International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 3
ISSN: 2088-5334

Comparison of Classification Algorithm and Language Model in Accounting Financial Transaction Record: A Natural Language Processing Approach

Bagas Adi Makayasa [a,1], Maria Ulfah Siregar [a,2], Bambang Sugiantoro [a], Agung Fatwanto [a]

[a] *Department of Informatics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia*
Corresponding author: [1]bagas13am@gmail.com; [2]maria.siregar@uin-suka.ac.id

Abstract—**The problem of financial recording not following the principles of accounting science has the potential to cause unnecessary problems. However, micro, small, and medium enterprises with their distinctive characteristics, though not all, still face many obstacles in writing financial reports. Even though there is already much financial software available, our study aims to investigate opportunities for implementing automation of accounting financial transaction records using the NLP approach, to interpret financial transactions based on text written on the transaction form into accounting journals (debits and credits). Experiments were carried out by comparing the performance of three classification algorithms, namely SVM, K-Nearest Neighbor, and Random Forest, with traditional (TF-IDF and BOW) and contextual (Word2Vec) Language Models. There are 200 financial transaction datasets consisting of ten classes. The data is divided into two parts, namely, the balance dataset and the imbalance dataset. The pair SVM and Word2Vec in the balanced dataset gave the highest accuracy (92.5%), precision (92.5%), recall/sensitivity (93.33%), and F1 score (92%). However, compared with the results of related semantic research (the average performance reaches 95%), the results obtained in this study are still lower. One point that may have a significant effect is the amount of data in the corpus, which is still lacking. Researchers suggest increasing the number of datasets and using a combination of other language models such as Glove, Bert etc. This study can also be used as a model for more complex financial transaction cases in future research.**

Keywords—**Financial recording; automation transaction record; financial transaction datasets; balance dataset; imbalance dataset.**

Manuscript received 18 Jun. 2023; revised 24 Mar. 2024; accepted 21 Apr. 2024. Date of publication 30 Jun. 2024.

I. INTRODUCTION

The Indonesian Ministry of Cooperatives recorded 64.2 million Micro, Small, and Medium Enterprises (MSMEs) in 2018 and 729 large-scale companies in 2022 [1]. The large number of MSMEs shows how vital the role of the people's economy is as one of the pillars of a country's economy. As a business institution, it cannot be separated from financial transactions, where companies must have financial documentation regarding the circulation of their money. In accounting science, there are several financial reports such as profit and loss reports, capital changes reports, accounts payable reports, and others [2]. Good and standardized financial reports are helpful for business planning, financial position information, cost control, investment funding considerations, business decision-making, tax calculations, and so on. However, financial records in MSMEs cannot be seen as large companies.

Organizationally, MSMEs are unlike large companies where the division of work is well organized according to each division. MSME actors can do many jobs, for example, in a sales, financial, or technical division, all at once in one day. The irregularity of this work system is undoubtedly one of the reasons why MSMEs do not overthink about making sound financial reports. Whereas accounting strictly emphasizes that every transaction must be recorded one by one and continuously [3]. So that no transaction history is lost because it will impact the financial reports produced later.

Researchers interviewed four accountant representatives and MSME actors to determine MSME behavior in financial transaction activities. The results show that MSMEs face obstacles in understanding accounting knowledge even though there is a lot of accounting software on the market [4]. Apart from that, each accounting software has a different User Interface (UI) and User Experience (UX), making it difficult for ordinary users to operate it [5].

From the existing problems above, researchers offer an alternative for recording financial transactions using the Natural Language Processing (NLP) approach for ordinary users or 'stupid users'—those who are not accountants and are untrained in accounting skills. The main goal of applying NLP in this research is to enable computers to understand information about financial transactions, such as the expertise of accountants [6]. It is as simple as when MSME actors say, "Buy a computer for two million rupiahs in cash," NLP makes a journal according to accounting standards. These journals will later be processed again to become financial reports.

