence Aware Dictionary and sEntiment Reasoner), the sentiment of each post was determined. The polarity score ranged from -1 (negative) to +1 (positive). Sentiment polarity is critical in determining how people feel about a stock, which can directly influence its price movement.

### Frequency of Mentions (Word Frequency):

We tracked how often specific stock symbols or keywords appeared in the posts. A higher frequency of mentions could indicate increased interest, which is relevant for stock movement prediction. Example: The frequency of mentions of "Tesla" or "AAPL" could give an insight into public interest or sentiment about these stocks.

### User Reputation (Optional Feature):

A feature based on user reputation (e.g., the number of karma points on Reddit) was also considered. Posts from users with higher karma were given more weight, assuming they could have more influence in the market.

**Relevance to Stock Movement:**

- **Sentiment polarity** directly correlates with market sentiment. Positive sentiments about a stock may indicate a potential price increase, while negative sentiments suggest the opposite.

- **Frequency of mentions** helps gauge public interest or buzz around specific stocks, a factor known to influence stock price movement.

- **User reputation** and **post popularity** provide context on the credibility and reach of stock discussions, further strengthening the accuracy of the sentiment analysis.

**Model Evaluation Metrics, Performance Insights, and Potential Improvements**

**Model Used:**

We used a combination of **Machine Learning Models** including **Naive Bayes** and **Logistic Regression** to predict stock movements. Sentiment scores (positive, neutral, negative) derived from the text data served as the primary features.

**Evaluation Metrics:**

- **Accuracy:** Measures the overall correctness of the model.

- **Precision:** The proportion of true positives out of the total predicted positives.

- **Recall:** The proportion of true positives out of the actual positives.

- **F1-Score:** Harmonic mean of precision and recall, providing a balanced measure of performance.

- **Confusion Matrix:** To visualize the performance in terms of true positives, true negatives, false positives, and false negatives.

**Suggestions for Future Expansions**

To enhance the project's scope and accuracy, the following future expansions are suggested:

1. **Integrating Multiple Data Sources**: Include data from diverse platforms like social media, financial blogs, and stock market trends.

2. **Advanced Models**: Utilize deep learning models such as BERT (Bidirectional Encoder Representations from Transformers) for improved contextual understanding.

3. **Real-Time Analysis**: Develop pipelines for real-time sentiment tracking to support decision-making processes.

These enhancements can improve the robustness and accuracy of predictions, making the solution more versatile and applicable in real-world scenarios.

**Github link** : https://github.com/nikhilrathod46/Stock-Movement-Analysis-Based-on-Social-Media-Sentiment

**Video link :** https://youtu.be/1XRmjgihCjU?si=I2mbSL3kh2YsfDpW