Multiclass Intent Classification for Chatbot Based on Machine Learning Algorithm

W. M. Amir Fazamin W. Hamzah
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
amirfazamin@unisza.edu.my

Mohd Kamir Yusof
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
mohdkamir@unisza.edu.my

Ismahafezi Ismail
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
ismahafezi@unisza.edu.my

Mokhairi Makhtar
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
mokhairi@unisza.edu.my

Hasnah Nawang
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
hasnah.nawang@kberang.mrsm.edu.my

Azwa Abdul Aziz
Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin
Terengganu, Malaysia
azwaaziz@gmail.com

***Abstract*— In recent years, the use of Chatbots has grown significantly in various industries, including support systems, education, health care, tourism, entertainment, and banking. Chatbot for education can provide instant feedback in interactions session with students. The responses from Chatbot can be based on machine learning algorithms or multiple heuristics techniques to select responses from the predefined library. The generated response depends on the user's intent of using the Chatbot. There are various classes or categories in the user's intent. However, the user's intent class cannot simultaneously belong to multiple classes. Therefore, this research proposed a multiclass intent classification for the Chatbot based on the machine learning algorithm. The findings of this research showed that Linear SVC is the best machine learning algorithm model for multiclass intent classification. The results of the analysis proved the accuracy of the prediction using Linear SVC.**

Keywords—chatbot, intent, classification, machine learning

I. INTRODUCTION

Today's advancement in digital technology presents a wide use of Chatbots over time [1][2]. Chatbot in education allows students to constantly interact with learning content. Using the Chatbot in education is an innovative way to overcome any barriers between technology and education. In addition, Chatbots can provide students with instant feedback during conversations [3]. The Chatbot can generate initial responses based on machine learning algorithms or select responses from a predefined library using several heuristics techniques [4][5]. The generated response depends on the user's intent when using the Chatbot.

There are various classes or categories in user's intent. Each intent can be classified into one of the classes. However, classes or categories for user's intent cannot belong to more than one class simultaneously. The same intent but categorised in many classes can cause inaccurate responses to Chatbot users.

Therefore, this research proposed a multiclass intent classification for Chatbot based on machine learning algorithm models to solve the problem of various classes or categories in the user's intent. This solution explains the systematic method and evaluation metrics of multiclass intent classification. The best model of machine learning algorithms will be used to predict the class of user intents. Multiclass intent classification helps Chatbot to make better sense of user intents and find patterns. Using these user's intent patterns offers greater insights into making more accurate responses.

II. RELATED WORK

A. Chatbot

Chatbot is a computer programme that simulates human language using a text-based dialogue system [1]. Users can communicate by text or voice input via a computer screen with text or audio/speech output. Intent refers to the goal of the Chatbot user when writing a question or comment. The Chatbot was initially developed using a simple keyword matching technique to find user input matches. After that, it is developed using different pattern matching algorithms to simulate fiction or true personality [2][3].

There are two models of Chatbot architecture commonly used for developing Chatbots; the generative model and the retrieval-based model [6][7]. Generative models are difficult to construct and train. Typically, generative models require millions of examples to train deep learning models to achieve excellent conversation quality, and users cannot be certain of their responses. Meanwhile, retrieval-based models are simpler to implement in Chatbot development, which has the potential to produce more predictable responses [8][9].

B. Classification

Classification is a predictive modelling problem in which an example of input data is used to predict a class label. From the point of view of modelling, classification needs a training dataset with many examples of inputs and outputs to learn from. Classification is broadly distinguished into binary classification, multiclass classification and multi-label classification [10][11][12].

Binary classification is when there are only two categories to classify data points. For example, detecting if a fruit is ripe (1) or not (0) or classifying whether COVID-19 testing is positive (1) or negative (0).

Multiclass classification is a type of classification task in machine learning that has more than two outputs, or classes. A sample can only belong to one class when it comes to multiclass classification. For example, the classification of news content into news categories. The categories can be a crime, sports, business, entertainment, politics, and current issues.

