Financial News as a Proxy of European Central Bank Interest Rate Adjustments

Davide Paris
Intesa Sanpaolo / Milan - Italy
University of Milano-Bicocca / Milan - Italy
davide.paris@intesasanpaolo.com

Elisabetta Fersini
University of Milano-Bicocca / Milan - Italy
elisabetta.fersini@unimib.it

Martina Menzio
Intesa Sanpaolo / Milan - Italy
University of Milano-Bicocca / Milan - Italy
martina.menzio@intesasanpaolo.com

Abstract

This paper examines the relationship between news coverage and the European Central Bank's (ECB) interest rate decisions. In particular, the hypothesis of a linear relationship between financial news and ECB indications regarding interest rate variations is investigated by leveraging state-of-the-art large language models combined with domain experts and automatically selected keywords. The analysis revealed two key findings related to how news contents can signal the ECB's decisions to raise or lower interest rates: (1) Sentence Transformer models, when combined with domain-specific keywords, exhibit a higher correlation with ECB decisions than state-of-the-art financial BERT architectures; (2) employing a grid search strategy to select subsets of informative keywords strengthened the relationships between news contents and ECB's decisions, highlighting how media narratives can anticipate or reflect central bank policy actions.[1]

1 Introduction

The European Central Bank (ECB) plays a pivotal role in shaping monetary policy for the Eurozone. Accurate prediction of ECB interest rates is crucial for financial markets, businesses, and policymakers. Traditionally, interest rate forecasts have relied on indicators such as inflation, GDP, unemployment rates, and consumer confidence indices. However, these metrics can be challenging to obtain and are often updated infrequently, as they are released at specific intervals rather than in real-time. Additionally, news outlets frequently feature advocacy from political leaders across individual Eurozone nations, who make statements aimed at influencing the ECB's decision-making process. This can complicate the prediction process. Recent advancements in machine learning, particularly through BERT-inspired models (Devlin, 2018), have revolutionized text analysis by enabling more sophisticated and nuanced interpretations of textual data. Previous research has shown that central bank communication plays a crucial role in shaping market expectations, particularly influencing bond yields and interest rates (Leombroni et al., 2021).

This paper examines the effectiveness of using natural language processing (NLP) models to analyze financial news in assessing ECB interest rate decisions over short-term (up to 6 weeks), medium-term (6 to 12 weeks), and long-term (12 to 18 weeks) periods. In particular, two main contributions are provided:

1. the combination of a Sentence Transformer model with domain-specific keywords supplied by experts to better correlate financial news with the increase/decrease of interest rates set by the ECB;

2. the implementation of a grid search approach to determine a sub-optimal set of keywords extracted from text to correlate ECB rate variations with financial news.

[1] The views and opinions expressed are those of the authors and do not necessarily reflect the views of Intesa Sanpaolo, its affiliates or its employees.

Proceedings of Recent Advances in Natural Language Processing, pages 907–914
Varna, Sep 8–10, 2025
https://doi.org/10.26615/978-954-452-098-4-104

2 Related Work

The study of interest rate trends has been extensively discussed in the literature until 2016 (Picault and Renault, 2017), leveraging the NLP tools available at that time, such as n-grams. For over six years, interest rates remained at zero, leading to a decline in scientific interest in the topic.

However, starting in 2022, the ECB decision to raise interest rates has reignited attention on the subject, making it a trending topic once again. Recent works, such as (Fanta and Horvath, 2024), employ modern large language models (LLMs) but focus on ECB speeches rather than news sources, and their datasets are limited to the period from 2008 to 2016. Other studies have developed innovative approaches to measure sentiment, such as using deep learning models to create indices from central bank narratives, further confirming the power of NLP in this domain (Niţoi et al., 2023). Notably, research has also demonstrated the predictive value of central bank sentiment, highlighting the informative content of ECB and Federal Reserve communications in forecasting monetary policy decisions (Hilscher et al., 2024). In this context, tools like CentralBankRoBERTa (Pfeifer and Marohl, 2023) have also emerged, enhancing the analysis of central bank communications using specialized transformers for language modeling. In parallel, projects like FedNLP (Lee et al., 2021) have concentrated on forecasting the Federal Funds Rate, covering the period from 2015 to 2020.

