Journal of Computer Science and Cybernetics, V.42, N.2 (2026), 167-181
DOI: 10.15625/1813-9663/22428

FINE-TUNED LANGUAGE MODELS FOR INTELLIGENCE AGGREGATION SYSTEMS IN THE BANKING SECTOR

ANH PHAN VIET[1,*], HIEU PHI MINH[2], HIEU NGUYEN QUOC[3], ANH NGUYEN DUC[2], LONG TRIEU HAI[2]

[1] *Le Quy Don Technical University,*
236 Hoang Quoc Viet Street, Nghia Do Ward, Ha Noi, Viet Nam
[2] *VNU University of Engineering and Technology*
144 Xuan Thuy Street, Cau Giay Ward, Ha Noi, Viet Nam
[3] *Viettel Post Joint Stock Coporation*
Duy Tan Street, Dich Vong Hau Ward, Ha Noi, Viet Nam

Abstract. This paper presents the design and implementation of a practical news aggregation system for decision making and risk management in the banking sector. The system gathers financial news from multiple sources and mines texts to provide insight into banks' professional activities such as policies, products, and financial performance. Multiple natural language processing (NLP) modules are integrated to address tasks such as topic classification and sentiment analysis. To identify the most suitable technologies for each component, we systematically evaluated a wide range of NLP techniques, including large language models (LLMs) and domain-specific pre-trained models. A Vietnamese financial corpus of 12,000 annotated samples was constructed to fine-tune models such as PhoBERT, ViT5, and BARTPho. Experimental results show that fine-tuned models significantly outperform general-purpose LLMs (e.g., LLaMA-3.1-8B, Vistral-7B) in both accuracy and computational efficiency. The fine-tuned models achieve a 7.15% accuracy improvement and reduce resource requirements. The study demonstrates a scalable and adaptable framework for building multi-source, text-based intelligent systems in the financial domain.

Keywords. News aggregation system, sentiment analysis, financial text, large language models, fine-tuned models.

1. INTRODUCTION

News aggregation systems play a significant role in enhancing banking system management by providing real-time insights and actionable intelligence [1]. These systems are typically composed of several components working together to extract meaningful knowledge from raw data. Key tasks include data crawling, scraping, storage, knowledge mining, and user interaction. By analyzing data from various sources, such as social media and financial reports, these systems assist banks and governments in enhancing decision-making, risk management, and policy enforcement. For instance, during a sudden market downturn, banks

*Corresponding author.
E-mail addresses: anhpv@lqdtu.edu.vn (A.P. Viet); 21020200@vnu.edu.vn (H.P. Minh); hieunguyenquoc1999@gmail.com (H.N. Quoc); 22022504@vnu.edu.vn (A.N. Duc); thlong@vnu.edu.vn (L.T. Hai).

...
Đóng góp cho thanh khoản cổ phiếu vua dồi dào trong năm qua phải kể đến sự bùng nổ của nhóm nhà băng nhỏ như PGB (gấp 11.9 lần), SGB (gấp 3.8 lần), NAB (gấp 3.1 lần), VBB (gấp gần 3 lần), EIB (gấp 2.3 lần), MSB (gấp hơn 2 lần).
Đáng buồn là vẫn có một số ngân hàng thu hẹp thanh khoản trong năm như BAB (giảm 89%), BVB (giảm 61%), KLB (giảm 51%), BID (giảm 41%)...
...

...
The significant contribution to the abundant liquidity of "blue-chip" stocks over the past year can be attributed to the boom of smaller banks such as PGB (up 11.9 times), SGB (up 3.8 times), NAB (up 3.1 times), VBB (up nearly 3 times), EIB (up 2.3 times), and MSB (up more than 2 times).
Unfortunately, some banks have experienced a contraction in liquidity during the year, including BAB (down 89%), BVB (down 61%), KLB (down 51%), and BID (down 41%),...
...

Figure 1: A piece of financial report (original Vietnamese text and English translation). The texts related to the indicator, positive and negative polarities are highlighted in orange, blue, and red, respectively.

can utilize provided information to adjust their investment portfolios or lending practices to reduce risks [2]. Similarly, governments can leverage these systems to track the performance of financial institutions and industry trends, enabling them to proactively revise policies and manage risks effectively [3].

In news aggregation systems, text mining tasks serve as foundational tools for extracting actionable insights from large amounts of unstructured data. For example, sentiment analysis on financial news is leveraged for stock market prediction [4, 5], forecasting market trends [6], developing investment strategies [6], and managing risks [7]. Techniques like Named Entity Recognition (NER) and Relation Extraction (RE) are essential for identifying key performance indicators (KPIs) and relationships in financial reports [8], extracting financial entities [9], analyzing financial events [10], and performing various downstream tasks [11]. Additionally, topic modeling and summarization help in recognizing trends and condensing essential information from lengthy financial documents such as earnings reports or regulatory filings [12]. These text mining techniques significantly enhance processes like risk management, decision-making, and compliance monitoring across the financial industry.

