JPPI Vol 11 No 2 (2021) 137 - 152

Jurnal Penelitian Pos dan Informatika

32a/E/KPT/2017

e-ISSN 2476-9266
p-ISSN: 2088-9402

Doi: 10.17933/jppi.v11i2.352

Ensemble Method Builds a Predictive by Integrating Several Models for Accurate Answer Prediction on Chatbot

Merry Anggraeni[1] Hillman Akhyar Damanik[2]

Fakultas Teknologi Informasi Universitas Budi Luhur[12]
Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, DKI Jakarta 12260[12] Indonesia

merry.anggraeni@budiluhur.ac.id

Received: 3 December 2021 ; Revision: 27 Desember 2021; Accepted: 27 Desember 2021

Abstract

Advances in machine learning algorithms that have a positive impact are Chatbots. A chat service which is actually a feature that has been very often used by tech-savvy people. But the difference is, the ones who reply in the chat process are robots or virtual characters. The chatbot will provide answers to the questions given to it which is basically the chatbot assigned to understand what context the user (user) means, then reply to it with the appropriate context. However, each context has a different input, human language has a very flexible way so that it is often found inaccuracies in the prediction of answers given by Chatbot. This could be due to the incorrect choice of algorithms for the classification of the context or the lack of training data provided. To overcome this, this study will focus on strengthening the prediction of chatbot answers with the method of the machine learning algorithm (Machine Learning) ensemble 5 classification method heterogeneously between basic classifiers and meta-algorithms and using maximum voting (Majority Vote) or Hard Voting on type ensemble. Classification is the process of finding a model or pattern that can describe and differentiate classes in a dataset. The goal is that the model can be used to predict objects with unknown class labels. It was found that the accuracy results were 86% for the data set with 6 classes, the avg Macro for each precision and recall was 92%, and the f1-score was 89%. Weighted avg for precision is 93% and each for recall and f1-score is 86%.

Keyword: Chatbot, Ensemble Technique, Text classification, Machine learning algorithm, meta-algorithm, answer prediction.

INTRODUCTION

Machine learning (ML) has been successfully applied to a variety of areas from information retrieval, data mining and speech recognition, to computer graphics, visualization and human-computer interaction. However, most users often treat machine learning models as black boxes because of their incomprehensible functions and unclear working mechanisms (Liu et al. 2017).

One type of machine learning that a chatbot needs is one that has accuracy and clarity in terms of classification, and this is an important thing in chatbots. The best classification accuracy is usually achieved by black-box machine learning models such as a Support Vector Machine, neural network or random forest, but usually requires a broad training set. On the other hand, machine learning methods whose predictions are easy to understand and interpret often have limited predictive capacity (inductive inference, linear regression, decision tree) or are flexible and not computer-complicated, such as explicit graphic models. These methods usually require less data to train, but provide lower classification accuracy. (Galitsky B., Goldberg S. 2019)

In the field of machine learning, classification techniques are often used for various things, including predicting student performance, classifying types of diseases, predicting fraud in credit card transactions and many more things that can be helped by using classification techniques (Y. Pristyanto, NA et al. 2017). Classification is the process of finding a model or pattern that can describe and differentiate classes in a dataset. The goal is that the model can be used to predict objects with unknown class labels. The model is based on training data analysis. The model from the classification results can be used to predict future data trends (T. M. Christian et al. 2014).

Improving the classification accuracy of text documents still needs further investigation, so that a classification system is developed which is a combination of several single classifiers (ensemble method). The results from each base classifier are combined to classify new data. Several studies conducted by Bauer and Kohavi (1999), Dietterich (2001) and Tsymbal (2002) show that the ensemble method can improve classification accuracy compared to the results of a single classification system, if the classifier is accurate and very diverse (Dietterich, 2001) .

The literature on text classification is very extensive, therefore only works that are closely related to this research are discussed. Various relevant studies have been carried out by previous researchers as follows:

1. M. Aryuni (2013) in his research: "Implementation of Ensemble Feature Selection and Feature Clustering in Text Document Classification". the static selection integration technique can increase the average accuracy by 10% and 10.66%, respectively. While the majority voting integration technique for the same trial can increase the average accuracy by 10% and 12%, respectively

