2022 2nd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 24-26, 2022

Classification of Customer Complaints: TF-IDF Approaches

Doni Purnama Alamsyah
Entrepreneurship Department
Bina Nusantara University
Jakarta, Indonesia
doni.syah@binus.ac.id

Toni Arifin
Computer Science Program
Universitas Adhirajasa Reswara Sanjaya
Bandung, Indonesia
toni.tfn@ars.ac.id

Yudi Ramdhani
Computer Science Program
Universitas Adhirajasa Reswara Sanjaya
Bandung, Indonesia
yudi.yrm@ars.ac.id

Febriansyah Achmad Hidayat
Computer Science Program
Universitas Adhirajasa Reswara Sanjaya
Bandung, Indonesia
febriansyahah1997@gmail.com

Leni Susanti
Entrepreneurship Department
Bina Nusantara University
Jakarta, Indonesia
leni.susanti@binus.ac.id

***Abstract*—Bank Rakyat Indonesia (BRI) is an Indonesian state-owned bank; the operations of the bank certainly aim to provide satisfaction to customers. This can be realized by quickly responding to the bank's response to complaints. The number of complaints received by BRI banks to date is large, but the categories of complaints are still mixed so each division of the bank has difficulties and must first filter complaints. This study aims to classify bank complaints using the Text Mining field to make it easier for BRI to provide complaints directly to each division. The classification method used is Neural Network and the data preprocessing method uses the TF-IDF method. The dataset used in the form of text is one million datasets with 5 classes, namely debit cards, credit cards, customer service, mobile banking, and loans. The results of this study indicate that the accuracy of the neural network method is better and becomes a recommendation to be implemented at BRI. This research is useful for banking in Indonesia or the service industry in managing customer relationship management.**

Keywords—Customer Complaints, Text Mining, TF-IDF.

I. Introduction

Bank is one of the business entities by collecting funds from the public in the form of savings and distributing them in the form of credit or other forms as in improving the standard of living of the community [1]. Banks in Indonesia are managed by a variety of agencies, ranging from government agencies to state-owned banks, or banks managed by private institutions. One of the most important things in the presence of a bank is the presence of customers. The more customers in a bank, the more the bank can progress [2]. Factors that influence customer arrivals are satisfaction and quality of bank services, so one form of improving service is to provide criticism and suggestions in the form of customer complaints. Every bank needs to provide a platform or platform for customers to submit complaints [3]. One of the banks that provides a platform to become a forum for customer complaints is Bank Rakyat Indonesia (BRI), which is owned by the Indonesian government.

Bank Rakyat Indonesia (BRI) has always paid attention to customer complaints, these complaints are divided into several categories such as debit card complaints, credit card complaints, BRI mobile banking complaints, customer service complaints and company complaints. The complaint will then be categorized manually by the complaint analysis section from an excel file, but the process experiences several obstacles such as inaccurate categorization, and ineffectiveness in terms of time in separating categories. This becomes an obstacle in the process of serving public complaints, so we need a system that can classify customer complaints automatically.

Sentiment analysis is a branch of research in text mining. Text mining is the application of data mining concepts and techniques to look for patterns in text. Sentiment analysis is the process of automatically understanding, extracting, and processing textual data to obtain information in the form of sentiments contained in opinion sentences. Sentiment analysis is carried out to see the opinion or tendency of a person's opinion on a problem or object, determine whether the opinion is positive or negative [4] [5]. Text mining is one of the fields of Deep Learning that can help with classification problems. The classification of complaints has been done, namely the classification of hotel customer complaints using the Particle Swarm Optimization (PSO) method with an accuracy of 89% [6], In addition, previous research has modeled customer complaints from the Financial Protection Bureau using the Latent Dirichlet Allocation (LDA) method with an accuracy of 75%, the dataset used is 200000 (two hundred thousand) datasets and the preprocessing processes are tokenizing and filtering. This accuracy is still considered lacking due to incomplete dataset processing [7], [8]. Linear Discriminant Analysis is a method in extracting the characteristics of an object. In extracting features, LDA divides several classes in the feature search, LDA has a very high accuracy [9]. In addition to these methods, deep learning methods can also solve current sentiment analysis problems. Deep learning is a relatively new branch of machine learning. The advantage of the deep learning method is that it has higher accuracy and can be trained and consistent when faced with a lot of data [5], [10]. One example of a method in deep learning is the Convolutional Neural Network (CNN). This method works well for image analysis and image classification because it can extract a feature area from global information and is able to consider the relationship between these

2022 International Conference on Intelligent Technologies (CONIT) | 978-1-6654-8407-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/CONIT55038.2022.9848056

features. Meanwhile, in sentiment analysis, CNN has a convolution layer to extract information with larger chunks of text. CNNs also require fewer connections and parameters making them easier to train. In addition, the CNN method combined with the word embedding feature generated from Word2vec can improve classification performance [5], [11].

