3666 ANLP

Predicting ECB Monetary Policy

Jonathan Fu
Fernando Espinosa
Shabnam Balamchi

Dec 16th/2023

Outline

- Project Overview
- Challenges, Complexities and Use Cases
- Process Flow
- Exploratory Data Analysis (EDA)
- Methods and Models
- Next steps

Project Overview: ECB Speech Analysis for Rate Change Prediction

Objective:

Predict ECB rate changes ("No Change," "Increase," or "Decrease") using speech text analysis.

- Total speeches: **3,434**
- Date range: **1999–2023**
- Language: Majority in **English**

Columns:

- Speaker details (e.g., who, date, speech title).
- Extracted text and term frequencies.
- Monetary policy rate changes (% changes, direction).

- **Key Question**:
  "How does speech content correlate with monetary policy decisions?“

Main Scope:

Data Analysis: Extraction and analysis of ECB speeches and press releases for predictive modeling.

Model Development: Building and fine-tuning a BERT-based vs LSTM vs Tree-based classification model for forecasting ECB rate changes.

Challenges, Approaches, and Use Cases

Challenge	Impact	Approach/Technique
Class Imbalance	Minority classes ("Increase" & "Decrease") were underrepresented. Leads to poor performance.	• **Class Weights**: Adjusted loss function. • **Manual Oversampling**: Balanced training data • **SMOTE:** Synthetic Minority Oversampling Technique in imbalance pipeline • **Tree-Based Models (XGBoost, LightGBM, CatBoost)**: Handled imbalance effectively with ensemble methods.
Semantic Complexity	Complex language made it challenging for simpler models.	• **BERT Fine-Tuning**: Extracted semantic-rich features. • **Keyphrase Extraction for LSTM**: Simplified inputs. • **Tree-Based Models**: Captured complex patterns using gradient boosting.
Overfitting Risks	Deep learning models overfit due to small dataset size.	• **Early Stopping**: Stopped training to prevent overfitting. • **Data Augmentation**: Paraphrased existing data. • **Voting Classifier**: Combined robust models to generalize better.

Examples

Corporate treasury management
(eg. lower borrowing costs)

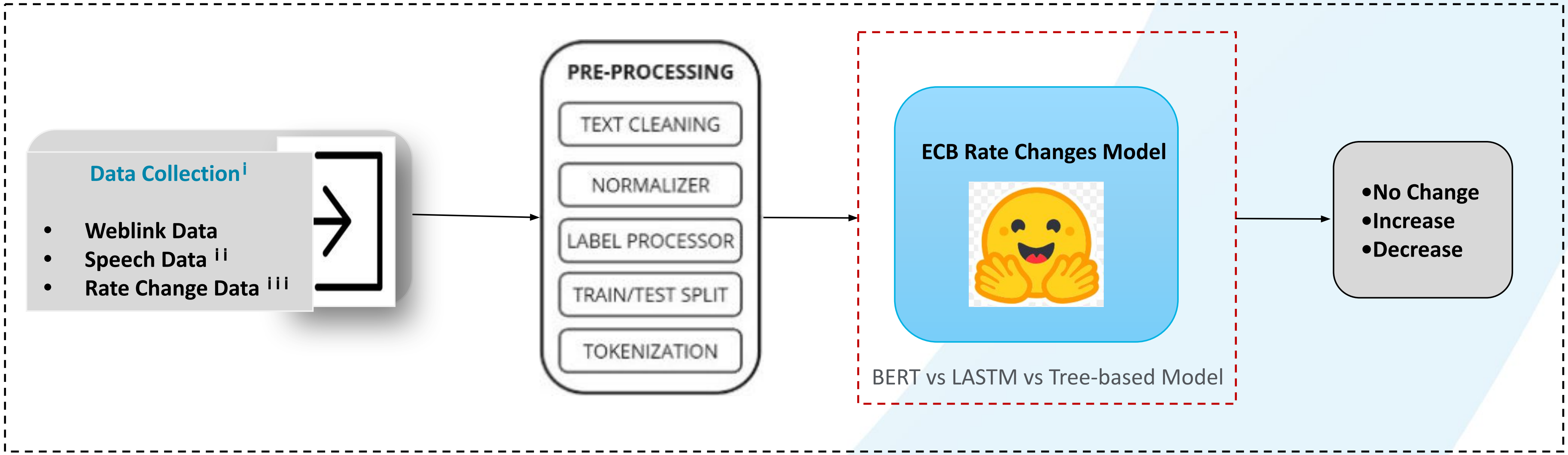
FX, currency trading/hedging and Sovereign debt analysis

Optimize investment strategy and portfolio management
(eg. asset allocation)

Better risk management
(eg. minimize the cost of interest rate hedging)

Process Flow

Title: *Methodology Overview for ECB Rate Change Prediction*



[i] Screen-scraped ECB speeches and press releases using BeautifulSoup; tagged articles by dates between monetary policy decisions.

