Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 24, No. 3, July 2025, pp. 407∼422
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v24i3.4729

Performance Evaluation of Artificial Intelligence Models for Classification in Concept Map Quality Assessment

Wahyu Styo Pratama[1], Didik Dwi Prasetya[1], Triyanna Widyaningtyas[1], Muhammad Zaki Wiryawan[1], Lalu Ganda Rady Putra[2], Tsukasa Hirashima[3]
[1]Universitas Negeri Malang, Malang, Indonesia
[2]Universitas Bumigora, Mataram, Indonesia
[3]Hiroshima University, Hiroshima, Japan

Article Info

Article history:

Received December 18, 2024
Revised January 11, 2025
Accepted March 05, 2025

Keywords:

Classification;
Concept Maps;
Machine Learning;
Quality Assessment.

ABSTRACT

Open-ended concept maps generated by students give better flexibility and present a complex analysis process for teachers. We investigate the application of classification algorithms in assessing open-ended concept maps, with the **purpose** of providing assistance for teachers in evaluating student comprehension. **The method** used in this study is experimental methods, which consists of data collection, preprocessing, representation generation, and modelling with Feedforward Neural Network, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Logistic Regression. Our dataset, derived from concept maps, consists of 3,759 words forming 690 propositions, scored carefully by experts to ensure high accuracy in the evaluation process. **Results of this study** indicate that K-NN outperformed all other models, achieving the highest accuracy and Receiver Operating Characteristic-Area Under the Curve scores, demonstrating its robustness in distinguishing between classes. Support Vector Machine excelled in precision, effectively minimizing false positives, while Random Forest showcased a balanced performance through its ensemble learning approach. Decision Tree and Linear Regression showed limitations in handling complex data patterns. Feedforward Neural Network can model intricate relationships, but needs further optimization. This research **concluded** that Artificial Intelligence classification enables a better assessment for teachers, enables the path for personalized learning strategies in learning.

Corresponding Author:

Didik Dwi Prasetya, +62 812-9700-0116,
Department of Electrical Engineering and Informatics, Faculty of Engineering,
State University of Malang, Malang, Indonesia,
Email: didikdwi@um.ac.id

How to Cite:
W. S. Pratama, D. D. Prasetya, T. Widyaningtyas, M. Z. Wiryawan, L. G. R. Putra, and T. Hirashima, ”Performance Evaluation of Artificial Intelligence Models forClassification in Concept Map Quality Assessment”, *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 24, no. 3, pp. 407–422, doi: 10.30812/matrik.v24i3.4729.

1. INTRODUCTION

Artificial Intelligence (AI) has arisen as a revolutionary element within computational science, empowering systems to scrutinize and interpret complex datasets [1]. The classification assessment constitutes a facet of AI, representing a methodology that entails data point allocations to established categories predicated on their intrinsic characteristics [2, 3]. This technique has demonstrated efficacy in numerous text-processing applications, where it aids in textual context categorization into several discrete subjects [4]. The classification procedure encompasses the training of algorithms on annotated datasets, enabling them to assimilate the fundamental patterns that differentiate various classes and promote precise predictions, augmenting decision-making processes across diverse domains [5, 6]. In the education sector, classification algorithms have positive implications. AI-driven classification for education enables a sophisticated student performance analysis to search for learning patterns and categorize students based on their strengths and weaknesses [7]. Insights given by classification allow teachers to tailor instructional strategies and interventions to meet each student's needs, fostering a personalized learning experience [8]. By classifying student assignments based on correctness, teachers can implement targeted support measures that enhance academic outcomes [9]. These algorithms can assist in automating administrative tasks, such as grading and course recommendations, allowing teachers to focus more on teaching and mentoring [10, 11]. As AI evolves, its role in transforming educational practices through effective classification can expand exponentially, paving the way for innovative approaches to the teaching and learning process [12, 13]. Concept maps are an innovative approach to the learning process that enables students to be creative.