In 2020, Mugisha and Paik conducted medical document retrieval research using a comparison between the TF-IDF Language Model, Global Vectors for Word Representation (Glove), and Bidirectional Encoder Representations from Transformers (BERT). Data cleaning techniques generally include changing text to lowercase, regular expressions and word replacement, punctuation marks, and removing non-alphanumeric characters (Stopwords Removal). In the training phase, the researcher divided the dataset by 90% as training data and 10% as test data. The final result shows that the combination of TF-IDF non-contextual Language Model (BOW &) achieves an accuracy of 80%, and BERT obtains the best accuracy of 98.2%. This study suggests further research to examine the problem of retrieving multilingual documents using broader query techniques and more extensive data to test performance. This study states that one of the drawbacks is that data does not have an even number of classifications (minority class is only represented by 29% of the data) for each class. Hence, it affects the prediction results. In addition, there is input with long sentences. Therefore, words must be simple (maximum 380 words) to make them easier to process. The results show that BERT performs better than the non-contextual Language Model [7].

Amin et al. [8] researched detecting tweets for sentiments of dengue fever and flu using Long Short Term Memory (LSTM) combined with word2vec with the Skip-gram (SG) technique and Word2Vec with Continuous-bag-of-words (CBOW). Before the data is processed, preprocessing is first carried out, which includes lowercase, stopword removal, stemming, and tokenization. This research uses 6.000 Twitter data obtained from scraping and already tagged. The dataset is divided into 3 parts: the trainset, validation dataset, and test dataset. Then, divide the training data and testing data with a ratio of 80:20. Based on the results carried out in this paper LSTM Word2Vec with CBOW, the confusion matrix shows the best results, namely 94% compared to LSTM with the Word2Vec SG Language Model technique. In this case, CBOW works efficiently for medium-sized text data, while the SG model works for large corpus. This research will try the CNN approach with different Language Models and optimization for future research.

Furthermore, Jayaratne and Jayatilleke in 2020 with open interview questions in case studies of NLP-based employee recruitment using a comparative model using seven languages such as TF-IDF and Latent Dirichlet Allocation (LDA). Using 4600 data, this study uses the SVM and Random Forest algorithms with tokenization preprocessing, stop word removal, lowercase, and lemmatization techniques that get the most optimal results of 87.83% is TF-IDF. This study uses a semantic approach and recommends using regression algorithms and Neural Networks in future work. It is also recommended to explore other types of features, such as the use of part of speech (POS), the use of emojis, etc., which will likely provide increased accuracy. In addition, information such as audio and video signals captured when job candidates answer questions can also be explored as signals to enhance text-based personality inference [9].

Until this research was written, research related to NLP and accounting that had been found was rare. One related study is entitled "Automated Interpretation of Accounting Data Based on Natural Language Processing" by Iswandi et al. [10]. However, Iswandi's research does not explain the form of the corpus, a systematic process from input to output, and the uses of algorithms. This research generally has the same ideas as Iswandi's but has different case studies, datasets, language models, and algorithms.

II. Materials and Methods

This research was conducted using a quantitative approach and a data collection process that was carried out through experiments [11]. The data is analyzed by a descriptive-arithmetic approach to calculate the performance of the sixteen types of combinations using classification algorithms and language models.

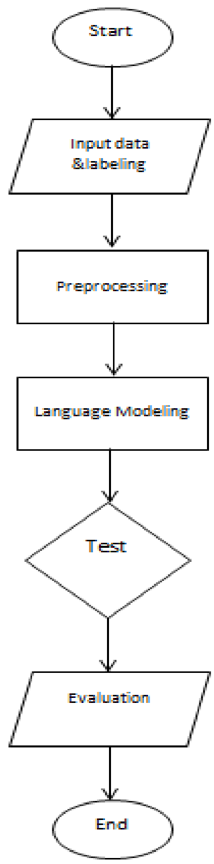

Fig. 1 Flowchart research method

In contrast, the performance evaluation instrument of this research uses the confusion matrix [12]. The research started with inputting corpus data in text interviews with MSME actors and the labeling process (Supervised Learning) by accountants as experts. Furthermore, the corpus data is tidied up through pre-processing before the language model stage is carried out. The data is then divided into two forms, namely 'training data' and 'testing data' before being tested using the SVM, KNN, and Random Forest algorithms [13]. The final stage is to evaluate the performance of the model that has been created using the confusion matrix. The programming language used to create NLP is Python, while the Graphical User Interface (GUI) uses the PHP programming language. The research method flowchart is shown in Figure 1.

System design begins with user input of financial transaction data in the form of text or voice. These data are put together into a corpus, which the NLP engine then works to determine the predicted results of accounting journals as the output. From these journals, data is flowed back to the user to become various kinds of reports such as General Ledgers, Financial Position Reports etc. [14], depicted in Figure 2.