2. Yoga Pristyanto (2019) in his research: "Application of the Ensemble Method to Improve Classification Algorithm Performance in Imbalanced Datasets". The results obtained for the accuracy of each classifier in the User Knowledge Modeling Dataset are 91.98%, 91.52%, 94.37%, respectively.
3. L. Rokach (2009) in his review research: “Ensemble-based classifiers”, reviews existing ensemble techniques and can be used as tutorials for practitioners who are interested in building an ensemble-based system.
4. L. Rokach (2009) in his review research: “Ensemble-based classifiers”, reviews existing ensemble techniques and can be used as tutorials for practitioners who are interested in building an ensemble-based system.
5. M. Chezian, et al. (2015) in his research: "Performance Evaluation of Machine Learning Techniques for Text Classification". examined machine learning techniques Naive Bayes, Support Vector Machine, K-Nearest Neighbor and Decision Tree for Text Classification and compared with each other on their performance, so that they can be used as literature reference for researchers who use machine learning to solve problems
6. Han Liu (2018) in his research: "Nature-inspired framework of ensemble learning for collaborative classification in granular computing context", examines the ensemble technique with the Bagging approach, to obtain an increase in overall accuracy in granular computing settings and the results obtained show good results.
7. Zu Hong et al. (2019) in his research: "An Effective Text Classification Model Based on Ensemble Strategy", proposed a new method called LAC_DNN to achieve text classification based on various approaches to feature representation and classifiers. LAC_DNN achieved superior performance with accuracy of 97.44% and 97.43% on the text dataset of Fudan and Netease news, respectively.
8. Ming Li et al. (2018) in his research: “Text classification based on ensemble extreme learning machine”. This research also develops a text classification framework that combines vector words and AE1-WELM. The experimental results show that the method used provides an accurate, reliable, and effective classification.

The contributions made in this study are: Applying the use of a classification ensemble with different data sets from previous researchers, a different technique approach, namely by using the Majority Voting Technique in the classification ensemble, then the classifier is with 5 different classifiers. heterogeneous consists of a basic classifier algorithm and a meta model algorithm, so it is hoped that the results of accuracy, sensitivity and g-mean will be better for the choice of answer predictions in this test case.

The definition of text classification is explained, besides that, it is also explained about ensemble learning and machine learning algorithms used in this study.

1. Text Classification

Grouping, classification, and categorization are the main techniques followed in text analysis (Vasa, 2016). This is the process of transferring, for example, documents to a specific class label (eg "History") among other available class labels such as "Education", "Medicine" and "Biology". Thus, text classification is a mandatory phase in knowledge discovery (Vasa, 2016).

In the field of machine learning, classification techniques are often used for various things, including predicting student performance, classifying types of diseases, predicting fraud in credit card transactions and many more things that can be helped by using classification techniques (Y. Pristyanto et al., 2017). Classification is the process of finding a model or pattern that can describe and differentiate classes in a dataset. The goal is that the model can be used to predict objects with unknown class labels. The model is based on training data analysis. The model from the classification results can be used to predict future data trends (T. M. Christian, 2014).

There are several classification algorithms that are often used in research, related to machine learning, namely Decision Tree (DT) (G. Gray et al., 2014), Neural Network (NN), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) (M. Mayilvaganan and D. Kalpanadevi, 2014).

Text classification is to classify text documents into a set of predefined categories. As one of the main technologies of text mining, text classification is widely used in the fields of information retrieval, search engines, inquiry systems, public opinion analysis and emotional analysis. With the rapid development of network technology, the number of web pages is growing exponentially. , and efficient and personal retrieval of information needs to develop a more accurate and effective text classification technology. Currently, according to Ming Li (2018) the most popular text classification methods include K-NN, Naive Bayes, decision trees, maximum entropy, support vector engines (SVM), neural networks, and fuzzy theory.

A. Process of Text Classification

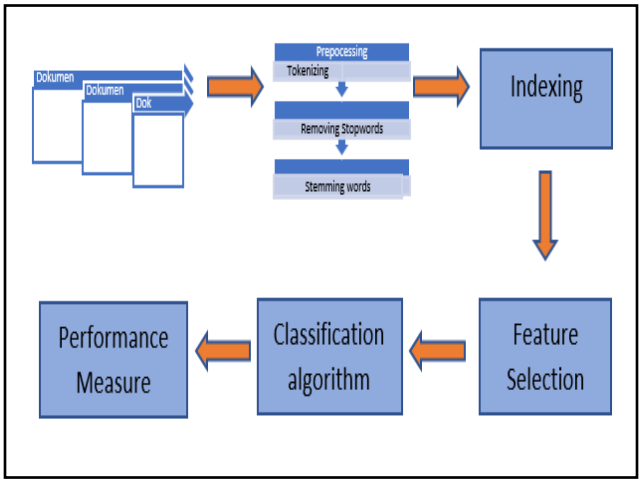

Figure1. Document Classification Process

2. Machine Learning Methods

The increase in data volume, speed and variety requires automation in text processing techniques including text classification. In some situations, defining a set of logical rules using knowledge engineering techniques and based on expert opinion to classify documents helps

automate the task of classification. Text classification can be divided into three categories: supervised text classification, unsupervised text classification, and semi-supervised text classification based on learning principles followed by data models (Korde & Mahender, 2012). According to Nitin H (2016) There are two types of machine learning techniques - supervised learning and Unsupervised learning: Supervised learning: Based on previous examples of prelabels, machines learn how to predict future test samples, based on the following categories:

- Classification: This is used when it is necessary to predict whether a test sample belongs to a class. If there are only two classes, this includes the problem of binary classification; otherwise it is a multiclass classification.
- Regression: This is used when it is necessary to predict continuous variables, such as the house price index and stock

  Unsupervised learning: When there is no labeled data and still need to predict class labels, this learning is called unsupervised learning. When it is necessary to group items based on similarities between items, this is called a grouping problem. Meanwhile, if it is necessary to represent high-dimensional data in a lower dimension, it is more of a dimensional reduction problem.

