IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 3, June 2025, pp. 1829~1838
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp1829-1838

The influence of sentiment analysis in enhancing early warning system model for credit risk mitigation

Angel Karentia[1], Derwin Suhartono[2]
[1]Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Aug 8, 2024
Revised Feb 4, 2025
Accepted Mar 15, 2025

Keywords:

Credit quality
Early warning system
Forward-looking
Model accuracy
Sentiment analysis

ABSTRACT

One important source of bank income is interest income from credit activities, another part of which is obtained from fee-based income. Rapid credit growth is directly proportional to an increase in potential credit risk (counterparty default). In addition to comprehensive credit assessment at the initial stage of credit initiation, banks need to monitor the condition of existing debtors. Empirically, difficulties in handling non-performing loans often occur due to delays in detection and preparation of action plans. In this case, losses due to non-performing loans can have implications for the bank's reputation and worsen its financial performance. This research aims to determine the effect of sentiment analysis (external sentiment prediction model [positive, neutral, and negative] with certain keywords) on the level of accuracy of the early warning system (EWS) model in predicting the credit quality of bank debtors in the coming months. This study found that upgrading EWS with sentiment analysis will give better accuracy levels compared to traditional EWS models. In addition, the predictive power of EWS (traditional and upgraded) is inversely proportional to the prediction period, the longer the target prediction time, and the less predictive power of the EWS model.

Corresponding Author:

Angel Karentia
Computer Science Department, BINUS Graduate Program-Master of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
Email: angel.karentia@binus.ac.id

1. INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic has had a massive impact where all countries have implemented activity restrictions (lockdowns) to prevent the rapid spread of the virus and the consequences of this lockdown have created challenges in the health, economic, and social sectors [1]. According to the World Bank in 2022, this emergency response creates new risks, namely increasing the debt ratio in the world economy, which can threaten the recovery process from the crisis because of the interconnection between the financial health of individuals/companies and financial institutions [2]. Banks as financial institutions have a critical role in the economy including the process of distributing funds to the public [3]. Banks are obliged to manage credit risk to avoid losses for the bank itself and its customers. However, the implementation of risk management for each bank can vary according to objectives, business policies, business size, and complexity, financial capabilities, supporting infrastructure, and human resource capabilities are taken into consideration and one of the applications is creating an early warning system (EWS) which is an innovation that was born as a result of the global crisis that occurred in 2007-2008.

The beginning of EWS in banking was made to monitor the overall stability of the financial system to prevent the occurrence of systemic risk where the economy declines so that the global financial crisis does not happen again by detecting crisis signals before the actual crisis occurs [4]. EWS used by many banks tends to produce false positives, the warnings given become meaningless and it is too late to carry out credit risk mitigation. Losses are inevitable because the data used is only based on traditional data such as financial data and money markets without considering external events that are happening. This was seen when the lockdown was implemented during the COVID-19 pandemic in 2020, economic activity decreased causing many debtors to be threatened with bankruptcy (default) which was detected too late by the EWS [5].

In the 21st century era, machine learning is widely used by the banking sector, especially risk management, to create more accurate risk modelling by identifying complexity and nonlinear patterns [6]. The potential of machine learning can be utilized to obtain data related to external events that have not been captured by the EWS by extracting sentiment related to current economic fluctuations [7]. Natural language processing (NLP) is used as an efficient way to research things that are important when not explicitly mentioned by Kim *et al.* [8]. A technique in NLP that was previously popularly used to obtain sentiment was the lexicon approach which creates a dictionary of positive and negative words, but this approach has the weakness of not being able to capture the context, so the sentiment results become inaccurate [9]. The discovery of transformers in the world of NLP can cover the shortcomings of this lexicon by coding a series of words into one part which is called a self-attention mechanism [10]. Transformers are proven to be the most frequently used architecture in NLP because their performance can exceed neural network architectures which are widely used to create models from unstructured data [11].

Earlier EWS research assessed overall financial stability [12]. Banks's monitoring credit risk can use developed internal models that are adjusted to the bank's strategy and conditions [13]. Other than that, EWS also heavily relied on financial reports using models that were easy to explain such as logistic regression [14] where [15] find the best using CatBoost algorithm. On the other hand, to identify credit risk using the sentiment from news already conducted to prove it can support the traditional method of assessing debtor risk. However, there is still no research about the integration of assessing debtor risk with the current sentiment news [16]. While the best method for extracting sentiment is use a transformer-based pre-trained model as [17] proved that IndoBERT surpass other methods in extracting Indonesian financial texts. This paper aims to create EWS using bank debtor's data and then add the results of news sentiment analysis to predict the credit quality payment (pass, special mention, and non-performing loan). The main contributions of this paper are i) improved traditional model accuracy by adding the results of news sentiment and ii) proving by adding the results of sentiment analysis data can help the credit risk mitigation process by providing warnings before debtors are threatened to default (forward-looking).