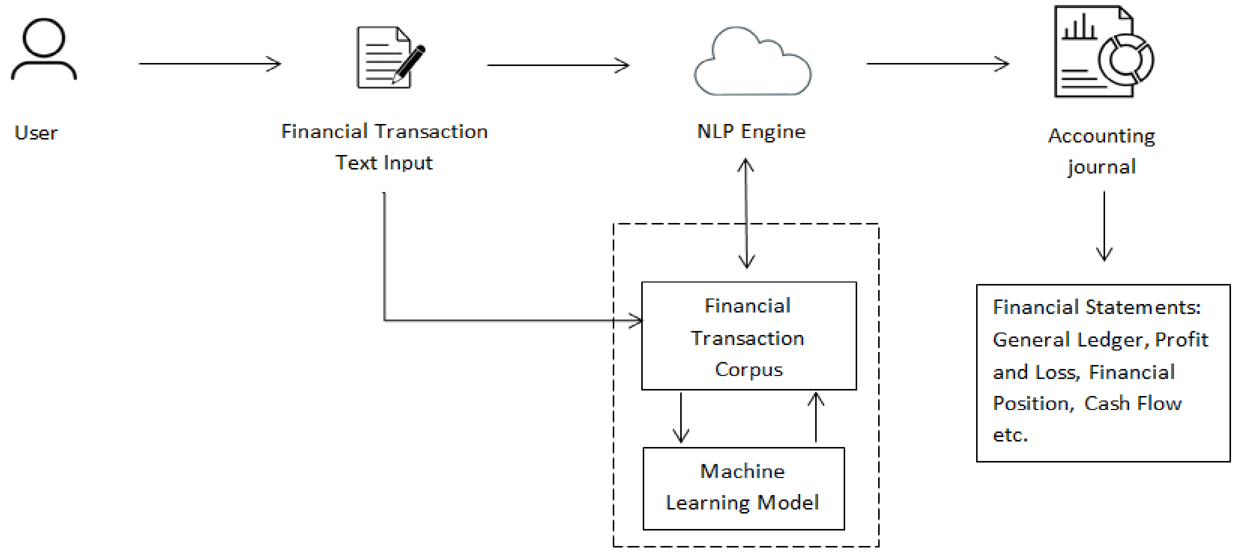

Fig. 2 System Design

A. *Input Data and Labelling*

Until this journal was written, no dataset related to the case studies was found. Researchers collected data by interviewing MSME actors about how they "talk" to conduct daily financial transaction activities. This data is labeled one by one by the accountant as an expert. [4], [14]. The collected dataset totals 200 data with output labeling and business sector categories (services, trade, and manufacturing) of MSMEs. From the dataset, there are ten classifications in debit (D) and credit (C), such as below [4]:

- Salary Expenses (D) - Cash (C)
- Other Expenses (D) - Cash (C)
- Prepaid Rent (D) - Cash (C)
- Cash (D) - Capital (C)
- Cash (D) - Income (C)
- Cash (D) - Receivables (C)
- Equipment (D) - Cash (K)
- Supplies (D) - Cash (C)
- Owner Returns (D) - Cash (C)
- Payables (D) - Cash (C)

Some business sectors have the same general financial transactions. For example, service, trading, and manufacturing companies both use the expression "paying employee salaries." Unlike the phrase "buying a laundry cupboard," which might be specifically meant in the laundry business sector, the sample corpus can be seen in Table 1.

TABLE I
EXAMPLE OF A LABELED CORPUS

No	Transaction Activity	Business	Journal Label
1	Paying employee salaries Rp. 2.500.000	Global	Salary Expenses (D) - Cash (C)
2	Purchased meeting meal consumption worth Rp. 100.000	Global	Other Expenses (D) - Cash (C)
3	Annual shop rent Rp. 5.000.000	Global	Prepaid Rent (D) – Cash (C)
4	Deposit Rp. 50.000.000 as initial capital	Global	Cash (D) - Capital (C)
5	Receiving orders for bridal make-up of Rp. 3.000.000	Salon	Cash (D) - Income (C)
6	Receive workshop service revenue of Rp. 350.000	Workshop	Cash (D) - Income (C)
7	Buy a laundry cupboard Rp. 1.000.000	Laundry	Equipment (D) - Cash (C)
8	Buy a travel car Rp. 120.000.000	Travel	Equipment (D) - Cash (C)
9	Pay for bed repairs	Lodging	Equipment (D) - Cash (C)
10	Providing services to customers Rp. 750.000	Global	Cash (D) - Receivables (C)