  Semi-supervised learning: This is classroom assignment and supervised learning techniques that also make use of unlabelled data for training. As the name suggests, it is more of a middle ground for supervised and unsupervised learning, where we use a small amount of labeled data and a large amount of unlabeled data to build a predictive machine learning model.

  Reinforcement learning: This is a form of machine learning in which agents can be programmed by rewards and punishments, without specifying how the task should be accomplished.

3. Ensemble Learning Method

The Ensemble method builds a predictive model by integrating several models, which can be used to improve predictive performance (Rokach, 2010). Ensemble classifier has been applied in various prediction system applications including protein structure prediction (Wu, 2010), breast cancer diagnosis (Huang, 2010), facial recognition (Yu, 2009), and document classification (Bennet, 2005) (Katakis, 2010).

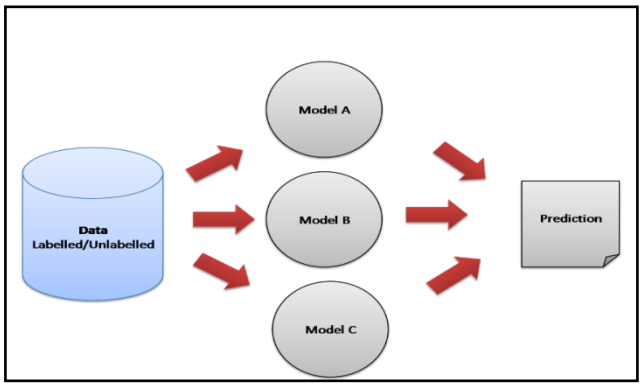

Figure2. Ensemble Concept

Table 1 some basic concepts that must be considered before going into detail, here are the types of ensembles according to (L. Rokach, 2009): Averaging: This is defined as taking the average of

the predictions from the model in case of a regression problem or when predicting the probability for a classification problem.

Model 1	Model 2	Model 3	Average Prediction
45	40	65	50

Majority Vote: Defined as taking a prediction by vote / maximum recommendation from several prediction models while predicting the outcome of a classification problem in table 2.

Model 1	Model 2	Model 3	Voting Prediction
1	0	1	1

$$class(x) = argmax \ (\sum g \ (yk(x), ci)$$
$$ci \in dom(y) \quad k$$

Weighted average: In this case table 3, different weights are applied to the predictions of several models then take the mean which means giving high or low importance to the output of a particular model.

	Model 1	Model 2	Model 3	Weight Average Prediction
Weight	0.4	0.3	0.3	
Prediction	45	40	60	48

4. Ensemble Method

The three most widely used techniques or methods for combining predictions from various models are:

1) Bagging: Bagging is also referred to as bootstrap aggregation. Bootstrapping is a sampling technique by selecting 'n' observations or rows from the original dataset of 'n' rows as well. But the key is that each row is selected with replacement from the original dataset so that each row is equally likely to be selected in each iteration. Let's say you have 3 rows numbered 1, 2 and 3.
2) Boosting: Boosting is a sequential technique in which, the first algorithm is trained on the entire dataset and subsequent algorithms are constructed by adjusting the residuals of the first algorithm, thereby giving higher weights to observations that were less predictable by the previous model.
3) Stacking: The generalized ensemble of Stacked is known as stacking (Wolpert 1992). It combines the decisions of different basic classifiers in a single ensemble architecture. In stacking multiple layers the machine learning models are placed one on top of the other where each model passes their predictions to the model in the layer above it and the top layer model takes decisions based on the output of the model in the layer below it.

5. Performance Evaluation

This is the last stage of text classification, where the evaluation of text classifiers is usually carried out experimentally, not analytically. Experimental evaluation of a classifier, rather than concentrating on the problem of Efficiency, usually tries to evaluate the effectiveness of the classifier, i.e. its ability to make appropriate categorization decisions. An important issue of Text categorization is how to measure the performance of the classifier.

For a Classifier, the prediction of the model is checked through the creation of a Confusion matrix and finally the model's f-score is calculated. The Confusion Matrix is nothing but a cross table between predicted and actual classes. It is a very powerful table which can calculate multiple live scores, Accuracy, Precisson, Recall, Specificity, F-score, etc, which can be used to check the predictions of the model being made.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

- True Positive is the number of positive records that are classified as positive.
- False positives are the number of negative records that are classified as positive.
- False negatives are the number of positive records that are classified as negative.
- True negative is the number of negative records that are classified as negative.

In this paper, to create a predictive model on the chatbot using the ensemble technique on the classifier, it is expected to increase the accuracy of predicting the chatbot's answers to the questions given by the user.

