JISTE (Journal of Information System, Technology and Engineering)
Volume 1, No. 3, pp. 89-94
E-ISSN: 2987-6117
http://gemapublisher.com/index.php/jiste
Received: July 2023
Accepted: August 2023
Published: September 2023

The Implementation of Multi Label K- Nearest Neighbor Algorithm to Classifying Essay Answers

John Cheno Lerian[1], Georgio Chenayan[2]

Batangas State University[1,2]

Correspondence Email: john.lerian@bsu.edu.ph[1]

Abstract

One way of assessing essays is to use the STAR (Situation, Task, Action, Result) method. This method aims to classify whether the essay already reflects the values of the situation, task, action, and outcome labels presented by the prospective driving teacher. Multi-label classification refers to a classification task in which each data sample can be classified into multiple categories or labels simultaneously. Various methods have been developed to perform multi-label text classification, such as MLKNN, MLTSVM, and MLP. This research aims to determine the value of the accuracy of the score. Therefore, the Multi-Label K-Nearest Neighbor, or MLKNN, algorithm can be implemented to classify essays with five labels: no star, situation, task, action, and result. With a total dataset of 110, which will be obtained from one of the assessors from the driving teacher. The MLKNN algorithm produces an accuracy rate with an average value of 15.6%.

Keywords: multi label classification, essay, multi label k-nearest neighbor.

INTRODUCTION

In various selection processes, one form of evaluation that is often used is an essay. In some cases, essays are used as one of the requirements in selection, such as registration for LPDP scholarships, teacher and lecturer certification, selection of teacher professional education, selection of driving teacher programs, and others (Muktamar et al., 2023). Essays are given as part of the selection process to provide opportunities for potential participants to express themselves, describe their experiences, and demonstrate their understanding and analytical skills on relevant topics (Rizkyanti et al., 2022). Mover Teachers are learning leaders who encourage the holistic, active, and proactive growth and development of students by developing other educators to implement student-centered learning, as well as being role models and agents of the transformation of the education ecosystem to realize the Pancasila Student Profile. One of the selection stages to become a driving teacher candidate is to take an essay test. The number of prospective driving teachers for each generation is increasing. According to one of the essay assessors, the number of assessors for class 10 is 1,426. With the number of prospective driving teachers in batch 10 of 55,000, it can be estimated that 1 assessor or assessors must evaluate the essays of 36 prospective driving teachers, where in 1 essay document there are 18 essays that must be assessed (Jefriyanto et al., 2023).

Selection is the process of choosing or determining a choice from a number of available options. This can include selecting individuals, ideas, alternatives, or other entities based on certain criteria or standards (Hasanah et al., 2021). The aim of selection is to select the best option that best fits the stated needs, objectives, or criteria. Selection is the activity of selecting and determining applicants who are accepted or rejected to become employees of the company. An essay is a piece of writing that contains an argument from the author (Ainun & Jefriyanto, 2023). An essay is divided into two types, namely formal and informal. A formal essay has a serious, objective, and professional character. Meanwhile, informal essays have a "personal element" and are more relaxed and freer. Essays originally meant prose essays in interesting language and ways. This essay usually discusses a problem in passing from the author's personal point of view. The keywords in the essay writing form are the factors of analysis, interpretation, and reflection. According to Hans Bague Jassin, an essay is an essay that talks about human and life issues. Some of the things discussed in an essay can be in the form of life lessons, responses, a collection of thoughts, or a philosophy of life that is compiled subjectively based on the thoughts and feelings of the author himself (Cupian et al., 2020).

In general, essays are subjective, which means they reflect the author's point of view, thoughts, and personal experiences. Essays are commonly used in various fields, for example, education, academics, business, and others (Magalhaes et al., 2023). Essay assessment is a process of evaluating the author's understanding, knowledge, analytical skills, critical thinking, writing skills, and communication skills in conveying messages clearly and effectively. Assessment of an essay is usually done manually; even though the method of assessment is the same, if the person assessing it is different, the results will be different. One method of evaluating essays in the selection process uses the STAR (Situation, Action, Task, Result) method. This method is usually used to assess a person's skills and experience in dealing with certain situations. This approach asks the candidate to describe the situation or task at hand, the actions taken, and the results achieved in response to the situation (Ramdani et al., 2023).

