Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

A Comparative Study on the Combination of Classification Algorithm and Language Model Implementation for Smart Accounting System

Bagas Adi Makayasa[1], Agung Fatwanto[2]

[1,2] Department of Informatics, Faculty of Science and Technology
Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia
E-mail: bagas13am@gmail.com[1], agung.fatwanto@uin-suka.ac.id [2]

ARTICLE INFO

History of the article:
Received March 9, 2023
Revised August 26, 2023
Accepted August 29, 2023

Keywords:
Accounting,
Classification Algorithm,
Financial Transaction,
Language Model,
Natural Language Processing

Correspondece:
E-mail: agung.fatwanto@uin-suka.ac.id

ABSTRACT

Micro, Small, and Medium Enterprises (MSMEs) normally dealing with financial documentation and reporting problems due to in sufficient budget for hiring professional accounting services. Although some of them might have utilized off the shelf accounting softwares, they still face many obstacles in compiling a proper financial documentation because the employed software do not have an automatic transaction classification capability to assist users in recording any transactions. This study was aimed to investigate the opportunity of implementing automatic transaction classification for accounting system by using a Natural Language Processing (NLP) approach to automatically interpret the suitable account for any financial transactions based on the text written on the transaction forms. An experiment was conducted to compare the performance of eight combinations comprising of four classification algorithms (i.e. SVM, KNN3, KNN5, and NB) with two language models (i.e. TF-IDF and BoW). The result showed that KNN5 and TF-IDF pair gave highest performance with accuracy 82,5%, precision 82,54%, recall/sensitivity 83,7%, specificity 92,06%, and F1 Score 81,5%.

INTRODUCTION

The Indonesian Ministry of Cooperatives recorded that there were 64.2 million Micro, Small and Medium Enterprises (MSMEs) in 2018, whilst the number of company institution were only 729 in 2022 (Hisnul, 2022), an interesting number to illustrate the potential financial transactions of MSMEs. A large number of MSMEs illustrates how vital the role of small businesses is as one of the pillars of a country's economy. As a business institution, MSMEs must have financial documentation as a summary of the money's circulation, no matter how simple the report is, such as profit and loss reports, capital change reports, accounts payable reports and others. All of these works are only accomplished by accounting as a science that focuses on financial recording.

By recording finances properly and standardizing, MSMEs will get benefit both internally and externally (Lulu, 2020). The financial reports are helpful for business planning, monthly financial position information, cost control, investment funding considerations, business decision-making, tax calculations, and so on. Based on accounting rules, every activity that involves incoming and outgoing money must be recorded, for example, "the owner of an MSME buys a computer worth two million rupias," this activity is recorded as a financial transaction. Every transaction must be recorded one by one and continuously, so there is no lost transaction history because it will have an impact on the financial reports that will be produced later. Therefore, like other business institutions, MSMEs are operationally preoccupied with many financial transactions in just a day, so the necessity to have access to accounting service is a must.

However, MSMEs have their own problems in utilizing accounting services, that is the neglected of understanding of accounting science. The statement comes from interviews with four representatives of accountants and MSME actors to determine MSME behavior and financial transaction activities (Sony et al., 2022). One of the conclusions is the accounting software user understanding of accounting principles is still lack, Figure 1. Beside that, each accounting software has a different User Interface (UI) and User Experience (UX), so that it is more or less inconvenience for ordinary people to operate.

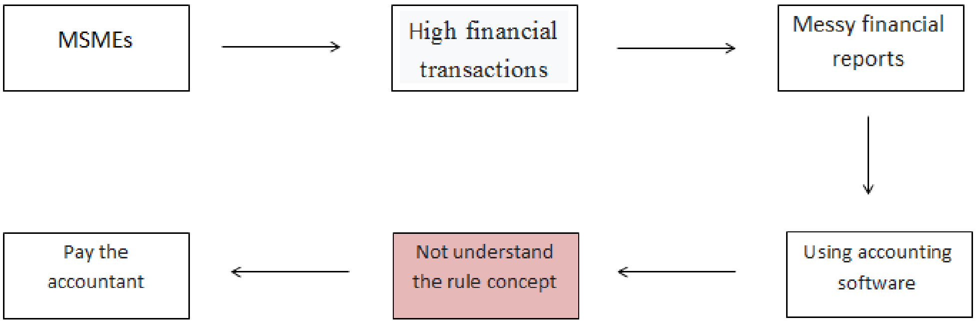

Figure 1. MSMEs behavior and financial transaction activities.