Based on the conceptual framework that has been put forward, the hypotheses that are expected to be generated are as follows is suspected that the application of the ensemble 5 method of machine learning algorithms to the classifier model can increase the accuracy and stability in predicting the results of chatbot answers.

METHODOLOGY

1. Research Method

The phase of activities in this research can be seen in Figure 3. In phase 1, problem identification, literature study and problem formulation are carried out. After that, data collection, data preprocessing, development of predictive models for chatbots, model testing and analysis and discussion were carried out. The following is an explanation of each phase.

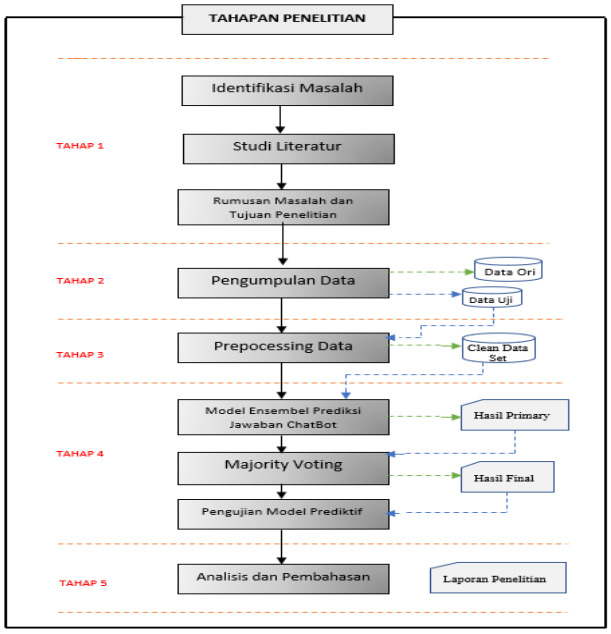

Figure 3. Process Ensemble Method Research

The data used in this study were in the form of questions and answers that were collected from various sources into the training set. This will help in improving the predictions of the chatbot, so that the chatbot can match new questions to the created

training set. The better the data collected, the better the ChatBot will respond.

The steps are as follows:

- The data set collected in this study was written in csv format.
- The data will be grouped based on similar questions into one class in the training set. so when the ChatBot gets a new question, it can easily classify it to one of the classes in the training data and provide an answer as text which is the usual answer to a question that belongs to this type of class.
- The second approach is to match the new question with all the questions in the training set and find the most similar questions in the training set. Thus the answer should also be the same for the user.

2. Data Visualization

Visualizing the data means understanding what types of data are held in this research. So because this research uses text, it requires knowledge of the most commonly used words in conversation.

3. Pre-processing of data

Data always has noise where there is additional meaningless information in it. So it is very important to clean it and save only the information needed to create the model. The pre-processing/data cleaning steps to be taken for this are unctuation removal, stop words removal, negation handling and POS based preprocessing.

4. Featured Selection

This phase is an important stage in text mining. One of the important functions provided by this stage of the process is to be able to choose what terms or words can be used as important representatives for the collection of documents to be analysed, this is:

TF-IDF word weighting

objective to give value/weight to the terms contained in a document. The weight given to the term depends on the method used. In this way, it is hoped that each class has a feature that indicates its uniqueness.

5. Ensemble Predictive Model Design

This model ensemble will work by training different models on the dataset and having each model make predictions individually. The predictions of these models are then combined in an ensemble model to make a final prediction.

The design of the ensemble learning framework for this research is as follows:

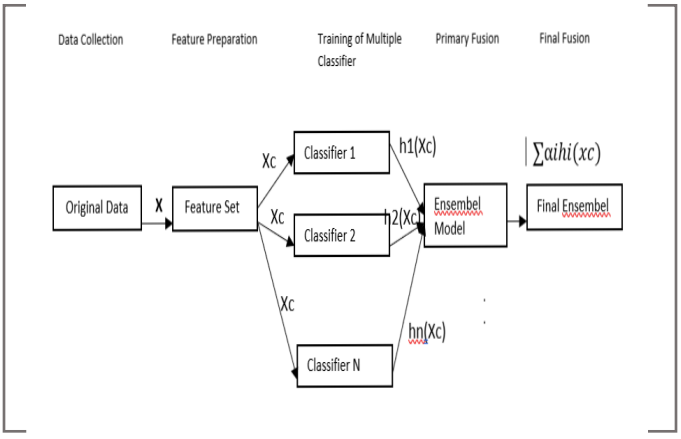

Figure 4 Design of the Ensemble Model Framework

6. Ensemble Method Analysis

The final result in this study will use the Voting Classifier and the calculation of the Hard/Majority voting technique in which the ensemble model makes predictions with the most votes. For example, if three models are used and predicted [1,

0, 1] for the target variable, the final prediction that will be made by the ensemble model is 1, because two of the three models estimate 1.