[ii] https://www.kaggle.com/datasets/robertolofaro/ecb-speeches-1997-to-20191122-frequencies-dm

[iii] https://www.ecb.europa.eu/stats/policy_and_exchange_rates/key_ecb_interest_rates/html/index.en.html

Exploratory Analysis

Frequency Analysis: Understanding key terms and bigrams in the speeches.

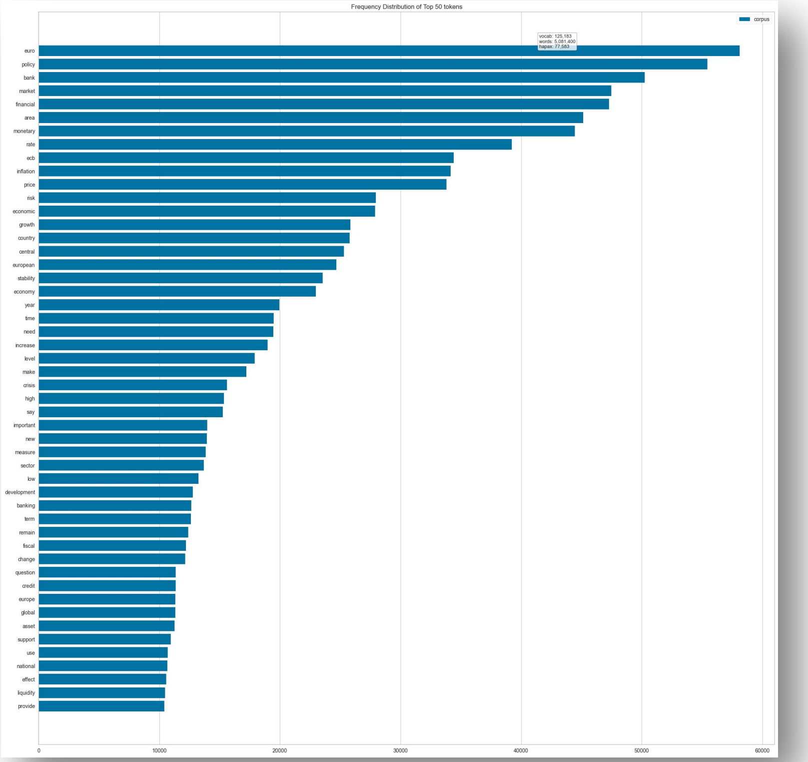
Distribution Insights: Temporal and categorical patterns of the data.

Topic Modeling: Extracting dominant topics and their keywords using LDA.

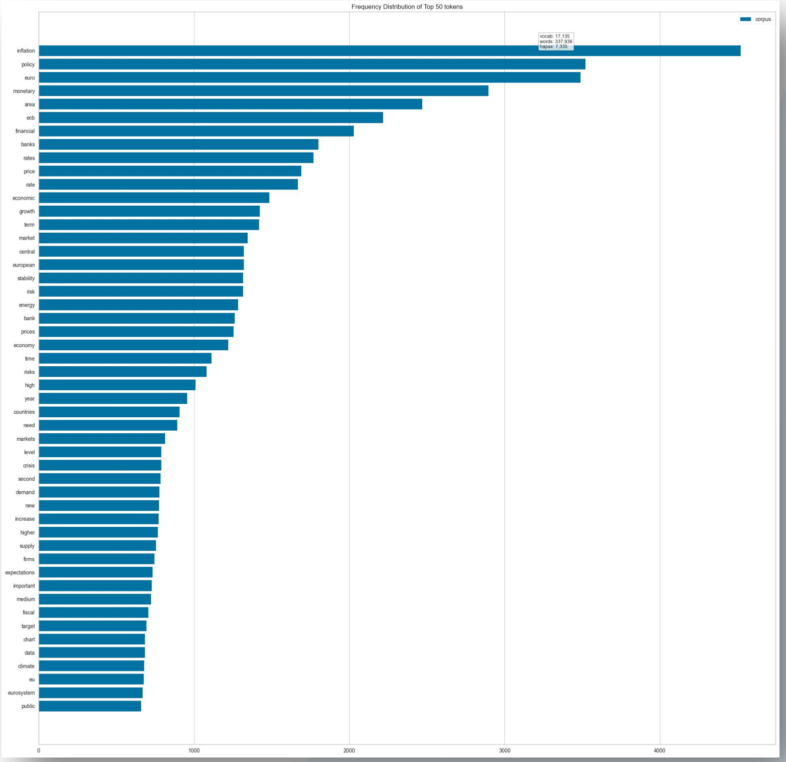
Clustering Analysis: Visualizing clusters and validating using silhouette scores.