2022 Seventh International Conference on Informatics and Computing (ICIC) | 979-8-3503-4571-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIC56845.2022.10006979

Multi-label classification allows for the classification of datasets with more than one target variable. In multi-label classification, several labels become the outputs for a given prediction. When making predictions, a given input may belong to more than one label. For example, when predicting a given food taste category, it may belong to sweet, salty, bitter, sour or all, simultaneously. In this example, the multi-labels can be assigned to a given food taste.

C. *Machine Learning*

Machine learning is related to the design of algorithms that allow computers to learn. Learning does not necessarily involve awareness, but learning is a matter of finding statistical equations or other patterns in the data. Therefore, an increased number of machine learning algorithms will hardly resemble how humans can approach learning tasks. However, learning algorithms can illustrate learning difficulties in different environments. Machine learning algorithms are structured into taxonomies, based on desired algorithm results. Common types of algorithms include supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning and transduction [12][13].

Supervised learning is a situation where input variables (x) and output variables (y) learn the mapping function through an algorithm from input to output $y = f(x)$ to get close to the mapping function so that when new input data (x) is available, predictions for output variables (y) can be made. Examples of supervised learning algorithms are Logistic Regression, K-Nearest Neighbours, Decision Trees, Random Forest, Naïve Bayes, and Support Vector Machine [13][14][20].

Each algorithm has its own advantages, such as the fact that Decision Trees do not require any assumptions about the linearity of the data and can therefore be utilised in situations where the parameters are non-linearly connected. Logistic Regression can forecast the likelihood of a target variable. The target or dependent variable is binary. This means that there will only be two classes. The categorical dependent variable is best predicted using logistic regression. The Random Forest is a set of Decision Trees. It is a form of an ensemble method that aggregates the outcomes of numerous predictors. The Random Forest also employs the bagging technique, which allows each tree to be trained on a random sample of the original dataset and gets the majority vote from the trees. The Naïve Bayes performs well with categorical input variables, converges faster, and requires less training data than other discriminative models like logistic regression. The Multinomial Naïve Bayes has the best classification performance on the training data and is more effective at classifying the time ahead data accurately. The Support Vector Machine determines the optimal way to classify data based on its position relative to a border between positive and negative classes. This is the hyperplane, which minimises the distance between data points of different classifications. Support Vector Machine, like Decision Tree and Random Forest, can be used in both classification and regression. The Linear SVC (linear support vector classifier) is for classification problems. [13][14][15][16][20].

However, all the selected algorithms must be tested to determine the best algorithms for classification. The test must be done to ensure its performance before making the final decision. The machine learning algorithm classifiers such as Logistic Regression, Random Forest Classifiers, Multinomial Naïve Bayes, and Linear Support Vector Classifiers (Linear SVC) are usually used to perform multiclass classification.

D. *Multiclass Classification Research*

A lot of research on using multiclass classification has been done before. Some research includes Text Classification, Image Classification, Malware Classification, Medical Diagnosis, and Social Media Content Analysis.

1) *Text Classification:* Moreo et al. have proposed word-class embeddings (WCEs) to improve multiclass classification accuracy [21]. While Parmar et al. have done Multiclass Text Classification and Analytics for Improving Customer Support Response through different Classifiers [22].
2) *Image classification:* The research of multiclass skin cancer image classification by convolutional neural networks was done by Maron et al. for better reflecting clinical differential diagnoses [23]. Vang et al. have proposed a deep learning framework for the multiclass breast cancer histology image classification problem [24].
3) *Malware classification:* Malware classification research by Verma et al. has proposed binary texture analysis over greyscale images created directly from their malware executables [25]. While Ghouti and Imam have done malware classification using compact image features and multiclass support vector machines [26].
4) *Medical diagnosis:* David et al. have proposed a deep convolutionary Neural Network (DCNN) based architecture to diagnose and classify brain tumours and assign grades to them [27]. Kuo et al., in their research, have identified the appropriate diagnosis code for type 2 diabetes mellitus patients by building a multi-class prediction model that is both parsimonious and possesses minimum features [28].
5) *Social media content analysis:* Mustafa et al. have proposed research on Multiclass Depression Detection in Social Media Based on Sentiment Analysis [29]. While Bouazizi and Ohtsuki have done multiclass sentiment analysis on Twitter. That research is about the feasibility of quantification and proposed an approach to perform it on a data set made of tweets for 11 different sentiment classes [30].