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'—the one who is not an accountant and untrained of accounting science. NLP is a derivative of the field of Artificial Intelligence which focuses on text or sound processing. The main goal of applying NLP is to make computers able to understand information like humans (S. Salloum et al., 2020).

Jayasurya et al. (2020) conducted research on the sentiment analysis covid-19 using Naive Bayes algorithm with the Bag of Words (BoW) and TF-IDF language modeling. The pre-processing stages consist of lemmitization, stop word removal, and stemming, the highest accuracy results were obtained by TF-IDF with 92%. The research also suggests another language modeling to check accuracy performance.

Furthermore, research on the document retrieval by (C. Mugisha and I. Paik, 2020) uses the comparation between TF-IDF language modeling and Bidirectional Encoder Representationsfrom Transformers (BERT). The preprocessing technique is only used tokenization and stemming. The final result shows that the highest accuracy is 85.5% using combination of TF-IDF and BERT, meanwhile the second position was obtained by TF-IDF 80%. The research also suggests on multilingual ducument retrieval using query expansion technique.

M. Jayaratne and B. Jayatilleke (2020) with the theme of open-ended interview question based on NLP using a comparative using seven language models such as TF-IDF, Latent Dirichlet Allocation. This research algorithm uses SVM and Random Forest with tokenization preprocessing techniques, stop word removal, lowercase, and lemmitization to get the most optimal results of 87.83% was TF-IDF. Chiramel recommend to try using regression algorithms and Nerual Networks on the futher works.

Until this journal was written, research related to NLP and accounting that had been found was a study entitled "Automated Interpretation of Accounting Data Based on Natural Language Processing" made by Irvan Iswandi et al. (2013). However, this research does not present a systematic process from input to output. The corpus that is applied is in the form of text and his study does not mention the algorithm in his work. In general, this research has the same ideas as Irvan Iswandi's research but with different case studies, datasets, language modeling, and algorithms.

RESEARCH METHODS

This research is conducted with a quantitative approach and data collection process was carried out through experiments. The data is analyzed by a descriptive-arithmetic approach to calculate the performance of the eight types of combinations using classification algorithms and language models, while the instrument of performance evaluation of this research uses the confusion matrix. The research start from the corpus data input in the form of text as well as a labelling process (Supervised Learning) by experts, in this case is accountants, then data pre-processing was carried out, then the data was divided into two, namely training data and test data before testing using the algorithm. The programming language used for making NLP is Python, while for the GUI display it uses the PHP programming language. The flowchart of the research method is shown in Figure 2.

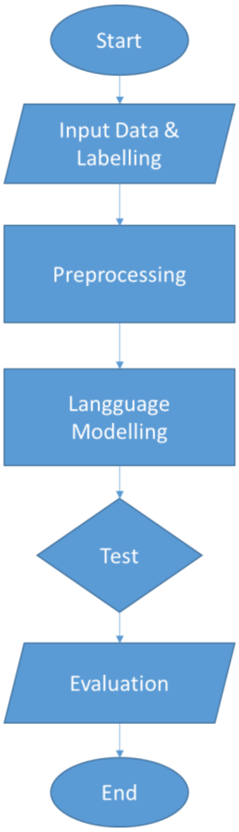

Figure 2. Flowchart research method

The system design starts with the input of data text or sound, and then the NLP works to determine the prediction results of accounting journals as the output, Figure 3.