In implementing multi-label text classification, there are several commonly used algorithms, such as Multi-Layer Perceptron (MLP), Multi Label Twin Support Vector Machine (MLTSVM), Naive Bayes, Random Forest, Multi Label K-Nearest Neighbor (MLKNN), and others. Multi-label text classification is done to predict whether there are points that cover the STAR method in each essay selection answer (Guterres et al., 2019). In a previous study, the hamming loss value was quite high, which was 0.0886, or around 91.14% of the data that was classified correctly. In another study, in a multi-label classification using the k-nearest neighbor (ML-KNN) multilabel method in cervical cancer, the results obtained were a hamming loss value of 3.59%, an accuracy of 93%, a precision weighted of 93%, a recall weighted of 96%, and a weighted f1-score of 94%. From the two studies above, it can be concluded that the multi-label classification of text objects obtains pretty good accuracy results (Fitriani & Yustanti, 2022). The objects used in the multi-label classification are quite varied, such as the classification of hadith texts, the classification of cervical cancer, and others.

METHOD

The author uses the literature study method by exploring related websites as well as collecting data and information contained in related journals and books needed for this research. The author conducted research using the CRISP-DM method. There are six phases in the CRISP-DM method: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. However, the authors conducted this research only up to the fifth phase. In the first stage, the authors carried out business understanding, which aims to determine research objectives and needs in detail. Then turn these goals and needs into a problem statement. Finally, prepare a strategy or formulation to achieve the goals that have been set. The second stage, in which the author performs data understanding, is the stage when the process of collecting data, evaluating data, and analyzing data determines whether it is in accordance with the needs of the author or not. The third stage is data preparation. The author aims to prepare data, determine which variables will be tested, and tidy up existing data to improve accuracy during modeling. The next stage is modeling, which involves implementing the algorithm on the data that has been prepared. If it is felt that the results of the implementation are still not appropriate, then you can return to the data preparation stage to get the data as you wish. The author chooses a multi-label algorithm to predict values according to research needs. The final stage is evaluation, which aims to conduct experiments related to the algorithm to be used. This is done by changing the value that becomes the parameter to get the desired result. The author conducted several scenarios to evaluate the results of modeling by changing the related value variables.

In this study, the authors used hardware in the form of an Acer Aspire v3-471 laptop with Intel Core i3 processor specifications, 8 GB of RAM, and a 120GB SSD. In this study, the laptop was used as a tool for implementing experiments. The software used in this research is Google Collaboratory, which uses Python. Google Collab is used to run the code and implement the K-Nearest Neighbor Multi-Label algorithm; the code repository is located on Google Drive. The author begins the research by looking for supporting data in relevant literature, such as data on the software used and knowledge references. Related documents also provide information for the author to understand the research topic, as well as citing these documents as the theoretical basis of this research. After the data is collected, the next step is the preprocessing step. The data obtained in the previous process can be described as raw data; that is, the data must be cleaned first. The process carried out at this stage is case folding (changing all data to lowercase, removing punctuation marks, and filtering non-alphabetical characters). Tokenizing (separating data into words or tokens), filtering (filtering words that are considered meaningless), and stemming (changing words back from their base). Then, after the data has been successfully cleaned, it can be processed at the term weighting stage using TF IDF (Term Frequency-Inverse Document Frequency). At this stage, the weight of a document is calculated based on the words contained in the document. Then the writer tests the data that has been collected and performs the training and testing process using the Multi-Label K-Nearest Neighbor algorithm. The tests are carried out with several scenarios; the writer chooses an odd number as one of the test scenarios because this research focuses on the classification process. If you use an even number of K values, then there is a possibility of a draw between two different classes.

RESULT AND DISCUSSION

Business understanding analysis: this stage is carried out to obtain an understanding equation based on a business perspective and converted into a data mining problem. Then determine the right solution to overcome these problems. The problem of this research is that there are a large number of driving teacher candidates who need to be assessed regarding essays that have been made using the STAR assessment method. As an example, in one candidate document there are around 18 questions, which means there are 18 paragraphs of essay answers that must be assessed and looked for to determine whether each of these paragraphs contains one, all, or none of the points from the STAR (Situation, Task, Action, Result) assessment method. The solution offered is the application of an algorithm using MLKNN (Multi Label K-Nearest Neighbor) to find the best accuracy, followed by an experiment by changing the neighbor value, or K value. The next stage is data understanding. The author collects essay data that has been given by the driving teacher assessor at the Ministry of Education and Culture to be used as training data in this study. There are five types of entities that will be developed to obtain accurate results: no STAR, situation, task, action, and result. The author collects 110 essays from 6 candidates; with 1 candidate, there are 18 related questions and 18 essay answers, which will be examined to determine whether there are points from the STAR assessment method. The following is an example of data to be tested in the system.