Concept maps serve as powerful pedagogical tools in educational practices, providing visual representations that help students organize and integrate complex ideas [14, 15]. It illustrated the relationships between concepts, enabling students to engage in deeper cognitive processes, fostering critical thinking, and enhancing their understanding of the subject matter. Concept mapping aids students' knowledge retention and facilitates collaborative learning environments where students can collectively construct meaning. Through its feature, concept maps facilitate meaningful learning by establishing connections between newly acquired information and existing cognitive structures [16]. This study uses an educational concept map with one specific approach, called the open-ended format [16]. This approach offers greater flexibility by allowing students to explore various connections between concepts without boundaries. Adaptability to an open-ended approach fosters creativity and critical thinking better than a closed-ended approach, as students can easily add, modify, or remove concepts as their understanding evolves [17]. This enables deeper reflection during connection creation within a concept map [18]. Nevertheless, several drawbacks and issues need more consideration regarding this approach. The open-ended format presents a challenge of complexity and unpredictability, shown in research conducted by [18]. While flexibility encourages better creativity and deeper understanding, it also complicates the analysis process for teachers. Teachers must conduct an evaluation of the broad hypotheses and patterns proposed by students manually to identify underlying themes, as stated in [19]. A student might establish a novel connection between seemingly unrelated topics, which teachers could overlook or misinterpret. Therefore, implementing a tool that can automatically classify students' understanding is needed [20]. One effective approach for achieving this is the application of classification methods in NLP.

Recent research in classification algorithms has explored concept maps or AI approaches to enhance accuracy and decision-making processes across different fields. The study [21] used concept maps created by 230 physics students, assessed using a 4-level rubric by two qualified raters. Using a pre-trained Support Vector Machine (SVM) results in a good performance of 80% accuracy. The research [17] exposed the knowledge relationships in open-ended concept maps using Extended Kit-Builder (EKB), which has a pattern on the connection between previous and new knowledge that facilitates meaningful learning; meanwhile, this study does not currently facilitate automated assessment. The study conducted by [20] to highlight AI in higher education, which gives insight into research trends in AI to foster decision-making. The research [15] only highlights the construction of concept maps for education, opening up the possibility of future experiments on applying artificial intelligence to concept mapping. Meanwhile, this research [22–27] carry out experimental measurements, and the results are accurate and reliable for different sectors, such as fake job postings, clinical datasets, rain prediction, and urban land use data. These previous methods imply that AI models enable the path for better assistance to many sectors. Nevertheless, minimal research still combines the strengths of AI models and open-ended concept maps to assist teachers in assessing student comprehension.

This study presents a **novelty** to classify open-ended concept maps using the advantages of NLP through the reliability comparison of several classification methods, such as Feedforward Neural Network (FNN), Random Forest (RF), Decision Tree (DT), SVM, Logistic Regression (LR), and K-Nearest Neighbor (KNN). Before previous studies were conducted, this study is focused on the performance of open-ended concept maps with the power of artificial intelligence. The classification process of our study uses text representation from Term Frequency-Inverse Document Frequency, which helps to improve concept map assessment [28]. This feature extraction acts to quantify the value of words within the framework of concept maps. We explore the reliability of every classification method using several metrics. Involved metrics derived from the confusion matrix, such as precision, recall, accuracy, and F1-score [29]. Every metric serves distinct **purpose**s in performance evaluation by measuring different aspects [30]. We also measure how well each classification can distinguish between each class using the Receiver Operating Characteristics – Area Under the Curve (ROC-AUC) [22, 31]. Metrics applied in this study align with our purpose to provide a better and improved assistance for teachers during assessments to evaluate student comprehension of a topic in a concept map [20].

2. RESEARCH METHOD

Our study employs an experimental method to evaluate the efficacy of classifier algorithms in concept mapping. The research method consists of five key steps: data collection, data preprocessing, representation generation, modelling, and performance evaluation of each model. Each step is meticulously designed to ensure that the model accurately categorizes the suggestions provided by students. This accuracy enables teachers to identify gaps in understanding and refine their teaching strategies accordingly. A detailed overview of each completed step is found in Figure 1.