In this study, 5 different machine learning models will be used to be included in the Voting Classifier derived from the basic classification algorithm and meta-algorithm:

1. Logistic Regression
2. Random Forest
3. Multinomial NB
4. Decission Tree
5. SVM (Support Vector Machine)

This phase, in order to create an ensemble learning model, it is necessary to have data collected and explored for preprocessing first, where Collecting Data - This is the most tedious part of this research model, collecting data from various sources and collecting it. But this is what will help in improving the predictions of this chatbot. The better the data collected, the better the ChatBot will respond.

The second approach in this study is to match the new questions with all the questions in the training set and find the most similar questions in the training set. Thus the answer must also be the same for them.

	question	answer
0	how are you doing today ?	greeting
1	have a nice day	goodbye
2	What for lunch	sandwich
3	kabar baik alhamdulillah	salam
4	nama panggilanmu siapa	nama

For this reason, this paper used a multiclass dataset consisting of 2 columns with 6 classes, namely:

1.2 question and answer columns (question and answer)

a) **Input:** questions

Example: "assalammu'alaykum"

b) **Output:** Kelas

Example: Salam

2.6 Class:

a) Greeting
b) Salam
c) Hari
d) Nama
e) Sandwich
f) Goodbye

By having multiple questions and answers in the training set, the ChatBot can match new questions to the training set. After understanding the problem and translating it into a technical manual, similar questions were grouped in this research training into one class. So now when a new question is obtained, it can be easily classified into one of the classes in the training data that is owned and provides answers as text which are the usual answers to questions belonging to this type of class.

Before entering into the machine training training model, you must first look at some examples and the number of questions and answers in each class that are visualized in table 2:

```
import pandas as pd
df = pd.read_csv('data.csv')
df.head()
```

Table 4. Example of text data in the Data Set class

Classifiers and learning algorithms cannot directly process text documents in their original form, because most classifiers and learning algorithms expect fixed size numerical feature vectors rather than raw text documents of variable length. Therefore, during the preprocessing step, the text is converted to a more manageable representation. The steps in this are:

1. Visualizing data - Once you have the data, the first thing to do is to understand what kind of data you have. In this study, the data possessed was text data so that we intuitively wanted to know what words were most commonly used in the existing questions. In this study:

a. Bag of word

One common approach to extracting features from text is to use a bag of words model: a model in which for each document, the narrative questions of the study, the presence (and their frequency) of words are considered, but the order in which they are happens to be ignored.

For this reason, wordcloud is first created to get a better understanding of the corpus data set that is owned, visualized in Figure IV.1 below:

From the data set:

Length of training set: 51

Figure 4 Display of word cloud results on the training data representation

The number of classes in the Data Set and the most words that come out are displayed so that they are clearly visualized, as illustrated in the following graphs in Tables 4 and 5:

```
Unique answers are {'goodbye',
'sandwitch', 'nama', 'greeting',
'salam', 'hari'} and number of
unique answers are 6
```

7. Text Classifications

After the text data input is ready, we can now start training this system model. To train supervised classifiers, first convert the "Question input narrative on the chatbot" to a numeric vector. This study explores its vector representation such as the TF-IDF weighted vector. After having a vector representation of this text, supervised classifiers were trained to train invisible "chatbot question input narrative" and predict "classes" that match the existing data sets.

After all of the above data transformations are carried out, the features and labels are owned, it is time to train classifiers, namely a number of algorithms used in this study.

After adjusting the training set, several predictions are made.

8. Model Selections

For classification, the characteristics of selecting the basic classifier model must be as accurate as possible and have as many errors as the class that is actually the majority class choosing a model for this essay is based on the results of the classification characteristics table in table 6 below:

TABLE 10.1. *Some characteristics of different learning methods. Key:* ● *= good,* ● *=fair, and* ● *=poor.*

Characteristic	Neural nets	SVM	Trees	MARS	k-NN, kernels
Natural handling of data of "mixed" type	●	●	●	●	●
Handling of missing values	●	●	●	●	●
Robustness to outliers in input space	●	●	●	●	●
Insensitive to monotone transformations of inputs	●	●	●	●	●
Computational scalability (large N)	●	●	●	●	●
Ability to deal with irrelevant inputs	●	●	●	●	●
Ability to extract linear combinations of features	●	●	●	●	●
Interpretability	●	●	●	●	●
Predictive power	●	●	●	●	●

RESULT AND DISCUSSION

Assessing Model Performance is done when the model is finished, as a measure of how well it is known how well the model is performing. For Classifiers, model predictions are checked through the creation of a Confusion Mactrix and finally the f-score is calculated from the model. The Confusion Matrix is nothing but a cross table between the predicted class and the actual class. It looks like a simple table but there are several predictive scores that can be calculated from it so it is a very strong table. In this matrix, several live scores are calculated for Accuracy, Precisson, Recall, Specificity, F-score. Which can be used to check the predictions of the model you have.

A. Testing Classifier with 6 input test questions (test data) on the ChatBot which represents the classes that belong to the data set. Testing was carried out to determine the performance of classifiers that have been trained by providing test data, with the following results which are described in the graphic images from Figures 6 to 11.