Frequency Distribution

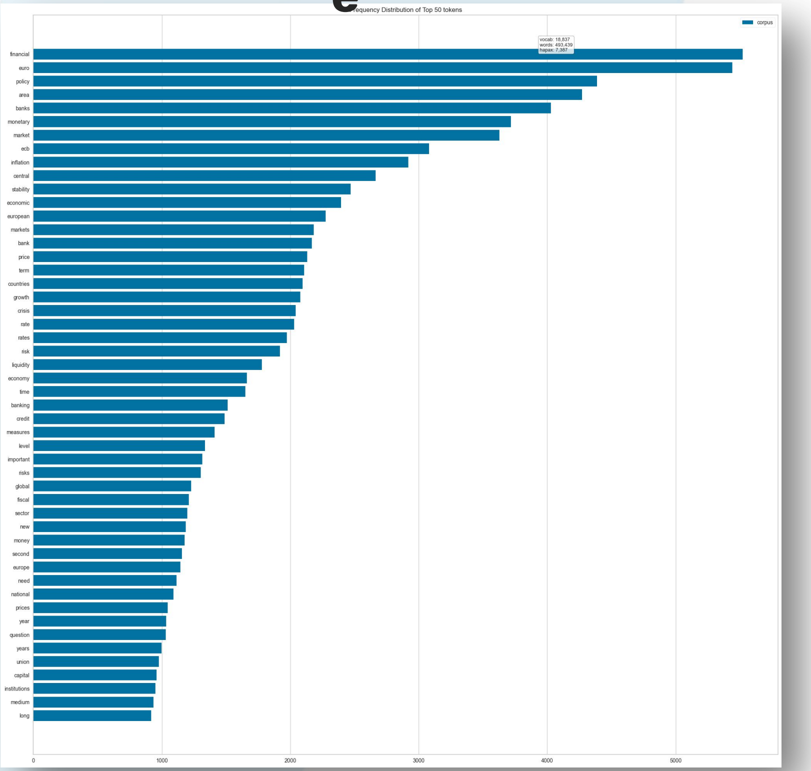
Overall



Increase

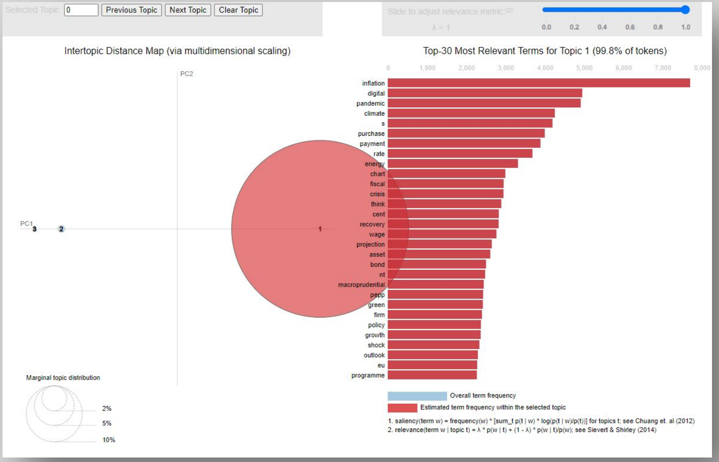


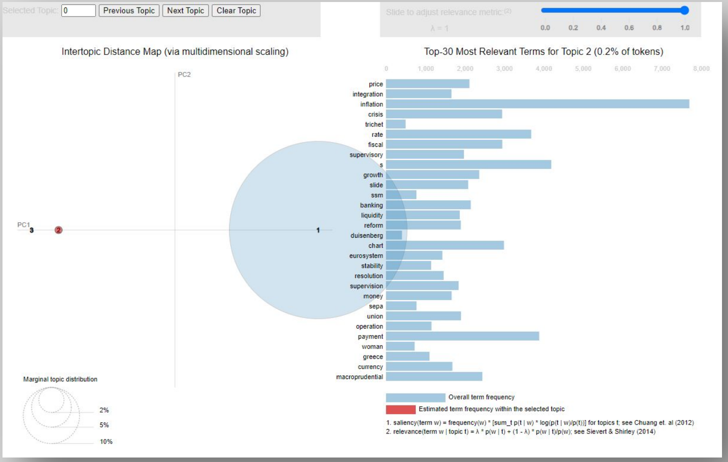
Decrease

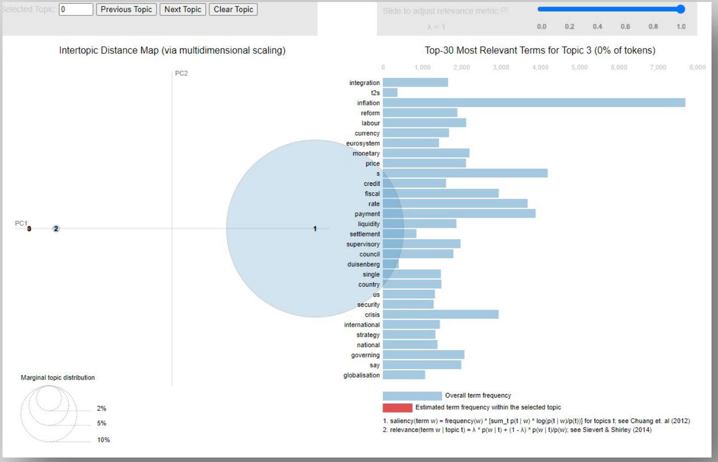


Key Topics in Speeches (Gensim LDA Visualization)

