Comparative Analysis of Macroeconomic Variables and Hungarian News Sentiment using Small-scale Large Language Models

Lívia Réka Ónozó[1,2], Frigyes Viktor Arthur[2], Bálint Gyires-Tóth[2]

[1] Central Bank of Hungary, Digitalization Technology Department
Krisztina krt. 55, 1013 Budapest, Hungary
E-mail: onozol@mnb.hu

[2] Budapest University of Technology and Economics,
Department of Telecommunications and Artificial Intelligence
Műegyetem rkp. 3, 1111 Budapest, Hungary
Email: {arthur,toth.b}@tmit.bme.hu

Abstract: We present a large language model-based method for aligning Hungarian macroeconomic indicators with the sentiment of business news. Our goal is to offer low-latency and detailed analyses of economic trends in an environment where high-frequency data are not available. By bridging the gap between human and artificial cognitive mechanisms, this work advances cognitive infocommunications. To train a deep learning-based sentiment classifier, we manually annotated a relatively small number of business-related sentences and used FastText- and transformer-based approaches (including small- and medium-scale BERT and GPT models). We then predicted the sentiment of economic news spanning over fifteen years, aggregated these sentiments into higher timeframes, and compared them with macroeconomic indicators such as GDP and PMI. Our results show that the most advanced classifier, coupled with our sentiment aggregation method, correlates strongly with the macroeconomic indicators. This finding underscores the potential of advanced language models for analyzing and understanding economic trends.

Keywords: GDP; PMI; NLP; sentiment classification

1 Introduction

Measuring public opinion on various aspects of the economy – such as specific markets, sectors, or the overall economic landscape – is critical for understanding market expectations and business confidence. Sentiment indices serve as valuable tools for analysts, investors, and policymakers alike. Traditionally, confidence and sentiment indices have been derived from surveys or other data collection methods.

However, alternative approaches – such as analyzing textual data from news articles and social media – have recently emerged as promising tools for short-term macroeconomic forecasting and approximating current economic activity. The implications of sentiment indices are significant: a positive index suggests optimism about the economy, potentially leading to increased business confidence, while a negative index may indicate pessimism among economic actors. By providing insights into prevailing economic sentiment, these indices serve as essential instruments for making informed business decisions in an ever-changing economic landscape.

Deep learning has become the primary technology in Natural Language Processing (NLP), achieving state-of-the-art results in various tasks, including sentiment analysis, machine translation, and question answering. In particular, sentiment analysis holds significant potential in finance and economics, as it can approximate market trends, public opinion, and the overall economic climate. Deep learning-based language models are capable of capturing complex linguistic patterns and can effectively learn from large amounts of data, making them well-suited for large-scale analyses of news articles. Furthermore, the data-driven methods can adapt to the evolving nature of the economy. However, one of the challenges in economic news sentiment prediction is the presence of domain-specific jargon and frequent ambiguity in news articles, which can lead to misclassification of sentiment. Additionally, the impact of economic news on macroeconomic indicators may differ across countries and time periods, necessitating the development of adaptable models.

CogInfoCom [1], an interdisciplinary field that merges cognitive sciences with infocommunications, plays a pivotal role in our research from two fundamental perspectives. First, the cognitive processes that underpin human interpretation are modeled by using advanced deep learning-based Natural Language Processing (NLP) techniques. These processes include the perception of sentiment and emotion, the formation of opinions, and the shaping of beliefs. By using deep neural networks, our goal is to replicate human cognitive functions and gain profound insights into the mechanisms driving sentiment in economic news. Second, our approach enables us to explore the relationship between the sentiments expressed in economic news and macroeconomic variables. By integrating the cognitive and emotional facets of human decision making, we aim to develop more effective decision-support systems that account for the intricate connections between sentiment and economic outcomes.

Our research is dedicated to exploring inter-cognitive communication, which refers to the exchange of information between humans and artificial cognitive systems. By leveraging advanced deep learning methods, we aim to narrow the cognitive gap between human capabilities and those of artificial systems, facilitating seamless collaboration. Moreover, our approach involves bridging representations, acknowledging that the sensory data perceived by humans and artificial systems differ significantly. To address this challenge, we implement a transformation

process that adapts and filters information, ensuring effective communication and cooperation between both entities.

Our research has implications for various stakeholders, including policymakers, businesses, and investors. By providing a more nuanced understanding of how sentiment in economic news influences decision making processes, our work can help these stakeholders make more informed choices and develop strategies that are better aligned with the prevailing economic sentiment. This can lead to more effective policy interventions, improved risk management, and enhanced investment decisions.