More recently, (Bernoth, 2025) applied a RoBERTa-based model to ECB communications (2019–2025), showing that the tone—classified as hawkish, dovish, or neutral—can significantly enhance forecasts of upcoming interest rate decisions.

Despite this progress, no current research has explored correlations between ECB interest rates and news-based datasets. This gap in the literature highlights the need for novel approaches that leverage recent advancements in NLP, particularly in the application of LLMs to real-time financial news data for the prediction of ECB interest rates.

3 Data

News articles were sourced from `reuters.com`. To identify relevant content, a selection process was implemented to filter articles based on specific keywords and phrases in their titles. The focus was on identifying news that mentioned the `European Central Bank (ECB)`, the `Eurozone` and related financial terms such as `interest rates` and `monetary policy`. Articles that included both references to the ECB and financial terms related to interest rates were selected for analysis. News articles published within 24 hours after the ECB's announcements regarding interest rate changes were excluded, as these typically focus on reporting the ECB's actions rather than providing forecasts or analyses of interest rates.

The dataset under investigation comprises 1,037 English-language news articles spanning from March 21, 2022, to April 17, 2025. The distribution of these articles on a monthly basis is illustrated in the figure 1.

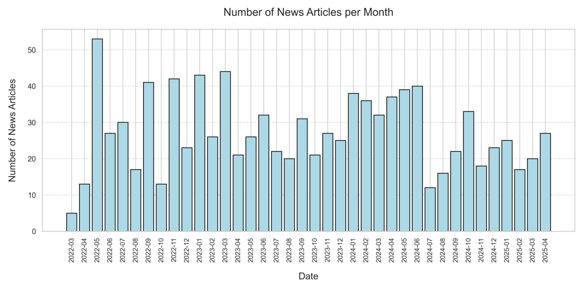

Figure 1: Number of news articles per month

The ECB makes decisions regarding adjustments to interest rates predominantly during the Governing Council meetings, which are convened approximately every six weeks. Accordingly, within this interval, we have analyzed a total of 27 such decisions. The interest rate over time considered in this study is illustrated in Figure 2.

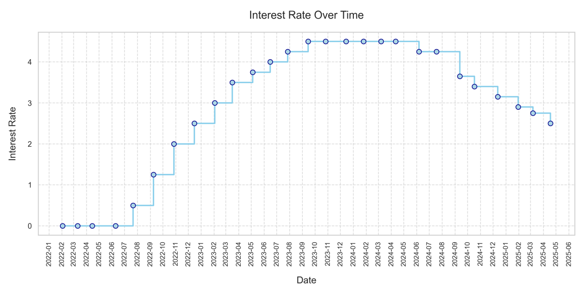

Figure 2: Trends in the ECB's Interest Rates Over Time

3.1 Preprocessing

News articles are preprocessed by removing stopwords[2], which are common but non-informative words. To further prevent overfitting, terms related

[2]NLTK library

Table 1: Example of Feature Table

Title	Body	Date	Δ_S	Δ_M	Δ_L	β_{SI}	β_{SD}	β_{MI}	β_{MD}	β_{LI}	β_{MD}
t1	b1	06-06-22	0	0.5	0.75	0.1	-0.1	0.4	-0.2	0.8	-0.3
t2	b2	07-06-22	0	0.5	0.75	-0.1	0.1	0.6	-0.3	0.7	-0.2
t3	b3	01-06-22	0.5	0.75	0.5	0.45	-0.2	0.6	-0.15	0.55	-0.35
...	...	...	...	...	...	...		...		...	
tN	bN	05-04-24	0	-0.25	0	-0.1	0.1	-0.1	0.35	-0.1	0.1

to days of the week, months, and numerical values are also excluded. This helps to reduce the introduction of overly specific patterns and improves generalization.

4 Baseline Models

To investigate the potential correlation between ECB's interest rate decisions and financial news, and consequently whether news could be used to predict future trends, a feature table has been created. The feature table includes the news title, body of the news, publication date, and the subsequent adjustments made by ECB, specifying the magnitude of each rate hike or cut over short-term (up to 6 weeks), medium-term (6 to 12 weeks), and long-term (12 to 18 weeks) periods. Finally, also a score reflecting the relevance of the news with respect to the ECB adjustments is reported also considering different periods at short-, medium- and long-term.