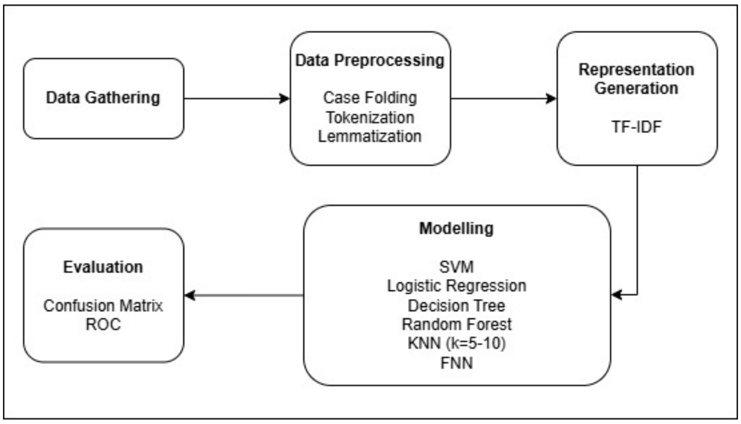


Figure 1. Research step

2.1. Data Gathering

Data gathered in our study were derived from concept maps created by Bachelor of Informatics students in a public university during a relational database course and saved in CSV format. This dataset contains 690 propositions with a total of 3.759 words created by 30 students [32]. Propositions in concept maps show each student's comprehension, which is represented by nodes and links between concepts. Every proposition is marked with scores between 0, as the lowest, and 3, as the most accurate possible. The quality assessment conducted by three experts serves as the ground truth for training and validating automated assessment models in subsequent stages of this research.

2.2. Data Preprocessing

Data preprocessing for the second step in our research has the goal of transforming a raw material from a dataset into a suitable format for processing and analysis in modelling [33, 34]. Since datasets may contain errors and inconsistencies that could affect the models' performance, this phase is improving the quality and dependability of the data. Preprocessing reduces problems that distort analytical results, such as redundant data and outliers [35]. By simplifying the dataset for analysis, efficient preprocessing improves model correctness. We use methods like lemmatization, tokenization, and case folding to prepare datasets for our models. Below is a description of each technique used in preprocessing.

a. Case folding
To ensure uniformity and remove inconsistencies brought on by variances in letter casing, case folding entails changing every character in a text string to a consistent case [36]. Textual data in the dataset often contains a mix of uppercase and lowercase letters, which can lead to inaccuracies during string comparisons and analyses. For example, the words "Benefit" and "benefit" would be treated as distinct entities without case folding, leading to skewed results in text classification [37]. Implementing case folding enhances reliability by allowing different representations of the same word to be treated as equivalents. Case folding facilitates more effective matching and searching within datasets.

b. Tokenization
Tokenization is a part of NLP that breaks down a sentence into smaller, manageable units known as tokens [38]. These tokens represent various elements, including words, phrases, sentences, or even characters, depending on the particular needs of the study [39]. The objective of tokenization is to convert unstructured text into a structured format that machines can easily analyze and interpret. Tokenization facilitates the identification of patterns and relationships within the data by employing a tokenizer. For example, a statement such as "Candidate key is an attribute that usually has a unique value" could be divided into the tokens ["Candidate," "key," "is," "an," "attribute," "that," "usually," "has," "a," "unique," and "value"].

c. Lemmatization
Lemmatization is a technique used to refine tokenized data by converting words into their base or dictionary form, hence improving lexical similarity metrics and ensuring consistency within a dataset [40]. Utilizing libraries such as NLTK, lemmatization processes each word's context and produces normalized forms of nouns, verbs, and adjectives. This process results in cleaner data, facilitating more accurate computations in subsequent analyses.
While both lemmatization and stemming aim to reduce words to a common base form, lemmatization offers significant advantages by preserving the semantic integrity of words [41]. Unlike stemming, which can arbitrarily truncate words and lead to a loss of meaning, lemmatization ensures that the output is a valid dictionary form. This accuracy enhances language interpretation, particularly in applications like NLP [42]. This method is preferred in scenarios requiring better language understanding, as it improves the clarity of responses in tasks such as concept mapping.
The goal of lemmatization is to reduce words to their dictionary form. Lemmatization is a complex procedure that takes the morphological structure and meaning of words, in contrast to stemming, which merely truncates words to their root forms without taking their context [41]. This method allows several inflected variants of a word to be grouped and examined as a single entity. Focusing on the context and grammatical role of words, lemmatization enhances the accuracy and clarity of text analysis [40]. The lemmatization process requires part-of-speech (POS) tagging, which assigns grammatical categories to words, such as nouns, verbs, or adjectives. This contextual information can determine the appropriate lemma for each word. Lemmatization reduces ambiguity and improves the performance of various NLP applications, including text classification.