This paper is an extension of our previous work [20]. Compared to our previous work, this study introduces more advanced models and assesses their performance in sentiment analysis together with their alignment with macroeconomic indicators. Altogether, our new results show superior performance to the prior work.

2 Related Work

Research in the field of textual data analysis and its impact on economic activities has been extensive. Studies have highlighted the critical role of textual information in influencing market volatility and uncertainty. For example, an analysis [2] of front-page articles from The Wall Street Journal revealed that periods of economic downturn are associated with increases in news-related implied volatility. Additionally, textual indicators have been recognized as potential precursors to financial crises [3], indicating their potential as early signals of financial system vulnerabilities. Further, it has been established that sentiments derived from news content can profoundly affect the valuation of assets, especially in recessionary times [4]. This underlines the cost-effectiveness and extensive application of text analytics within the financial sector, leading to the prevalent use of sentiment indices [5]. The primary methodologies employed in text analytics include dictionary-based and machine-learning approaches.

The emergence of the word2vec model [6] represented a significant advancement in the field of natural language processing (NLP), changing the paradigm by allowing words to be represented as dense vectors within a multidimensional space. Such representations, including the skip-gram and continuous bag-of-words models, are adept at encapsulating semantic and syntactic information. FastText [7], an evolution of word2vec, specializes in learning subword unit representations, thereby efficiently capturing morphological nuances and addressing out-of-vocabulary words.

The introduction of transformer-based architectures [8] marked a further pivotal development in NLP, with the innovation of the multi-head self-attention mechanism enabling the effective capture of contextual information and long-range

dependencies. The foundational paper on transformers notably highlighted their impressive training speeds. Transformer models exhibit scalability relative to the size of the dataset and the model's parameters, as outlined in [9]. BERT [10], leveraging the transformer framework, introduced an innovative pre-training method that combines masked language modeling with next-sentence prediction, facilitating the learning of deep, bidirectional context for each word within a sentence. Having been trained on extensive datasets, BERT achieved unparalleled performance across a variety of NLP benchmarks. Subsequent models like GPT [11], including its iterations GPT-2 [12], GPT-3, and GPT-4, have been distinguished by their autoregressive language modeling prowess, supported by training on diverse and substantial datasets, thus heralding the era of large language models (LLMs) with human-comparable performance across numerous NLP tasks.

Transformer architectures are particularly effective in discerning the subtle sentiment nuances embedded within economic news, offering insights into their potential effects on macroeconomic indicators. Their adeptness at contextual understanding is invaluable in the analysis of financial texts, where the sentiment attributed to a specific term can vary significantly based on its context. This capability is crucial for navigating the inherent ambiguity of language in economic discourse.

In the evolving landscape of natural language processing (NLP) and its applications to economic analytics, specific advancements like Claude and Mixtral have pushed the boundaries further, offering nuanced insights into financial text analysis. Claude, a multifaceted foundational AI developed by Anthropic, excels in advanced reasoning, vision analysis, code generation, and multilingual processing. Its unique capacity for complex cognitive tasks allows it to interpret and synthesize vast arrays of economic data and textual information, making it an invaluable tool for economic forecasting and analysis.

Recent innovations, including models like Mixtral, extend the capabilities of NLP further into the economic sphere. Mixtral, with its Mixtral-8x7B [19] configuration, represents a leap forward with its sparse mixture of experts model, delivering exceptional performance and efficiency. Notably, Mixtral surpasses comparable models in benchmark tests while offering significantly faster inference speeds, making it an ideal choice for processing and analyzing economic texts with complex contextual dependencies and nuances.

Despite these technological strides, challenges persist with the deployment of LLMs, including substantial computational demands, environmental considerations, and the need for interpretability. The "black-box" nature of such models complicates their application in scenarios requiring transparency and clear explanation of AI-driven predictions, highlighting areas for ongoing research and development in the field of AI and NLP.

In this study, we present transformer-based medium resource-intensive solutions, together with a low-resource model based on FastText CBOW embeddings and

classifier multilayer perceptron network. We leveraged pre-trained “smaller-scale” language models (huBERT, PULI GPTrio) for classification of economic sentiment, a task that put on a significant challenge even for experts in finance or economics. With the sheer volume of news articles' sentences, efficiency remains a critical aspect of our research. Our computational workload was carried out on two A100 GPUs, each with 80 GB of GPU RAM.