The structure of the paper is as follows: section 2 shows the literature review of this research. Section 3 provides how the research is conducted. Section 4 presents the analysis of the research result. Section 5 concludes this research.

2. LITERATURE REVIEW

2.1. Theoretical foundation

2.1.1. Early warning system

The first research related to EWS is predicted debtor bankruptcy using rough sets and decision trees to find problematic credit patterns so that they can predict problematic credit before customers become debtors. The rough set accuracy was able to reach 87.42%, superior to the decision tree whose accuracy was 83.33% [18]. Second research made a comparison of several classification algorithms in the case of predicting bankruptcy of corporate segment companies using 2014-2016 financial report data of bankrupt companies in France. As a result, CatBoost's performance can surpass 8 other popular algorithms, namely discriminant analysis, logistic regression, support vector machine, neural network, random forest, gradient boosting, deep neural networks, and XGBoost [15].

Third research used a logistic regression algorithm to detect bank bankruptcy using financial report data published by the bank and macroeconomic data. Logistic regression produces an accuracy of 89.76%, where the crisis signal that was successfully captured (true positive (TP)) was 67.64%, the crisis signal that was lost (false negative (FN)) was 32.35%, and the false crisis signal (false positive (FP)) of 7.52% [14]. The fourth and fifth research predicts debtor bankruptcy using a neural network algorithm that optimized weight and threshold with a cross-over of 0.4 and mutation of 0.2 and accuracy reached 97% with error reduced by 55.8% while comparing backpropagation neural networks with models that are usually used to predict future credit risk, namely autoregressive models and logistic regression, it was found that backpropagation neural networks increase the accuracy of future-looking credit risk management [19], [20]. Sixth research modelling a comprehensive EWS for Chinese Enterprises by utilizing quantitative and qualitative using fusion-logistic algorithm reached accuracy 89.7% in predicting special treatment companies [21].

2.1.2. Sentiment analysis

First research related to sentiment analysis applied in the financial sector was conducted using IndoBERT, which to identify news related to stock price movements, obtained an accuracy of 68%, surpassing the performance of other algorithms such as logistic regression, discriminant linear analysis, k-nearest neighbors, decision tree, support vector machine, random forest, XGboost, naïve Bayes, long short-term memory, and multi-layer perceptron [17]. The second research used a lexicon approach to analyze sentiment from German tweets related to companies by combining SentiWortschatz (German) and SentiWordNet 3.0 (English) enhanced with GermaNet resulting in an accuracy of 59.19% and surpassing random classification performance [22]. Third research has created a model to identify credit risk through financial news by extracting sentiment from electra-based target entity sentiment (TES) which analyses news related to companies that are in the process of applying for credit. This model produces an F1-score of 77% where the best performance is in the organization tag [23].

Fourth research created a model for identifying credit risk from news that combines source-latent Dirichlet allocation (LDA), named entity recognition (NER), and target aspect-based sentiment (TABSA) using bidirectional encoder representations from transformers (BERT). This combination model proves that sentiment scores strengthen traditional predictions that only use structured financial variables and increase the predictive power of companies leaving banks other than due to bankruptcy [24]. The fifth research identifies credit risk using news making the Risky News Index by aggregating news scores related to credit, investment strategies, and funding based on the word vector distance resulting from word embeddings from financial news created using the Word2Vec method to obtain score aggregation. News is positively correlated with classic financial reports [16].

2.2. Theoretical gaps

2.2.1. Inaccuracy of EWS model made with structured financial variables only

The COVID-19 pandemic has made global economic conditions uncertain. The use of EWS at that time became less accurate because it did not consider the external events that were happening. An innovative approach is needed where EWS and external sentiment do not run individually.

2.2.2. Specific terms in the financial domain and language structures vary between texts in different languages

Various NLP studies have been conducted to improve sentiment analysis specifically for the financial domain. However, if the language of the text used is different, a model specific to that language is needed. Also, the unavailability of datasets specific to the financial domain is still a big challenge. An appropriate approach is needed so that the resulting sentiment analysis can be useful.

2.2.3. Sentiment analysis research in identifying credit risk limited to assess new debtor

Sentiment analysis research in identifying credit risk only analyzed debtors who will apply for credit and proved that there was a correlation with traditional financial data. Due to the shortcomings of these studies, it is necessary to have a method for detecting signals of deterioration from existing debtors. It can capture future-looking credit risks by creating an EWS that combines traditional data from financial reports with the latest financial news so it can reflect the current economic situation.