The text classification process is used by several methods, but the Convolutional Neural Network method has never been used in classifying customer complaints so that in this study the CNN method will be applied and TF-IDF is applied as the pre-processing text. CNN is considered good in classifying text and TF-IDF is good for preprocessing text such as research that performs malware classification based on the API-System using CNN and TF-IDF methods resulting in an accuracy of 93.71% [12]. This shows that CNN has a good performance, especially with the help of TF-IDF. TF-IDF is a preprocessing which is part of word embedding, where at this stage the conversion of words into a matrix will be carried out. In addition to using TF-IDF, other preprocessing is needed, as for text preprocessing other than TF-IDF in this study, namely tokenizing, stemming, normalizing and stop word removal. The dataset used in this study amounted to 1,000,000 (one million) datasets with the number of classes, namely 5 (five) classes including debit cards, credit cards, BRI mobile banking, customer service and corporate.

II. METHODS

Research is one of the activities carried out with the intention of finding the truth and solving an existing problem, research can also be interpreted as one of the scientific activities in obtaining correct knowledge about a problem. The research itself consists of facts, concepts, generalizations, and theories that can be understood by users to solve the problems at hand [13]. The next thing that will be faced, is usually carried out in general methods commonly used either by individuals or in combination with four methods such as action research, experiments, case studies, surveys. Then in this study using experimental research methods with a qualitative approach [14].

A. Dataset

At this stage, determine the data to be studied. Integrate all data into the data set, including determining the necessary variables. The dataset used in this study was taken directly from BRI Bank which is private data and this data consists of 1,000,000 records with 3 attributes which can be seen in Table 1.

TABLE I. DESCRIPTION OF DATASET ATTRIBUTE

No.	Attribute	Description
1	Complaint Narration	Contents of customer complaints
2	*Product*	Products of customer complaints
3	*Sub Product*	Sub contents of customer complaints
4	Problem	Problem of customer

Next is requires exploration or deepening of the Consumer Narrative Dataset. Exploration is carried out with the aim of showing that all attributes and classes in the dataset are valid, so that they can be used for good research objects. Therefore, the aim is to find out the best classification results from the Consumer Narrative Dataset. The Consumer Narrative Dataset has 4 attributes with 1,000,000 records. However, in working on text mining using the Neural Network Algorithm, it can be done optimally if the variable or attribute has a categorical value instead of continuous. The Consumer Narrative Dataset has mixed variables such as discrete or categorical variables and continuous variables. So that this dataset requires a data transformation process in the early stages of data processing. This process uses several stages, namely filtering, stemming and stopword removal. For testing the model used, the data will be divided into two parts, including training data and testing data using Split Validation. The amount of data sharing is 70% for training data and 30% for testing data. This training data is for model development and this testing data is for model testing.

B. Model Proposed

Consumer Narrative Dataset is secondary data that is ready to be processed in text mining. The model used in this process is a model that has never been done by previous researchers in the use of Consumer Narrative datasets. The model that will be used in this research is the Neural Network classification method. The way the Neural Network works is as follows.