Ensuring the accuracy of text mining solutions is critical. Misleading or incorrect data can lead to poor decision-making. Extracting knowledge from financial news faces many challenges that require advanced NLP techniques to retrieve the relevant content and understand the context and sentiment. A single document may mention various topics, and each topic contains many indicators of institutions. Figure 1 shows a piece of a financial report that presents banks' stock liquidity on both positive and negative sides. To obtain the indicator for each bank, NLP techniques must locate the paragraph, segment complex sentences, and analyze sentiment. Additionally, banking-related text involves complex financial jargon, industry-specific terminology, and nuanced economic factors. As a result, general-purpose text mining techniques may not be sufficient to extract relevant information. Thus, fine-tuned language models in a financial context are necessary [13].

This paper presents a real information aggregation system for the banking sector. The processing flow consists of the following steps: 1) scraping financial news from multiple sources, 2) processing and storing data, 3) extracting business indicators like policies, products, and financial performance for each institution, 4) generating useful reports, and 5) interacting with users. The highlights of the system are scalability and flexibility. Since components are designed to perform separate tasks asynchronously, the system can be scaled efficiently without performance bottlenecks. The extracted data are organized hierarchically, allowing for reporting from overview to detail level.

The processing and extraction modules handle the preparation of information for generating useful reports by converting each news item into structured data. For each news item, the first step is to classify it as either financial or non-financial text. The long texts are then segmented into smaller sections, each containing information relevant to a specific indicator for an institution. As shown in Figure 1, a resulting paragraph could be "*Unfortunately, some banks have experienced a contraction in liquidity during the year, including KLB (down 89%)*.". Then, different elements are extracted, including the indicator type, entity name, and sentiment, i.e., liquidity, KLB, and negative, respectively.

We prepared an annotated dataset to evaluate solutions and fine-tune language models to the financial domain. Original data is collected from financial news from six prestigious sites in Vietnam. Each news item is processed and separated into segments as described above. After filtering non-financial, we obtain 12,000 segments known as samples. Each sample was annotated by volunteers and verified by our team as positive, neutral, and negatives semantically compatible with specific topics. We publish the dataset to facilitate further research. [a]

For sentiment analysis, we have found that advanced methods commonly used for general texts may not be well-suited to the financial domain. While LLMs typically perform well across various NLP tasks, they exhibited low accuracy on the annotated dataset. In our experiments, we tested two LLMs including LLama-3.1-8B [14] and Vistral-7B [15], using different prompt engineering techniques. Additionally, we fine-tuned several pre-trained models, such as PhoBERT [16], ViT5 [17], and BartPho [18], on a portion of the dataset. The results have shown that fine-tuned models significantly outperform the LLMs. The substantial improvement of fine-tuning language models for the financial domain has been confirmed in other studies [13].

In summary, the key contributions of our research are as follows:

- A real-world banking-focused news aggregation system capable of multi-source scraping, structured storage, indicator extraction, and hierarchical reporting.
- Fine-tuned language models for financial sentiment analysis, demonstrating that domain adaptation significantly improves accuracy, reduces memory usage, and increases inference speed, making them more practical than generic LLMs.
- A 12,000-segment Vietnamese financial sentiment dataset, manually annotated for training and evaluation.

The rest of the paper is organized as follows: Section 2 reviews related studies on the applications of news aggregation systems and financial text analysis. Section 3 provides an in-depth description of the system, including its components, design, and text mining techniques. Section 4 details the annotated dataset and outlines the experimental setup. Finally, Section 5 discusses the experimental results and analyzes the model's predictions.

[a] The dataset is publicly available at https://huggingface.co/datasets/iaiuet/banking_sentiment_vietnamese

2. RELATED WORK

2.1. The applications of news aggregation systems

News aggregation systems have become indispensable tools in various domains, particularly for their ability to process and deliver real-time information from vast, dynamic sources. These systems are widely applied in a wide range of areas such as journalism, cybersecurity [19], military [20], social media [21], e-commerce, and finance to enhance decision-making and provide actionable insights [22]. In journalism, news aggregation assists in delivering up-to-date information from diverse sources, enabling journalists and readers to follow trends and developments from multiple perspectives [23]. Similarly, in e-commerce, aggregated insights from consumer feedback and competitor information allow companies to adjust strategies, refine product offerings, and manage reputational risks effectively [24]. In finance, these systems facilitate data-driven decision-making and risk management [25].

In the financial sector, news aggregation has seen significant application due to the critical role of information in financial decision-making and risk management. Banks and investment firms utilize news aggregation to monitor financial markets, interpret market sentiment, and anticipate fluctuations. Sentiment analysis of aggregated news content is increasingly used for stock price prediction and volatility forecasting, where algorithms analyze public sentiment around financial markets, companies, and global economic events to provide early warnings of market changes. In this way, financial institutions can adapt their strategies in response to potential risks or opportunities emerging from current events.

For banks specifically, news aggregation systems enable the tracking of financial indicators, such as regulatory changes, competitor activities, and economic conditions, to manage investment portfolios and lending policies. For example, during sudden economic downturns, banks can analyze aggregated news data to adjust lending criteria and manage liquidity [2]. Similarly, news aggregation assists in monitoring compliance by keeping up with regulatory developments in different jurisdictions, ensuring that banks and financial institutions adapt their policies to maintain compliance with evolving standards.

In addition, governments benefit from news aggregation systems in overseeing financial markets and banks' activities, as they help regulators track market trends and banks' compliance with regulatory frameworks. Aggregated insights allow governments to proactively address systemic risks, ensuring financial stability by anticipating and managing economic impacts more effectively [3]. These systems further facilitate the monitoring of financial institutions' performance by analyzing market data alongside news, thereby enabling regulatory bodies to detect early signs of financial distress or malpractices within the sector.