However, the research on multiclass intent classification is relatively less done. Therefore, this research is worth doing to solve the problem of various classes or categories in the user's intent of Chatbot.

III. METHOD

Multiclass intent classification is the process of precisely labelling an input in the form of a natural language utterance from a predetermined set of intents. The machine learning model is trained to output a predicted classification for a given intent. There are various methods or approaches to implement multiclass intent classification. The selection of methods is usually based on the dataset's type and the number of data. Figure 1 shows the process for the systematic method and evaluation metrics of multiclass intent classification based on the machine learning algorithm. The process includes data collection, data pre-processing, dataset splitting, building a model, evaluation metrics and prediction [17].

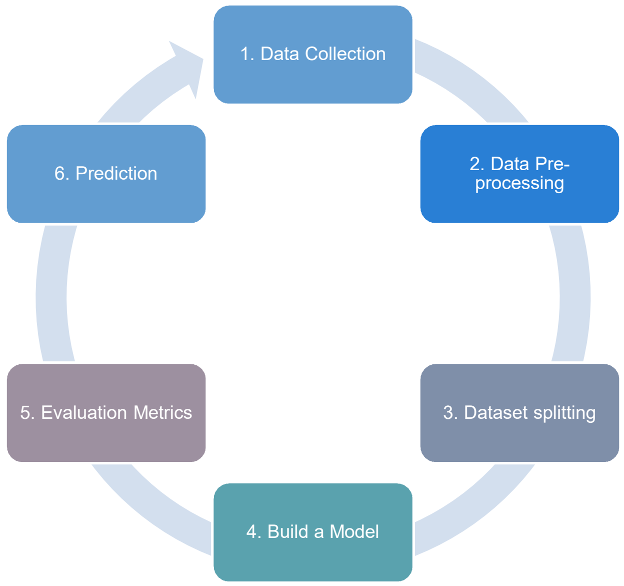

Fig. 1. Multiclass Intent Classification Process

A. Data Collection

The dataset used is a predefined library of Chatbot user's intent for the Web Application Development subject. The intent contents are related to HTML, CSS, JavaScript, Java Server Pages (JSP), Servlet and MySQL. Creating a dictionary object encodes the intents classes as integer values.

B. Data Pre-processing

Data pre-processing is an important stage in developing a machine learning model, and its success is dependent on how effectively the data has been pre-processed. The data in this research is the user's intent who use Chatbot, which is text-type data. Therefore, text pre-processing is used in this research. Text pre-processing is the method of cleaning and preparing text data in natural language processing (NLP). It changes text into a form that is easier to understand so that algorithms for machine learning can work better [18][19]. The text pre-processing steps that are involved are Stemming, Removing Stop Words, Lower Casing, Tokenization and Eliminating Unnecessary Characters.

1) *Stemming:* Stemming is the process of reducing a word to its stem or root. It eliminates the word's affixes, leaving only the root.
2) *Removing Stop Words:* The most prevalent words in any language are stop words. However, they contribute little to the text's clarity. Stop words consist of conjunctions, pronouns and articles. Eliminating stop words will allow the model to concentrate on training-relevant terms.
3) *Lower Casing:* The dataset is converted to lowercase.
4) *Tokenization:* Separating sentences into smaller word pieces is known as tokens. By examining the word tokens, this method allows the model to comprehend sentences.
5) *Eliminating unnecessary characters:* The text dataset may contain unnecessary characters that do not provide value to the model. Eliminating these characters allows the model to concentrate on essential information.