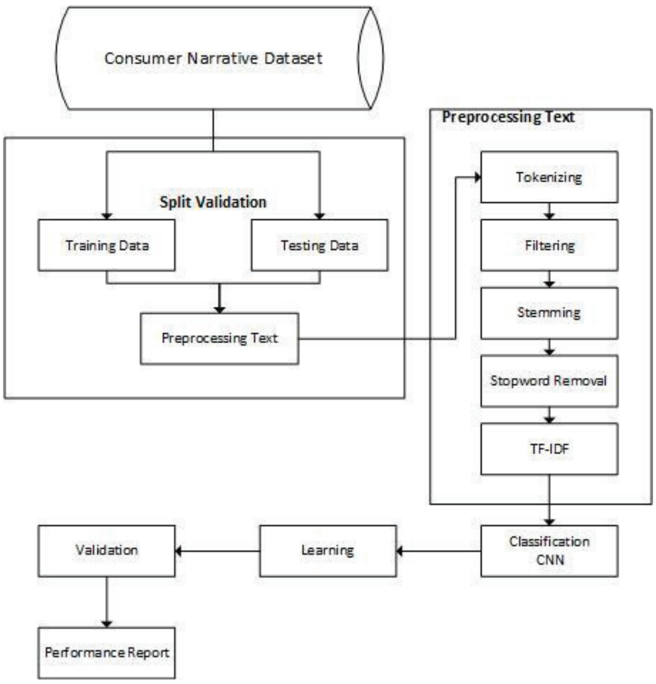

Fig. 1. System Overview

1. Prepare the dataset for processing.
2. Split the dataset, then cross-validation will be carried out, namely changing the portion or division of each dataset in the training and testing process. Suppose 70:30, or 80:30.
3. Do preprocess text to clean the data because the dataset is still dirty. Preprocessing carried out is tokenizing, filtering, stemming and stop word removal.

4. After preprocessing the text, the classification process is carried out using the CNN method with the output looking for the weight value of each class.
5. After the learning process is carried out, a performance evaluation will appear

C. Experimentation and Model Testing

In this stage, experiments are carried out on the data model that will be processed using the proposed method using python. Experiments in this study were carried out on the classification of the Consumer Narrative Dataset, after which the dataset would be divided into two parts, namely training data by 70% while testing data by 30%. The amount in the training data and testing data is determined by changing the split method, which is initially relative to absolute.

D. Evaluation and Validation

In this stage, an evaluation of the model to be used is carried out. This evaluation process is carried out using the Confusion Matrix and ROC (Receiver Operating Characteristic) curve. While the validation process will be carried out using Split Validation, namely data that has been divided into training data and testing data. Evaluation stage using tools and python libraries while the process is running.

III. Result and Discussion

In the implementation of the Convolutional Neural Network and TF-IDF models for this customer complaint classification case, supporting software is needed. In this study, a simulation will be made to test the Convolutional Neural Network model to produce good accuracy using the Python programming language and library as shown in Table II.

TABLE II. TECHNOLOGY USED

No	Library	Function
1	NumPy	Read dataset, process dataset
2	NLTK	Preprocessing (sample: tokenizer)
3	Hard	Library for word CNN
4	Tensor flow	Library to implement on CNN
5	Satrawi	Library for stemming process where this library stores basic words according to Indonesian.
6	Seaborn	Library for graphics design
7	Python 3.6	Python environment that supports the required libraries
8	Sckit Learn	Library to implement on TF-IDF

A. Data Input

The input data is the dataset used in this study stored in Comma Separated Values (CSV) format and totaling 1,000,000 datasets with 5 classes as shown in Table III. and as an example, some of the datasets are shown in Table IV.

TABLE III. DATASET CLASS

No	Class
1	Complaints for Debit Card
2	Complaints for Credit Card
3	Complaints for Customer Service
4	Complaints for Mobile Banking
5	Complaints for Loan

TABLE IV. DATASET SAMPLE

Id	Text	Class
1348286903527768065	"How many times have you withdrawn cash at the ATM, the balance is deducted, the money doesn't come out"	Debit Card
1348302168248340481	"It's weird why my credit card always leaks and I always have to refute transactions"	Credit Card

B. Preprocessing Stage

This stage is done to clean the data. The dataset that we take from the source with the original form of course is not necessarily ready to be used for classification. The dirty data includes empty text, duplicate text, and words that have multiple interpretations. The following is the preprocessing stage of the text in this research.