Table 1 provides an example of the feature table. Δ_S, Δ_M, and Δ_L represent adjustments in the short-, medium-, and long-term, respectively. The coefficient β_{SI}, β_{SD}, β_{MI}, β_{MD}, β_{LI} and β_{LD} refer to the scores for each time period (S stands for short-, M for medium- and L for long-term) focusing on decreasing and increasing adjustment (I stands for increase, while D for decrease). The table lists several news articles in the rows. For instance, in June 2022, we were approaching the first interest hike, while in April 2024, we anticipated a cut.

Various language models have been used to estimate the β score. For detailed information on how these are calculated, please refer to sections 4.1, 4.2, 5.1, 5.2, and 5.3. This score, along with the value of the subsequent decision of interest from ECB, is used to calculate the Pearson correlations. The Pearson correlation was specifically chosen for its effectiveness in measuring the linear relationship between the calculated score coefficients and future changes in interest rates. This comparison allows us to evaluate how well each model aligns with observed data over various time horizons. The subsequent sections will provide the details of the investigated approaches, distinguishing between baselines and proposed models.

4.1 FinBERT

In this study, FinBERT [3] (Picault and Renault, 2017), a pre-trained model for financial sentiment analysis, was used as a baseline. The news articles were divided into sentences [4], and the sentiment score for each sentence has been taken from the probability distribution of the class label given by the FinBERT model. In particular, if the predicted sentiment was positive, then the corresponding probability has been considered as it was, otherwise such a probability was kept with a negative sign. Given a news i, the corresponding score $\beta^i_{(\cdot)}$ is computed as the average sentiment probability across all the sentences. In this specific case, the score $\beta^i_{(SI)} = \beta^i_{(SD)} = \beta^i_{(MI)} = \beta^i_{(MD)} = \beta^i_{(LI)} = \beta^i_{(LD)}$ since the estimation does not depend on the considered time period. Based on scores $\beta^i_{(\cdot)}$ for all the news sources and the upcoming predictions of ECB over three time ranges (Δ_S, Δ_M and Δ_L), the Pearson correlation coefficient was calculated.

4.2 Mixtral

With the advent of Large Language Models (LLMs), we decided to incorporate a LLM as an additional baseline in our study, specifically utilizing Mixtral (Jiang et al., 2024).

The following prompt was used to obtain the scores $\beta^i_{(\cdot)}$, distinguishing between increase and decrease policies for each news:

[3] https://huggingface.co/ProsusAI/finbert

[4] The sent_tokenize function from the NLTK library was used

LLM Prompt

Give a score from 0 to 1 about the prevision that interest rate will be increase or decrease in the short middle or long period.
The results will be formatted like CSV:
Score increase short:[0-1], Score increase middle:[0-1], Score increase long:[0-1], Score decrease short:[0-1], Score decrease middle:[0-1], Score decrease long:[0-1], Motivation: [Max one or two sentences].
Where short is in the next 6 weeks, middle between 6 to 12 weeks and long from 12 to 18 weeks.
From this news of date: [DATE] Title: [TITLE] Text: [TEXT]

To extract the relevant scores from Mixtral's output, we implemented a parsing routine. However, it is important to note that Mixtral does not consistently adhere to the prescribed output format. As a result, we had to employ regular expressions to accurately extract the scores from the output, ensuring that formatting inconsistencies did not affect the retrieval process. As explained in the previous paragraphs, the coefficients $\beta_{(\cdot)}$ extracted from Mistral's output have been compared with the upcoming ECB decisions over three distinct time periods (Δ_S, Δ_M and Δ_L) to calculate Pearson correlations.

5 Proposed Models

5.1 Sentence Transformer with Domain Keywords (ST+DK)

The FinBERT model is specifically designed for sentiment analysis within financial texts. However, rather than capturing the full semantic meaning of the sentences, FinBERT primarily focuses on classifying sentiment, which may result in a limited understanding of the nuanced content conveyed in the text. To overcome this limitation, a Sentence Transformer (Reimers and Gurevych, 2019) has been chosen. This type of model leverages the architecture of BERT (Bidirectional Encoder Representations from Transformers), but it is further optimized for producing high-quality sentence embeddings. These embeddings are crucial for tasks such as semantic similarity, clustering, and information retrieval, as they capture the nuanced meaning of sentences in a dense vector space.