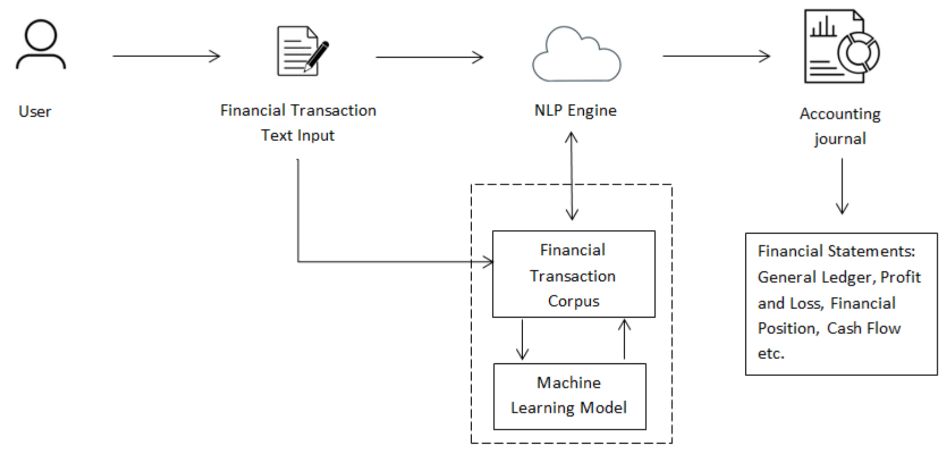

Figure 3. System design

1. Input data and labeling

At the time this research was conducted, no corpus for the case study was found. Researchers collaborated with lecturers and students of the Master of Accounting Program at Gadjah Mada University for the data collection process. The corpus of financial transaction data is recorded manually which eventually collects 200 (with all the limitations during data collection) text-based-datasets with several MSMEs business sector categories, such as service, trade, and manufacture. Each dataset needs to be determined as one of three classification of financial activities, there are Receipts, Expenditures, and Receipts and Expenditures (Sony et al., 2022).

These classifications could be applied generally and vice versa through any sector of business, whether it is service, trade, or manufacture sector. The financial transaction of paying employee salary that is classified as Expenditures, for instance, could be applied to service, trade, or manufacture MSMEs business. But, the transaction of "buy a laundry cupboard" is only specifically applied for the MSMEs in service business sector. The corpus sample is shown in Table 1. In this research, the corpus contains seven MSMEs service business sectors, namely: Salon, Workshop, Laundry, Lodging, Car Wash, Printing & Publishing, and Tour & Travel.

Table 1. Sample of labeled corpus

No	Input	Journal Output	Bussiness	Class
1	Membayar gaji karyawan Rp 2.500.000	Beban Gaji (D) Kas (K)	Umum	Pengeluaran
2	Membeli konsumsi makan rapat senilai Rp. 100.000	Beban Lain-lain (D) Kas (K)	Umum	Pengeluaran
3	Membeli tunai bahan bakar motor Rp. 50.000	Beban Transportasi (D) Kas (K)	Umum	Pengeluaran
4	Menyetorkan Rp. 50.000.000 sebagai modal awal	Kas (D) Modal (K)	Umum	Penerimaan
5	Menerima order rias pengantin sebesar 3.000.000	Kas (D) Pendapatan (K)	Salon	Penerimaan
6	Menerima pendapatan jasa bengkel 500.000	Kas (D) Pendapatan (K)	Bengkel	Penerimaaan
7	Beli lemari laundry 10.000.000 kredit	Peralatan (D) – Utang (K)	Laundry	Penerimaan & Pengeluaran
8	Membeli kendaraan travel baru 120.000.000	Peralatan (D) Kas (K)	Tour Travel	Penerimaan & Pengeluaran
9	Bayar perbaikan tempat tidur 90.000	Beban Lain-lain (D) Kas (K)	Penginapan	Pengeluaran
10	Memberikan layanan kepada pelanggan Rp. 250.000 bayar bulan depan	Piutang (D) Pendapatan (K)	Umum	Penerimaan

http://dx.doi.org/10.35671/telematika.v16i2.2258

2. Preprocessing

Before testing, the corpus was pre-processed by tidying up the dataset to get the more optimal NLP model (C. Mugisha et al., 2022). There are several stages of preprocessing as follows:

a. Lowercase

This is the process to change all words of sentence(s) into the Lowercase mode. The purpose of this process is to make words present simple.

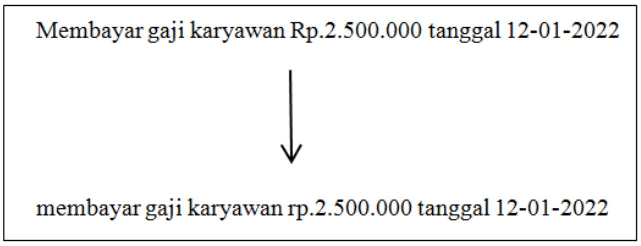

Figure 4. *Lowercase sample*

b. Tokenizing

Break sentences into words (tokens) and remove spaces (whitespace). As an example of the sentence "paying employee salaries" then the sentence is tokenized into 3 words, namely "paying", "employee", and "salaries".