2.3. Conceptual model

When the bank gives credit, ideally the debtor will pay according to the payment schedule. If the credit payment follows the payment schedule, then the credit quality is 'pass'. If the payment is delinquent by 1-90 days from the payment schedule, then the credit quality is 'special mention'. Meanwhile, if the payment is delinquent by more than 90 days from the payment schedule, then the credit quality is a ‘non-performing loan’. Credit quality in the coming months will be the target of EWS prediction with independent variables from the internal model. In addition, from those EWS studies, logistic regression is widely used because of its ease of explanation [25]. After other more advanced techniques were found, such as rough set, neural network, and boosted model, their accuracy outperforms the logistic regression, but their explanation is still not as easy as logistic regression.

The emergence of transformer has made it easier for the NLP community to develop pre-trained models called large language models (LLM) which use a lot of text in its creation process [26]. LLM can simplify the model creation process where you do not have to train from scratch because it just needs to fine-tune/transfer learning from LLM to the dataset to extract features from the dataset. Financial news in the Indonesian language is still a challenge due to the unavailability of datasets specific to that domain [27]. This limitation can be overcome by fine-tuning/transfer learning the LLM model with a specific dataset on financial news to be able to capture the context, so the result of sentiment related to economic fluctuations reflects the external events that are currently occurring. This has been implemented in FinBERT by fine-tuning BERT using a dataset of English financial texts which is proven to improve the performance of regular BERT on financial texts [28], [29].

2.4. Innovation

The main contribution of this paper is to make a sentiment analysis model from the Indonesian language LLM which is specifically for the economic domain by continuing the learning process of the original IndoBERT using Indonesian financial news. IndoBERT, where this model can surpass the multilingual performance of BERT and MalayBERT in extracting sentiment from general Indonesian language text [30]. The upgraded IndoBERT is used to extract features from the financial news sentiment analysis dataset.

After the sentiment is generated, it will be added to the bank debtor dataset to predict the credit quality in the next 2 years using the CatBoost algorithm. CatBoost is used as an algorithm because it has features that prevent overfitting and is efficient when the inference process is carried out. In the end, the result of decision tree is symmetrical [31], [32]. This model will be the EWS.

EWS testing is divided into 2 modes where EWS without sentiment results and EWS with sentiment results. The aim is to compare the performance of the traditional model with the other model that has been added with sentiment analysis results. In addition, the EWS that was created investigated whether it uses sentiment analysis results to create forward-looking using Shapley additive explanations (SHAP) so that it can create a better credit risk mitigation process by providing warnings before the debtor goes bankrupt.

3. METHOD

3.1. Data collection

The data collection is carried out by web scraping the CNBC Indonesia news portal with the Scrapy package using the Python programming language in Google Collaboratory (Python 3.9 version). The collected news is the news that was published from 2018-2022. The results are saved every month in a file in CSV format. After obtaining the news data, it was randomized and then annotated into 3 labels (negative, neutral, and positive) based on the detailed content of the news article with 500 instances of each label, so the total sentiment analysis dataset instances are 1,500 instances. Meanwhile, EWS data uses corporate segment credit data from one of the private banks in Indonesia from the 2018-2022 period. Figure 1 shows the whole flow of the proposed method.