The third stage is the data preparation or data processing stage. The author performs existing data processing in three stages: text preprocessing, term weighting, and split test and training data. These stages function to clean the text by weighting and dividing the data to be tested so that the results have a higher level of accuracy. The text preprocessing flow includes several stages; the first is to remove special characters such as ",", "=", etc. Then the dataset is converted into lowercase or case-folding. Then proceed with tokenization, which separates the text into small units such as words or phrases. Then, in the filtering process, the data will be filtered and deleted if it is included in the stop word category. The next process is stemming, which removes the nature of the word affixes and changes the words in the dataset to become basic words. For example, in one of the existing essay datasets, namely "It happened 3 months ago when accepting new students, the situation I faced was that new students who entered school did not have a track record of learning styles, interests, talents, and literacy and numeracy abilities that they possessed, even though in planning learning that focuses on students this must have been known earlier", the text preprocessing process will be carried out. Case folding Converts all letters in a document that are capitalized to lowercase. Tokenizing is the stage of cutting words in sentences into small parts by using spaces as delimiters, which will make tokens in the form of words and remove punctuation marks and also characters that are not needed. The filtering stage is the word filtering stage after tokenizing. Words that are considered insignificant or have no meaning are called stop words. Stemming is the process of returning a word to its basic form by removing initial and final affixes. The author uses the literary module in Python to do stemming in Philippines words.

Term weight: at this stage, word weighting is done with the aim of giving word weight to all data so that each word has a value that describes its effect on the document. The procedure at this point is to load the previously held dataset and then perform a weight calculation using the IDF TF. In this process, each word is given a weight in the document and the entire document. Word weighting is done after the completion of the text preprocessing process; the weighting will be used as a calculation in the MLKNN algorithm. The word weighting process using the IDF TF is divided into two parts, namely TF and IDF. Term Frequency calculates the frequency of occurrence of words in a document by assigning a weight to each word based on the number of occurrences. The purpose of calculating TF is to emphasize words that appear more frequently in a document because they can provide more relevant or important information. IDF calculates the reciprocal of the frequency of occurrence of a word or term in the entire document collection. Words that appear rarely in the entire document collection will have a high IDF value, while words that appear frequently will have a low IDF value. Next, multiply the term frequency value with the inverted document frequency value.

Train-test-split is a process that begins by loading the dataset from Google Drive. The next step is to divide the dataset into training data (the training set) and test data (the test set). The dataset is divided randomly into a certain proportion: 80% for training and 20% for testing. The data set can be used as a benchmark for the accuracy of the model. The fourth stage is modeling, namely the implementation of the algorithm to be tested. The writer chooses the MLKNN algorithm to get multi-label predictions from the candidate's essay answers. The basic stages of MLKNN are the calculation of prior probability, euclidean distance, membership vector, posterior probability, and maximum a posterior (MAP). The final stage, namely evaluation, is a stage to assess the level of accuracy of the algorithm in several scenarios. The author chooses a scenario to change the neighbor value or the value of the 'K' variable and determines which 'K' value gets the highest accuracy value. The author determines that the 'K' values to be tested are 3, 5, 7, 11, 13, and 15. After obtaining the accuracy value of each existing scenario, the author will make comparisons to find the highest accuracy value.

After experimenting with 7 scenarios of K values, the predicted results of multi-label classification were obtained in the selection essay text. The evaluation that the writer did was calculating the validation score, accuracy score, precision score, recall, f1 score, and hamming loss. Multi-label classification using the K-Nearest Neighbor Multi Label algorithm obtains the highest accuracy, which is 0.2727 in scenario K = 5, and the lowest is 0.0455 in scenario K = 15. Meanwhile, the lowest hamming loss value is found in scenario 4 with a value of 0, and the highest in scenarios 1, 3, and 6 with a value of 0.2636.