3 Dataset

We employ a dataset of financial news articles collected from two prominent Hungarian news platforms. Each article includes the publication date, URL, and textual content, which we use for sentiment prediction. It should be noted that the news data cover the period of 2000-2020, however, due to the alignment with macroeconomic variables, the beginning of the data series has been truncated to articles beginning in 2005.

3.1 Target Variables

This research investigates the implications of two principal economic indicators, Gross Domestic Product (GDP) and Purchasing Managers' Index (PMI), as target variables. GDP represents a comprehensive quantification of all goods and services produced domestically over a specific timeframe. It is a pivotal gauge of a nation's economic vigor and gives an extensive overview of its economic activities. Given its quarterly publication, GDP data often provide retrospective rather than instantaneous insights, reflecting past rather than present economic conditions. On the other hand, the PMI, calculated from monthly surveys conducted among private sector firms, responds more swiftly to immediate economic fluctuations. PMI can presage economic trends as a leading economic indicator, offering precursory signs of growth or decline. While GDP and PMI share a correlation, delineating this relationship reveals a complex interplay influenced by numerous economic dynamics without a straightforward cause-and-effect linkage. Notably, PMI primarily mirrors the performance of the manufacturing and service industries and may only encompass some of the economic landscape.

The hypothesis posited by this study is that the sentiments extracted from the news articles provide a timely snapshot of the economic climate, potentially enabling predictions about PMI changes. This model, in turn, might allow for the extrapolation of GDP trends based on PMI data.

3.2 Two-Decade Hungarian Economic News Dataset

Owing to the lack of publicly available archives for Hungarian news articles, we developed a custom database while strictly adhering to legal data acquisition standards. We selected two prominent Hungarian news outlets that reach over 60% of the country's online audience and are renowned for their long-standing reputation. Confidential agreements with these outlets prevent us from revealing their identities. We collected data through web scraping without an API, ensuring rigorous compliance with legal requirements. Content that was not relevant to our research was excluded. The web scraping process began in January 2021, and the data distribution is detailed in Table 1 and illustrated in Figure 1.

3.3 Economic Sentiment Dataset

A subset of articles from the dataset was selected to create a sentiment analysis dataset. We used this particular subset of data to train, validate, and test our sentiment analysis models. The selection was randomized using a pseudorandom generator to ensure uniform distribution. Following selection, articles underwent preprocessing to segment text into sentences, which researchers manually annotated for sentiment, assigning positive, neutral, or negative labels based on perceived sentiment. The final sentiment dataset comprises 414 negative, 342 positive, and 675 neutral sentences. The first article that was used

Table 1
Descriptives of the text corpora between the period 2005-01-01 and 2020-12-31

	Number of articles	**Average number of articles by month**
Medium 1 (grey)	232,837	807
Medium 2 (blue)	323,156	1,041
Total	555,993	1448

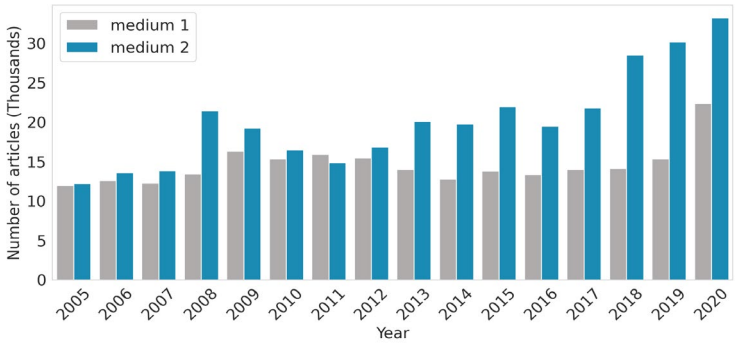


Figure 1
Number of economic news articles by two Hungarian sites between the years of 2005-2020

4 Methods

Based on the sentiments expressed in the economic news articles, we aim to construct a time series. To accomplish this goal, we utilized a multi-layer perceptron and three transformer-based approaches, both of which will be discussed in greater detail below.

4.1 Sentiment Classifier

This study presents advanced deep learning methodologies for sentiment classification tasks. It encompasses a multi-layer perceptron (MLP) classifier that employs FastText embeddings, as well as finetuned transformer models, including a large language model containing 7 billion parameters. All language models underwent pre-training, and the research involved an extensive series of experiments using various configurations, including different pre-trained models, embedding sizes, and training methodologies.