1. *Enter* Question = *"hi how are you today?"*

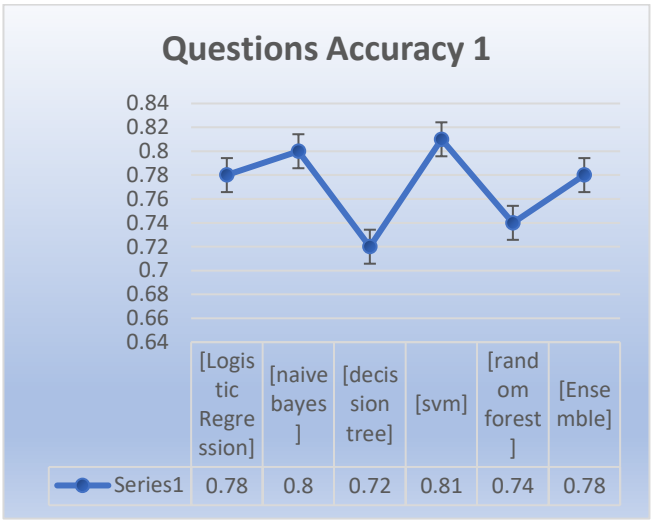

Figure 6 Accuracy Graph 5 basic classifiers and Ensemble MV in question 1

2. Enter Question = “would you like to have fried rice for your breakfast?”

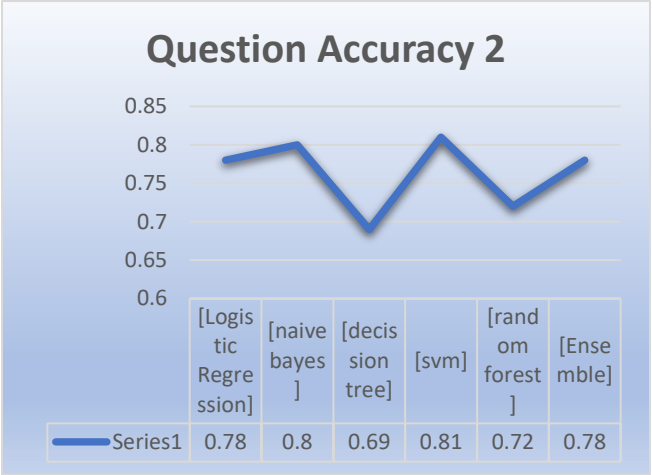

Figure7 Accuracy Chart of 5 basic classifiers and Ensemble MV in question 2

3. Enter Question =” kapan kamu mau main ke rumahku?”

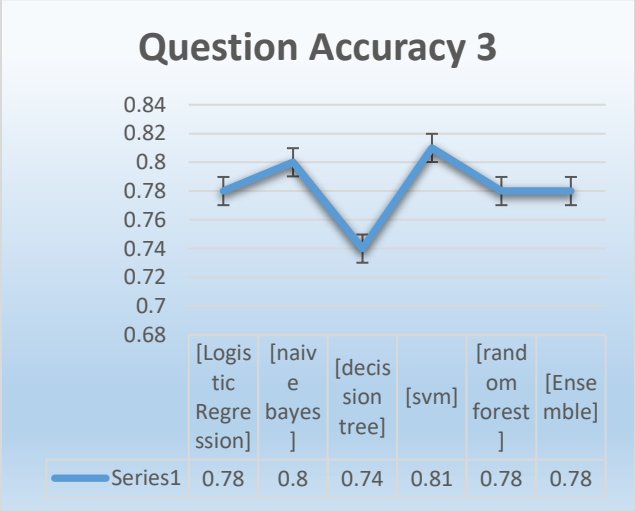

Figure 8 Accuracy Graph 5 basic classifiers and Ensemble MV in question 3

4. Enter Question = “ok its time for me to go home now, see you later”

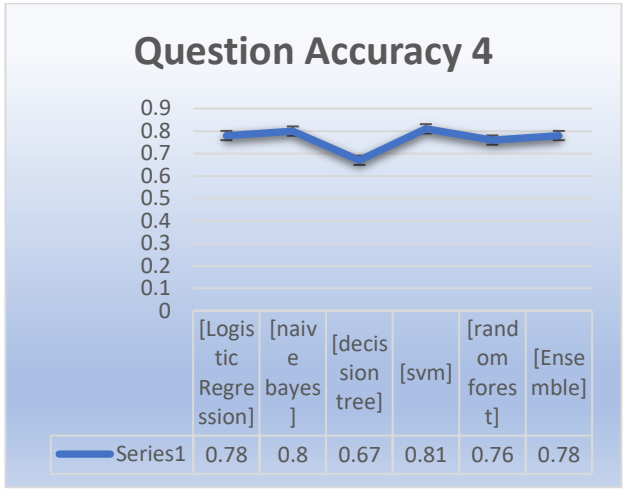

Figure 9 Accuracy Graph 5 basic classifiers and Ensemble MV on question 4

5. Enter Question = “hari minggu besok libur, mau bersepeda bersamaku?”

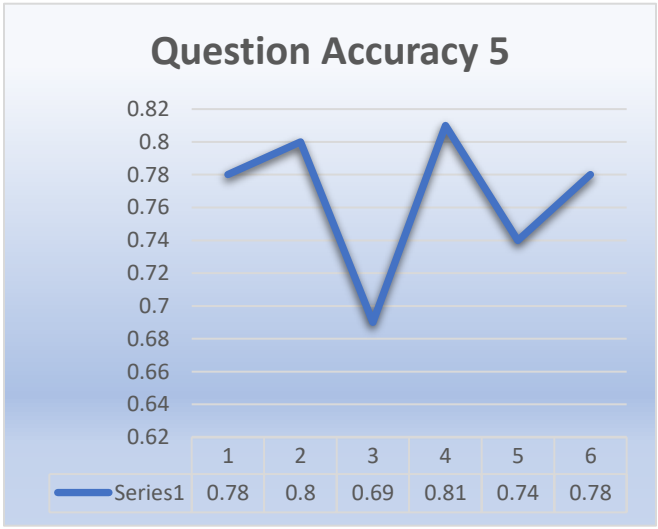

Figure 10 Accuracy Graph 5 basic classifiers and Ensemble MV on question 5

6. Enter Question =”apakah kamu punya binatang kesayangan?”

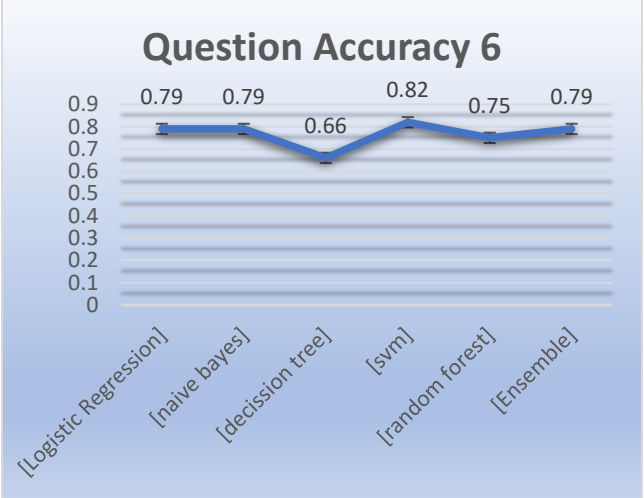

Figure 11 Accuracy Graph 5 basic classifiers and Ensemble MV on question 6

Selected Class Based on Ensemble MV		
Input	**Class**	**Accuray MV**
hi,how are you today?	greeting	0.78
would you like to have fried rice for your breakfast	goodbye	0.78
kapan kamu mau main ke rumah	salam	0.78
ok its time for me to go home now, see you later	goodbye	0.78
hari minggu besok libur, mau bersepeda bersamaku?	hari	0.78
apakah kamu punya binatang kesayangan?	nama	0.79