Table 1. Comparison of Evaluation Values

	Validation Score	Acc. Score	Precision Score	Recall	F1 Score	Hamming Loss
K = 3	0.1745	0.2273	0.4762	0.3571	0.4028	0.2636
K = 5	0.1275	0.2727	0.55	0.3929	0.4583	0.2364
K = 7	0.1392	0.1364	0.4783	0.3929	0.4314	0.2636
K = 9	0.1136	0.1818	0.6875	0.3929	0.5	0.2
K = 11	0.0828	0.0909	0.6923	0.3214	0.439	0.2091
K = 13	0.1392	0.1364	0.4783	0.3929	0.4314	0.2636
K = 15	0.0667	0.0455	0.0556	0.1786	0.2703	0.2445

CONCLUSION

The K-Nearest Neighbor Multi-Label Algorithm model in this study uses 110 essays. The data used to train the model is 86, and 22 are used for testing the model. The accuracy value has increased and decreased significantly; the higher the K value, the lower the accuracy value. The level of accuracy in conducting multi-label classification in essays gets an average value of 15.59%, with the highest score in scenario 2 with a value of 27.27% and the lowest score in scenario 7 with an accuracy rate of 4.55%. Based on the results of these studies, the authors draw several conclusions: Accuracy values have increased and decreased significantly. The accuracy value is relatively low, with the highest value being 27.27%. The K value with the highest accuracy is found at K=5. The higher the K value, the lower the accuracy value obtained. This research still has many shortcomings; therefore, the authors hope that this research can be developed further. Based on the research that the authors have done, they can provide the following suggestions: It is necessary to increase the existing datasets, and the labeling performed by the assessors must be greater than 1 so that the accuracy of multi-label classification increases. It is necessary to try a comparison between multi-label classifications in order to get the best algorithm for multi-label classification. The text scale needs to be reduced first to per sentence so that the training data can be better.

REFERENCES

Jefriyanto, J., Ainun, N., & Ardha, M. A. A. (2023). Application of Naïve Bayes Classification to Analyze Performance Using Stopwords. *Journal of Information System, Technology and Engineering*, *1*(2), 49–53.

Cupian, C., Zaky, M., Nurjaman, K., & Kurnia, E. (2020). Analysis of the Implementation of Recruitment, Selection and Placement Based on the Perspective of Islamic Human

Capital. *Komitmen: Jurnal Ilmiah Manajemen*, 1(1), 50–63. https://doi.org/10.15575/jim.v1i1.8289

Ramdani, H. T., Ainun, N., & Muktamar B, A. (2023). Implementation of Progressive Web App on Dropship Data Management Application to Anticipate Product Order Errors. *Journal of Information System, Technology and Engineering*, *1*(2), 38–42.

Fitriani, E. E., & Yustanti, W. (2022). Comparison of the Performance of the Problem Transformation-KNN Method and the Adaptation-KNN Algorithm on the Multi-Label Classification of Question Boxes. *Journal of Emerging Information Systems and Business Intelligence*, 3(3), 122–129.

Magalhaes, A. D. J., Sopwandin, I., & Bakri, A. A. (2023). Online Training Application Design with Website-Based Blended Learning System Method. *Journal of Information System, Technology and Engineering*, *1*(2), 43–48.

Guterres, A., Gunawan, & Santoso, J. (2019). Tetun Language Stemming Using Rule Based Approach. *Teknika*, 8(2), 142–147. https://doi.org/10.34148/teknika.v8i2.224

Ainun, N., & Jefriyanto, J. (2023). Development of Kirchoff's Law Drawing Tools to Improve Student's Science Skills in Learning Process of Direct Flow Circuits. *Journal of Information System, Technology and Engineering*, *1*(2), 32–37.

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementation of the CRISP-DM Model Using the Decision Tree Method with the CART Algorithm for Prediction of Potential Flood Rainfall. *Journal of Applied Informatics and Computing*, 5(2), 103–108. https://doi.org/10.30871/jaic.v5i2.3200

Muktamar B, A., Lumingkewas, C. S., & Rofi'i, A. (2023). The Implementation of User Centered Design Method in Developing UI/UX. *Journal of Information System, Technology and Engineering*, *1*(2), 26–31.

Rizkyani, E., Ernawati, I., & Chamidah, N. (2022). Multi-Label Classification Using the K-Nearest Neighbor (Ml-Knn) Multi-Label Method in Cervical Cancer. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika),* 7(4), 1281–1293. https://doi.org/10.29100/jipi.v7i4.3260