4.1.1 Multi-Layer Perceptron Classifier Utilizing FastText

FastText [7] uses 300-dimensional embedding vectors and incorporates subword information. This embedding approach leverages 5-character long n-grams and a window size of 5 to capture contextual information. In our methodology, FastText embeddings were generated for each word in a sentence, and the mean of these embeddings was used as input to a Multi-Layer Perceptron (MLP), with the corresponding sentiment label serving as the output. We employed a two-stage training process for the MLP model: initially, the model was trained on an external user reviews database [13] for up to 1000 epochs with a batch size of 32 and a dropout rate of 0.6. In the second stage, the model was further trained on the sentiment dataset described in Subsection 3.3. Cross-entropy was used as the loss metric. In addition to dropout, early stopping regularization was applied in both stages, with a patience of 50 and a minimum change in the monitored improvement of 0.001. The MLP was implemented with three hidden layers, containing 128, 64, and 16 units respectively. Hyperparameter optimization was performed manually.

4.1.2 Transformer-based Sentiment Prediction

In our research, we modified an existing model, PULI GPTrio (developed by Yang et al. [14] and [15] in 2021 and [16] in 2023), and used advancements in transformer architecture to perform sentiment classification on finance and business-related content, as introduced in the Dataset section. A substantial amount of Hungarian text data was employed to train the selected pre-existing models. Initially, [13] was designed for named entity recognition. We propose adapting this transformer model for sentiment classification of economic news, based on the assumption that it is sufficiently versatile for this task. [14] is trained specifically for sentiment classification tasks, while [15] is a trilingual model encompassing Hungarian,

English, and Chinese languages, with 7.67 billion parameters [16]. This model was chosen based on the idea that exposing a language model to diverse linguistic contexts, such as English phrases, can enhance its robustness and adaptability. Subsequently, our dataset was used to finetune the pre-trained model, which was meticulously preprocessed to meet the requirements of the three transformer models. In each instance, to customize the model for our specific needs, the final classification layer was modified. This new layer categorizes sentiments into three classes: negative, neutral, and positive. A dropout layer was added before the classification layer to prevent overfitting and promote generalisation.

Particular attention was given to tokenization, using the tokenizer provided by the transformer library to efficiently handle this crucial task. In languages like Hungarian, tokenization is especially important due to its complex morphological structure and extensive use of compound words. Effective tokenization is vital for accurately capturing these linguistic nuances [16]. We employed the tokenizer associated with the transformer-based models for this purpose. Following industry standards, the Hugging Face transformer library was used for tokenizing transformer-based models. These tokenizers were selected because they are optimized to work seamlessly with the huBERT and PULI GPTrio models, ensuring that the input data is properly formatted and compatible with the models' architecture. Additionally, the tokenizer ensures uniform sequence lengths for efficient batch processing by padding and truncating sequences according to the model's maximum input lengths.

PULI GPTrio is built on GPT-Neox and was developed for training large-scale language models. To fit this model within our GPU RAM constraints, we employed low-rank adaptation using the LoRA configuration during training. The rank of the trainable matrices was set to 64, and a dropout probability of 10% was applied to the LoRA layers, similar to the settings in the original QLoRA paper [18]. By using 4-bit quantization, we further enhanced efficiency. A learning rate scheduler was implemented to dynamically adjust the model's learning rate, ensuring an effective training process. Our methodology also incorporates an early stopping mechanism to prevent unnecessary computational costs and overfitting by halting training once the model's performance on the validation set plateaued. The validation data comprised 15% of the training data (1,287 sentences), and the test set included 144 sentences. Both the MLP and Transformer models were evaluated using the same test set.

We identified the optimal values for parameters such as batch size, learning rate, early stopping patience, weight decay, and dropout probability through hyperparameter optimization. The parameter space was effectively explored using various distributions and iterative refinement. Implementing the best-performing model led to significant improvements in sentiment classification accuracy. After hyperparameter optimization, the transformer-based sentiment models achieved their best performance within 100 epochs, using batch sizes of 32 and learning rates of 1e-5. To regularize the model, a dropout rate of 0.6 was applied along with a

weight decay rate of 0.01. Overfitting was prevented through early stopping with a patience of 5.