B. *Preprocessing*

Before testing, the corpus is pre-processed by tidying up the dataset so that NLP modeling is more optimal [15]. There are several stages of preprocessing as follows: lowercase, tokenizing, stopwords removal, and stemming,

C. Language Model

It is easy for humans to understand the linkage of words in linguistic terms, but computers are not as simple as that [16]. For example, humans understand words like "king" and "queen," "man" and "lady," and "tiger" and "lion" to have a certain kind of relationship between them. However, the computer needs a range of actions to figure that out, and this is where the language model comes into play in natural language processing (NLP). A language model is a core component of modern NLP which applies a statistical approach to analyze human language patterns in word prediction [17]. The language model works by determining the probability of the next word by analyzing the role of the word in a sentence. These models interpret words by inputting them into the algorithm. Then, the algorithm creates rules to translate the intent of the entered text. This study uses non-contextual language model (BOW and TF-IDF) and contextual language model (Word2Vec), while the explanation is as follows:

1) Bag of Words: Bag of Words (BOW) can be identified as one of the most straightforward language models in representing words numerically in the form of true or false [18]. For example, in the sentence, "I like to play football on the field on weekends", BOW will break (separated by spaces) the sentence into 8 words (unique words) out of 9 words. Each word will be repeated as many times as there are words; if the first word is "me" and matches the word order, it will be marked with 1, and the rest are zeros [1, 0, 0, 0, 0, 0, 0]. Then proceed with the word "like," it will produce encoding [0, 1, 0, 0, 0, 0, 0], and so on until the last word. One of the weaknesses of BOW is not understanding the relationship between words or standing alone (syntax) [19].

2) Term Frequency – Inverse Document Frequency (TF-IDF): Term Frequency – Inverse Document Frequency (TF-IDF) is a weighting technique that is carried out by calculating the amount (*frequency*) emergence of words (*term*) in sentences [20]. Suppose there is a document **j** and want to look for tokens **i**, then the number of occurrences i in th j $\boldsymbol{tf}_{\boldsymbol{i,j}}$ multiplied by the logarithm (log) total document **N** divided by the number of documents it contains i $\boldsymbol{df}_{\boldsymbol{i}}$. Formula 1 summarized it.

$$W_{i,j} = tf_{i,j} \times \log \frac{N}{df_i} \qquad (1)$$

3) Word2ec: One of the semantic language models is Word2Vec. For example, suppose Word2Vec is trained using a fairly complete corpus (Fig. 3). In that case, the vector representing the word "Indonesia" will be adjacent to the vector "Vietnam" in the context of the country, just as the vector "Cat" will be adjacent to the vector "Rabbit" in the context of animals. A computer may do this if the representation of the vector value is "Vietnam," for example, 0.9234, and "Indonesia," 0.85234. While "Cat" is 0.2342 and "Rabbit" is 0.3878. Word2Vec uses a Neural Network to get these vectors.

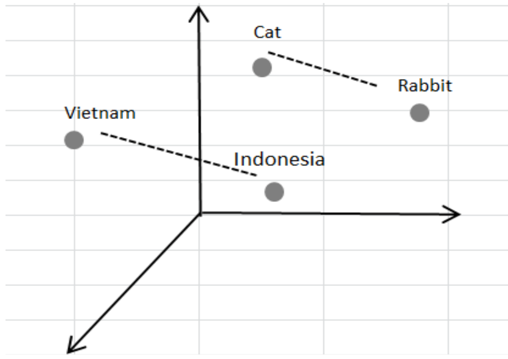

Fig. 3 Example of Word2Vec representation

Word2vec architecture consists of three layers: Input, Projection (hidden layer), and Output [21]. Input to Word2vec is one hot encoding vector with a length equal to the number of unique words in the training data. There are two types of neural network architectures from Word2Vec, namely Skip-gram and Continuous Bag of Words (CBOW), shown in Figure 4.