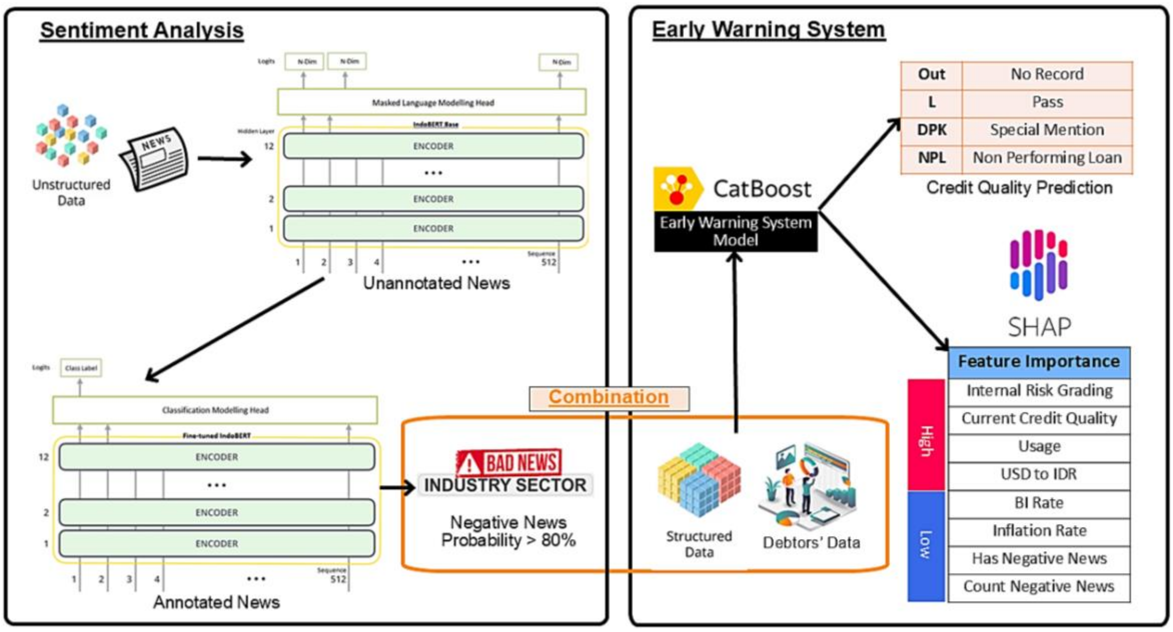

Figure 1. Proposed method flow

3.2. Pre-processing

Start from pre-processing the text data by case-folding to equate all characters used in news articles to lowercase using the str.lower() function and removing meaningless repetitive sentences using the replace() function from the Pandas package (2.2.0 version). The next step is removing numbers and punctuation using the Regex package (2023.12.25 version). After that, correcting non-standard words into standard words using an Alay language dictionary from the research [33]. After correction, continued to the stopwords removal to remove meaningless words such as conjunctions using the Indonesian stopword dictionary from the NLTK package (3.8.1 version). The last step is extracting the root words from words in news articles using the Sastrawi package (1.0.1 version). Pre-processing the dataset for the language model does not require stemming to maintain paragraphs that can be understood by humans.

Pre-processing on the EWS dataset was carried out by adding sentiment analysis results into the EWS dataset on RStudio (R 4.2.0 version). The process of aggregating sentiment analysis results in the EWS dataset using the dplyr (1.0.9 version) to filter negative sentiment results which have a probability prediction above 80% and str_detect function from tidyverse packages (1.3.1 version) to detect debtor's industrial sector. Table 1 shows the details of EWS variables.

Table 1. Variable of EWS dataset

Variable	Data type	Explanation
Internal risk grading	Numeric	Low risk-high risk: 1-10
Current credit quality	Categorical	Delinquent 0 day: 'Pass' 1-90 days: 'Special Mention' > 90 days: 'Non-performing Loan'
Usage	Numeric	Credit outstanding divided by credit plafond
USD to IDR	Numeric	Exchange rate from USD to IDR
Bi rate	Numeric	The prevailing interest rate of the central bank in Indonesia
Inflation rate	Numeric	Indonesia's inflation rate
Has negative news	Categorical	Debtors have negative news based on their industry sector If any: 'Yes' On the contrary: 'No'
Count negative news	Numeric	Total count of debtor's negative news based on its industry sector

3.3. Modelling

The modelling process starts with the unannotated news data to create an IndoBERT model that is specific to the economic domain by fine-tuning an existing IndoBERT model (indobenchmark/indobert-base-p2) provided by IndoBenchmark on the hugging face portal into a new language model. Next, the new language model became a model that extracts features from the sentiment analysis dataset and the results became a model for predicting news sentiment. Both processes are carried out using the simple transformers package (0.64.3 version).

The data testing method used for modelling news sentiment analysis uses a percentage split test method of 90% where 1,350 instances as training data are taken, 75 instances of which are taken as validation data and 150 instances as test data. Hyperparameter tuning was also conducted on the sentiment analysis using learning rate, epoch, and regularization values. The batch size and max sequence length values are made fixed where the batch size used is 16 while the max sequence length is 512 which is the maximum capacity of a BERT-based architecture to receive sequential data. The optimizer that is used is Adam W. In addition, early stopping is used to prevent overfitting by monitoring the increase of Mathews correlation coefficient (MCC) value which has been proven to be reliable for assessing the model proportionally for each class element in the dataset [34]. All parameters are summarized in Table 2.