How ECB prioritizes its messaging in response to economic changes





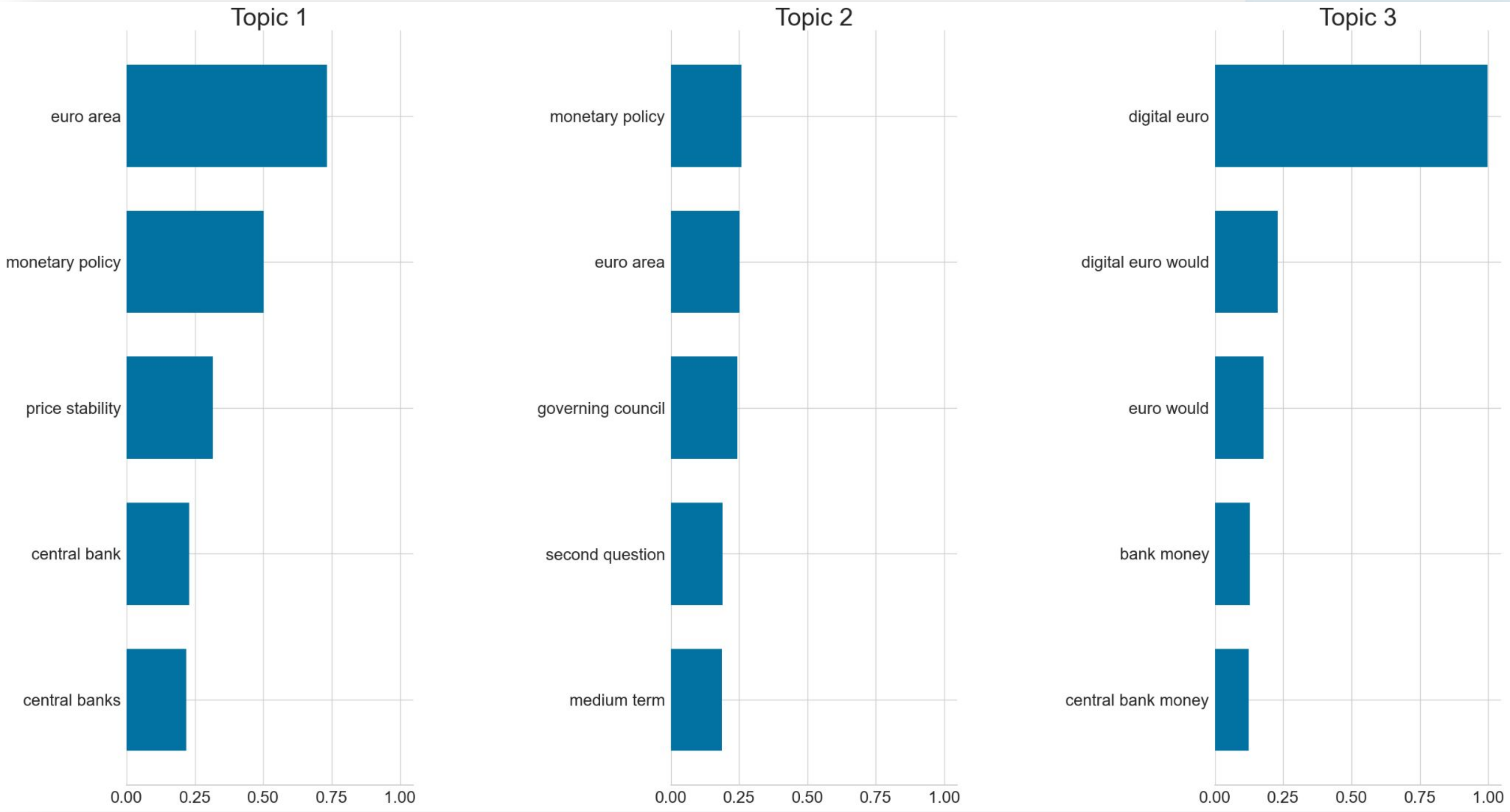


Speeches cluster around 3–4 major themes:

- Topic 1: Monetary policy, inflation, and price stability dominate discussions.
- Topic 2: Digital euro and fiscal policy have emerged in recent years.
- Topic 3: Broader economic growth and crisis-related topics.