2.2. Language models for financial text analysis

Language models (LMs) play a crucial role in mining knowledge from unstructured financial text, particularly for tasks such as sentiment analysis and key entity detection. Early studies applied general-purpose models such as BERT and GPT-2 to financial text with moderate success. However, these models often lacked domain-specific financial knowledge and failed to capture subtle terminology and industry-specific nuances. BERT-based sentiment analysis has been widely adopted in financial text mining [26]. Sousa et al. [27] further fine-tuned BERT for sentiment analysis of financial news articles, demonstrating its effectiveness in providing decision-relevant information for stock market analysis. In addition, Jia Miao combined GPT and RoBERTa for investor sentiment analysis on Chinese interactive investment platforms by jointly modeling news articles, investor questions, and corporate secretary responses.

To overcome the limitations of general-purpose LMs, fine-tuning on domain-specific financial corpora has emerged as an effective approach to improve accuracy and contextual understanding. Empirical evidence consistently shows that models adapted to the financial domain significantly outperform generic LMs in capturing financial language intricacies and semantic relationships. Fine-tuned models such as FinBERT and FinGPT have demonstrated strong performance in sentiment classification, financial event extraction, and market prediction. FinBERT, for instance, is trained on large-scale financial text and achieves higher accuracy in tasks including earnings call sentiment analysis, stock market forecasting, and risk assessment.

In sentiment analysis, domain-specific fine-tuned models provide substantial advantages by accurately identifying sentiment polarity and intensity, which are critical for investment decision-making, credit risk assessment, and portfolio management. Recent studies report that such models can achieve up to 20% higher accuracy in financial sentiment tasks compared to general-purpose LMs.

Beyond sentiment analysis, fine-tuned financial LMs have also enhanced knowledge extraction from financial reports and regulatory filings, which often contain dense jargon and complex structural information. Models such as Financial DistilBERT have shown effectiveness in named entity recognition (NER) and relation extraction (RE), enabling the identification of relationships among entities such as companies, financial instruments, and economic indicators. These capabilities support automated modeling of financial relationships and facilitate deeper analytical insights for financial analysts and decision makers.

3. A SOLUTION TO NEWS AGGREGATION SYSTEMS

This study develops an end-to-end news aggregation system to support decision making and risk management in the banking sector. The system, as shown in Figure 2, is structured into five main components: (1) Data Scraping, (2) Processing and Storing, (3) Extracting Structured Information, (4) Dataset Construction, and (5) Application. The data scraping module continuously collects articles from reputable Vietnamese financial news sources such as kinhtedothi.vn, vietstock.vn, and vneconomy.vn. The processing and storing module filters economic-related content and organizes it for downstream analysis. The extraction module applies NLP techniques to segment each article into meaningful paragraphs, identify triplets of (bank, indicator, and sentiment), and recognize sentiment polarity, institutional entities, and financial indicators. These structured outputs are stored in a relational database and subsequently used to build an annotated dataset for model training and evaluation. The final dashboard module retrieves and visualizes the aggregated information, offering dynamic insights into institutional performance and macroeconomic trends. This section will describe in detail the implementation of the components.

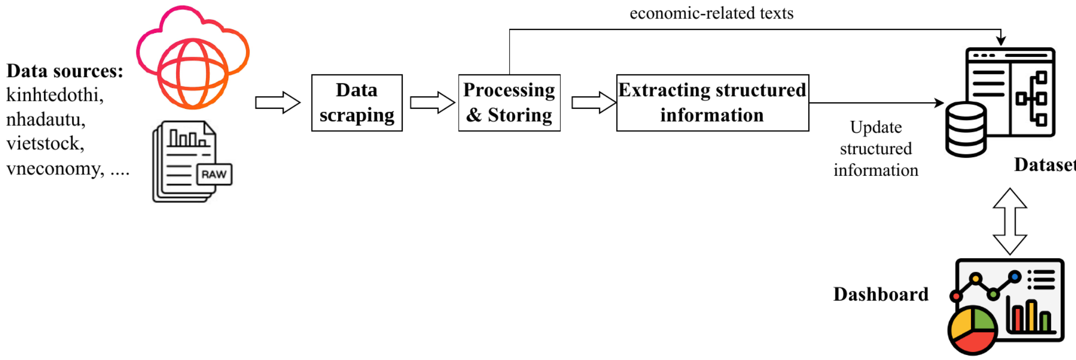

Figure 2: The news aggregation system pipeline

3.1. Database design

We design a relational database to manage, organize, and analyze collected data. In the news aggregation system, the database also ensures cohesive operation among pipeline components (Figure 2). It consists of four main tables: **News**, which stores economic-related articles; **Banks**, which contains detailed information about financial institutions; **Topics**, which manages financial topics along with metadata and sentiment-oriented keywords; and **Sentiment**, which records sentiment scores at the paragraph level, where each article is segmented by bank and topic.

3.2. Data scraping

Data were collected from nine websites that provide in-depth analysis and commentary on business and financial markets. Using the Scrapy framework, we developed a tool to regularly update news from sources including *cafef*, *dantri*, *kinhtedothi*, *nhadautu*, *vietnambiz*, *vietnamnet*, *vietstock*, *vneconomy*, and *vnexpress* [b]. Each website analyzes the financial market according to different aspects such as market structure, economic indicators, market performance, and so on.