Table 2. Parameters of sentiment analysis modelling

Parameter	Value
Batch size	16
Optimizer	Adam W
Early stopping	MCC (delta=0.01)
Max sequence length	512
Learning rate	5e-5-1e-5
Epoch	2-5
Regularization	0-0.2

After the sentiment analysis model is created, the EWS dataset starts to be pre-processed by adding variables from the sentiment analysis results. EWS modelling is carried out using the CatBoost algorithm with a percentage spilt testing method of 80% and 100 iterations for predicting the debtor's credit quality for the next 3, 6, 9, 12, 15, 18, 21, and 24 months. The process of dividing the dataset for modelling used the scikit-learn (sklearn) package (1.0.2 version) and the CatBoost algorithm from the Catboost package (1.2.3 version) on the Jupyter Notebook (8.2.0 version).

3.4. Performance measurement

The process of measuring language model can use log probability and perplexity as evaluation metrics where x^t is a token at t-position and p is the probability [35]. The sentiment analysis model and the EWS model use the confusion matrix to obtain precision, recall, and accuracy values for a classification model. Table 3 shows the confusion matrix of 3 classes.

$$Log\ Probability = \sum_{t=N}^{1} log_2\, p(x_t|x_{<t}) \tag{1}$$

$$Perplexity = P(x_1, x_2, \dots, x_n) = 2^{-\frac{1}{N}\sum_{t=N}^{1} \log_2 p(x_t|x_{<t})} \tag{2}$$

Table 3. Confusion matrix of 3 classes

Actual	Prediction		
	Class 1	Class 2	Class 3
Class 1	True C1	False C2a	False C3a
Class 2	False C1a	True C2	False C3b
Class 3	False C1b	False C2b	True C3

$$Precision = \frac{\frac{True\ C1}{True\ C1+False\ C1a+False\ C1b}+\frac{True\ C2}{False\ C2a+True\ C2+False\ C2b}+\frac{True\ C3}{False\ C3a+False\ C3b+True\ C3}}{3} \tag{3}$$

$$Recall = \frac{\frac{True\ C1}{True\ C1+False\ C2a+False\ C3a}+\frac{True\ C2}{False\ C1a+True\ C2+False\ C3b}+\frac{True\ C3}{False\ C1b+False\ C2b+True\ C3}}{3} \tag{4}$$

$$Accuracy = \frac{True\ C1 + True\ C2 + True\ C3}{True\ C1 + False\ C1a \dots C1b + True\ C2 + False\ C2a \dots C2b + True\ C3 + False\ C3a \dots C3b} \tag{5}$$

To have a better understanding of how the EWS model works whether it utilizes the results of sentiment analysis or not, a search for feature importance is carried out to know how the model makes decisions using TreeSHAP from the SHAP package. TreeSHAP is designed to extract Shapley values from specific tree-based models so that it is suitable for the CatBoost algorithm [36], [37].

4. RESULTS AND DISCUSSION

While earlier research successfully modeled the classic financial reports to catch the signal of deterioration, and the financial news has a positive correlation with it, no research has combined the financial reports with the financial news. Start by making a sentiment analysis model from financial news then the result is added to the financial reports. After it is added, it will be the dataset of enhanced EWS.

4.1. Sentiment analysis model

From web scraping from the CNBC Indonesia news portal, there are 258,678 news collected, and it was used as a dataset to fine-tune IndoBERT into a new language model. As shown in Table 4, the results of the new language increase log probability by 7.04 and reduce perplexity by 58,275.74, indicating that the results of fine-tuning IndoBERT into a new language model can recognize financial news text better than the original IndoBERT model without fine-tune. The sentiment analysis created from the new language model produces the best results on hyperparameters learning rate=3e-5, regularization=0.1, and epoch=5 with precision of 80.89%, recall of 80.67%, and accuracy of 80.67% on test data. This result shows that fine-tuning the new language model can achieve better results compared to the original language model.

As shown in Table 5, this result is higher compared to other similar previous research from Anderies *et al.* [17] which obtained an accuracy of only 68% on test data. This model is used to generate news sentiment and add it to the EWS dataset. However, this good result can be obtained because the proposed method continued the process of original language model learning, so it has a better understanding of the dataset. The proposed method also only uses one source of financial news which can have the same writing style. Future research may add more resources of data to give more variation to give a better generalization.

Table 4. Result of fine-tune IndoBERT into language model

Fine-tune status	Log probability	Perplexity
Before	-10.97	58,326.80
After	-3.93	51.06

Table 5. Comparison to others similar research

Researcher	Result (Accuracy) (%)
Anderies *et al.* [17]	68
This Research	81

4.2. EWS model

A summary of the performance of the EWS model can be seen in Table 6. From the results of the EWS modelling, the precision, recall, and accuracy values after adding the sentiment analysis results are mostly higher than before adding the sentiment analysis results so the EWS performance after adding the sentiment analysis results produces better performance. However, over time the predicted credit quality still shows poor results. In addition, EWS before and after adding sentiment analysis results always has higher precision compared to recall, indicating that EWS has described the information provided even though the level of information retrieval is quite low.