Tokenizing. Tokenizing is an effort to prepare datasets for classification. This process is to separate each word in a sentence with a separator which can be a space, comma, semicolon or period. In this study, the delimiter used is like space. **Filtering**. The filtering stage is a stage of taking important words according to the tokens in the dictionary. The dictionary taken must be in Indonesian. After the next tokenization, the results after the text are filtered. **Stemming**. Stemming is one of the most important efforts in text preparation. Stemming is the step of removing affixes or returning a word to its base word. In this study, to do stemming required a library with the name Sastrawi where this library contains all the basic words and affixes. When a word with an affix is detected, but if it reaches stage two it has not yet found the basic word, so the combination process will be carried out again. With the base word that has been generated it will end up being combined with the affix in twelve configurations. **Stop word removal**. Stop word removal is a filtering process, by selecting important words from the token results, namely what words have been used to represent documents. In this study, the reference for stop word removal is a dictionary taken from the Sastrawi library, namely the stop word list in Indonesian. This stage will do a word search and remove the words that are in the stop word removal. Then in NLP (Natural Language Processing) stop words, one of the words that are ignored in processing, these words will usually be stored in stop lists. Then the main characteristic in choosing a stop word is usually a word that usually has a high frequency of occurrence, for example connecting words such as "and", "or", "but", "will" and others. **TF-IDF**. Term Frequency — Inverse Document Frequency is an algorithmic method that aims to calculate the weight of each word that is often used. By using this method, it is known to be efficient, easy, and has very accurate results, of course. The method then continues to calculate the value, namely Term Frequency, with Inverse Document Frequency, in time until each token "word" from each document in the corpus. Then in simple terms, the TF-IDF method is generally used to find out how often a word that appears very often appears in the document. Next is Term Frequency (TF), that is, there are several stages of the type of formula that can be used:

a) Binary TF (binary TF), that is, only by paying attention to whether a word or term is present or not in the document, then if it exists it will be given a value of one, continue while if it is not it will be given a value of zero.

b) Pure TF (raw TF), a TF value is given based on the number of occurrences of one of the terms in the document. For example, if only five appeared, then it would be times then the word would be worth five too.

c) TF normalization, namely, which uses a comparison between frequency and a term with a maximum weight value, namely the whole or also a collection of frequencies so that the term is in a document.

d) Logarithmic TF is in the form of this which aims to avoid dominance with documents that currently contain at least one term in the query, but if they already have a very high frequency.

In order to simplify the classification, the text must be changed to a one-hot matrix form, this stage is done because each word basically has a different dimension. Each array is the length of the dictionary, and every value in the dictionary that is not a token value is represented by the number 0 while the token value is represented by the number 1.

C. Classification

The first stage of this research is classification by using the Convolutional Neural Network algorithm. This feature, which has been determined previously, is then used as input for the calculation, namely the Convolutional Neural Network, which is used to classify documents. Furthermore, at this stage, the training document will be used as a document with input, the algorithm for the Convolutional Neural Network, which is as shown in Figure 2.

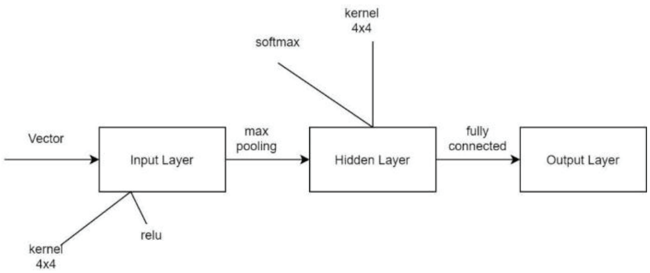

Fig. 2. Process of Convolutional Neural Network

D. Output

After the proposed model is created and training is carried out according to Figure 2. Then a validation split is carried out and changes in the number of epochs are made, then there are results from the performance of the model made in Table V and Table VI.

TABLE V. PERFORMA MODEL NON TF-IDF

No	Sum of Epoch	Split	Accuracy
1	10	70:30	72%
2	10	80:20	74%
3	15	70:30	82%
4	15	80:20	83%
5	20	70:30	84%
6	20	80:20	85%
7	25	70:30	85%
8	25	80:20	85%

TABLE VI. PERFORMA MODEL WITH TF-IDF

No	Sum of Epoch	Split	Accuracy
1	10	70:30	73%
2	10	80:20	75%
3	15	70:30	81%
4	15	80:20	82%
5	20	70:30	82%
6	20	80:20	89%
7	25	70:30	89%
8	25	80:20	89%

Based on Table VI. In the training process, changes were made to the number of epochs and the composition of the split dataset and the best accuracy results were obtained at the number of epochs of 20 and the composition of splits of 80:20. Meanwhile, when the epoch is increased to 25, the accuracy remains the same, so there is no need to add the number of epochs. In evaluation and testing, precision and recall values will be sought. This value is obtained from the number of True Positive (TP), False Positive (FP), True Negative (TN) and also False Negative (FN). The following is also the number of TP, FP, TN and FN in Table VII.

