Project Proposal

Dataset:

- https://drive.google.com/file/d/1BdGlZbBoZzTYewA3E2w1oTHj9IrJID5h/view?usp=drive_link

Description:

Variables

1. Date: The trading day.
2. close: The closing price of S&P 500 on the given date.
3. prev_close_1 to prev_close_20: Historical closing prices for the 20 days preceding the current date, with each variable representing the closing price 1 day ago, 2 days ago, and so on, up to 20 days ago.
4. Decayed_Compound: The sentiment scores of FOMC meeting statements score * (counts of inflation in the statements - counts of economic in the statement) with the decayed effects.
5. pct: The percentage change or return, representing the daily return based on the closing price.

Code:

- https://drive.google.com/file/d/1E2C2vWCz3kXhRYAmC5ItTcGqJDJPMg9L/view?usp=sharing

Background:

The decisions made by the Federal Reserve (Fed) are among the most influential factors affecting the stock market. As the central bank of the United States, the Fed's primary mandate includes managing inflation, supervising the nation's banking system, and maintaining financial stability. Its decisions, especially regarding interest rates and monetary policy, play a pivotal role in shaping investor sentiment and market dynamics.

Questions for the dataset?:

- Tentative question 1: What are the good models for estimating the stock price based on conference sentiment score?
- Tentative question 2: How effectively can Support Vector Machine (SVM) models, utilizing historical closing prices and sentiment scores derived from FOMC meeting statements, predict the daily return of the S&P 500?
- Tentative question 3: To what extent do historical closing prices (from 1 to 20 days prior) influence the predictive accuracy of the SVM model for the S&P 500 returns?
- Tentative question 4: How significantly does the inclusion of the 'Decayed_Compound' sentiment score, based on FOMC statements, enhance the model's forecasting accuracy?

Modeling Approach:

- **Data Preprocessing:**
    - Data Cleaning: Handle missing values, outliers, and data anomalies.
    - Feature Engineering: Standardize historical closing prices and sentiment scores to ensure uniformity in the model's input scale.

- **Model Implementation:**

    - SVM Kernel Selection: Choose the appropriate SVM kernel (linear, polynomial, RBF, etc.) based on preliminary analysis.
    - Parameter Tuning: Utilize potentially techniques like cross-validation and grid search to find the optimal hyperparameters for the SVM model.
    - Test the effectiveness of other machine learning models.

- **Model Evaluation:**
    - Use metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared to evaluate the model's performance.
    - Compare the models; performance with and without certain features to assess its impact.

- **Backtesting:**
    - Employ backtesting strategies to evaluate the model's effectiveness over different historical periods to ensure robustness and consistency in predictions.