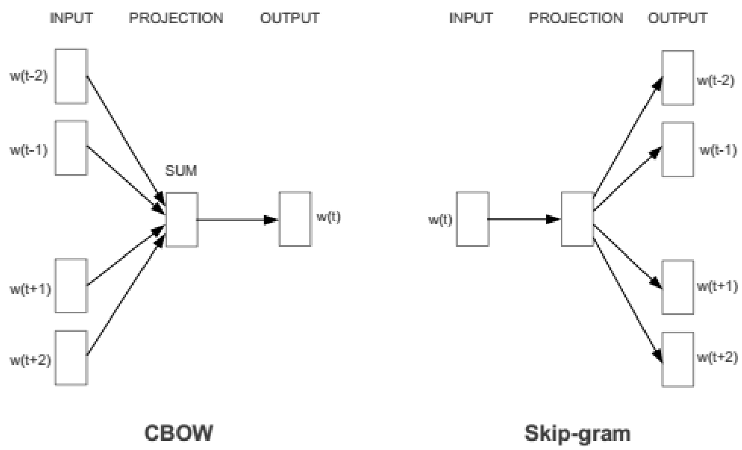

Fig. 4 Word2Vect Architecture

D. Supervised Learning Algorithm

Based on the case studies, this research uses supervised learning with a classification approach, where datasets are labeled before predictions are made.

1) Support Vector Machine: Support Vector Machine looks for a hyperplane (separator function between classes) in a support vector (the two closest data from different classes) with the max-margin or the most significant distance. The SVM model for this study was formed using the kernel Radial Basis Function (RBF) for non-linear data classification [22].

2) K-Nearest Neighbour: KNN is an algorithm that classifies output based on the quantity k nearest neighbour. This study uses cosine similarity as a distance calculation. $A\_(\ i)$ shows the term number i in document A, $B\_(\ i)$ is the term at i in document B, and n is the number of unique terms in the dataset [23] (Formula 2).

$$similarity(A,B) = \frac{A.B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}} \quad (2)$$

3) Random Forest: Random Forest is a collection of decision trees or decision trees to carry out the selection process, where the decision tree will be divided recursively based on data in the same class [24]. This algorithm combines each tree from the decision tree, which is combined into one model. Determination of classification with the Random Forest is carried out based on the voting results of the tree formed. Making predictions with a random forest is closely related to using the Gini index. Formula 3 decides how the nodes in the decision tree branches.

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2 \quad (3)$$

E. Evaluation

Evaluation of the performance of this study uses a multiclass confusion matrix [25].

- True Positive (TP)
  The amount of data in the actual class is correct, and the prediction results are positive (relevant).
- True Negative (TN)
  The actual class has the correct amount of data, but the predicted results are negative (irrelevant).
- False Positive (FP)
  The amount of data in the actual class is wrong, but the prediction results are positive (irrelevant).
- False Negative (FN)
  The amount of data in the actual class is wrong, and the prediction results are negative (relevant).

Furthermore, calculations are carried out for accuracy, precision, recall or sensitivity, specificity, and F1 Score from the confusion matrix table. Table 2 reports these results.

TABLE II
TRIAL RESULT

Algorithm	Language Model	Imbalance Dataset 200 Data				Balance Dataset 100 Data				Balance Dataset 200 Data			
		Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
RF	BoW	47.06	37.54	37.54	35.61	70	69.79	73.12	64.67	85	84.15	80.24	81.27
	TF-IDF	55.88	43.86	41.93	42.07	75	70.83	75.62	67.66	82.5	81.33	79	78
	W2Vec CBOW	52.94	35.26	44.29	37.14	65	60	66.67	61.63	82.5	86.71	86.67	84.54
	W2Vec Skip Gram	52.5	32.64	39.83	35.4	75	79.63	72.22	73.33	87.5	87.67	87.25	85.9
K-NN 3	BoW	29.41	18.16	18.95	18.17	65	51.04	66.88	52.17	57.5	71.9	59	59.22
	TF-IDF	38.24	23.86	29.47	24.65	75	75	76.25	71.46	80	79.5	77.74	76.86
	W2Vec CBOW	44.12	32.87	37.78	33.62	75	62.96	72.22	66.67	82.5	87.92	84.83	83.81
	W2Vec Skip Gram	50	36.48	43.06	38.31	75	70.37	75.93	66.67	82.5	81.67	83.06	78.81
K-NN 5	BoW	32.35	17.19	21.58	18.95	65	51.04	66.88	52.17	55	61.39	44	45
	TF-IDF	41.18	28.6	32.11	29.12	75	75	76.25	71.46	70	75.28	72.32	68.98
	W2Vec CBOW	41.18	29.51	36.75	28.21	65	66.67	66.67	64.81	85	87	82.5	82.26
	W2Vec Skip Gram	60	47.67	51.82	47.44	70	57.92	60	56.81	87.5	81.67	81.67	79.33
SVM	BoW	55.88	42.11	44.56	42.25	70	69.79	73.12	64.67	82.5	80.5	79	78.65
	TF-IDF	61.76	40	41.93	40.06	75	70.83	75.62	67.66	82.5	86.33	87.74	83.07
	W2Vec CBOW	50	52.86	54.56	48	75	57.5	67.5	60.14	92.5	93.5	93.42	92.29
	W2Vec Skip Gram	60	50.9	51.97	48.62	70	65.93	80.56	71.11	92.5	92.5	93.33	92