3.2.1. Economic news recognition

Upon retrieval, the text data undergoes NLP-based classification to determine its relevance to economic topics. We utilize three methods to identify the most effective approach: Support Vector Machine (SVM), Random Forest, and deep learning through fine-tuning PhoBERT. Term Frequency-Inverse Document Frequency is used for feature extraction in SVM and Random Forest, while PhoBERT leverages its pretrained language model architecture. This step ensures that only data relevant to economic topics is retained for further processing and analysis.

The dataset used to train and test the classifier was created from the data scraped using the aforementioned method. Labels were assigned based on predefined categories available

[b] https://cafef.vn/, https://dantri.com.vn/, https://kinhtedothi.vn/, https://nhadautu.vn/, https://vietnambiz.vn/, https://vietnamnet.vn/, https://vietstock.vn/, https://vneconomy.vn/, https://vnexpress.net/

Table 1: Performance comparison of classifiers

Model	Accuracy	Precision	Recall	F1
SVM	0.80	0.80	0.80	0.80
PhoBERT	0.79	0.80	0.79	0.79
Random Forest	0.78	0.78	0.78	0.78
XGBoost	0.76	0.76	0.76	0.76

on certain websites, such as Business, Entertainment, Real Estate or Health, etc. However, classification remains crucial, as some sources lack these clear categorizations, necessitating the use of automated methods to identify relevant content.

The performance of the classification models was thoroughly evaluated using the test set, with detailed results presented in Table 1. Based on the results, we selected SVM as the primary method for classifying economically relevant data, as it outperformed Random Forest and XGBoost in accuracy while being significantly more lightweight than PhoBERT. The data identified through SVM is systematically stored in the **News** table for subsequent in-depth analysis.

3.3. Extracting business indicators

The business indicators extraction process systematically handles news articles to extract precise and relevant information for analysis. It consists of four main steps: text segmentation, topic identification, bank entity recognition, and segment filtering.

3.3.1. Text segmentation

The process begins by retrieving news articles from the **News** table in the database and performing text segmentation. This step involves breaking each article into smaller, coherent segments, specifically dividing the text into paragraphs that represent meaningful units of information.

3.3.2. Topic identification

Following text segmentation, the system identifies topic(s) discussed within each segmented paragraph. This is achieved through keyword extraction techniques, leveraging a predefined list of topics stored in the **Topics** table in the database. By aligning extracted keywords from each paragraph with entries in the topic repository, the module determines the primary themes and subtopics relevant to the text.

3.3.3. Bank entity recognition

Within each segmented paragraph, the system employs bank entity recognition to identify mentions of specific banks. This process utilizes name-based recognition techniques, referencing data from the **Banks** table in the database. The module extracts bank names, abbreviations, and aliases from the database, ensuring accurate identification of banking institutions mentioned in the text.

3.3.4. Segment filtering

To refine the analysis further, the system filters paragraphs to retain only those segments that mention specific banks within the context of the identified topic(s). This ensures that subsequent analysis, including sentiment evaluation, is confined to the most relevant and

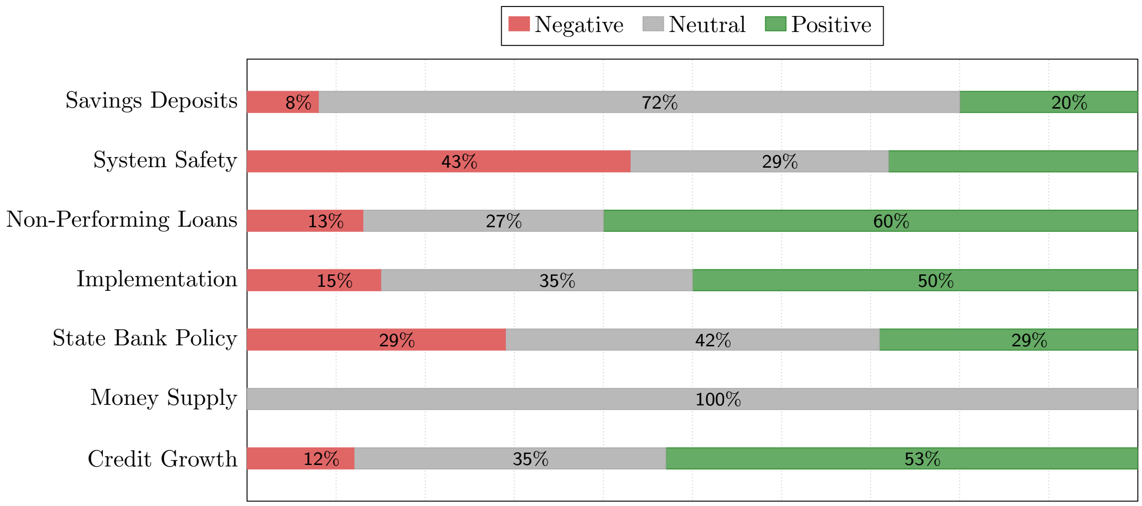

Figure 3: Sentiment distribution by topic of a bank

focused content. By filtering at this level of detail, the workflow maintains a sharp focus on discussions pertinent to the banking sector, improving the accuracy and relevance of the insights generated. The filtered segments are then passed to the sentiment analysis module for in-depth evaluation, as described in the following sections.