4.2 Sentiment Prediction for Two Decades

We used the economic news articles and sentiment database provided by the Central Bank of Hungary (MNB), see Subsection 3.2. The comprehensive database consists of 3 decades of financial and economic online news, collected and annotated by the experts of the institution. The task of analyzing this vast data presented both a challenge as well as an opportunity. Economic insights at the sentence level made it possible to detect dynamics in the Hungarian macroeconomic indicators.

For the Multi-Layer Perceptron framework, we standardize, tokenize and batch each sentence in the corpus. For the three transformer models, preprocessing includes splitting into sentences, and then using the corresponding tokenizers.

The final layer of the models transforms the confidence values of each sentence into positive, negative, or neutral sentiment values. The scores are then weighted with their corresponding signs and summed into monthly frequency. As a result of this approach, we are able to measure the impact of the number of positive or negative articles in a given period of time. After scaling the obtained values, a sentiment index can be constructed and compared with target variables, such as the year-over-year growth in the Gross Domestic Product (GDP).[1]

This rigorous approach enabled us to create a quantitative sentiment index using MNB's extensive text corpus. This type of index illustrates how large-scale text analysis in economics is able to provide information about the trajectory of the economy over the past two decades.

5 Result

Below, we summarize in this section the conclusion of the analysis of the sentiment classifiers and the confusion matrices and corresponding evaluation matrices on the economic news sentence that served as a test database. Also, inference results will be presented for the unlabeled sentences and aggregated into time series.

5.1 Sentiment Classifier

We have found definitive evidence that transformer architectures outperform baseline multilayer perceptron (MLP) models using FastText embeddings for sentiment classification tasks. Different evaluation metrics (accuracy, precision, recall) show a strong performance of the applied transformer-based models.

[1] It must be noted that GDP is published quarterly but is estimated by economists of the Central Bank of Hungary to a higher frequency.

A notable disparity is observed between the transformer models and the MLP approach (Figure 3) regarding prediction accuracy and recall metrics. Specifically, the MLP model achieved recall rates of 9.1% for negative, 66.7% for neutral, and 27.03% for positive sentiments. In contrast, the transformer-based named entity recognition (NER) model attained significantly higher recall rates of 69.84% for negative and 54.55% for positive sentiments. Additionally, transformers reached a 70.27% recall rate in the neutral category, slightly surpassing the MLP approach. The superior performance of the transformer-based model is particularly impressive considering the limited resources and the linguistic complexity of the Hungarian language.

Compared to huBERT named entity recognition (trained on the NerKor corpus [21]), huBERT sentiment (trained on the HTS5 corpus [22]) model achieved similar accuracy overall, but with a large improvement in the number of positive sentences incorrectly predicted as negative. The HTS5 model was pre-trained specifically for sentiment classification tasks, but not on financial or economic corpus, and achieved a weighted average precision of 64.8% after being finetuned on our labeled dataset. Macro accuracy (considering all three categories) is 63.9% (see Table 2). This model has the advantage of significantly reducing the proportion of "large errors", when the model is changed from negative to positive or from positive to negative. In fact, this rate decreased from 38.6% to 13.5% with the HTS5 compared to its NerKor counterpart, where the ground truth was positive, but the model predicted negative sentiment. In the neutral category, the HTS5 model was slightly underperforming. The nature of economic articles tends to favor neutral emotions, which is why we attempted to track this proportion with our test database. Consequently, our model could already achieve 43.75% accuracy if everything were categorized as neutral.

Introducing the PULI GPTrio model was a remarkable accomplishment due to its ability to improve the overall accuracy while maintaining neutral dominance across all categories. Compared to the previous approach the PULI GPTrio transformer model (Figure 3) showed significant improvement in accuracy and recall measurements. This model achieved 63.63% in ‘negative’, 80.95% in ‘neutral’ and 59.46% in ‘positive’ recall rates. In terms of precision it means 73.68, 68 and 70.96 percent respectively. The overall accuracy scored 70.14%. In various natural language processing tasks, transformer architectures have been increasingly recognized as effective.