III. RESULT AND DISCUSSIONS

The data is divided into three parts: an imbalanced dataset of 200, a balanced dataset of 100, and a balanced dataset of 200. The imbalance dataset determines the model's ability to deal with unbalanced data. In contrast, the balanced dataset is differentiated to determine the model's ability to influence the amount of data. The trial begins by selecting the type of language model and machine learning algorithms used and then viewing the evaluation results through the confusion matrix, Table 2. The results show that: The balanced dataset has lower results than the balanced dataset, both 100 and 200 data. The 200-balance dataset gets the highest performance. In the balanced dataset, the combination of BOW and KNN=5 using 100 data has the lowest accuracy, 51.04%. In the balanced dataset, the combination of Word2Vec Skip Gram and SVM using 200 data has the highest accuracy of 92.5%. The BOW model has the lowest performance compared to TF-

IDF and Word2Vec. The Word2Vec Language Model has the best performance.

The following are several factors that might influence the results of this study: Data input in the test, not in the corpus, causes prediction errors. Preprocessing increases accuracy by about 10% compared to data without preprocessing. The greater the number of classes, the lower the level of accuracy because more label variations are formed [26]. The balance dataset performs about 35% better than the imbalance dataset. Related to the dataset imbalance affecting performance significantly, the way to overcome it is done by resampling technique [27]. Generally, models built using 200 data show better performance than 100 data. This indicates that the larger the data, the more robust the model is likely to be. The Word2Vec Language Model generally performs better than BOW and TF-IDF because it can learn the context of sentences [28]. Word2Vec Skip Gram has more optimal performance than Word2Vec CBOW. The size of the BOW corpus follows the number of unique words from the entire document. If there are new unique words, the size of the corpus will also increase. This affects the computation time needed when we train machine learning models. BOW produces a lot of zero vectors, which is usually called a sparse matrix [29]. This is ineffective because the model only finds little information in large data sizes. In general, the TF-IDF Language Model is better than BOW because of its ability to prioritize words that are considered "important" in sentences based on the number of occurrences.

Some of the limitations encountered during the implementation of the research are as follows: Data on financial transactions and labeling are done by inputting one by one because datasets are not yet available. The dataset, consisting of 200 items, is still lacking; for now, it only contains service business types. The distribution of the existing datasets is classified as imbalance data where the distribution of classes is uneven. It should be possible to try to provide a resample treatment using the oversampling technique (increasing the size of the dataset randomly) and the undersampling technique (decreasing the size of the dataset randomly). The journal accounts used for the labeling process are still global, while the variations in financial transactions are numerous and unique for each type of business. For example, financial transactions in trading and manufacturing businesses are much more complicated and complex than in service businesses [30].