C. Dataset Splitting

Splitting the data into train and test sets. The data needs to be pre-processed to be fed to the classification algorithm. The original data was split into features (X) and targets (y), which were then split into train (75%) and test (25%) sets. A reasonable rule of thumb is to test with 25% of dataset. So, the algorithms were trained on one dataset and tested on a completely different dataset.

D. Build a Model

Train the machine learning algorithm classifiers models such as Logistic Regression, Random Forest Classifiers, Multinomial Naïve Bayes, and Linear SVC. Plot each model's performance to view the accuracy. Then, compare "mean accuracy" and "standard deviation" to determine the best model.

E. Evaluation Metrics

1) *Use the best model and make predictions:* Train the best model to predict the test data. Find the most correlated n-gram (unigrams and bigrams). The n-gram model counts characters or word sequences to provide rich pattern finding in text.
2) *Model evaluation and classification report:* The model performances are based on precision, recall and F1-score. If all classes are balanced, accuracy is an excellent starting point. When classes are uneven, precision and recall become increasingly critical. Aim for more precision if false positive forecasts are worse than false negatives. Aim for increased recall if erroneous negative predictions are worse than false positives. The F1-score combines precision and recall. Classification reports are made to obtain more insights into model performance. The confusion matrix conveys the model's right and wrong predictions on data.

F. Prediction

Predict unseen data. Predicting is done using the best model.

IV. RESULT AND DISCUSSION

The dataset used contains 941 Chatbot intents (in the form of text statements) related to HTML, CSS, JavaScript, Java Server Pages (JSP), Servlet and MySQL. The intent was divided into three classes: 'explanation', 'solution' and ‘code’ as shown in Table 1. 'Class' refers to the categories in the dataset, 'User’s intent' refers to the user's questions when using the Chatbot, and 'category id' refers to the user's intent classes in integer values. There was a class imbalance that may be a property of the problem domain. The ‘explanation’ class was seen to dominate the question in the dataset. However, this did not affect the classification process.

TABLE I. INTENT CLASSES

No.	Class	User’s Intent	Category id
0	Explanation	What is JSP?	0
1	Explanation	Tell me what is JSP?	0
2	Explanation	What is the meaning of JSP?	0
3	Explanation	Explain what is JSP.	0
4	Explanation	What is Java Server Page?	0
…	…	…	…
695	Solution	How to resolve pass control from one JSP page ...	1
696	Solution	How to mitigate pass control from one JSP page...	1
697	Solution	How to stop pass control from one JSP page to ...	1
698	Solution	How to defend pass control from one JSP page t...	1
699	Solution	How to get secured pass control from one JSP p...	1
…	…	…	…
936	Code	Code example of JSP Implicit Objects	2
937	Code	Code of JSP Implicit Objects	2
938	Code	Give me some sample code of JSP request implicit...	2
939	Code	Code example of JSP request implicit object	2
940	Code	Code of JSP request implicit object	2

The data pre-processing performed involved stemming, removing stop words, lower casing, tokenization and eliminating unnecessary characters. The dataset was divided into a train set (75%) and a test set (25%). Then, a model performance comparison was performed to determine the best machine learning algorithm model for multiclass intent classification. The box plot in Figure 2 displays the accuracy of each machine learning algorithm. There were various box plot shapes and positions. Logistic Regression, Random Forest Classifiers, Multinomial Naïve Bayes (Multinomial NB) and Linear SVC have different centres. Linear SVC was entirely above other machine learning algorithms.