In particular, MiniLM-L6-v2[5] has been selected for its ability to balance efficiency and efficacy. This model is particularly well-suited for practical applications where computational resources and response times are crucial factors. MiniLM-L6-v2 has been used to create the embedding vector of textual spans, i.e. sentences within the collected news and pseudo-sentences composed of domain expert keywords that reflect the concepts related to policy decisions and duration.

[5]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Table 2: Keywords from domain experts to potentially can correlate with policy changes (decrease or increase) and time periods (short, middle, or long)

Policy	Time Period	Keyword
increase	short-middle-long	hike
increase	short-middle-long	increase
increase	short-middle-long	raises
increase	short-middle-long	policy tightening
increase	short-middle-long	rise
decrease	short-middle-long	cut
decrease	short-middle-long	decrease
decrease	short-middle-long	lowers
decrease	short-middle-long	policy easing
decrease	short-middle-long	reduction
increase-decrease	short-middle-long	interest
increase-decrease	short-middle-long	rate
increase-decrease	short-middle-long	rates
increase-decrease	short	short-term
increase-decrease	short	immediate
increase-decrease	short	this month
increase-decrease	short	near-term
increase-decrease	short	start sooner
increase-decrease	short	very soon
increase-decrease	middle	mid-term
increase-decrease	middle	in coming months
increase-decrease	middle	medium-term
increase-decrease	middle	expected soon
increase-decrease	middle	over the next few months
increase-decrease	long	long-term
increase-decrease	long	future
increase-decrease	long	over the next year
increase-decrease	long	not in a hurry
increase-decrease	long	be patient

More specifically, with the support of domain experts (three individuals with 5-years of experience in the financial sector), a set of keywords was identified to describe policies related to increases and decreases, as well as the corresponding time periods (short, medium, and long term). These keywords are listed in Table 2. The keywords are concatenated according to the following rule: first, the policy decision *(e.g., hike, increase, raises)*, followed by the object *(interest rate, rates)*, and finally, the term decisions *(e.g., short-term, immediate, this month)*, to generate a phrase that reflects both the policy decision and the associated duration.

Sentence embeddings were calculated for each sentence in every news article, and the cosine similarity was determined between these embeddings and the embeddings of concatenated keywords related to future interest rate changes over the short, medium, and long term. Since the score was calculated as the average cosine similarity between each news article i and the set of keywords that depends on a given time period and policy, such a score can be derived as β^i_{SI}, β^i_{SD}, β^i_{MI}, β^i_{MD}, β^i_{LI} and β^i_{LD}. Given a time period, each score vector has been used to estimate the Pearson correlation with the corresponding Δ_S, Δ_M and Δ_L to estimate the magnitude of a potential linear relationship.

5.2 Sentence Transformer with Narrow Keywords (ST+NK)

The main issue related to the approach presented in Section 5.1 is that concatenating numerous keywords can lead to repetition and produce embeddings that do not generalize well for the keywords that define specific policies and time period (e.g. increase-short). In order to overcome this limitation, a selection approach was employed, considering all possible combinations of up to three keywords from the initial list. The limit of three keywords was chosen to balance the comprehensiveness of the search with computational time constraints. Subsequently, keywords were selected based on those that achieved a higher Pearson correlation coefficient.

Similarly to what has been done in the previous section, we determined the embeddings of each sentence in every news article, and the cosine similarity between these embeddings and the embeddings of the subset of the obtained keywords has been estimated. The score was determined by averaging the similarities for each news article, which was then used to compute the Pearson correlation with future interest rate changes over the short, medium, and long term. Table 3 presents the selected keywords.

Table 3: Optimal keywords obtained by limiting the composition of the final keywords to contain at most 3 original terms from the original list reported in Table 1.

Policy	Time Period	Keyword
increase	short-middle-long	hike
increase	short-middle-long	increase
increase	short-middle-long	raises
decrease	short-middle-long	cut
decrease	short-middle-long	reduction
increase-decrease	short	immediate
increase-decrease	short	very soon
increase-decrease	middle	mid-term
increase-decrease	long	be patient

5.3 Grid Search of optimal keyword set (GS)

The main limitation of the approaches presented in Section 5.1 and 5.2 is that the keywords employed were neither exhaustive nor fully representative of journalistic financial jargon. In order to overcome this limitation, an iterative algorithm is proposed to identify a list of keywords without relying on any domain expertise.

