Project Summary

Our project explores how Federal Reserve Chair Jerome Powell's language during press releases influences financial markets. Because hawkish and dovish monetary policy stances significantly affect the S&P 500, the tone of Federal Reserve communications can provide an early signal of the stance being adopted.

We will build a machine learning pipeline that:

1. **Collects Data** → Federal Reserve press conference transcripts (2020-2025) aligned with market data (S&P 500 returns)

2. **Analyzes Sentiment** → Uses NLP models to assign a score to Powell's tone (0 = dovish and 100 = hawkish)

3. **Finds Patterns** → Links sentiment scores with next day S&P 500 reactions to quantify the relationship between tone and market moves.

4. **Builds Predictive Models** → Trains models to forecast whether the S&P 500 will move up or down after a press conference, as well as the magnitude of the move.

5. **Delivers a Web APP** → A user facing tool where new or past Powell statements can be analyzed. It will output a sentiment score, show historical reaction patterns, and provide a forecast of likely market response.

Choice of Dataset

For this project, we will use two datasets. The first one is the "Jerome Powell Press Release Q&A" from Kaggle, which contains txt files of the Federal Reserve press conference transcripts. These transcripts will be analyzed using NLP techniques to generate sentiment and tone scores that reflect whether the language is hawkish or dovish.

The second dataset is financial market data of S&P 500 collected through Yahoo Finance using the yfinance library. These two datasets allow us to test whether language sentiment can predict stock market performance.

Methodology

1 - Data Preprocessing

The Jerome Powell FED Press Conference Transcripts (2020–2025) dataset is feasible for NLP and financial modeling tasks. It contains structured transcripts from all FOMC press conferences, covering 5 years of consistent policy communication, making it good for studying the relationship between language tone and market reactions. Additionally, daily market data (S&P 500) are publicly available from sources like Yahoo Finance and FRED, allowing easy alignment between Powell's speeches and market movements.

2 - Machine Learning Model

We want to first assign a score to the speech based on sentimental analysis and then second make the forecast of the future market growth. We will apply the **finBERT** model for processing a financial context such as JP press conference script, which gives the probabilities for positive/negative/neutral tone.

Alternative models:

option1 : random forest/XGBoost/Logistic regression
Pros: small dataset, simple, easy to train
Cons: missing language text

option2 : LSTM
Pros: capture the importance of time dependence
Cons: hard to train on long text

3 - Evaluation Metric

confusion matrix and accuracy: classification
ROC-AUC: market direction separatio
Correlation coefficient: between sentiment and return
SHAP analysis: interpret which words/tone drive predictions
Maybe other evaluation metrices (ask TPM)

Application

1 - User Input

Input Type: Text
The user will paste or upload the transcript (or part of it) from a Federal Reserve press conference.

Interface
A text box for manual input or a file upload button for .txt or .csv transcripts.
For demonstration purposes, a dropdown menu will let users choose from past Powell speeches preloaded from our dataset.

2 - Output

Market Prediction
Predicted Market Movement: Upward or Downward

Sentiment Visualization
A “hawkishness vs dovishness” score based on keyword frequencies (e.g., “inflation”, “uncertainty”, “growth”)

Keyword Heatmap
Highlight key words or phrases in the transcript that contributed most to the model’s decision (via SHAP values or attention scores).