The prediction results from Table 7 basic classifiers that have been previously trained using the data train, which are given input test data as many as 6 questions which are considered to represent each class in the data set. From the results obtained for the 6 test data input questions, there were class prediction results that were found to be different in some classifiers. This is where the role of the Majority Voting Ensemble is seen in determining the final results of class selection according to the test data

B. Model Testing with a Confusion Matrix

Test the ability of the model created in predicting the given test data.

Table 8 Multiclass Confusion Matrix (6 classes)

	precision	**recall**	**f1-score**	**support**
greeting	0.5	1	0.67	1
sandwitch	1	1	1	1
salam	1	1	1	1
nama	1	1	1	1
hari	1	1	1	1
goodbye	1	0.5	0.67	2
accuracy			0.86	7
macro avg	0.92	0.92	0.89	7
weighted avg	0.93	0.86	0.86	7

From the Confusion Matrix table obtained, the following information is obtained:

Accuracy = 86%
Avg macro for precision and recall respectively = 92%, f1-score = 89%
Weighted avg for precision = 93% and respectively for recall and f1-score: 86%

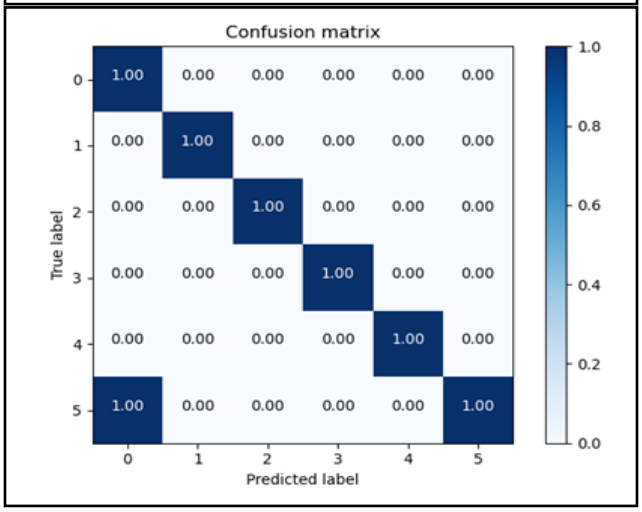

Figure11. Confusion Matrix graphs for test data from 6 classes

CONCLUSION

Ensemble is a tried and tested method to greatly improve the performance of machine learning models. The Majority Voting (MV) method is generally used for classification problems.