2.3. Representation Generation

As the third step in this study, representation generation is proposed to interpret and classify textual data accurately. Models can perform better with generated representations and provide a more thorough knowledge of text data. This study employs TF-IDF, a statistical technique for assessing a word's significance in a document within a corpus [43]. TF component counts how often a term appears in a document. Meanwhile, IDF is the evaluation process of how rare a term is throughout the corpus [44]. This approach identifies frequently occurring terms that are unique to a given document. The process involved is through text conversion into a sparse matrix, where a document is represented by a row and a term's weight is represented by a column to improve the model's prediction ability at the final evaluation stage [44]. TF-IDF is calculated through Equations 1, 2 and 3 [45].

$$tf(w) = \frac{n(w)}{\sum_j^{n(w)}} \tag{1}$$

$$idf(w) = log\left(\frac{\sum_D}{D_w}\right) \tag{2}$$

$$tfidf(w) = tf(w) * idf(w) \tag{3}$$

Where, w is a word, $n(w)$ is a number of word occurrences. $\Sigma_j n(w)$ is the total words that appeared in the document, ΣD is the number of documents. D_w is the number of documents with the appearance of the word w.

2.4. Modelling

Classification modelling is the fourth part of this research that involves creating mathematical representations of phenomena [46]. Various modelling techniques exist suited to different data types and specific analytical goals. Effective modelling can help to understand underlying trends and correlations better. Classification models can provide valuable insights that inform strategies and drive innovation [47]. Six different models are used in this work, including FNN, SVM, DT, KNN, RF, and LR, as explained below.

a. SVM
SVM is a supervised approach used to find the best hyperplane in an N-dimensional feature space for classifying data points [48]. Known as support vectors, this hyperplane is designed to optimize the margin, the gap between the hyperplane and the closest data points from either class. Both linear and nonlinear classification problems can be handled using SVM. When dealing with linearly separable data, SVM determines the optimum way to divide the classes: a hyperplane in more dimensions or a straight line in two dimensions. SVM uses kernel functions to convert the original feature space into a higher-dimensional space while working with non-linear data. The resilience of this model against overfitting has been demonstrated, particularly in high-dimensional spaces [49]. SVM is constructed through Equation 4 [24].

$$f(x) = sign(\sum_{i=1}^{n} a_i y_i (x^T y_i) + b) \tag{4}$$

Where, a_i is Lagrange coefficient, y_i is the class label of the training sample. x is the input feature vector to be predicted and b is bias parameter.

b. DT
The method is designed as a tree-like model with nodes, branches, and leaves to calculate a function's $f(x)$ result is called a DT [23]. Internal nodes in DT indicate decision points based on particular features, and the root node represents the complete dataset [25]. Every branch represents the result that leads to a leaf or more nodes. DT implements a recursive partitioning technique, which then constructs the tree by dividing the dataset into subsets according to the best attribute at each node [26]. To identify the most informative decision-making features, they use various criteria for dividing nodes, such as information gain or Gini impurity.

c. RF
RF handles classification tasks using an ensemble learning technique. It builds several decision trees during training by employing a technique known as bootstrap aggregating or bagging, which uses a random subset of the input [50]. By minimizing overfitting, RF improves model robustness and accuracy since various trees can identify distinct patterns in the data [26]. Combining the outputs of these trees using averaging for regression or majority voting for classification can lead to the final forecast. This model extracts data from the "wisdom of crowds", in which the combined judgment of several unrelated models produces better forecasts than any one decision tree, demonstrating its efficacy [51]. RF is constructed through these steps [52]. The first step is to choose trees with a k-value less than m for every attribute. Selected N random samples from the dataset for each tree, then randomly selected a subset of predictors up to m ¡ p predictor variables. It is necessary to repeat the k-tree procedure's second and third steps. Each tree's prediction is based on the majority class of the classification findings. The overall prediction is based on the categorization results of most of the trees.

d. KNN
Using the classes of its closest neighbors, the KNN algorithm forecasts the class or value of a new data point [53]. This approach finds a predefined number of neighbors, denoted by k. KNN calculates the distance between each new and existing data point in the training dataset using metrics such as Euclidean distance [27]. The new point is allocated to the majority class of the k nearest neighbors when they have been determined. Plotting the right value affects how well it performs [54]. Plotting error rates against different values of k is a common empirical technique used to identify the ideal k. Since they successfully balance bias and variance and produce reliable predictions without being overly sensitive to noise, values between 5 and 10 are frequently used.

e. LR
LR is part of statistical techniques that can be applied to binary classification issues, in which the result variable is categorical and might have two alternative values, such as 0 and 1 [55, 56]. The probability that a particular input point belongs to a particular category is estimated in logistic regression. This is done by converting any real integer into a decimal value between 0 and 1 using the logistic function, also known as the sigmoid function. To get the highest likelihood parameter, logistic regression coefficients are estimated using the maximum likelihood estimation (MLE) technique [57].
One major advantage of logistic regression is that it is interpretable. Odds ratios, which shed light on the connection between the predictor factors and the result, can be used to interpret the calculated coefficients [58]. If the odds ratio is more than 1, it means that the probability of the outcome being one increases as the predictor grows, but if it is less than 1, the odds decrease. The interpretability of logistic regression makes it a popular choice of selected predictor variables in several fields, including marketing, the social sciences, and medical.