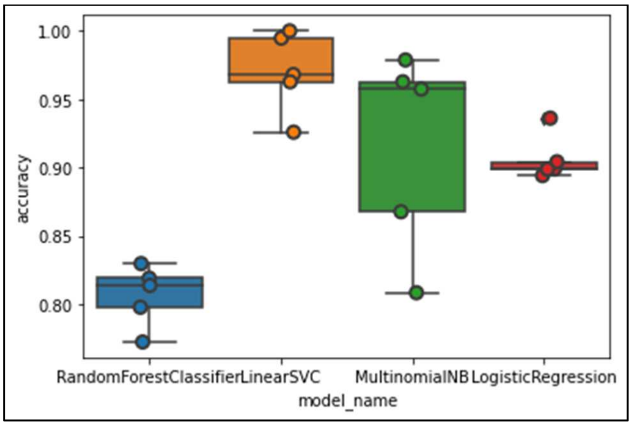

Fig. 2. Box plot of machine learning algorithm models

Table 2 illustrates the data comparison of ‘mean accuracy' and ‘standard deviation’ for each machine learning model. From the table, the accuracy of the Support Vector Machine classifier model, which is Linear SVC (0.970292), outperformed all other machine learning algorithms. Meanwhile, Random Forest Classifier (0.806625) was the machine learning algorithm with very low accuracy. Therefore, the Linear SVC was used for training and making predictions.

TABLE II. COMPARISON OF "MEAN ACCURACY" AND "STANDARD DEVIATION"

Model Name	Mean Accuracy	Standard Deviation
Linear SVC	0.970292	0.029611
Logistic Regression	0.906496	0.016967
Multinomial NB	0.915034	0.073687
Random Forest Classifier	0.806625	0.022289

Fitting the Linear SVC model to the training data is essentially the modelling process's training. Generating text using the n-gram model counts characters or word sequences of intents. The most correlated n-gram (unigrams and bigrams) with each defined intent class is shown in Table 3. Only the two most correlated n-grams are displayed in the table.

TABLE III. THE MOST CORRELATED N-GRAM (UNIGRAMS AND BIGRAMS

Class	Top correlated unigrams	Top correlated bigrams
Code	code	code jsp
	tag	code example
Explanation	mean	function components
	components	jsp syntax
Solution	defend	requestdispatcher servlet
	mitigate	define filters

The performance of the Linear SVC model is displayed through the classification report and confusion matrix. Table 4 shows the classification report. Precision, recall and F-1 score were used to evaluate model performance. All the categories yielded better classification results. A clean dataset contributed to this good result. According to the confusion matrix in Figure 3, the Linear SVC model has done a good job as it correctly predicted all the classes; ‘explanation’ (152), ‘solution’ (27) and ‘code’ (57).

TABLE IV. CLASSIFICATION REPORT

	precision	recall	F1-score	support
Explanation	1.00	1.00	1.00	152
Solution	1.00	1.00	1.00	27
Code	1.00	1.00	1.00	57
Accuracy			1.00	236
Macro avg	1.00	1.00	1.00	236
Weighted avg	1.00	1.00	1.00	236

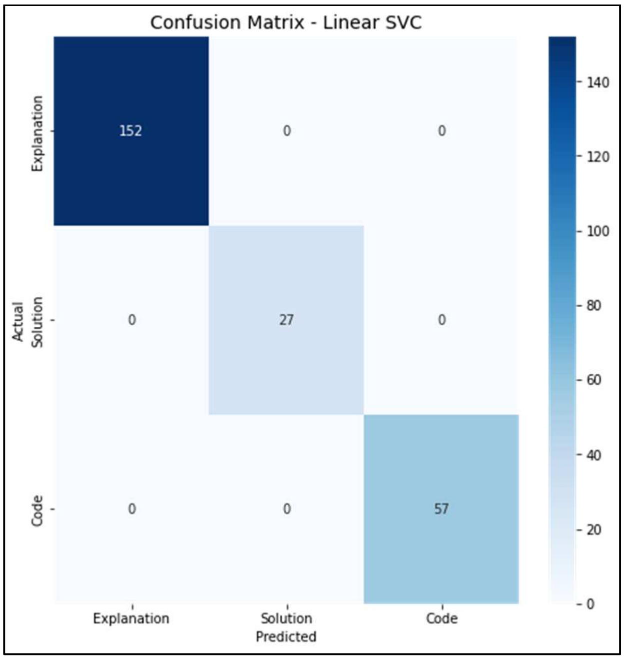

Fig. 3. Confusion Matrix

After Linear SVC was trained, the model can be used to make predictions. Figure 4 displays the prediction of unseen data using the Linear SVC model. The class prediction that has been done is as follows.