5.2 Sentiment Prediction for Two Decades

Gross Domestic Product and Purchasing Managers’ Index are compared in the present section with the predicted sentiment path. We visualized the inference results of sentiment predictions (see Figure 2) and calculated the mean absolute error (MAE) of the different models in use (see Table 2). Several key observations can be drawn from the experiments:

- There is a wide distribution of sentiment scores in the **MLP model**, with the average being 3.48. Generally, this model is not quite well-suited for predicting macroeconomic variables' dynamics.
- In analyzing the predictions of the **huBERT NerKor transformer model**, we find that the average sentiment is negative with a score of -5.661. This model is sensitive to economic downturns, such as the financial crisis of 2008 and the COVID pandemic in 2020.
- It appears that the transformer model's predictions are more aligned with the real-world indicators such as GDP change to the similar period of the previous year (year-on-year, yoy) and PMI indices, supporting its use in predicting economic downturns.
- All three transformer models were able to foresee the 2008 economic crisis in terms of the overall activity of the economy (that is the change in GDP compared to the same period of the previous year) within three to four months.
- Among transformer-based models PULI GPTrio was the most capable in capturing the dynamics in both GDP and PMI indices.
- It is expected that as the complexity of the models increases, the inference predictions become more accurate, i.e. the average absolute error will decrease.
- Economic indices are closely aligned with the prediction of the Transformer models, compared to Multi-Layer Perceptron.

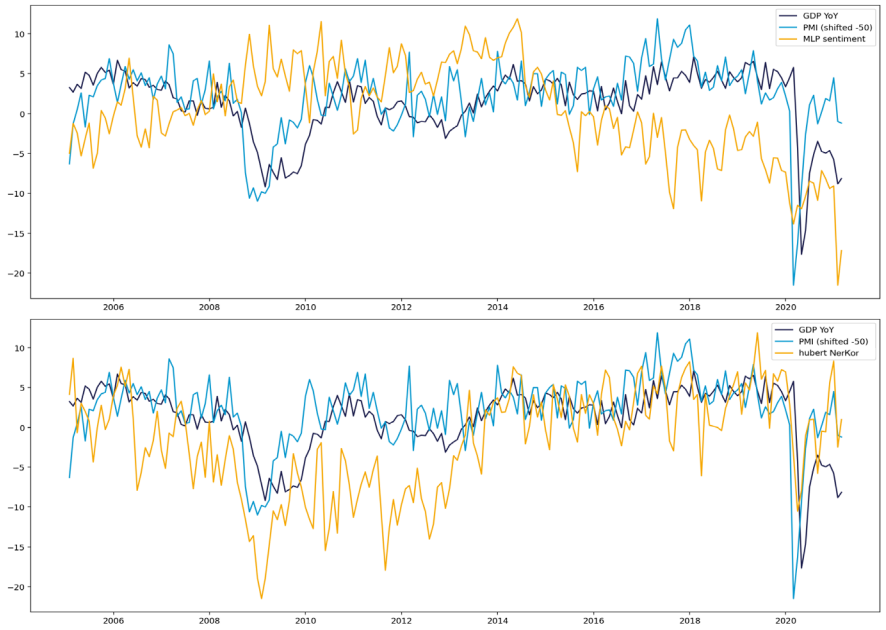
Table 2

Accuracy of the multilayer perceptron and transformer-based models (PULI GPTrio, huBERT hts5 and nerkor) on the test dataset and the mean absolute error (MAE) on Gross Domestic Product and Purchasing Managers' Index

Method	**Macro Accuracy**	**Precision**	**Recall**	**F1-scores**	**MAE GDP**	**MAE PMI**
PULI GPTrio	0.7014	0.7008	0.7019	0.7012	2.701	3.216
HuBERT HTS5	0.639	0.635	0.651	0.638	3.545	3.816
HuBERT NerKor	0.65	0.65	0.65	0.65	3.893	4.980
MLP	0.338	0.34	0.34	0.32	8.520	7.705

6 Discussion

The results of our research demonstrate that the transformer-based models outperformed the FastText with MLP approach in capturing the sentiment of economic news. We have expanded our previous work [x20] by including two additional transformer-based models: the huBERT sentiment classifier (trained on HTS5 dataset) with 110 million parameters and the much larger PULI GPTrio with 7.67 billion parameters. The huBERT model (pre-trained to sentiment classification, and finetuned to our business-related new sentiment dataset) showed more reliable results, by achieving almost the same accuracy, while significantly decreasing the classification errors between positive and negative cases, compared to the previous Named Entity Recognition (NER)-based model. The updated huBERT model reduced the number of errors in misclassifying positive sentiment as negative. As expected, the PULI GPTrio, with its 7.67 billion parameters, achieved the best performance among all the models tested. The inclusion of the large PULI GPTrio model significantly increased the time required for inference when generating time series predictions. To address the challenges posed by the 7 billion parameter model, we used the Low Rank Adaptation (LoRA) method. LoRA is a technique that reduces the number of trainable parameters while maintaining the model's performance.