Table 2. Tokenizing sample

Input	Membayar gaji karyawan rp.2.500.000 tanggal 12-01-2022
Token - 1	membayar
Token - 2	gaji
Token - 3	karyawan
Token - 4	rp.2.500.000
Token - 5	pada
Token - 6	tanggal
Token - 7	12-01-2022

c. Stopwords Removal

This stage ignores punctuation and meaningless words where the presence or absence of these words has no effect on the sentence. As seen in Figure 7, the words, "rp.2500,000" were changed to "2500000", then the words "pada" and "tanggal" were removed.

Table 3. Stopwords Removal sample

membayar	membayar
gaji	gaji
karyawan	karyawan
rp.2.500.000	2500000
pada	
tanggal	
12-01-2022	12-01-2022

d. Stemming

Taking the basic word of a word, is done to increase the potential for similarity. Examples of the words accept, to accept, accepted, accepting, both have the basic word accept.

Table 4. Stemming sample

membayar	bayar
gaji	gaji
karyawan	karyawan
rp.2.500.000	
pada	
tanggal	
12-01-2022	

3. Language Modeling

Language Modeling is the process of converting text data types into vectors, this process is needed because computers are only capable of processing numeric data types (P. Danenas and T. Skersys, 2022). Language modeling has a big influence on the algorithms later, this study uses several techniques as below:

a. Bag of Words

Bag of Words or BoW is a simple weighting technique by counting the occurrences of words with a value of 1 (there are) and 0 (no) in a sentence (A. Radhakrishnan et al., 2020)., as shown in Table 5.

Table 5. BOW ilustrasion

bayar gaji karyawan	
bayar	[1 0 0]
gaji	[0 1 0]
karyawan	[0 0 1]

b. 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 (A. Radhakrishnan et al., 2020). Suppose there is a document **j** and want to look for tokens **i**, then the number of occurrences i in th j $\boldsymbol{tf}_{i,j}$ multiplied by the logarithm (log) total document **N** divided by the number of documents it contains i $\boldsymbol{df}_{i}$.

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

4. Supervised Learning Algorithm Experiment

This study uses several supervised learning algorithms as follows:

a. Support Vector Machine (SVM)

Support Vector Machine works by looking for a hyperlane (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 (C. Mugisha and I. Paik, 2022).

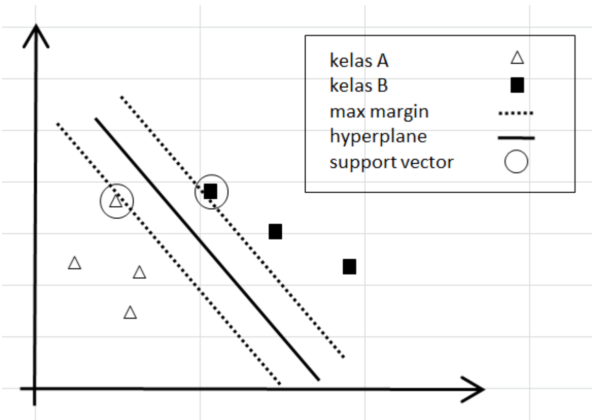

Figure 5. SVM algorithm ilustration

b. K-Neareset Neighbour

KNN is an algorithm that classifies output based on quantity **k** nearest neighbor Figure 6, this study uses cosine similarity as a distance calculation. $\boldsymbol{A}_i$ show the *term* number i in document A, and $\boldsymbol{B}_i$ is the *term* at i at document B, and **n** is the number of unique terms in the dataset (H. A. Ahmed et al., 2021).