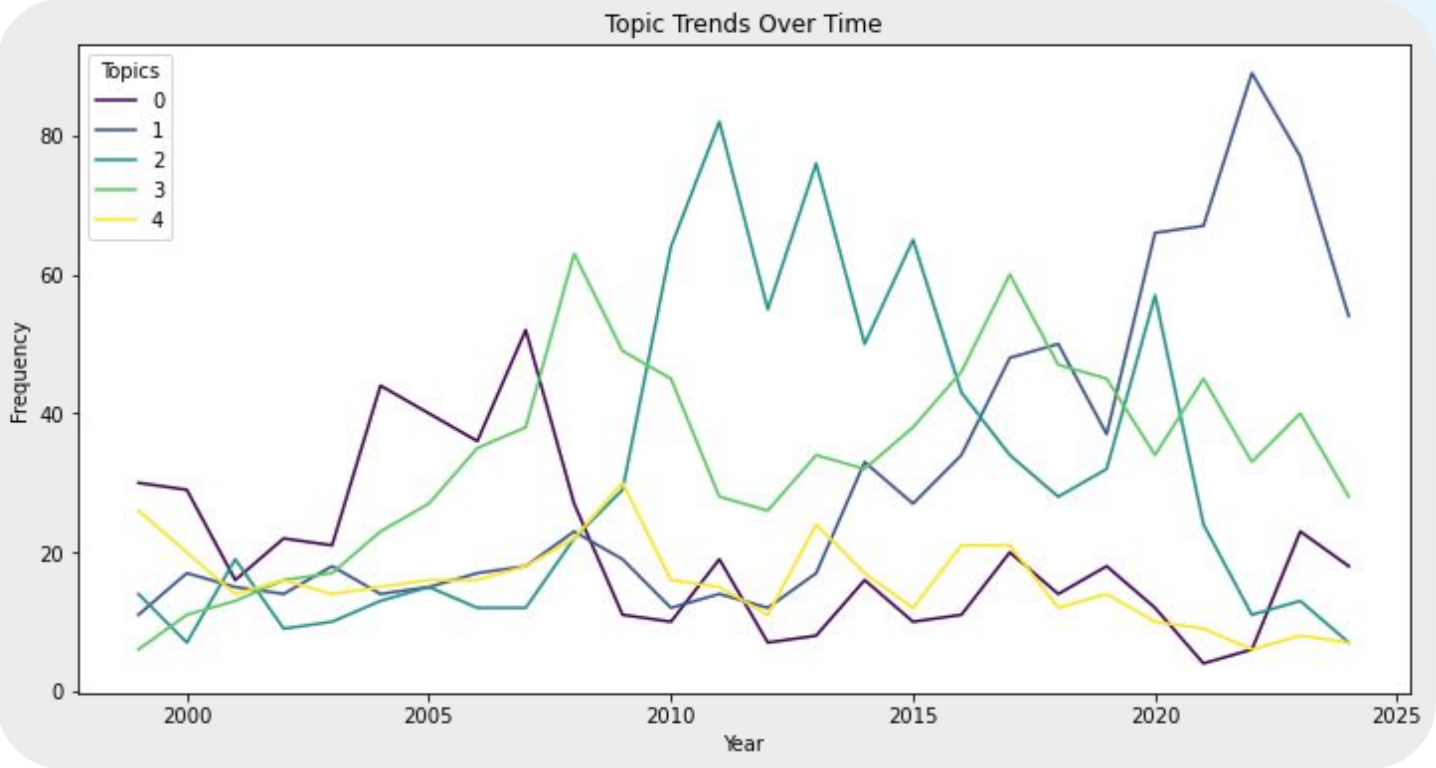
N-Grams

No significant difference between the overall topics vs. increase/decrease classes



Temporal Trends in Topics

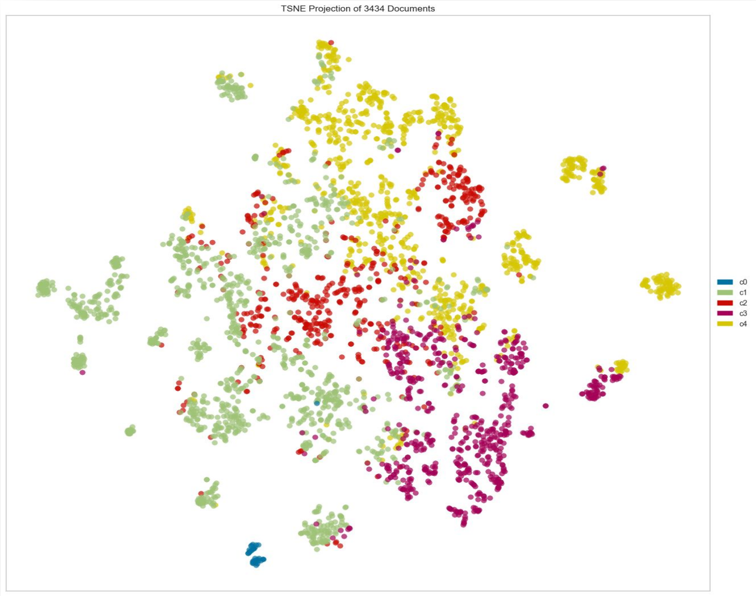
- Digital euro and fiscal policy discussions have risen sharply since 2015.
- Inflation and monetary policy remain consistently discussed across years.
- This demonstrates how the ECB adapts its messaging to evolving economic conditions.