IV. Conclusion

Based on the research that has been done, the following conclusions are obtained. In developing the model, it is necessary to pay attention to the condition of the dataset, where, in this study, the balance of the dataset significantly affects the performance of the machine learning model. By the twelve combinations tested, the pair of SVM algorithms with the Word2Vec CBOW Language Model has the best performance with an accuracy rate of 92.5%, 93.5% precision, 93.42% recall, and 92.29% F1 score. In general, the Word2Vec Language Model is more optimal than TF-IDF and BOW because of its ability to understand the context of sentences. Then, TF-IDF is more optimal than BOW because the TF-IDF model utilizes information about the frequency of occurrence of each word in a sentence in the corpus. At the same time, the BOW model only uses information about the presence or absence of words (terms) in sentences in the corpus. Based on the results obtained from this study, statistically, the performance of the model is lower than that of similar semantic language model research, where the average performance is above 95%. The factor that might influence the results significantly is the amount of data in the corpus that is used, which is still meager (200 data with 10 classes). This study can also be used as an initial model for more complex financial transaction cases and other NLP language models in future research.

Acknowledgment

The researchers thank Mr. Sony Warsono, MAFIS., Ak., CA., Ph.D. as a lecturer in Accounting at the Faculty of Business Economics, Gadjah Mada University, who has given many opportunities and told many stories about accounting science so this accounting research could go this far.

References

[1] Kurnia Rahayu, S., Budiarti, I., Waluya Firdauas, D., & Onegina, V, "Digitalization and informal MSME: Digital financial inclusion for MSME development in the formal economy", Journal of Eastern European and Central Asian Research (JEECAR), 10(1), 9-19, 2022, doi: 10.15549/jeecar.v10i1.1056.

[2] Harahap, S. S., Halim, A., & Prayoga, Y, "The Role of Financial Statements On Increasing Income In SMEs", International Journal of Community Service, 2(2), 157–164, 2022, doi: 10.51601/ijcs.v2i2.80.

[3] D. Simanjuntak, S. Nurjanah, Willy, and I. Muda, "Historical cost vs current cost accounting method", Braz. J. Develop., vol. 9, no. 12, pp. 31828–31840, Dec. 2023, doi: 10.34117/bjdv9n12-085.

[4] Nurjannah, D., Wardhana, E. T. D. R. W., Handayati, P., Winarno, A., & Jihadi, M, "The Influence of Managerial Capabilities, Financial Literacy, and Risk Mitigation On Msmes Business Sustainability". Journal of Law and Sustainable Development, 11(4), e520, 2023, doi: 10.55908/sdgs.v11i4.520.

[5] Al Hashfi, R., Zusryn, A., Khoirunnisa, N., & Listyowati, A, "Online Payment: Individual Characteristics and Digital Financial Inclusion in OIC Countries", Journal of Islamic Monetary Economics and Finance, 6(4), 767 – 788, 2020, doi: 10.21098/jimf.v6i4.1148.

[6] S. Salloum, T. Gaber, S. Vadera and K. Shaalan, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," in IEEE Access, vol. 10, pp. 65703-65727, 2022, doi: 10.1109/access.2022.3183083.

[7] C. Mugisha and I. Paik, "Comparison of Neural Language Modeling Pipelines for Outcome Prediction from Unstructured Medical Text Notes," in IEEE Access, vol. 10, pp. 16489-16498, 2022, doi:10.1109/access.2022.3148279.

[8] S. Amin et al., "Recurrent Neural Networks With TF-IDF Embedding Technique for Detection and Classification in Tweets of Dengue Disease," in IEEE Access, vol. 8, pp. 131522-131533, 2020, doi:10.1109/access.2020.3009058.

[9] M. Jayaratne and B. Jayatilleke, "Predicting Personality Using Answers to Open-Ended Interview Questions," in IEEE Access, vol. 8, pp. 115345-115355, 2020, doi: 10.1109/access.2020.3004002.

[10] Iswandi, Irvan, et al. "Penelitian Awal: Otomatisasi Interpretasi Data Akuntansi Berbasis Natural Language Processing." Sriwijaya Journal of Information Systems, vol. 5, no. 2, Oct. 2013.

[11] G. G. Jayasurya, S. Kumar, B. K. Singh and V. Kumar, "Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter," in IEEE Transactions on Computational Social Systems, vol. 9, no. 4, pp. 1101-1111, Aug. 2022, doi: 10.1109/TCSS.2021.3122439.

[12] S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," in IEEE Access, vol. 9, pp. 95608-95621, 2021, doi: 10.1109/ACCESS.2021.3093563.

[13] M. U. Siregar, I. Setiawan, N. Z. Akmal, D. Wardani, Y. Yunitasari and A. Wijayanto, "Optimized Random Forest Classifier Based on Genetic Algorithm for Heart Failure Prediction", Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia, 2022, pp. 01-06, 2022, doi:10.1109/ICIC56845.2022.10006987.