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

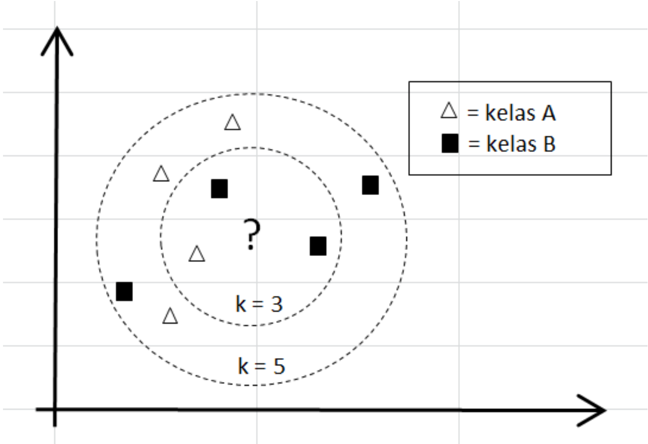

Figure 6. KNN algorithm ilustration

c. Naive Bayes

Naive Bayes is an algorithm based on Bayes' theorem where the probability of an event will occur (A) with return probability (P) from other events that have occurred (B) (H. A. Ahmed et al., 2021).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

5. Evaluation

Evaluation of the performance of this study uses a multiclass confusion matrix as shown in Table 6 (Aziz et al., 2020).

Table 6. Confussion Matrix

		Actual	
		Positive	Negative
Prediction	Positive		
	Negative		

a. True Postitive (TP)

The amount of data in the actual class is correct and the prediction results are positive (relevant).

b. True Negative (TN)

The amount of data in the actual class is correct but the predicted results are negative (irrelevant).

c. False Positive (FP)

The amount of data in the actual class is wrong but the prediction results are positive (irrelevant).

d. False Negative (FN)

The amount of data in the actual class is wrong and the prediction results are negative (relevant).

Furthermore, from the confusion matrix table calculations are carried out for accuracy, precision (positive relevant ratio compared to all positive results), recall or sensitivity (positive relevant ratio compared to all relevant results), specificity, and F1 Score Figure 7.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{F1 Score} = (2\ X\ Recall\ X\ Precision) / (\text{Recall} + \text{Precision})$$

Figure 7. Calculation of accuracy, precision,recall, specificity, and F1 Score.

RESULTS AND DISCUSSION

Testing activity is done by dividing the dataset into training data by 80% and test data by 20%. First of all the user is asked to input a sentence (text or voice) of a financial transaction and select the wanted type of algorithm, and finally assess whether the predicted results given are appropriate or not. The Graphical User Interface (GUI) is built using the PHP programming language while the machine learning uses Python, both of which are linked by an API. GUI shown in Figure 8 and the predicted output of the journal is shown in Figure 9.