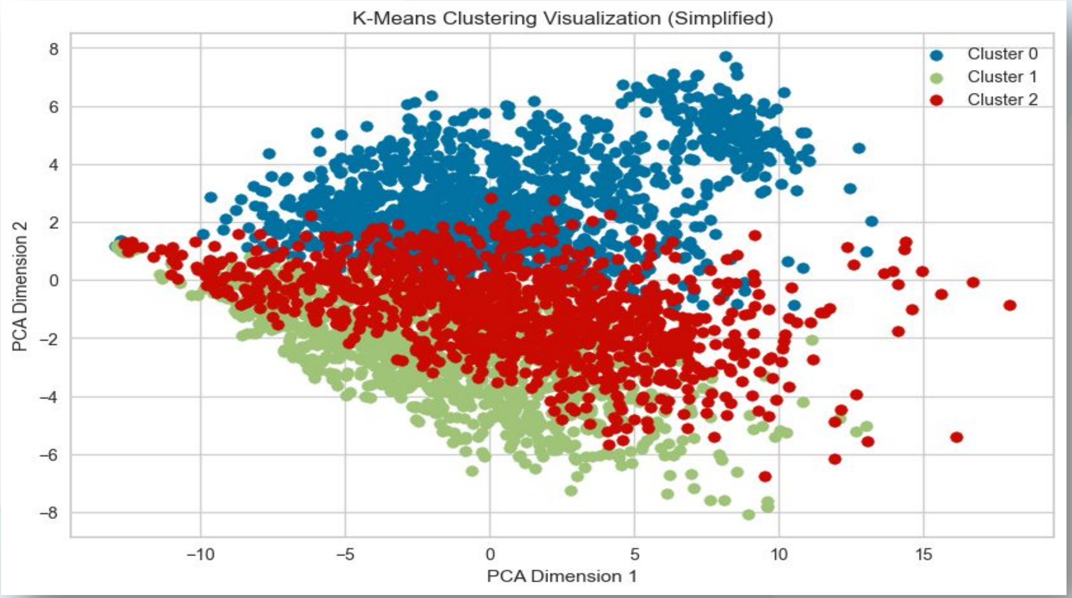


- Topic 0 (financial growth and markets) saw a rise around 2010 but declined in prominence after 2015.
- Topic 3 (economic and fiscal policy) peaked during the late 2000s, possibly reflecting responses to the 2008 financial crisis.
- Topic 4 (central banking and risk management) had a steady focus, indicating its importance across different timeframes.

Unsupervised Cluster Analysis/t-SNE visualization

- **Best Clustering:** The silhouette score is highest at **k=2**, indicating two dominant groupings or themes in ECB speeches.
- **Cohesion & Separation:** k=2 provides the best balance between cohesion (within-cluster similarity) and separation (between-cluster distance).
- **Diminishing Returns:** Lower scores for higher **k** suggest limited benefits from additional clusters, though finer nuances might still be captured.





Methods and Models

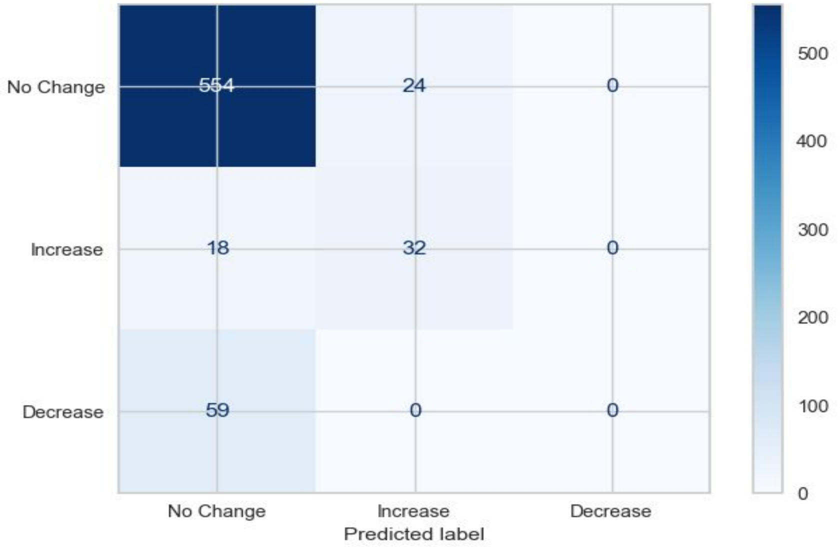
- Metrics: F1-Score (Macro and Weighted), Precision and Recall, Confusion Matrix
- Train/Eval split: 80/20 Test split: 20% held-out data to evaluate the final model.
- BERT, LSTM, and Voting Classifier models
- Classification Task: multiclass
  - Labels: No Change/Increase/Decrease