In this research, we discussed some of the basic ideas behind the ensemble. It is worth mentioning here that the ensemble can be joined together to make a more complex ensemble. This can help, for the most part the performance drops. Combining several models trained using a stochastic learning algorithm.Combining several models with different hyperparameters and The limitation of the voting ensemble is that it treats all models equally, which means that all models contribute equally to predictions. This is a problem if some models are good in some situations and bad in others.

REFERENCES

(Galitsky B., Explainable Machine Learning for Chatbots. In: Developing Enterprise Chatbots., 2019).

Kuncheva, L. and Whitaker, C., Measures of diversity in classifier ensembles, Machine Learning, 51, pp. 181-207, 2003

Sollich, P. and Krogh, A., Learning with ensembles: How overfitting can be useful, Advances in Neural Information Processing Systems, volume 8, pp. 190-196, 1996.

Brown, G. and Wyatt, J. and Harris, R. and Yao, X., Diversity creation methods: a survey and categorisation., Information Fusion, 6(1), pp.5-20, 2005.

Adeva, J. J. García; Cerviño, Ulises; Calvo, R. "Accuracy and Diversity in Ensembles of Text Categorisers" (PDF). CLEI Journal. 8 (2): 1–12.

Ho, T., Random Decision Forests, Proceedings of the Third International Conference on Document Analysis and Recognition, pp. 278-282, 1995.

Gashler, M.; Giraud-Carrier, C.; Martinez, T. "Decision Tree Ensemble: Small Heterogeneous Is Better Than Large Homogeneous" (PDF). The Seventh International Conference on Machine Learning and Applications. 2008: 900–905..

Y. Pristyanto, N. A. Setiawan, and I. Ardiyanto, “Hybrid Resampling to Handle Imbalanced Class on Classification of Student Performance in Classroom,” in The First International Conference on Informatics and Computational Sciences (ICICoS 2017), 2017, pp. 215–220.

T. M. Christian and M. Ayub, “Exploration of classification using NBTree for predicting students’ performance,” in Proceedings of 2014 International Conference on Data and Software Engineering, 2014, pp. 1–5.

Liu, Han and Haig, Ella, “Nature Inspired Framework of Ensemble Learning for

Collaborative Classification in Granular Computing Context", 2019/10/01

G. Gray, C. McGuinness, and P. Owende, "An application of classification models to predict learner progression in tertiary education," 2014 4th IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 549–554, 2014.

M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," Commun. Netw. Technol. (ICCNT), 2014 Int. Conf. Comput. Intell. Comput. Res., pp. 113–118, 2014

Ming Li, Peilun Xiao,and JuZhang," Text classification based on ensemble extreme learning machine", Natural Science Foundation of China(No.61672488), Ministry of science and Technology, 2018.

Zhu Hong1, Jin Wenzhen1 and Yang Guocai1, "An Effective Text Classification Model Based on Ensemble Strategy", Journal of Physics: Conference Series, Volume 1229, 2019 3rd International Conference on Machine Vision and Information Technology (CMVIT 2019) 22–24 February 2019, Guangzhou, China.

Fan Huimin, Li Pengpeng, Zhao Yingze, Li Danyang, "An Ensemble Learning Method for Text Classification Based on Heterogeneous Classifiers",2018 International Conference on Sensor Network and Computer Engineering (ICSNCE 2018)

Lior Rokach, "Ensemble-based classifiers", 19 November 2009, Springer Publisher, Science Business Media B.V. 2009

Mediana Aryuni, "Penerapan Ensemble Feature Selection Dan Klasterisasi Fitur Pada Klasifikasi Dokumen Teks", ComTech Vol.4 No. 1 Juni 2013: 333-342

Vandana Korde and C Namrata Mahender, "Text Classification And Classifiers: A Survey", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012.

R.Manicka chezian dan C.Kanakalakshmi, "Performance Evaluation of Machine Learning Techniques for Text Classification", Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications, 27th March 2015.

Nitin Hardeniya, Jacob Perkins, Natural Language Processing: Python and NLTK, Packt Publishing, Copyright ©2016

Yoga Pristyanto, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset", Jurnal TEKNOINFO, Vol. 13, No. 1, 2019, 11-16, ISSN: 2615-224X

Wolpert DH (1992) Stacked generalization. Neural Netw 5(2):241–259.

https://doi.org/10.1016/S0893-6080(05)80023-1

Mccallum A, Nigam K. A comparison of event models for Naive Bayes text classification. In: AAAI-98 workshop on 'Learning for Text Categorization'; 1998.

C.M. Bishop, "Pattern recognition and machine learning" (information science and statistics) Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006).