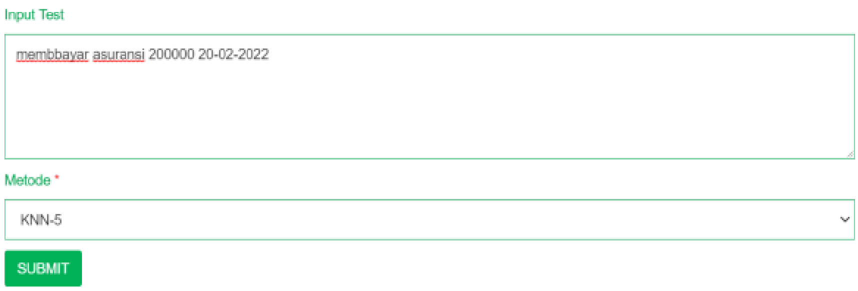

Figure 8. GUI form input

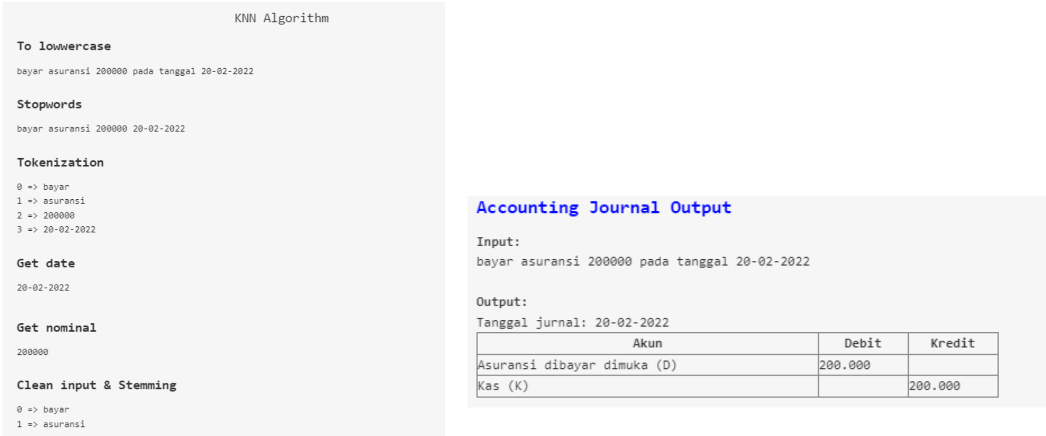

Figure 9. Journal output prediction

From Figure 9, it can be seen that the system flow starts from input, preprocessing, and output in the form of journal predictions. The journal output is streamed to the general ledger so that it becomes various kinds of reports automatically, because the main key to accounting is the data documented in the journal (Syamsul, 2022). At the preprocessing stage, the date and nominal are taken, if these two datas are not found, so the process cannot be continued or is considered an input error. The way to retrieve the nominal is to convert text into the number data type while for dates by converting text into a date data type. The pure input that the system gets is meaningful other than date and nominal. The built script does not handle stop word removal in the form of abnormal characters or emojis, bearing in mind that real users will be MSME owners and the data is in the form of financial transactions which can be said to be sensitive, not include autocorrect input that is not appropriate (typo).

The testing process was carried out by an accountant (not corpus input) with a test scheme twice using 20 kinds of financial transactions, the first using 100 datasets and the second with 200 datasets using 40 kinds of financial transactions. This is done to determine the effect of the amount of data on the performance of the machine learning model that has been made.

In an experiment of 100 datasets, the best performance was obtained from a combination of the KNN 5 algorithm and the TF-IDF language model with an accuracy rate of 80%, precision 80.73%, recall 81.81%, specificity 92.65%, and F1 score 81.16 %. While the lowest performance for 100 datasets was obtained from a combination of the Naive Bayes algorithm and the BoW language model of 45%, 44.37% precision, 44.4% recall, 60.88% specificity, and 44.13% F1 score.

Then in the 200 dataset experiment, the best performance was also obtained from the combination of KNN 5 and the TF-IDF language model with an accuracy rate of 82.5%, 82.54% precision, 83.7% recall, 92.06% specificity, and F1 Score 83.12%. However, the lowest performance for 200 datasets was obtained from a combination of the Naive Bayes algorithm and the TF-IDF language model with an accuracy rate of 65%, precision 62.69%, recall 62.63%, specificity 81.19%, and F1 score 62.69. All combinations of algorithms and language models experienced an increase in accuracy in the 200 dataset except for Naive Bayes and TF-IDF pairs.

Based on the two test scenarios using 100 and 200 datasets, the lowest performance was obtained when using the Naive Bayes algorithm. This shows that the algorithm's performance is unsuitable for the case study raised. Meanwhile, based on two test scenarios using 100 and 200 datasets, the results obtained were that the combination of the KNN 5 algorithm and the TF-IDF language model produced the best performance of the eight types of combinations tested. This shows that the combination of KNN and TF-IDF modeling has relatively better performance for the case studies raised compared to the other seven combinations. The complete test results are shown in Table 7.

Table 7. Test result

Algorithm	Language Modeling	100 Dataset					200 Dataset				
		Acc	**Prec**	**Rec**	**Spec**	**F1**	**Acc**	**Prec**	**Rec**	**Spec**	**F1**
Naive Bayes	BOW	45	44,73	44,4	60,88	44,13	70	69,84	69,44	85,13	62,66
	TF-IDF	70	70	70,29	84,78	70,65	62,5	62,69	62,63	81,19	62,69
KNN 3	BOW	60	60,11	60,67	74,36	60,51	72,5	72,22	72,77	86,55	72,5
	TF-IDF	75	75,91	76,13	86,4	76,18	80	80,15	81,96	91,16	81,05
KNN 5	BOW	65	65,71	65,97	77,24	64,84	75	75	74,19	87,82	74,59
	TF-IDF	80	80,73	81,81	92,65	81,16	82,5	82,54	83,7	92,06	83,12
SVM	BOW	65	65,3	65,69	82,47	64,19	65	65,08	64,58	82,74	64,83
	TF-IDF	65	64,72	65,1	80,94	65,24	70	60,84	70,14	85,12	69,99

