Predicting Interest Rate Decisions

Ben Nam-Soo Weihrauch

Machine Learning for Quantum Scientists

July 31, 2026

Objective & Hypothesis

Problem

- Central bank interest rates drive global financial markets.
- Goal: Predict interest rate decision (Cut, Hold, Hike)
- Inputs:
  1. Official FOMC statements (Text).
  2. Consumer Price Index (CPI) & Nonfarm Payrolls (NFP) (Tabular Data).

Objective & Hypothesis

Problem

- Central bank interest rates drive global financial markets.
- Goal: Predict interest rate decision (Cut, Hold, Hike)
- Inputs:
  1. Official FOMC statements (Text).
  2. Consumer Price Index (CPI) & Nonfarm Payrolls (NFP) (Tabular Data).

Hypothesis

- Multimodal model with sentiment from text and economic data should outperform unimodal (text-only or tabular-only) baselines.

Dataset & Look-Ahead Bias

Dataset Overview

- $N = 205$ FOMC meetings (2000 – Present).
- Very imbalanced target classes: Mostly “Hold”.

Dataset & Look-Ahead Bias

Dataset Overview

- $N = 205$ FOMC meetings (2000 – Present).
- Very imbalanced target classes: Mostly "Hold".

Look-Ahead Bias

- Important: Macroeconomic data (especially jobs numbers) gets revised frequently
- Use unrevised data from the time of the decision
- Evaluated using expanding window walk forward validation (not random train-test splits)

Evaluating Economic Data: XGBoost vs. DNN

To test the macroeconomic data I compared a Deep Neural Network (DNN) against Extreme Gradient Boosting (XGBoost).

What is XGBoost?

- An ensemble of simple decision trees.
- Each new tree tries to predict the errors of the last one.
- Effective without needing the massive datasets required by neural networks.

Evaluating the Text: FinBERT

To test the FOMC statements in isolation, I fine-tuned FinBERT.

What is FinBERT?

- A language model (based on BERT) pre-trained on financial texts.
- FinBERT knows cut isn't physical damage here.
- To combat overfitting, I froze the bottom 11 layers and only trained the final layer.

Multimodal Fusion & Evaluation Metric

The Multimodal Fusion Network

- Used an embedding vector from FinBERT and an embedding vector from the DNN.
- Used a gated fusion layer to combine them.
- Allows model to weigh the textual and economic branches before the final classification.

Multimodal Fusion & Evaluation Metric

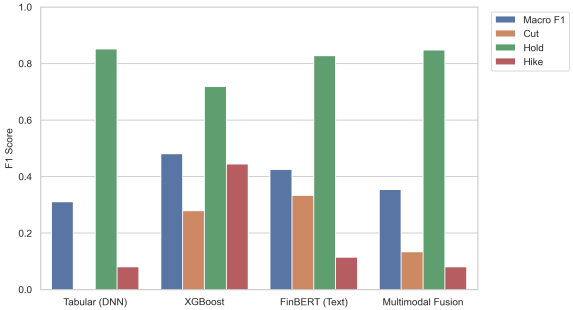
The Multimodal Fusion Network

- Used an embedding vector from FinBERT and an embedding vector from the DNN.
- Used a gated fusion layer to combine them.
- Allows model to weigh the textual and economic branches before the final classification.

Evaluation: Macro F1-Score

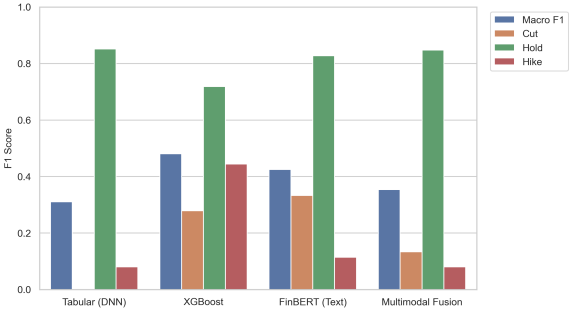
- Because the dataset is heavily skewed, a model could achieve high accuracy just by predicting "Hold" every time.
- *Macro F1* calculates the F1-score for each class (Cut, Hold, Hike) independently, then averages them.
- This equally penalizes models that fail to predict the rare, but crucial, rate cuts and hikes.

Results



- **1st: XGBoost (0.48)**
- **2nd: FinBERT (0.43)**
- **3rd: Fusion (0.35)**
- **4th: Tabular DNN (0.31)**

Results



- **1st: XGBoost (0.48)**
- **2nd: FinBERT (0.43)**
- **3rd: Fusion (0.35)**
- **4th: Tabular DNN (0.31)**

Note:

- FinBERT missed many hikes and cuts
- XGBoost falsely predicted many hikes and cuts
- similar scores for different reasons

Why Fusion Failed & Next Steps

Why did the Multimodal Network underperform?

- The tabular DNN branch was too weak for this low-dimensional low sample size data.
- The fusion gate failed to filter out the bad tabular representation, which dragged down the better text signal from FinBERT.
- *Takeaway:* Despite trying to combat overfitting using class weights, dropout and parameter freezing in FinBERT there was just not enough data for the more complex models to show their strengths.

Why Fusion Failed & Next Steps

Why did the Multimodal Network underperform?

- The tabular DNN branch was too weak for this low-dimensional low sample size data.
- The fusion gate failed to filter out the bad tabular representation, which dragged down the better text signal from FinBERT.
- *Takeaway:* Despite trying to combat overfitting using class weights, dropout and parameter freezing in FinBERT there was just not enough data for the more complex models to show their strengths.

Outlook

- Use FinBERT to generate a Sentiment Score.
- Feed this sentiment score as a third tabular variable directly into the XGBoost model alongside CPI and NFP.
- Also add other macroeconomic indicators like treasury yields.