3.3.5. Module output management

The output of this module is stored in the **Sentiment** table, where each extracted segment is linked to its topics and identified bank entities. At this stage, the *sentiment* field remains empty and will later be filled by the sentiment analysis module discussed in the following sections. Null values simply indicate that the sentiments of these segments have not yet been evaluated.

3.4. Sentiment analysis

Our sentiment analysis module employs language models to classify sentiment in banking related texts. We evaluate two prompt-based large language models, LLaMA-3.1-8B and Vistral-7B, alongside fine-tuned Vietnamese pre-trained models, including PhoBERT, ViT5, and BARTPho. The LLMs are used solely in inference mode with prompt-based strategies, while the smaller models are fine-tuned to capture domain-specific linguistic characteristics. Detailed comparative results are reported in Sections 4, and 5. After inference, sentiment-labeled paragraphs are stored in the **Sentiment** table together with their corresponding classifications.

3.5. Report generation

Our system employs sophisticated data visualization techniques to transform processed sentiment data into interactive and insightful charts and graphs. The visualization process begins by extracting relevant data from the **Sentiment** table in the database. This data is then filtered based on the desired criteria, such as specific banks, topics, and time periods.

Once filtered, the sentiment scores (positive, negative, and neutral) are aggregated for each bank, topic, and time period to facilitate comprehensive trend analysis.

To create these visualizations, we utilize the advanced data visualization library Matplotlib. Our system supports many different chart types, including bar charts and pie charts, in this report, stacked bar charts will be taken as an example. This chart effectively illustrates the distribution of sentiment across different categories. As shown in Figure 3, the stacked bar chart includes the following elements: the X-axis represents categories such as banks, topics, or time periods, while the Y-axis shows the percentage of sentiment (positive, negative and neutral). Each stacked bar represents the proportion of each sentiment category, with a legend indicating the different sentiment types. Tooltips are incorporated to provide detailed information about each data point, enhancing user understanding and engagement.

The system is designed to be highly interactive, allowing users to filter the data by selecting specific banks, topics, or time periods. Users can also drill down to view sentiment distribution at a more granular level, such as individual articles or specific dates. Hover effects are implemented to display additional information when users hover over data points, offering deeper insights into the sentiment data.

Customization options are available to users, enabling them to adjust the appearance of the charts, including colors, fonts, and styles. Additionally, the system provides options to export the charts in various formats, such as PNG, JPEG, or PDF, for reporting and presentation purposes. Users can also export the underlying data in Excel (xlsx) format, allowing for further analysis using tools like Excel or PowerBI.

4. EXPERIMENTAL SETUP

4.1. Dataset development

To effectively analyze sentiment in the banking sector, we developed a comprehensive dataset using the aforementioned system. The dataset consists of 10,000 training samples and 2,000 test samples, automatically collected from multiple online financial news sources. Each data instance was generated through a structured pipeline of data scraping, text segmentation, topic identification, and filtering as described in Section 3.

The samples were organized by topic and independently labeled by three trained student annotators. Annotating followed the detailed guidelines designed by lecturers from the Banking Academy. Each sample was assigned one of three sentiment categories including positive, negative, or neutral, based on topic-specific sentiment nuances. After annotation, our research team conducted consensus validation to assess inter-annotator agreement. In cases of disagreement, the final sentiment label was determined through majority voting with manual verification by the research team.

The financial news articles were published between May 20, 2018, and July 1, 2024, covering both historical and contemporary trends in the Vietnamese banking industry. The dataset encompasses 13 major topics reflecting key aspects of the banking sector. A detailed distribution of samples across topics and the train–test split is shown in Table 2.

This dataset is meticulously designed for sentiment analysis within the banking sector, offering deep insights into public opinion across a broad spectrum of banking topics. By covering 13 distinct aspects of the banking industry, the dataset provides a highly practical and comprehensive foundation for training and evaluating sentiment analysis models. Its

Table 2: Distribution of data across different topics in the dataset

Topic	Train	Test	Topic	Train	Test
Banking Services	1,495	311	Savings Deposits	793	142
Non-Performing Loans	1,381	286	Business Efficiency	635	134
Credit Growth – Money Supply	1,198	212	Loan Products	554	124
Liquidity	949	196	Management Quality	390	72
Digitalization of Banking Operations	859	146	System Safety	353	82
Bank Capital	846	168	Role of Regulatory Tools	292	67
			State Bank Policy Implementation	255	60
Total			**Train: 10,000**	**Test: 2,000**	

extensive coverage ensures that the models are well-aligned with the diverse and nuanced realities of the financial domain, making the dataset exceptionally relevant for real-world applications.

4.2. Settings

To evaluate sentiment analysis in the banking domain, we conducted experiments with both large and small language models. Large language models, including LLaMA-3.1-8B and Vistral-7B, were evaluated purely in inference mode, in both full-precision and 4-bit quantized settings. We designed several prompt variants to assess sensitivity to contextual information, external domain knowledge, and example-based reasoning, ranging from prompts without additional context to prompts enriched with banking-specific sentiment cues, as well as zero- and few-shot prompts containing up to five labeled examples. For few-shot inference, topic-consistent examples were dynamically retrieved from the training set using a FAISS-based vector database. In contrast, smaller language models (BARTPho, PhoBERT, and ViT5) were fine-tuned to capture domain-specific linguistic patterns across 13 banking-related topics. All fine-tuned models and quantized LLMs were evaluated on an NVIDIA V100 GPU (16 GB), while the full-precision LLaMA-3.1-8B and Vistral-7B models were deployed on an NVIDIA A5000 GPU (24 GB).