Note: Acc (Accuracy), Prec (Precision), Rec (Recall), Spec (Specificity), F1 (F1 Score)

Several estimates of factors that are affect the results of this study:

1. Data input sentences in the test that are not in the corpus cause prediction errors (M. F. Mridha, 2020).
2. The TF-IDF and BoW algorithms could not recognize the context of sentences or synonym that contain in the corpus, for instance, the text input of "buying a laundry cupboard". The word “cupboard” is supposed to be identified as “equipment”, based on the entry of the corpus, but the algorithms stuck on their limit of integration to the corpus. Therefore, the result of the text input considered as an incorrect input (M. F. Mridha, 2021).
3. Language modeling that cannot understand context also causes problems, for example, the sentence “memberikan layanan kepada pelanggan seharga Rp250.000 bayar bulan depan”. This sentence means getting income for services (Receipts), but because the word "memberikan" means giving to other, so sometimes the algorithm recognizes it as an activity that costs money (Expenditures).
4. Long and wordy text input causes prediction errors (X. Chen et al., 2022). For example, the input, "Received a consumption bill of IDR 500,000 from the HaTo restaurant (*halalan toyiban*) to purchase food and snacks for 10 days. Bills are paid immediately in cash,” causes error result. This error could be solved in the way of simplifying the text input to ease the machine to process data that is, “paying a consumption bill of Rp. 500,000”.
5. Related to point 6, the researcher tries to test the performance of NLP google voice search with long text input "KFC near me that I can reach in five minutes by walking or by bike and still open" compared to short input " KFC near me”, concise and clear. The result is as in Figure 10 and Figure 11.

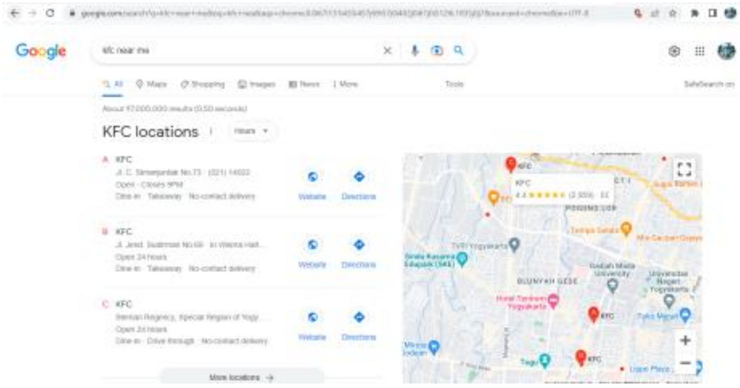

Figure 10. Google voice search, short and clear input show accurate results

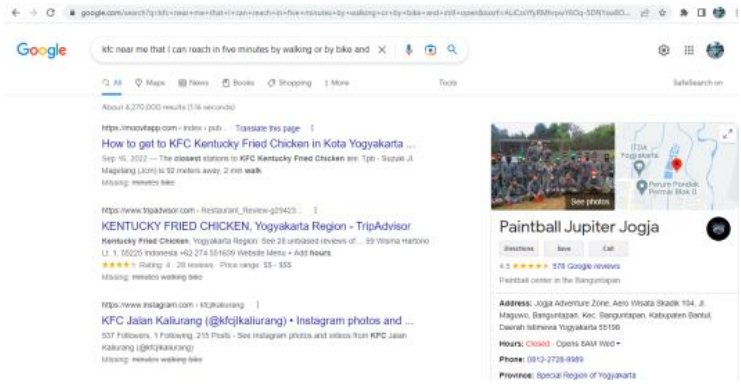

Figure 11. Google voice search, verbose input show inaccurate results

It can conclude that verbose input is not necessarily a shortcoming of the developed NLP model, but rather the user's understanding in operating the tool. Besides that, long text has its own study and focus on the topic of text summarization (J. Jiang et al., 2021). However, it must also be recognized that rambling input patterns can reduce the accuracy in this study.