Table 6. Result of fine-tune IndoBERT into language model

Credit quality next	Without sentiment analysis result			With sentiment analysis result			Difference
	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)	(Accuracy) (%)
3 Months	87.40	68.37	97.51	88.39	69.24	97.56	0.05
6 Months	84.09	63.29	95.56	85.38	63.60	95.69	0.13
9 Months	77.46	55.82	93.59	78.64	55.28	93.62	0.03
12 Months	79.95	53.94	91.89	80.14	52.65	91.92	0.03
15 Months	72.47	44.33	89.87	72.95	43.62	89.89	0.02
18 Months	74.80	41.76	88.44	72.06	43.11	88.46	0.02
21 Months	71.10	41.46	87.31	74.78	43.11	87.55	0.24
24 Months	71.31	38.30	86.30	75.14	42.90	86.96	0.66

The EWS model added with the sentiment analysis results was then analyzed using TreeSHAP to assess the feature importance of the sentiment analysis results based on the size of the Shapley value. Figure 2 shows the movement of the average Shapley value of the entire EWS dataset as the predicted period increases. From the figure, the highest average Shapley value is predicting credit quality for 12 months which means that EWS is most effective in predicting credit quality for the next 12 months.

The variable of whether there is a presence or absence, and the total of negative news based on the debtor's industrial sector has a small average Shapley value when compared to other variables. It means that the sentiment analysis result is not the main component of the EWS model to predict the credit quality in the future where it is clear from the graph that the average Shapley value of the usage and internal risk grading variables is greater than the other variables. However, the average Shapley value of the variable whether there is presence or absence and the amount of negative news based on the debtor's industrial sector tends to increase, which means the importance of the variable resulting from sentiment analysis increases as the predicted period increases, which indicates that EWS leverages the results of sentiment analysis from news related to economic fluctuations to asses future conditions so that it can support a better credit risk mitigation process by providing warnings before the debtor’s credit quality start to deterioration. These results prove that adding sentiment analysis results from financial news to financial reports data can create a forward-looking EWS. Future research may study another method to produce EWS with reliable performance of information retrieval over time the predicted credit quality.

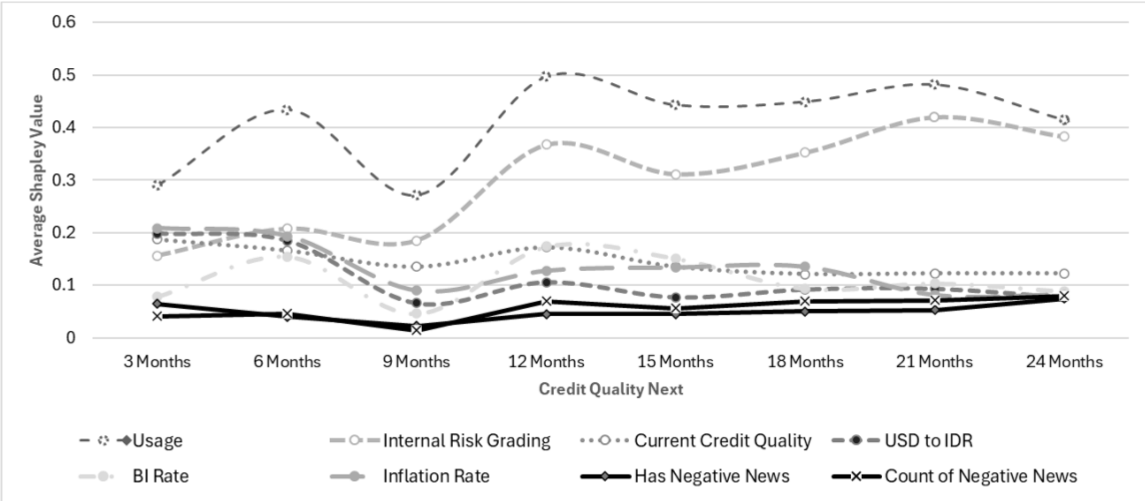

Figure 2. Average Shapley value of EWS model after adding sentiment analysis results

5. CONCLUSION

This research successfully improved EWS model accuracy by adding the results of news sentiment indicated by the mostly higher precision, recall, and accuracy values than before adding the sentiment analysis. Apart from that, the average Shapley value of the sentiment analysis result increases as the predicted period increases. It means that the EWS utilizes the results of sentiment analysis from news related to economic fluctuations to see future conditions that make the EWS forward-looking. The future direction of this research would use big data technology to simplify the process of combining financial reports with the financial news in real-time to accelerate credit risk mitigation.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Angel Karentia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Derwin Suhartono		✓				✓	✓			✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

DATA AVAILABILITY

News data are available from the corresponding author, [AK], upon reasonable request. The debtor data supporting the findings of this study cannot be published due to the strictness of Indonesian banking laws and regulations, especially regarding the protection of personal data and confidentiality of customer data. Furthermore, we have conducted a screening process of the debtor data/information used in this research and ensured that the data/information does not violate the debtor's privacy/personal rights and compliance with banking regulations in Indonesia.