- Predictions of class for the question 'Give me information about iteration in JSP?' is 'explanation'.
- Predictions of class for the question 'Give me some sample code of add user in JSP?' is 'code'.
- Predictions of class for the question 'How to get secured against upload a file using Servlet?' is 'solution'.

All the prediction results showed the correct prediction of the class for each 'question' of the intent class. The number of classes affects the predicted results. This is because a small class has a high probability of correct prediction and is easier to predict.

Fig. 4. Predict unseen data

V. CONCLUSION

In conclusion, this research has implemented the systematic method and evaluation metrics of multiclass intent classification based on machine learning algorithm models. This approach can be used to solve class issues that are more than one class of intents simultaneously. The research results revealed Linear SVC as a machine learning algorithm with high accuracy of 0.970292. Therefore, Linear SVC was selected and used in doing multiclass intent classification. The performance of Linear SVC based on the classification report and confusion matrix showed high values for 'precision', 'recall,' and 'F1-score' represented by 1.00. Predictions made using Linear SVC showed that the predicted 'classes' were accurate based on the 'questions' asked. This prediction result is significant for understanding user intents and discovering learning patterns. This research also succeeded in finding the most correlated n-grams (unigrams and bigrams) with each defined intent class.

ACKNOWLEDGEMENT

The Ministry of Higher Education Malaysia (MOHE) funded this research through the Fundamental Research Grant Scheme (FRGS), with the project reference code: FRGS/1/2020/ICT06/UNISZA/02/3. Special thanks to the Centre for Research Excellence and Incubation Management (CREIM) at Universiti Sultan Zainal Abidin (UniSZA) for their assistance in carrying out this research.

REFERENCES

[1] D. Zumstein and S. Hundertmark, “Chatbots: an interactive technology for personalized communication and transaction,” *Int. J. WWW/Internet*, 2018.

[2] A. Ho, J. Hancock, and A. S. Miner, “Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot,” *J. Commun.*, 2018, doi: 10.1093/joc/jqy026.

[3] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, “Chatbot for university related FAQs,” in *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI* 2017, doi: 10.1109/ICACCI.2017.8126057.

[4] Y. Wu, W. Wu, C. Xing, C. Xu, Z. Li, and M. Zhou, “A sequential matching framework for multi-turn response selection in retrieval-based chatbots,” *Comput. Linguist.*, 2019, doi: 10.1162/coli_a_00345.

[5] Y. Wu, W. Wu, C. Xing, Z. Li, and M. Zhou, “Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots,” in ACL 2017 - *55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference* (Long Papers), 2017, doi: 10.18653/v1/P17-1046.

[6] A. Bartl and G. Spanakis, "A retrieval-based dialogue system utilizing utterance and context embeddings," in *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA* 2017, doi: 10.1109/ICMLA.2017.00011.

[7] Y. Wu, Z. Li, W. Wu, and M. Zhou, "Response selection with topic clues for retrieval-based chatbots," *Neurocomputing,* 2018, doi: 10.1016/j.neucom.2018.07.073.

[8] W. M. A. F. Wan Hamzah, I. Ismail, M. K. Yusof, S. I. Mohd Saany and A. Yacob, "Using Learning Analytics to Explore Responses from Student Conversations with Chatbot for Education", *Int. J. Eng. Ped*., vol. 11, no. 6, pp. 70-84, Dec. 2021.

[9] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, 2020.

[10] de Carvalho, André CPLF, and Alex A. Freitas, "A tutorial on multi-label classification techniques." *Foundations of computational intelligence*, volume 5, pp.177-195, 2009.

[11] Wang, Ran, Robert Ridley, Weiguang Qu, and Xinyu Dai, "A novel reasoning mechanism for multi-label text classification." *Information Processing & Management,* 58, no. 2, 102441, 2021.

[12] Sarker, Iqbal H, "Machine learning: Algorithms, real-world applications and research directions." *SN Computer Science*, vol. 2, no. 3, pp. 1-21, 2021.