6. The same type of journal label can consist of several types of input, for example the output label "equipment (debit) and cash (credit)" can be made with the input "buy a cash table", "paid for a table purchase", "buy a table to support operational". If buying the table is done on credit, the output label will be "equipment (debit) and debt (credit)". Here it can be seen that the key to the accuracy of NLP is the richness of the data corpus (R. Devika et al., 2021).
7. Not always the resulting journal output has a ratio of one debit and one credit (H. S. Nawaz, 2022). Sometimes it is found that the prediction is correct but the journal preparation is wrong because of the various input variations, for example "buying 500,000 tables at the Maju Jaya store, paying 300,000 in cash and the rest on credit". The input only mentions two nominal lifts but the labeling produces three types of accounts, namely "equipment (debit), debt (debit), and cash (credit).
8. To predict results, the scripts that are built take an average of 0.03 seconds on both the 100 and 200 datasets. Further research is needed to test the comparison of processing time by increasing the number of datasets (H. S. Nawas et al., 2022).

Some of the limitations encountered during the implementation of the research are as follows:

1. Data on financial transactions and labeling is done by inputting one by one since there is no dataset available yet. The dataset which consists of 200 items is still lacking and for the time being it only contains types of service businesses.

2. The journal accounts used for the labeling process are still global in nature, 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 difficult and complex than in service businesses (Sony, 2018).

CONCLUSIONS AND RECOMMENDATION

Based on the experimental data, we can conclude that: (i) Of the eight combinations tested, the pair of KNN5 algorithm with TF-IDF language model has the highest performance with an accuracy rate of 82.5%, precision rata of 82.54%, re all/sensitivitas rata of 74.19%, specifically rata of 92.06%, and F1-score of 83.12%. (ii) In general, the KNN algorithm has the highest level of performance since KNN is basically a multi-class classifier algorithm (which is fit for this research case), while SVM was originally designed as a binary classifier algorithm. Although NB basically also a multi-class classifier algorithm, it considers that each feature as part of the input is mutually independent from one another so it is not suitable for classifying natural language texts where the meanings of words in sentences are interdependent (word order in a sentence affects meaning). In particular, KNN 5 is better than KNN 3 because a higher value of “k” reduces classification noise (although the consequences can further blur the boundaries between labels). (iii) In general, the TF-IDF language model 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, while the BoW model only utilizes information about the presence or absence of words (terms) in sentences within the corpus.

ACKNOWLEDGEMENT

The researcher would like to 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/BIBLIOGRAPHY

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.

Aziz Husain, Fadhila Tangguh Atmojo, and Erma Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah", Jurnal Technocom, Vol. 19, No. 3, Agustus 2020: 286-294.

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.

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.

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.

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.

Hisnul, Pompong Budi Setiadi, Sri Rahayu, " UMKM di Masa Pandemi COVID 19 Berdampak Pada Teknologi Dan Ddigitalisasi Pada Pusat Oleh-Oleh Rahma di Desa Kendalrejo”, Jurnal Ekonomi dan Bisnis, Vol. 11 No. 1 Juli 2022.

Interview with Anam dan Antariksa, UMKM players, at 12 Oktober 2022

Interview with Sony Warsono, Nadya Windi, and Anggit Firmansyah, Lecturer and Sutdent of Magister Akuntansi Universitas Gadjah Mada, pada 04 Oktober 2022

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.

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.

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.

Nurul, Lulu, “Pentingnya Penyusunan Laporan Keuangan UMKM Bagi Para Pengusaha Bakery, Cake, dan Pastry (BCP) di Kota Blitar”, Jurnal Graha Pengabdian, Vol 2 No 2, 2020.

P. Danenas and T. Skersys, "Exploring Natural Language Processing in Model-To-Model Transformations," in IEEE Access, vol. 10, pp. 116942-116958, 2022, doi: 10.1109/ACCESS.2022.3219455.

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.

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.

Syamsul, “Analisis Pencatatan dan Pelaporan Keuangan UMKM di Kota Palu”, Jurnal Keuangan dan Bisnis, Vol 10 No 1, 2022.

Warsono, Sony, “Dasar-Dasar Akuntansi: Tes Potensi Akuntansi”, Yogyakarta, ABPublisher, 2018.

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.