Assignment 3: NLP Analysis and Forecasting using the Federal Reserve Meeting Minutes

NLP Assignment_3 – Do not distribute

Assignment 3: NLP Analysis of Federal Reserve Meeting Minutes and Economic Indicators

Background

Natural Language Processing (NLP) plays a crucial role in extracting insights from institutional communications, particularly in the financial sector. The Federal Reserve's meeting minutes contain valuable information about monetary policy, economic outlook, and market sentiment. Financial analysts, economists, policymakers, and investors use NLP to analyze trends in central bank communications to anticipate policy changes and economic shifts.

Organizations utilize central bank communication analysis for:

● Market Predictions – Financial institutions monitor central bank language for signals about future interest rate changes and economic outlook.

● Policy Analysis – Economists track shifts in central bank rhetoric to understand evolving monetary policy frameworks.

● Sentiment Tracking – Tracking the sentiment in monetary policy statements helps gauge the central bank's confidence in economic conditions.

● Risk Assessment – Financial institutions use central bank communication analysis to identify potential economic risks and policy uncertainties.

● Economic Forecasting – Analyzing the relationship between central bank language and economic indicators helps improve economic forecasts.

This assignment will guide you through collecting, processing, and analyzing Federal Reserve meeting minutes alongside economic indicators to extract meaningful insights about monetary policy communication and economic relationships.

Project Tasks

Step 1: Build a Dataset Using Federal Reserve Meeting Minutes- Notebook 1

- ● Scrape Federal Reserve meeting minutes from: https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm ○ Collect transcripts from 2015-2025 (approximately 60 transcripts) ● For each transcript, extract: ○ Meeting Date ○ Meeting Time (if available) ○ Full Text Content ○ Any other available metadata ● Ensure comprehensive data collection with proper error handling

Step 2: Classical NLP Approach - Sentiment Analysis - Notebook 2

● Use VADER sentiment analysis tool: ○ Extract sentiment scores (positive, negative, neutral, compound) for each transcript ○ Create visualizations of sentiment trends over time ○ Plot sentiment scores against economic indicators ● Create word clouds: ○ Generate a 9x9 subplot of word clouds showing top 20 sentiment words for each year (2015-2025) ○ Analyze changes in sentiment vocabulary over time ● Document the narrative revealed by your analysis

Step 3: Classical NLP Approach - Topic Modeling - Notebook 2

● Use Gensim for LDA (Latent Dirichlet Allocation) topic modeling: ○ Extract key topics from Federal Reserve communications ○ Track topic evolution over time (2015-2025) ○ Visualize topic distributions and their changes ○ Overlay topic distributions with sentiment analysis ● Document the narrative revealed by combining topic and sentiment analysis

Step 4: Modern NLP Approach - Transformer-Based Sentiment Analysis - Notebook 3

● Create **text noise analysis graphs and clean data** - do that for step 2 and step 3 as well

Use FinBERT for zero-shot sentiment analysis: ○ Apply Hugging Face transformers and pipeline library ○ Extract sentiment scores for each transcript ○ Create visualizations of sentiment trends over time ● Generate word clouds: ○ Compare with classical approach results ● Document the narrative revealed by this modern approach

Step 5: Modern NLP Approach - Extractive Summarization - Notebook4

● Use FinBERT for extractive summarization: ○ Apply FinBERT with summarization techniques ○ Generate concise summaries of Federal Reserve transcripts ○ Evaluate summary quality and information retention ○ Compare summaries across different time periods

Step 6: Modern NLP Approach - Topic Extraction - Notebook 5

● Apply BERTopic for topic modeling: ○ Extract topics using a transformer-based approach ○ Create visualizations of topic trends over time ○ Compare with classical LDA topic modeling results ● Document insights from this modern approach to topic modeling

Part 2: Presentation (9-Slide Deck)

Your presentation should cover:

1. Introduction – Importance of central bank communication analysis & economic value
2. Data Collection – Source, methodology, dataset overview
3. Draw out your entire flow in picture end to end on how you built this application
4. Exploratory Data Analysis – Key statistics & visualizations of both text and economic data
5. Text Preprocessing – Tokenization, cleaning, lemmatization impact
6. Sentiment Analysis – Classical vs. Modern approaches, key findings
7. Topic Modeling – Topic evolution and relationship with economic indicators
8. Economic Insights – How the analysis of Fed communications correlates with economic trends
9. Conclusion – Key takeaways & future improvements

Submission Requirements

● Jupyter Notebook (.ipynb) – Well-documented code

● Code notebook as pdf

● CSV datasets – The datasets you created

● PowerPoint Presentation (.pptx or Google Slides) – No PDFs