TABLE VII. VALUE OF TP TF TN AND FN

	Total
TP	880231
FP	119769
TN	721231
FN	278769

Based on the calculations in Table VII, it is known that the precision value is 0.880231 and the recall value is 0.7594. The findings of this study explain the importance of the TF-IDF method in increasing the value of classification accuracy on banking customer complaints. This research was conducted using a quantitative survey method, the survey was carried out to mobile payment e-wallet users in the city of Bandung. Considering the city of Bandung as one of the cities in Indonesia that has the most users on mobile payment e-wallet besides the city of Jakarta. In addition, the city of Bandung is known as a student city and culinary city, so the possibility of transactions using mobile payment e-wallet is quite high among students. Questionnaires were distributed to respondents online for one month and 227 respondents were found who filled out valid questionnaires, that is, they have made transactions via mobile payment e-wallet. Data from respondents is then carried out data processing and analysis based on the research hypothesis design.

IV. CONCLUSIONS

As the results of testing with analysis that have been carried out on the BRI bank customer complaint classification program system, it can be concluded that the application of the Convolutional Neural Network method to this study resulted in a fairly high accuracy of 85% in epoch 25. TF-IDF is very helpful for Convolutional Neural Network to get the best accuracy in text mining problems with a comparison if not using TF-IDF the accuracy obtained is 84% while the TF-IDF increases to 89%. This research has limitations because the study is limited to the study of methods that still need to be implemented in the current system. So the suggestion for further research is to build a system for implementation in customer relationship management.

REFERENCES

[1] N. Ahmad and S. Haron, “Perceptions of Malaysian corporate customers towards Islamic banking products & services,” *International Journal of Islamic Financial Services*, vol. 3, no. 4, pp. 1–16, 2002.

[2] I. Ayu, A. Rai, and N. K. Purnawati, “FAKTOR-FAKTOR YANG MEMPENGARUHI KREDIT PADA BANK UMUM SWASTA NASIONAL (BUSN) DEVISA,” vol. 6, no. 11, pp. 5941–5969, 2017, [Online]. Available: www.bi.go.id

[3] Ayu Azhari Meliana, “Faktor-Faktor Yang Mempengaruhi Kepuasan Nasabah Terhadap Pelayanan Di Bank Safir Kota Bengkulu (ditinjau Dari Produk terhadap pembiayaan),” Bengkulu, 2017.

[4] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends® in information retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[5] A. Pambudi and S. Suprapto, “Effect of Sentence Length in Sentiment Analysis Using Support Vector Machine and Convolutional Neural Network Method,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, pp. 21–30, 2021.

[6] S. Khomsah, “Jurnal Penelitian Pos dan Informatika Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews Optimasi Naive Bayes Classifier Pada Sentiment Analysis Komentar Pelanggan Hotel,” vol. 10, no. 2, pp. 157–168, 2020, doi: 10.17933/jppi.2020.100206.

[7] K. Bastani, H. Namavari, and J. Shaffer, “Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints,” vol. 127, pp. 256–271, 2019, doi: 10.1016/j.eswa.2019.03.001.

[8] Y. Ramdhani, “Komparasi Algoritma LDA Dan Naïve Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear,” *Jurnal Informatika*, vol. 2, no. 2, 2015.

[9] F. Brilian, A. F. Huda, and I. Taufik, “Sistem Pengenalan Wajah Dengan Menggunakan Metode Linear Discriminant Analysis (LDA),” *Konferensi Nasional Sistem Informasi*, pp. 1941–2355, 2014.

[10] Q. T. Ain *et al.*, “Sentiment analysis using deep learning techniques: a review,” *Int J Adv Comput Sci Appl*, vol. 8, no. 6, p. 424, 2017.

[11] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, “Sentiment analysis using convolutional neural network,” in *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*, 2015, pp. 2359–2364.

[12] M. Schofield *et al.*, “Convolutional Neural Network for Malware Classification Based on API Call Sequence,” pp. 85–98, 2021, doi: 10.5121/csit.2021.110106.

[13] N. Alfianika, *Buku ajar metode penelitian pengajaran bahasa Indonesia*. Deepublish, 2018.

[14] H. A. Bischoff-Ferrari *et al.*, “Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials,” *Bmj*, vol. 339, 2009.