Given n as the number of keywords and k as the number of combinations, the total number of subsets of k elements is given by the binomial coefficient, which we denote as n_k:

$$n_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

As the parameter k increases, the number of possible combinations grows exponentially, making it impractical for any subsequent activity based on the possible keyword combination. Consequently, it was decided that, for increasing values of k, a decreasing value of n is necessary. In Table 4, the values of k, n and the number n_k of combinations used in this investigation are shown. In particular, for each value of k, the parameter n has been set such that the number of combinations n_k is as close as possible to a given threshold $\alpha = 3000$ (the α value has been arbitrarily selected in order to guarantee a feasible computational time). This allows us on one hand to limit the number of combinations to be explored and on the other hand to evaluate the larger space of possible keyword combinations.

Comb. k	**N° keywords** n	**Keyword comb.** n_k
1	1000	1000
2	77	2926
3	27	2925
4	17	2380
5	14	2002
6	13	1716
7	13	1716
8	13	1287
9	13	715
10	13	286

Table 4: Number of keywords and keyword combinations according to k combinations

In order to quantify the contribution of each individual keyword m belonging to the set of candidate keywords CS (with respect to each time period), we estimate the average Pearson correlation. In particular, given a news article i composed of z sentences and a keyword combination q that belongs to the set of keyword combinations $Q^k_{(\cdot)}$, where $Q^k_{(\cdot)}$ can be Q^k_{SI}, Q^k_{SD}, Q^k_{MI}, Q^k_{MD}, Q^k_{LI} or Q^k_{LD} according to the considered time periods and policies, we followed the subsequent steps:

1. We embed the z sentences and the keyword combination q using the Sentence Transformers.

2. We compute the cosine similarity between each sentence embedding and the keyword combination embedding, and the average of these similarities is what we refer to as the score $\beta^i_{(\cdot)}$ (where $\beta^i_{(\cdot)}$ can be β^i_{SI}, β^i_{SD}, β^i_{MI}, β^i_{MD}, β^i_{LI} or β^i_{LD} according to the considered time periods and policies).

3. The vector $\beta_{(\cdot)}$ of such scores is used coupled with the vector of future interest rate $\Delta_{(\cdot)}$ to calculate the Pearson correlation c_q.

4. We repeat steps 1-3 for all combinations of keywords belonging to the set $Q^k_{(\cdot)}$.

5. Given a keyword m and the set $Q^{km}_{(\cdot)}$ that denotes those combinations in $Q^k_{(\cdot)}$ containing the keyword m, we estimated the average Pearson correlation as follows:

$$c_m = \frac{1}{|Q^{km}_{(\cdot)}|} \sum_{q=1}^{|Q^{km}_{(\cdot)}|} c_q \tag{1}$$

The coefficients $\tilde{c}_m$ will be subsequently used to finally determine the optimal set of keywords.

In particular, the process reported in **Algorithm 1** has been followed. In order to have an initial set of candidate keywords to be considered for the keywords combinations ($Q^k_{(\cdot)}$), we estimated for all the terms of the vocabulary the corresponding TF-IDF value. The initial set CS of candidate keywords is composed of the top-1000 terms with the highest TF-IDF coefficient. This initial selection is particularly important to identify the most relevant terms that encapsulate the core themes of the text. We then initialize a few sets of variables to track the best combination of keywords and their correlation with future interest rates (lines 1-2). When processing the keywords related to a given value of k, we initially create the set $Q^k_{(\cdot)}$ as composed of all the possible combinations of keywords belonging to the candidate set CS (line 4).

Then, for each combination q of keywords belonging to the set $Q^k_{(\cdot)}$, the algorithm encodes q (line 6) and calculates the average cosine similarity $\beta^i_{(\cdot)}$ with the encoded sentences of a given news articles (lines 8-9). The vector $\beta_{(\cdot)}$ composed of similarity coefficients $\beta^i_{(\cdot)}$ is then correlated with the vector of future interest rate $\Delta_{(\cdot)}$ to identify the best performing keyword combination (line 11). If a higher correlation is found, the algorithm updates the best combination and its corresponding correlation value (lines 12-14). Then the average Pearson correlation c_m is estimated and stored in the corresponding vector $\vec{c}_m$, and subsequently used to create a new (reduced) set of candidate keywords (line 20). In particular, the subsequent candidate set CS is created by extracting the n_{k+1} keywords that have the highest $\vec{c}_m$. This approach ensures that the most relevant keywords are prioritized in the subsequent iterations. The list of relevant keywords extracted according to the Grid Search strategy is reported in Table 5.