Model	Architecture	Size	Challenges Addressed
BERT	Transformer	Base	Semantic Complexity, Class Imbalance
LSTM	RNN	Base	Class Imbalance, Small Data Handling
Voting Classifier	**Voting Classifier**	**Voting Classifier**	**Voting Classifier**

** The Voting Classifier outperformed all other models by combining strengths of XGBoost, LightGBM, and CatBoost through weighted soft voting*

BERT

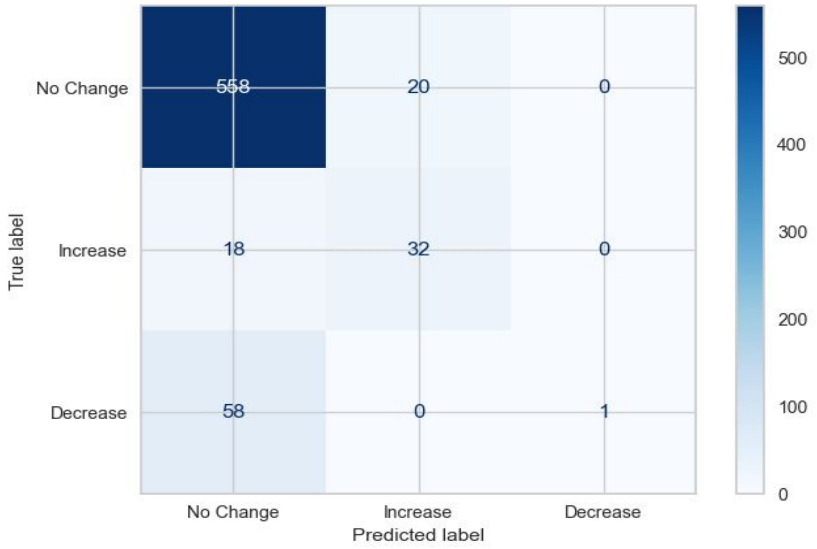
```
Validation loss: 0.403
Accuracy: 0.85

Classification Report:
              precision    recall  f1-score   support

   No Change       0.88      0.96      0.92       578
    Increase       0.57      0.64      0.60        50
    Decrease       0.00      0.00      0.00        59

    accuracy                           0.85       687
   macro avg       0.48      0.53      0.51       687
weighted avg       0.78      0.85      0.81       687
```



BERT - First 10 Epochs

BERT - 14 Additional Epochs

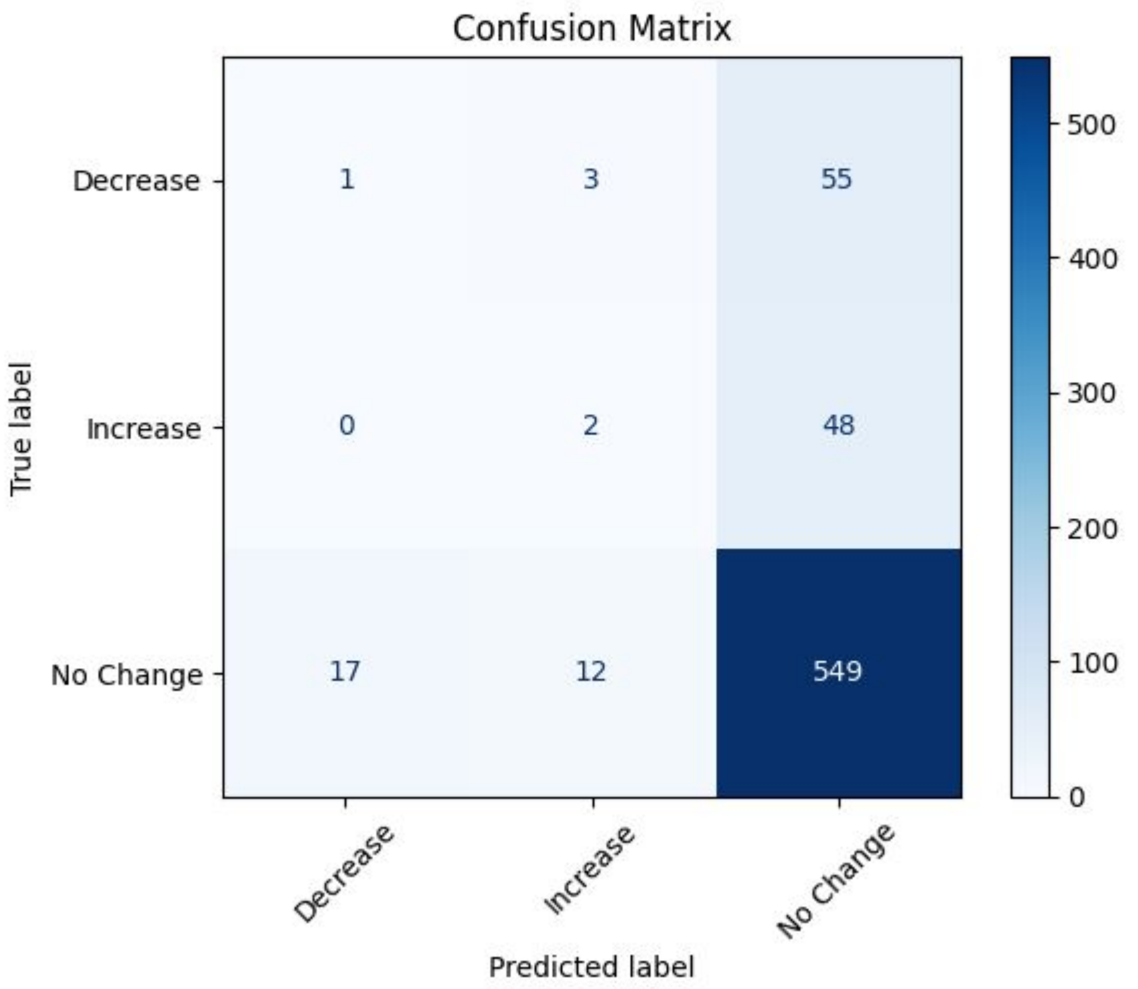
```
Validation loss: 0.363
Accuracy: 0.86

Classification Report:
              precision    recall  f1-score   support

   No Change       0.88      0.97      0.92       578
    Increase       0.62      0.64      0.63        50
    Decrease       1.00      0.02      0.03        59

    accuracy                           0.86       687
   macro avg       0.83      0.54      0.53       687
weighted avg       0.87      0.86      0.82       687
```