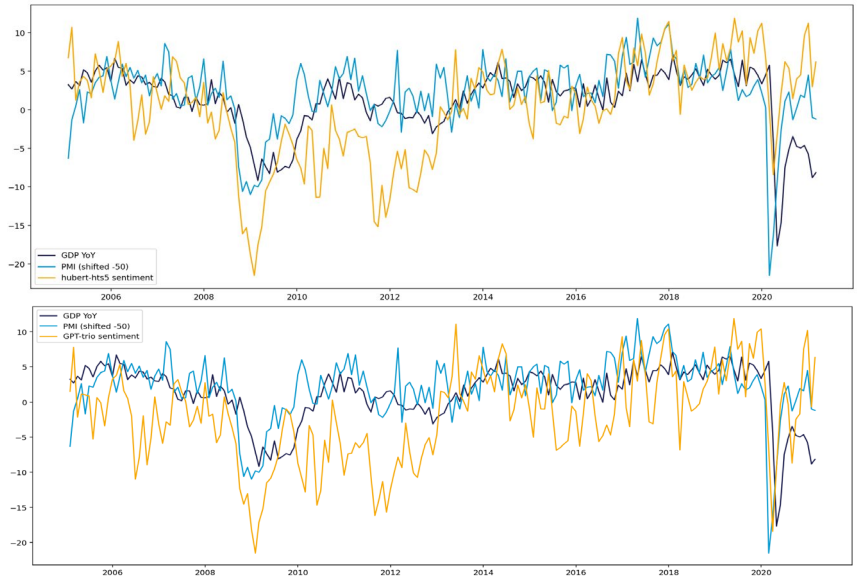


Figure 2

Comparison of Sentiment time series of Transformer and MLP predictions of economic news indicators. Visualization presents our four distinct machine learning approaches – Multilayer Perceptron, huBERT NerKor Named Entity-Recognition, huBERT HTS5 sentiment, and PULI GPTrio – plotting sentiment predictions against GDP year-on-year change and PMI index.

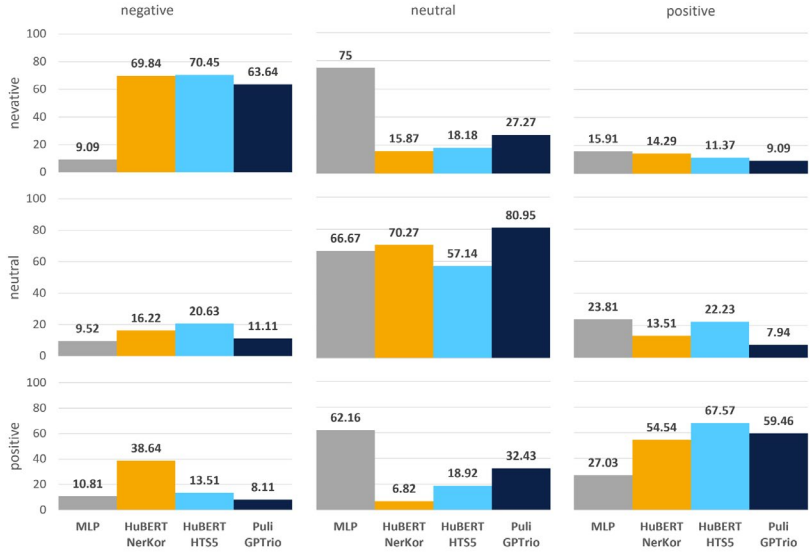


Figure 3

Confusion matrix for the trained MLP, HuBERT Named Entity Recognition (NerKor), HTS5 and Puli GPTrio transformer networks on the test dataset

The research presented in this paper reveals that sentiment analysis can offer meaningful insights into short-term economic fluctuations, emerging trends, and the dynamics of economic variables. However, it's important to recognize the limitations of these predictions, particularly regarding data timeliness. News articles are often published after events have transpired, which may make the ability to assess market sentiment in real-time impossible. Analysing macroeconomic conditions using low-frequency data – such as Gross Domestic Product (GDP) and Purchasing Managers' Index (PMI) figures that are released with a one- to two-month lag – is inherently challenging. Consequently, market participants rely on high-frequency data sources like time series analyses and monthly surveys to capture and interpret underlying economic processes. Yet, survey-based business indicators face their own obstacles, including delays in data collection and processing and the potential for selection bias. The results of these studies may provide useful tools for central bankers and macroeconomic analysts, which may enhance their forecasting capabilities.