[13] Nurshahira Endut, W. M. Amir Fazamin W. Hamzah, Ismahafezi Ismail, Mohd Kamir Yusof, Yousef Abu Baker and Hafiz Yusoff. "A Systematic Literature Review on Multi-Label Classification based on Machine Learning Algorithms", *TEM Journal*, 11(2), 658-666, 2022.

[14] Punia, Sanjeev Kumar, Manoj Kumar, Thompson Stephan, Ganesh Gopal Deverajan, and Rizwan Patan, "Performance analysis of machine learning algorithms for big data classification: Ml and ai-based algorithms for big data analysis." *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 12, no. 4, pp. 60-75, 2021.

[15] Ibrahim, Ibrahim, and Adnan Abdulazeez. "The role of machine learning algorithms for diagnosing diseases.", *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 10-19, 2021.

[16] Elmogy, Ahmed M., Usman Tariq, Mohammed Ammar, and Atef Ibrahim. "Fake reviews detection using supervised machine learning." *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[17] Bujang, Siti Dianah Abdul, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, and Nor Azura Md Ghani. "Multiclass prediction model for student grade prediction using machine learning.", *IEEE Access*, vol. 9, pp. 95608-95621, 2021.

[18] Anandarajan, Murugan, Chelsey Hill, and Thomas Nolan. "Text preprocessing." In *Practical Text Analytics*, Springer, Cham, pp. 45-59, 2019.

[19] Keerthi Kumar, H. M., and B. S. Harish. "Classification of short text using various preprocessing techniques: An empirical evaluation." In *Recent findings in intelligent computing techniques*, Springer, Singapore, pp. 19-30, 2018.

[20] H. Nawang, M. Makhtar. and W.M.A.F. Wan Hamzah, "Comparative analysis of classification algorithm evaluations to predict secondary school students' achievement in core and elective subjects", *International Journal of Advanced Technology and Engineering Exploration*, *9*(89), p.430, 2022.

[21] Moreo, A., Esuli, A., and Sebastiani, F, Word-class embeddings for multiclass text classification. *Data Mining and Knowledge Discovery*, *35*(3), 911-963, 2021

[22] Parmar, P. S., Biju, P. K., Shankar, M., and Kadiresan, N, Multiclass text classification and analytics for improving customer support response through different classifiers. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* IEEE, pp. 538-542, 2018.

[23] Maron, R. C., Weichenthal, M., Utikal, J. S., Hekler, A., Berking, C., Hauschild, A., and Thiem, A., Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *European Journal of Cancer*, *119*, 57-65, 2019.

[24] Vang, Y. S., Chen, Z., and Xie, X, Deep learning framework for multi-class breast cancer histology image classification. In *International conference image analysis and recognition*, Springer, Cham, pp. 914-922, 2018.

[25] Verma, V., Muttoo, S. K., and Singh, V. B, Multiclass malware classification via first-and second-order texture statistics. *Computers & Security*, *97*, 101895, 2020.

[26] Ghouti, L., and Imam, M, Malware classification using compact image features and multiclass support vector machines. *IET Information Security*, *14*(4), 419-429, 2020.

[27] David, D. S., Saravanan, D., and Jayachandran, A, Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification system. *Solid State Technology*, *63*(6), 3599-3623, 2020.

[28] Kuo, K. M., Talley, P., Kao, Y., & Huang, C. H, A multi-class classification model for supporting the diagnosis of type II diabetes mellitus. *PeerJ*, *8*, e9920, 2020.

[29] Mustafa, R. U., Ashraf, N., Ahmed, F. S., Ferzund, J., Shahzad, B., and Gelbukh, A, A multiclass depression detection in social media based on sentiment analysis. In *17th International Conference on Information Technology–New Generations (ITNG 2020)*, Springer, Cham, pp. 659-662, 2020.

[30] Bouazizi, M., and Ohtsuki, T., Multi-class sentiment analysis in Twitter: What if classification is not the answer. *IEEE Access*, *6*, 64486-64502, 2018.