5. RESULTS AND DISCUSSION

Table 3 presents the performance of the language models in terms of accuracy, inference time, and GPU usage. PhoBERT achieves the highest accuracy (68.90%), closely followed by ViT5 (68.20%) and BARTPho (67.80%). Although tested with few-shot prompts or external knowledge, LLMs exhibit lower accuracy. The best performing LLM configuration, LLaMA-3.1 with a 5-shot prompt, reaches an accuracy of 61.75%, which is 7.15% lower than PhoBERT. The results confirm the effectiveness of domain adaptation for Vietnamese financial sentiment analysis, where linguistic nuances and topic-specific expressions are highly specialized.

In terms of computational efficiency, fine-tuned models are substantially more practical for real-world deployment. PhoBERT achieves the best trade-off between accuracy and efficiency, requiring only 0.5 minutes of inference time and around 0.2 GB of GPU memory. In contrast, LLaMA-3.1-8B and Vistral-7B demand extensive computational resources, exceeding 9 GB of GPU memory and 30-100 minutes of inference time. High computational resource demands limit the feasibility for large intelligence systems.

Table 3: Comparison of models according to Accuracy, GPU usage, and Inference Time. The best LLMs for each scenario of external knowledge integration and few-shot prompt are selected.

Model	Accuracy (%)	Inference Time (minutes)	GPU Usage (GB)
Llama_5-shot	61.75	39.5	9.414
Llama_E_4-shot	57.25	40.7	9.256
Llama_Q_3-shot	54.15	16.76	20.377
Llama_Q_E_5-shot	61.60	29.19	22.332
Vistral_5-shot	48.45	97	11.97
Vistral_E_5-shot	48.00	105.8	12.268
Vistral_Q_ 0-shot	48.20	22.41	15.955
Vistral_Q_E_5-shot	45.95	93.04	20.615
BARTPho	67.80	0.77	0.889
ViT5	68.20	1.45	1.482
PhoBERT	**68.90**	0.50	0.214

Table 4: Evaluation of full (F) and quantized (Q) versions of LLama-3.1-8B

External	Prompt Type	Accuracy		Inference Time		GPU Usage	
		Q	F	Q	F	Q	F
No	0-shot	55.10	51.50	24	11.59	7.816	20.54
No	1-shot	52.65	53.70	30	16.93	8.148	20.38
No	2-shot	56.70	53.80	32.6	16.77	8.424	20.38
No	3-shot	58.70	54.15	35.5	16.76	8.844	20.38
No	4-shot	59.50	54.00	36.7	16.92	9.158	20.38
No	5-shot	**61.75**	53.70	39.5	16.78	9.414	20.38
Yes	1-shot	49.85	54.00	31.6	19.12	8.172	21.01
Yes	2-shot	52.50	57.95	37	21.02	8.41	22.66
Yes	3-shot	55.50	59.20	37.9	23.15	8.902	22.69
Yes	4-shot	57.25	60.75	40.7	25.58	9.256	21.93
Yes	5-shot	56.65	**61.60**	41	29.19	9.492	22.33

Table 5: Evaluation of full (F) and quantized (Q) versions of Vistral-7B

External	Prompt Type	Accuracy		Inference Time		GPU Usage	
		Q	F	Q	F	Q	F
No	0-shot	**53.75**	48.20	28.7	22.41	6.738	15.96
No	1-shot	35.40	36.55	45.8	60.88	6.816	6.01
No	2-shot	35.40	36.25	57	36.68	9.053	6.01
No	3-shot	40.00	35.90	68.8	36.72	9.443	6.01
No	4-shot	43.35	36.35	81.2	36.99	8.33	6.01
No	5-shot	48.45	36.60	97	36.71	11.97	6.01
Yes	1-shot	41.05	36.25	66	44.66	7.104	16.40
Yes	2-shot	37.15	36.90	78	55.24	10.484	19.79
Yes	3-shot	40.95	39.90	78	66.68	9.67	20.37
Yes	4-shot	43.50	43.05	92	78.27	11.184	19.74
Yes	5-shot	48.00	45.95	105.8	93.04	12.268	20.62