Conclusions

This paper introduced a new approach to nowcasting macroeconomic indicators by applying deep learning-based sentiment analysis. Initially, we developed an economic sentiment classifier using various techniques, including multilayer perceptron-based and transformer models. We then used these classifiers on a substantial corpus of economic news articles sourced from two major Hungarian websites. The aggregated sentiment predictions were compared with the macroeconomic indices under examination. Our findings indicate that transformer-based architectures are particularly effective at modeling sentiments expressed in economic news articles, especially concerning GDP and PMI indicators. These results could assist economic analysts in evaluating the current state of the economy when high-frequency or real-time data is unavailable. Furthermore, the methodology presented in this research has the potential to serve as a decision-support tool for market forecasters, central bank officials, and policymakers.

Acknowledgement

The work reported in this paper carried out as a collaboration between the Central Bank of Hungary and BME. Project no. TKP2021-NVA-02 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme. We gratefully acknowledge the support of NVIDIA with the donation of the A100 GPU used for this research.

References

[1] P. Baranyi, A. Csapo, and G. Sallai, *Cognitive infocommunications (coginfocom)* Springer, 2015

[2] A. Manela and A. Moreira, “News implied volatility and disaster concerns,” *Journal of Financial Economics*, Vol. 123, No. 1, pp. 137-162, Jan. 2017

[3] R. Nyman, S. Kapadia, and D. Tuckett, "News and narratives in financial systems: Exploiting big data for systemic risk assessment," *Journal of Economic Dynamics and Control*, Vol. 127, p. 104119, Jun. 2021

[4] D. Garcia, "Sentiment during recessions," *The Journal of Finance*, Vol. 68, No. 3, pp. 1267-1300, May 2013

[5] C. Kearney and S. Liu, "Textual sentiment in finance: A survey of methods and models," *International Review of Financial Analysis*, Vol. 33, pp. 171-185, May 2014

[6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'13. Red Hook, NY, USA: Curran Associates Inc., 2013, pp. 3111-3119

[7] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135-146, Dec. 2017

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017

[9] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess

[10] R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," 2020, arXiv preprint arXiv:2001.08361

[11] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy

[12] M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019, arXiv preprint arXiv:1907.11692, 364

[13] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI*, 2018 [Online] Available: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

[14] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI, 2019 [Online] Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[15] M. Miháltz, "Opinhubank," in *Proceedings of Hungarian Conference on Computer Linguistics*. University of Szeged, 2013, pp. 343-345

[16] Z. G. Yang and T. Váradi, "Training language models with low resources: Roberta, bart and electra experimental models for hungarian," in *Proceedings of 12th IEEE International Conference on Cognitive Infocommunications*. IEEE, 2021, pp. 279-285

[17] Z. G. Yang and L. J. Laki, “Improving Performance of Sentence-level Sentiment Analysis with Data Augmentation Methods,” in 12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021), 2021, pp. 417-422

[18] Z. G. Yang and L. J. Laki and T. Váradi and G. Prószéky, “Mono- and Multilingual GPT-3 Models for Hungarian” 2023, 10.1007/978-3-031-40498-6_9

[19] A. Kornai, P. Halácsy, V. Nagy, C. Oravecz, V. Trón, and D. Varga, “Web-based frequency dictionaries for medium density languages,” in *Proceedings of the 2nd International Workshop on Web as Corpus*, 2006 [Online] Available: https://aclanthology.org/W06-1701.pdf

[20] T. Dettmers, A. Pagnoni, A. Holtzman and L. Zettlemoyer (2024) QLoRA: Efficient Finetuning of Quantized LLMs. Advances in Neural Information Processing Systems 36

[21] Jiang, Albert Q., et al. "Mixtral of experts." arXiv preprint arXiv:2401.04088 (2024)

[22] Arthur, F. V., Gyires-Tóth, B., Debreczeni, M. I., & Ónozó, L. R. (2023 September) Language of the Market: NLP-Driven Sentiment Analysis of Hungarian Economy. In 2023 14th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 93-98)

[23] Simon, Eszter; Vadász, Noémi. (2021) Introducing NYTK-NerKor, A Gold Standard Hungarian Named Entity Annotated Corpus. In: Ekštein K., Pártl F., Konopík M. (eds) Text, Speech, and Dialogue. TSD 2021, Lecture Notes in Computer Science, Vol. 12848, Springer, Cham.