LSTM

Classification Report:

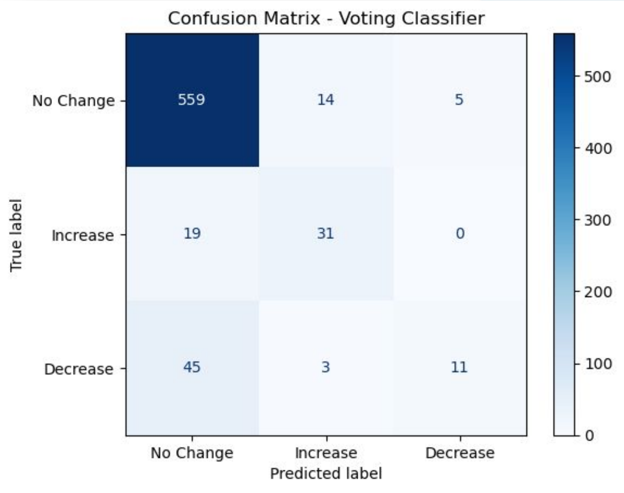
	precision	recall	f1-score	support
Decrease	0.06	0.02	0.03	59
Increase	0.12	0.04	0.06	50
No Change	0.84	0.95	0.89	578
accuracy			0.80	687
macro avg	0.34	0.34	0.33	687
weighted avg	0.72	0.80	0.76	687



Voting Classifier: The Best Performing Model

Model	Macro Precision	Macro Recall	Macro F1-Score	Increase F1-Score	Decrease F1-Score	Comments
Logistic Regression	0.57	0.68	0.61	0.57	0.39	Good recall but lower precision.
XGBoost	0.86	0.50	0.55	0.57	0.15	High precision, poor recall for "Decrease".
LightGBM	0.76	0.57	0.61	0.65	0.25	Balanced but weak "Decrease" recall.
CatBoost	0.61	0.62	0.60	0.58	0.32	Best "Decrease" F1-score so far.
Voting Classifier	**0.79**	**0.59**	**0.62**	**0.63**	**0.29**	**Best overall balance across metrics.**

```
              precision    recall  f1-score   support

   No Change       0.90      0.97      0.93       578
    Increase       0.65      0.62      0.63        50
    Decrease       0.69      0.19      0.29        59

    accuracy                           0.87       687
   macro avg       0.74      0.59      0.62       687
weighted avg       0.86      0.87      0.85       687
```



Challenges Addressed:

- **"Improved recall for minority classes (Increase/Decrease) and handled class imbalance effectively."**

"Voting Classifier achieves the best trade-off between overall accuracy and minority class performance."

Why Tree-Based Models Performed Better? (possibilities)

- **Handling Imbalanced Data**:
  - Tree-based models use **class weights** and ensemble methods to manage class imbalance effectively.
- **Small Dataset Suitability**:
  - BERT and LSTM require large datasets; tree-based models generalize better with limited data.
- **Efficient Training**:
  - Tree-based models are **faster to train** and less computationally intensive.

Next Steps

- **Enhance Minority Class Detection**:
  Adjust **decision thresholds** and explore precision-recall trade-offs for **Increase** and **Decrease** classes.

- **Newer Transformer Models**:
  Test models like **FinBERT**, **RoBERTa**, and **DeBERTa** for improved text understanding.

- **Improved Sampling**:
  Use **ADASYN** and **Cluster-Based Sampling** for better handling of class imbalance.

Q&A