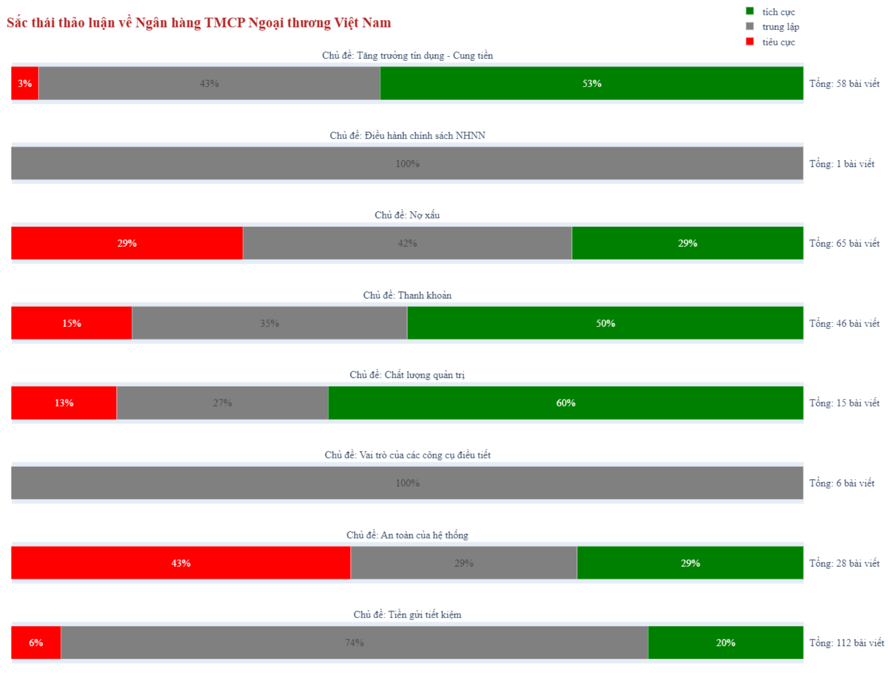

Figure 4: The accuracy of the large language models with few-shot prompting on several topics such as Non-Performing Loans (NPL), Digitalization of Banking Operations (DBO), Role of Regulatory Tools (RRT), and Bank Capital (BC).

The impact of few-shot prompting on model accuracy. The results from both LLMs demonstrate a nuanced relationship between few-shot learning and accuracy. For LLama-3.1-8B, increasing the number of in-context examples consistently improves the quantized model's accuracy, peaking at 61.75% in the 5-shot without external knowledge and at 61.60% in the 5-shot with external knowledge. This indicates that LLama effectively leverages the provided examples to better understand the context. In contrast, Vistral-7B's quantized version shows high instability, with accuracy dropping from 53.75% (0-shot) to 35.40% (1-shot and 2-shot). It suggests that Vistral-7B is more sensitive to the quality and quantity of retrieved examples.

Effect of external knowledge on the prompts. An unexpected observation is that providing banking-specific phrases of sentiment cues reduces accuracy for LLMs (Tables 4, 5). In financial discourse, the polarity of sentiment expressions depends heavily on topic context. For instance, "increased investment costs" is negative in the Digitalization of Banking Operations (DBO) topic, whereas "increased experiences" is positive. Similarly, "increased rapidly compared to the previous period" signals negative sentiment in Non-Performing Loans (NPL). Consequently, embedding external phrases into prompts may introduce contextual conflicts that confuse model interpretation.

Figure 4 shows how accuracy varies across topics when additional contextual information is provided. The results indicate that few-shot prompting enhances performance more significantly in topics where sentiment cues align consistently with their polarity. For example, in Role of Regulatory Tools (RRL), positive signals such as "stable operation" and "appropriate implementation" maintain clear semantic consistency. Conversely, in topics where sentiment terms exhibit contextual ambiguity or polarity reversal, such as NPL or DBO, the benefits of few-shot prompting diminish. For instance, in NPL, phrases like "increased dramatically" or "increased rapidly compared to the previous period" convey negative sentiment despite the presence of positive lexical patterns. These findings confirm that banking sentiment analysis requires topic-specific contextual grounding rather than surface-level keyword associations.

Computational efficiency and practical implications. From a deployment perspective, the fine-tuned small models are more practical for integration into real-time banking systems. In contrast, large-scale models such as LLaMA-3.1-8B require substantial GPU memory and extensive inference time, which hinders scalability. Fine-tuned models achieve superior accuracy while maintaining computational efficiency. They use less than 2 GB of GPU and infer under 2 minutes. For production environments, latency, resource optimization, and cost-effectiveness are critical considerations. In summary, the experiments reveal that (1) even when provided with additional contextual information such as few-shot examples or sentiment cue phrases, LLMs continue to face challenges in interpreting domain-specific sentiment; and 2) fine-tuning small models yields significantly better performance in both predictive accuracy and computational efficiency.

6. CONCLUSION

In this study, we developed a practical and scalable news aggregation system designed to support decision-making and risk management in the banking sector. The system integrates multiple NLP modules such as topic classification and sentiment analysis into an end-to-end framework. We systematically evaluated various NLP approaches, including fine-tuned Vietnamese models (PhoBERT, ViT5, and BARTPho) and large language models (LLaMA-3.1-8B and Vistral-7B), to identify the most effective solutions for domain-specific sentiment analysis. The findings highlight the inherent challenges of topic-dependent sentiment interpretation, where sentiment cues may not consistently align with polarity across different banking contexts. Experimental results demonstrate that fine-tuned, domain-specific models significantly outperform general-purpose LLMs. Specifically, PhoBERT achieves a 7.15% improvement in accuracy while requiring substantially less computational time and GPU resources. The proposed system demonstrates potential for real-world use in banking and provides practical guidance for choosing appropriate NLP technologies for similar applications.

REFERENCES

[1] M. N. Ashtiani and B. Raahemi, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review," *Expert Systems with Applications*, vol. 217, p. 119509, 2023.