Algorithm 1 *KeywordsGenerations(CS)*

1: $OptimalKeywordCombination \leftarrow []$
2: $BestCorr \leftarrow 0$
3: **for** $k = 1$ **to** 10 **do**
4: $\quad Q^k_{(\cdot)} \leftarrow Combination(k, CS)$
5: $\quad$ **for** $q \in Q^k_{(\cdot)}$ **do**
6: $\quad\quad h_q \leftarrow SentenceTransformer(q)$
7: $\quad\quad$ **for** $i \in N$ **do**
8: $\quad\quad\quad h_z \leftarrow SentenceTransformer(z)\ \forall z \in Z_i$
9: $\quad\quad\quad \beta^i_{(\cdot)} \leftarrow \frac{1}{|Z_i|} \sum_{z=1}^{|Z_i|} CosineSimilarity(h_q, h_z)$
10: $\quad\quad$ **end for**
11: $\quad\quad c_q \leftarrow PearsonCorrelation(\beta_{(\cdot)}, \Delta_{(\cdot)})$
12: $\quad\quad$ **if** $c_q > BestCorr$ **then**
13: $\quad\quad\quad OptimalKeywordCombination \leftarrow q$
14: $\quad\quad\quad BestCorr \leftarrow c_q$
15: $\quad\quad$ **end if**
16: $\quad$ **end for**
17: $\quad$ **for** $m = 1$ **to** —CS— **do**
18: $\quad\quad c_m \leftarrow \frac{1}{|Q^{km}_{(\cdot)}|} \sum_{q=1}^{|Q^{km}_{(\cdot)}|} c_q$
19: $\quad$ **end for**
20: $\quad CS \leftarrow ExtractTop(n_{k+1}, CS, \vec{c}_m)$
21: **end for**

To explore a broader space, a trivial variant was introduced. If in CS are present keywords for which the cosine similarity between their embeddings and the embeddings of other keywords is greater than 0.8, then only the keyword with the highest c_m is maintained.

Table 5: Keywords extracted according to the GS strategy

Policy	Time Period	Remove similar	Keywords
increase	short	False	higher, scale, strength, weigh, hiking, pointed, soaring, additional, lift, accounts
increase	short	True	hikes, weigh, senior, strength, soaring, boost, pointing, accounts, lift
increase	middle	False	weigh, scale, higher, boost, soaring, pointing, hike, lift
increase	middle	True	weigh, soaring, boost, higher, scale, forwards, additional, lift, pointed, exchange
increase	long	False	gain, forwards, accounts, hiking, account, exchange, imported, scale, deposit
increase	long	True	gain, account, imported, global, turn
decrease	short	False	dropped, part, sharp, close, cutting, make, cut, made, closed
decrease	short	True	cut, part, make, sharp, drop, close, fell, made, sensitive, closed
decrease	middle	False	fell, cuts, cutting, sensitive, part, dropped, closed
decrease	middle	True	cut, part, make, closed, fell, sensitive
decrease	long	False	cuts, cutting, fell, part, fallen, close, sharp, reductions, closed
decrease	long	True	cuts, reductions, part, close, fell, sharp, made

Table 6: Comparison of Pearson Correlation Coefficients between: FinBERT, Mixtral, Sentence Transformer with domain keywords (ST+DK), Sentence Transformer with Narrow Keywords (ST+NK), Grid Search (GS) and Grid Search with similar removal (GS+SR)

Time Period	Policy	FinBERT	Mixtral	ST+DK	ST+NK	GS	GS+SR
Increase	Short	0.2164	0.3514	0.2894	0.3896	0.5709	0.5675
Increase	Middle	0.2804	0.5037	0.3040	0.4041	0.5438	0.5200
Increase	Long	0.2657	0.4487	0.2492	0.3550	0.5363	0.5412
Decrease	Short	-0.2164	-0.3222	-0.3317	-0.4751	-0.6201	-0.6322
Decrease	Middle	-0.2804	-0.4419	-0.4119	-0.5257	-0.5906	-0.6062
Decrease	Long	-0.2657	-0.4087	-0.4258	-0.5633	-0.6510	-0.6350

6 Results

We report in Table 6 the Pearson correlation coefficients for all the approaches presented above, distinguishing between different time periods and policies. In particular, other than FinBert and Mixtral, we refer to Sentence Transformer with Domain Expert Keywords (ST+DK), Sentence Transformer with Narrow Keywords (ST+NK), Grid Search (GS) and Grid Search with Similar Removal (GS+SR). The analysis of the FinBert model reveals a moderately weak correlation, particularly with respect to short-term policies. An important consideration is that the model does not offer immediate interpretability; thus, the use of an explainer is necessary for further clarification.