[14] Kusuma, H., Muafi, M. and Kholid, M.N, "Pro-Environmental MSMES Performance: The Role of Green it Adoption, Green Innovative Behavior, and Financial Accounting Resources", Journal of Law and Sustainable Development. 11, vol 4 (Aug. 2023), e673, 2023, doi: 10.55908/sdgs.v11i4.673.

[15] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/access.2021.3129786.

[16] X. Chen, P. Cong and S. Lv, "A Long-Text Classification Method of Chinese News Based on BERT and CNN," in IEEE Access, vol. 10, pp. 34046-34057, 2022, doi: 10.1109/access.2022.3162614.

[17] J. Jiang et al., "Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization," in IEEE Access, vol. 9, pp. 123660-123671, 2021, doi:10.1109/access.2021.3110143.

[18] R. Devika, S. Vairavasundaram, C. S. J. Mahenthar, V. Varadarajan and K. Kotecha, "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data," in IEEE Access, vol. 9, pp. 165252-165261, 2021, doi:10.1109/access.2021.3133651.

[19] H. S. Nawaz, Z. Shi, Y. Gan, A. Hirpa, J. Dong and H. Zheng, "Temporal Moment Localization via Natural Language by Utilizing Video Question Answers as a Special Variant and Bypassing NLP for Corpora," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6174-6185, Sept. 2022, doi:10.1109/TCSVT.2022.3162650.

[20] A. Radhakrishnan, D. Mahapatra and A. James, "Consumer Document Analytical Accelerator Hardware," in IEEE Access, vol. 11, pp. 5161-5167, 2023, doi: 10.1109/access.2023.3237463.

[21] H. A. Ahmed, N. Z. Bawany and J. A. Shamsi, "CaPBug-A Framework for Automatic Bug Categorization and Prioritization Using NLP and Machine Learning Algorithms," in IEEE Access, vol. 9, pp. 50496-50512, 2021, doi: 10.1109/access.2021.3069248.

[22] R. Sonbol, G. Rebdawi and N. Ghneim, "The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review," in IEEE Access, vol. 10, pp. 62811-62830, 2022, doi: 10.1109/access.2022.3182372.

[23] A. A. Wazrah and S. Alhumoud, "Sentiment Analysis Using Stacked Gated Recurrent Unit for Arabic Tweets," in IEEE Access, vol. 9, pp. 137176-137187, 2021, doi: 10.1109/access.2021.3114313..

[24] S. Lyu, X. Tian, Y. Li, B. Jiang and H. Chen, "Multiclass Probabilistic Classification Vector Machine," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 3906-3919, Oct. 2020, doi: 10.1109/TNNLS.2019.2947309.

[25] S. Sugriyono, and M. U. Siregar, "Prapemrosesan Klasifikasi Algoritme kNN Menggunakan K-means dan Matriks Jarak untuk Dataset Hasil Studi mahasiswa", Jurnal Teknologi dan Sistem Komputer, vol. 8, no. 4, pp. 311-316, Oct. 2020, doi:10.14710/jtsiskom.2020.13874

[26] Husni, F. H. Rachman, I. O. Suzanti and M. K. Sari, "Word Ambiguity Identification using POS Tagging in Automatic Essay Scoring," 2022 IEEE 8th Information Technology International Seminar (ITIS), Surabaya, Indonesia, 2022, pp. 140-144, doi:10.1109/ITIS57155.2022.10009034..

[27] M. Khushi et al., "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," in IEEE Access, vol. 9, pp. 109960-109975, 2021, doi:10.1109/access.2021.3102399.

[28] I. Ashrafi et al., "Banner: A Cost-Sensitive Contextualized Model for Bangla Named Entity Recognition," in IEEE Access, vol. 8, pp. 58206-58226, 2020, doi: 10.1109/access.2020.2982427.

[29] S. Singh and A. Mahmood, "The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architectures," in IEEE Access, vol. 9, pp. 68675-68702, 2021, doi:10.1109/access.2021.3077350.

[30] J. M. Pérez et al., "Assessing the Impact of Contextual Information in Hate Speech Detection," in IEEE Access, vol. 11, pp. 30575-30590, 2023, doi: 10.1109/access.2023.3258973.