REFERENCES

[1] S. Naseer, S. Khalid, S. Parveen, K. Abbass, H. Song, and M. V. Achim, "COVID-19 outbreak: impact on global economy," *Frontiers in Public Health*, vol. 10, 2023, doi: 10.3389/fpubh.2022.1009393.

[2] World Bank, "World development report 2022: finance for an equitable recovery," *World Development Report*, Washington: United States, vol. 1, 2022.

[3] A. N. Berger, P. Molyneux, and J. O. S. Wilson, *The Oxford handbook of banking,* Oxford, United Kingdom: Oxford University Press, 2014, doi: 10.1093/oxfordhb/9780199688500.001.0001.

[4] E. J. Casabianca, M. Catalano, L. Forni, E. Giarda, and S. Passeri, *An early warning system for banking crises: from regression-based analysis to machine learning techniques*, Padova, Italy: Marco Fanno Working Papers, 2019.

[5] Galytix and PwC, "Banks must act on their early warning systems or risk ROE downturn in association with," *Galtix Limited,* London, England, 2022.

[6] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: a literature review," *Risks*, vol. 7, no. 1, 2019, doi: 10.3390/risks7010029.

[7] M. E. K. Agoraki, N. Aslanidis, and G. P. Kouretas, "U.S. banks' lending, financial stability, and text-based sentiment analysis," *Journal of Economic Behavior and Organization*, vol. 197, pp. 73–90, 2022, doi: 10.1016/j.jebo.2022.02.025.

[8] S. Kim, S. Choi, and J. Seok, "Keyword extraction in economics literatures using natural language processing," *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN),* Jeju Island, Korea, Republic of, 2021, pp. 75-77, doi: 10.1109/ICUFN49451.2021.9528546.

[9] B. Liu, *Sentiment analysis: mining opinions, sentiments, and emotions, second edition*. Cambridge, United Kingdom: Cambridge University Press, 2020.

[10] A. Rahali and M. A. Akhloufi, "End-to-end transformer-based models in textual-based NLP," *AI*, vol. 4, no. 1, pp. 54–110, 2023, doi: 10.3390/ai4010004.

[11] T. Wolf *et al.*, "Transformers: state-of-the-art natural language processing," *EMNLP 2020 - Conference on Empirical Methods in Natural Language Processing, Proceedings of Systems Demonstrations*, Association for Computational Linguistics, pp. 38–45, 2020, doi: 10.18653/v1/2020.emnlp-demos.6.

[12] A. Costea, "An early-warning system for financial performance predictions," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 56, no. 2, pp. 5–20, 2022, doi: 10.24818/18423264/56.2.22.01.

[13] A. Siddique, M. A. Khan, and Z. Khan, "The effect of credit risk management and bank-specific factors on the financial performance of the South Asian commercial banks," *Asian Journal of Accounting Research*, vol. 7, no. 2, pp. 182–194, 2022, doi: 10.1108/AJAR-08-2020-0071.

[14] A.-I. Străchinaru, "Early warning systems for banking crisis and sovereign risk," *Journal of Financial Studies and Research*, pp. 1–9, 2022, doi: 10.5171/2022.441237.

[15] S. Ben Jabeur, C. Gharib, S. Mefteh-Wali, and W. Ben Arfi, "CatBoost model and artificial intelligence techniques for corporate failure prediction," *Technological Forecasting and Social Change*, vol. 166, pp. 1-19, May. 2021, doi: 10.1016/j.techfore.2021.120658.

[16] P. Labonne and L. A. Thorsrud, *Risky news and credit market sentiment*, Oslo, Norway: BI Norwegian Business School, 2023.

[17] Anderies, R. Rahutomo, and B. Pardamean, "Finetunning IndoBERT to understand indonesian stock trader slang language," *Proceedings of 2021 1st International Conference on Computer Science and Artificial Intelligence, ICCSAI 2021*, pp. 42–46, 2021, doi: 10.1109/ICCSAI53272.2021.9609746.