In contrast, the Sentence Transformer coupled with domain expert keywords produced results that were notably lower for the increase-long time period compared to the FinBert method, yet yielding higher results for the decrease of interest rate. This highlights a preference for the Sentence Transformer approach, because of its ability of capturing the entire sentence meaning.

Turning to the Sentence Transformer with reduced domain expert keywords, the outcomes were quite promising, with the exception of the long-term increase decisions. The Pearson correlation coefficients reported in Table 6 are significantly higher than those observed in the previously discussed methods, suggesting that a sub-optimal selection based on a grid search approach can lead to better capture the relationship with the ECB decisions.

The Pearson correlation coefficients for Mixtral indicate a moderate to strong relationship. However, it is essential to note that Mixtral operates as a large language model (LLM), which typically demands substantial computational resources.

For what concerns the proposed *Grid Search* approach for optimal keyword set, even when considering its variant where similar keywords are removed, we can point out significantly high magnitudes of the Pearson coefficient, suggesting a strong correlation.

It is important to highlight that the proposed approach for generating keywords stands out due to its efficiency and effectiveness; it does not necessitate the application of separate explainability techniques. The keywords extracted automatically, as demonstrated in Table 5, highlight the coherence and reliability of the proposed grid search approach. For instance, in the context of policy increase, the terms *scale, strength, gain* is prominently featured, while in discussions of policy decrease, the word *cuts, fell, drops* appears frequently.

7 Conclusions

In this paper, we introduced a novel technique utilizing sentence transformers to extract keywords with the highest correlation. While it is well-established that the Pearson coefficient is not a measure of causality, the extracted keywords indicate a notable presence of terms related to the rise or fall of interest rates when considering ECB decisions. This suggests a potential linkage, albeit without implying direct causation.

Future work will first focus on evaluating alternative robustness measures beyond the Pearson correlation to ensure the reliability of the findings. Once validated, these insights will be leveraged to develop a predictive model to forecast trends based on real-time news data. As part of this, we could create a benchmark dataset by tagging sentences with labels: 'increase', 'neutral' or 'decrease'. This dataset would be useful for training a classification model and evaluating its performance. Furthermore, applying this algorithm to generic classification datasets could provide a valuable comparison with established benchmarks in classification tasks, thereby enhancing our understanding of its efficacy and versatility.

References

Kerstin Bernoth. 2025. Analyzing ecb communications improves forecasting of interest rate decisions. *DIW Weekly Report*, 15(16/17):100–105.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Nicolas Fanta and Roman Horvath. 2024. Artificial intelligence and central bank communication: the case of the ecb. *Applied Economics Letters*, pages 1–8.

Jens Hilscher, Kyle Nabors, and Alon Raviv. 2024. Information in central bank sentiment: An analysis of fed and ecb communication. *Available at SSRN 4797935*.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Jean Lee, Hoyoul Luis Youn, Nicholas Stevens, Josiah Poon, and Soyeon Caren Han. 2021. Fednlp: an interpretable nlp system to decode federal reserve communications. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2560–2564.

Matteo Leombroni, Andrea Vedolin, Gyuri Venter, and Paul Whelan. 2021. Central bank communication and the yield curve. *Journal of Financial Economics*, 141(3):860–880.

Mihai Niţoi, Maria-Miruna Pochea, and Ştefan Constantin Radu. 2023. Unveiling the sentiment behind central bank narratives: A novel deep learning index. *Journal of Behavioral and Experimental Finance*, 38:100809.

Moritz Pfeifer and Vincent P. Marohl. 2023. Centralbankroberta: A fine-tuned large language model for central bank communications. *The Journal of Finance and Data Science*, 9:100114.

Matthieu Picault and Thomas Renault. 2017. Words are not all created equal: A new measure of ecb communication. *Journal of International Money and Finance*, 79:136–156.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.