BoE - Notebooks' Discoveries

Lara Cadel

Table of contents:

The series of notebooks provided crucial insights into the dynamics of financial communication, sentiment, and macroeconomic indicators, revealing nuanced relationships, data patterns, and modeling challenges.

---

1. Notebook 1: Foundational Discoveries

1. **Speech and Sentiment Dynamics**
    - The *all_speeches* dataset showed that financial speeches often feature complex language, which can dilute strong sentiment.
    - By applying the *sentiment_word_list*, key sentiment categories (e.g., *Positive*, *Negative*, *Uncertain*) were mapped to the speech text, uncovering a lexicon-centric measure of financial tone.
2. **Economic Indicators as Contextual Drivers**
    - The *economic_indicators* dataset revealed trends in GDP, inflation, labor markets, and wages. These trends provided a macroeconomic context essential for interpreting financial sentiment, showing clear alignments between key metrics and periods of heightened speech activity.
3. **Data Preparedness for Analysis**
    - Cleaning and preprocessing improved data consistency, enabling the creation of integrated datasets like *loughran_mcdonald_sentimentscore.csv*, setting the stage for sentiment and correlation analyses.

---

2. Notebook 2: Sentiment and Correlation Analysis

1. **Sentiment Analysis Patterns**
    - **512 Tokens Truncation Approach:** Captured a wider sentiment range, with 12.9% Negative and 1.1% Positive scores.
    - **Statement-by-Statement Approach:** Showed a stronger skew toward neutrality (94.7%), reflecting how narrower contexts dampen emotional tone.
    - **Key Discovery:** Broader speech segments yield richer sentiment variability compared to isolated statements.
2. **Macroeconomic Correlations**
    - While GDP and CPI showed weak correlations with speech sentiment, **redundancies (lag 2)** had a significant negative relationship, suggesting layoffs dampen financial optimism.
    - The VAR model demonstrated limited short-term interplay between macroeconomic variables and sentiment, underscoring the complexity of these relationships.
3. **Speech Tone Dynamics**
    - Neutral sentiment dominated, highlighting the cautious and measured tone prevalent in financial communication, particularly in policy contexts.

---

3. Notebook 3: Predictive Modeling Insights

1. **Random Forest Findings**
    - Feature importance analysis highlighted GDP rate and CPI as key predictors, but the model struggled with accuracy, emphasizing the complexity and non-linearity of sentiment dynamics.
2. **SARIMA Model Discoveries**
    - **GDP Rate:** Emerged as a significant driver of sentiment, with a positive influence (p-value = 0.001).
    - **CPI (3-Month Average):** Showed a negative association, indicating that inflationary pressures correspond to lower sentiment.
    - **Key Limitation:** SARIMA captured long-term trends but lacked precision for short-term sentiment variations, highlighting challenges in addressing context-driven fluctuations.

---

4. Key Cross-Notebook Discoveries

1. **Sentiment Neutrality**
    - Neutral sentiment dominated financial speeches, particularly in statement-level analyses, suggesting cautious rhetoric by policymakers.
2. **Macro-Sentiment Relationships**
    - While GDP showed a consistent influence, short-term metrics like CPI and earnings had weaker impacts. Layoffs emerged as a notable short-term sentiment driver.
3. **Modeling Challenges**
    - Traditional models like Random Forest and SARIMA struggled with the context-dependent nature of financial sentiment, pointing to the need for more adaptive methods, such as hybrid or deep learning models.

These discoveries underscore the intricate and measured nature of financial communication, the subtle influence of macroeconomic variables on sentiment, and the challenges of modeling this complexity effectively.