[2] A. Ashta and H. Herrmann, "Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance," *Strategic Change*, vol. 30, no. 3, pp. 211–222, 2021.

[3] H. K. Duan, M. A. Vasarhelyi, M. Codesso, and Z. Alzamil, "Enhancing government accounting information systems using social media information: An application of text mining and machine

learning," *International Journal of Accounting Information Systems*, vol. 48, p. 100600, 2023.

[4] A. E. Khedr and N. Yaseen, "Predicting stock market behavior using data mining techniques and news sentiment analysis," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 7, pp. 22–31, 2017.

[5] C. T. Thuy and D. V. Thanh, "New resolution for analyzing Vietnams stock market." *Journal of Computer Science and Cybernetics*, vol. 24, no. 2, pp. 107–118, 2008.

[6] S. W. K. Chan and M. W. C. Chong, "Sentiment analysis in financial texts," *Decision Support Systems*, vol. 94, pp. 53–64, 2017.

[7] C. Nopp and A. Hanbury, "Detecting risks in the banking system by sentiment analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015, pp. 591–600.

[8] L. Hillebrand, T. Deußer, T. Dilmaghani, B. Kliem, R. Loitz, C. Bauckhage, and R. Sifa, "KPI-BERT: A joint named entity recognition and relation extraction model for financial reports," in *International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 606–612.

[9] Y. Zhang and H. Zhang, "Finbert–mrc: Financial named entity recognition using bert under the machine reading comprehension paradigm," *Neural Processing Letters*, vol. 55, no. 6, pp. 7393–7413, 2023.

[10] G. Jacobs and V. Hoste, "Sentivent: Enabling supervised information extraction of company-specific events in economic and financial news," *Language Resources and Evaluation*, vol. 56, no. 1, pp. 225–257, 2022.

[11] S. Kaur, C. Smiley, A. Gupta, J. Sain, D. Wang, S. Siddagangappa, T. Aguda, and S. Shah, "Refind: Relation extraction financial dataset," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 3054–3063.

[12] S. Garca-Mndez, F. de Arriba-Prez, A. Barros-Vila, F. J. Gonzlez-CastaΛo, and E. Costa-Montenegro, "Automatic detection of relevant information, predictions and forecasts in financial news through topic modelling with latent dirichlet allocation," *Applied Intelligence*, vol. 53, no. 16, pp. 19 610–19 628, 2023.

[13] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, "Finbert: A pre-trained financial language representation model for financial text mining," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2021, pp. 4513–4519.

[14] A. Dubey and A. Jauhri, "The llama 3 herd of models," *arXiv e-prints*, pp. arXiv–2407, 2024.

[15] V. N. Chien and N. Thuat, "Vistral-7b-chat: Towards a state-of-the-art large language model for Vietnamese," 2023.

[16] D. Q. Nguyen and A. T. Nguyen, "Phobert: Pre-trained language models for Vietnamese," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, 2020, pp. 1037–1042.

[17] L. Phan, H. Tran, H. Nguyen, and T. H. Trinh, "Vit5: Pretrained text-to-text transformer for Vietnamese language generation," in *Proceedings of the 2022 NAACL-HLT Student Research Workshop.* Association for Computational Linguistics, 2022, pp. 136–142.

[18] N. L. Tran, D. M. Le, and D. Q. Nguyen, "Bartpho: Pre-trained sequence-to-sequence models for Vietnamese," *arXiv preprint arXiv:2109.09701*, 2021.

[19] M. Schafer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti, and V. Lenders, "Blackwidow: Monitoring the dark web for cyber security information," in *International Conference on Cyber Conflict (CyCon)*, 2019, pp. 1–21.

[20] L. Ball, "Automating social network analysis: A power tool for counter-terrorism," *Security Journal*, vol. 29, pp. 147–168, 2016.

[21] C. E. H. Chua, V. C. Storey, X. Li, and M. Kaul, "Developing insights from social media using semantic lexical chains to mine short text structures," *Decision Support Systems*, vol. 127, p. 113142, 2019.

[22] J. R. G. Evangelista, R. J. Sassi, M. Romero, and D. Napolitano, "Systematic literature review to investigate the application of open source intelligence (osint) with artificial intelligence," *Journal of Applied Security Research*, vol. 16, no. 3, pp. 345–369, 2021.

[23] F. Marconi, *Newsmakers: Artificial intelligence and the future of journalism.* Columbia University Press, 2020.

[24] F. Wang, Y. Yang, G. K. F. Tso, and Y. Li, "Analysis of launch strategy in cross-border e-commerce market via topic modeling of consumer reviews," *Electronic Commerce Research*, vol. 19, pp. 863–884, 2019.

[25] H. A. Javaid, "Ai-driven predictive analytics in finance: Transforming risk assessment and decision-making," *Advances in Computer Sciences*, vol. 7, no. 1, 2024.

[26] L. Zhao, L. Li, X. Zheng, and J. Zhang, "A bert-based sentiment analysis and key entity detection approach for online financial texts," in *IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD).* IEEE, 2021, pp. 1233–1238.

[27] M. G. Sousa, K. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "Bert for stock market sentiment analysis," in *IEEE International Conference on Tools with Artificial Intelligence (ICTAI).* IEEE, 2019, pp. 1597–1601.

Received on February 18, 2025
Accepted on January 27, 2026