[18] N. Luangnaruedom and S. Prakancharoen, "Loan repayment status prediction," *International Journal of Innovation, Creativity and Change,* vol. 15, no. 4, pp. 133-153, 2021.

[19] Y. Wang, "Research on commercial bank risk early warning model based on dynamic parameter optimization neural network," *Journal of Mathematics*, vol. 2022, 2022, doi: 10.1155/2022/9754428.

[20] Y. Cheng, Q. Yang, L. Wang, A. Xiang, and J. Zhang, "Research on credit risk early warning model of commercial banks based on neural network algorithm," *Procedia Computer Science*, vol. 199, pp. 1168–1176, May 2024, doi: 10.1016/j.procs.2022.01.148.

[21] H. Wei and X. Wang, "Financial risk management early-warning model for Chinese enterprises," *Journal of Risk and Financial Management*, vol. 17, no. 7, 2024, doi: 10.3390/jrfm17070255.

[22] A. Mengelkamp, K. Koch, and M. Schumann, "Creating sentiment dictionaries: process model and quantitative study for credit risk," *9th European Conference on Social Media, ECSM 2022*, pp. 121–129, 2022, doi: 10.34190/ecsm.9.1.167.

[23] N. Ahbali, X. Liu, A. A. Nanda, J. Stark, A. Talukder, and R. P. Khandpur, "Identifying corporate credit risk sentiments from financial news," *NAACL 2022 - 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Industry Papers*, pp. 362–370, 2022, doi: 10.18653/v1/2022.naacl-industry.40.

[24] J. C. Duan and X. Yao, "Media sentiments for enhanced credit risk assessment," *NUS Credit Research Initiative*, pp. 1-26, 2022.

[25] F. Gambarota and G. Altoè, "Ordinal regression models made easy: A tutorial on parameter interpretation, data simulation and power analysis," *International Journal of Psychology*, vol. 59, no. 4, pp. 1263-1292, 2024, doi: 10.1002/ijop.13243.

[26] A. Dash, *The LLM advantage: how to unlock the power of language models for business success: a practical guide to leveraging AI for innovation, growth, and competitive edge*, 1st ed. Thane West, India: Grazing Minds Publishing, 2023.

[27] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of sentiment analysis in finance: from lexicons to transformers," *IEEE Access*, vol. 8, pp. 131662–131682, 2020, doi: 10.1109/ACCESS.2020.3009626.

[28] D. Araci, "FinBERT: financial sentiment analysis with pre-trained language models," *M.Sc. Thesis,* Faculty of Science, University of Amsterdam, Amsterdam, Netherlands, 2019.

[29] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, "FinBERT: A pre-trained financial language representation model for financial text mining," *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20): Special Track on AI in FinTech*, pp. 4513–4519, 2020, doi: 10.24963/ijcai.2020/622.

[30] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: a benchmark dataset and pre-trained language model for indonesian NLP," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.

[31] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00369-8.
[32] L. Owen, *Hyperparameter tuning with Python*. Birmingham, United Kingdom: Packt Publishing, 2022.
[33] M. O. Ibrohim and I. Budi, "Multi-label hate speech and abusive language detection in indonesian twitter," *Proceedings of the Third Workshop on Abusive Language Online*, 2019, pp. 46–57, doi: 10.18653/v1/w19-3506.
[34] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, 2020, doi: 10.1186/s12864-019-6413-7.
[35] Y. Pruksachatkun, M. Mcateer, and S. Majumdar, *Practicing trustworthy machine learning: consistent, transparent, and fair AI pipelines*, 1st ed. California, United States: O'Reilly Media, Inc, 2023.
[36] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.
[37] C. Molnar, *Interpretable machine learning: a guide for making black box models explainable*, 2nd ed. North Carolina, United States: Lean Publishing, 2020.

BIOGRAPHIES OF AUTHORS

Angel Karentia is a graduate student of computer science from Bina Nusantara University, Indonesia. Currently, she is working as a data analyst specializing in risk management in the financial industry in Indonesia. Her research interest includes artificial intelligence, machine learning, and deep learning. She can be contacted at email: angel.karentia@binus.ac.id.

Derwin Suhartono is a faculty member of Bina Nusantara University, Indonesia. He got his Ph.D. degree in computer science from Universitas Indonesia in 2018. His research fields are natural language processing. Recently, he has been continually doing research in argumentation mining and personality recognition. He is actively involved in the Indonesia Association of Computational Linguistics (INACL), a national scientific association in Indonesia. He has his professional memberships in ACM, INSTICC, and IACT. He also takes the role of reviewer in several international conferences and journals. He can be contacted at email: dsuhartono@binus.edu.