f. FNN
FNN is part of the Neural Network architecture that is distinguished by the unidirectional flow of data from input nodes to output nodes via hidden layers [59]. This model is easier to train and has a simpler structure since this algorithm has no cycles or loops. The main purpose of FNN is to learn from a collection of input-output pairs to approximate complex functions. Each neuron in the network applies a weighted sum of its inputs, followed by a non-linear activation function to give the model non-linearity and enable it to learn complex patterns [60]. The FNN training procedure utilizes backpropagation, which computes the gradient of the loss function concerning each weight using the chain rule. This enables effective weight updates using gradient descent optimization algorithms. Prediction accuracy is increased by this process, which reduces the discrepancy between expected and actual outputs. FNNs have proven their adaptability and efficiency in managing a broad range of tasks in natural language processing [61]. This makes FNN a part in creating increasingly complex architectures through their capacity to learn hierarchical representations of input.

2.5. Evaluation

To assess the effectiveness of binary classification models, our study employs an evaluation methodology using ROC-AUC [22]. ROC-AUC summarizes the model's ability to distinguish between classes by contrasting the True Positive Rate (TPR) and False Positive Rate (FPR) metrics at various threshold levels [31, 62]. This measures the model's general reliability with values between 0 and 1. Perfect discrimination between classes is indicated if the ROC-AUC value is 1. Equations 5, 6, and 7 display the TPR, FPR, and ROC-AUC formulas.

$$ROC - AUC = \sum_{i=1}^{n-1} \frac{(FPR_{i+1} - FPR_i) \times (TPR_{i+1} - TPR_i)}{2} \tag{5}$$

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + FN} \tag{7}$$

ROC-AUC metrics such as TPR and FPR are based on the confusion matrix. This method is also employed in evaluation to determine how effective the analysis was [29]. The confusion matrix aggregates true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) to give insight into the model's performance [30]. This makes it easier to understand the models' capacity to categorize the dataset by allowing researchers to compute important performance metrics, including accuracy, precision, recall, and F1-score. The following Equations 8, 9, 10, and 11 display the F1-score, recall, accuracy, and precision [48].

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{10}$$

$$Accuracy = \frac{total\ correct\ predictions}{total\ predictions} \tag{11}$$

3. RESULT AND ANALYSIS

Section 3 delves into the findings of each phase of our research, starting with data collection, preprocessing, representation generation with TF-IDF, and modelling with six categorization models. Every action is intended to create the best possible categorization analysis of the concept map. The first step of data collection was obtained, which contained formed propositions and labels indicated by numbers ranging from 0, 1, 2, and 3. Example data from 690 propositions is represented through Table 1.

Table 1. Dataset example

Proposition	Score
bahasa disebut SQL	0
bahasa disebut formal	1
relasi memiliki kardinalitas	3
QUEL adalah SQL	1
istilah-istilah yaitu relasi	3
bahasa contoh QUEL	2

The data obtained from the initial step of our analysis is then carried forward to the preprocessing stage. During this phase, we implement several techniques to enhance our dataset. These techniques include case folding, which standardizes the text by converting all characters to lowercase; tokenization that breaks down the text into individual units or tokens; and lemmatization, which reduces words to their base or root form. The outcome of the preprocessing step is a clean and prepared dataset for analysis [35]. Visual representation of the preprocessing stage can refer to Figure 2.

Figure 2. Preprocessing step

The dataset that has been prepared in the second step is then applied to a text representation method using the TF-IDF. This transforms the textual proposition data into a numerical format reflecting each term's importance within the entire dataset. Following this transformation, the resulting TF-IDF representations are utilized to train and categorize the data into six distinct classification models. Each model employs different algorithms and approaches to classify the text effectively. To evaluate and compare the effectiveness of these classification models, we have compiled their performance characteristics in Table 2. This table provides a comprehensive overview of each model's performance, highlighting key metrics such as accuracy, precision, recall, and F1-score. Analyzing these performance metrics gives valuable insights into the strengths and weaknesses of each classifier, thereby informing future decisions regarding model selection and optimization. This version elaborates on the TF-IDF method, categorizing into classification models, and emphasizes the importance of evaluating model performance.

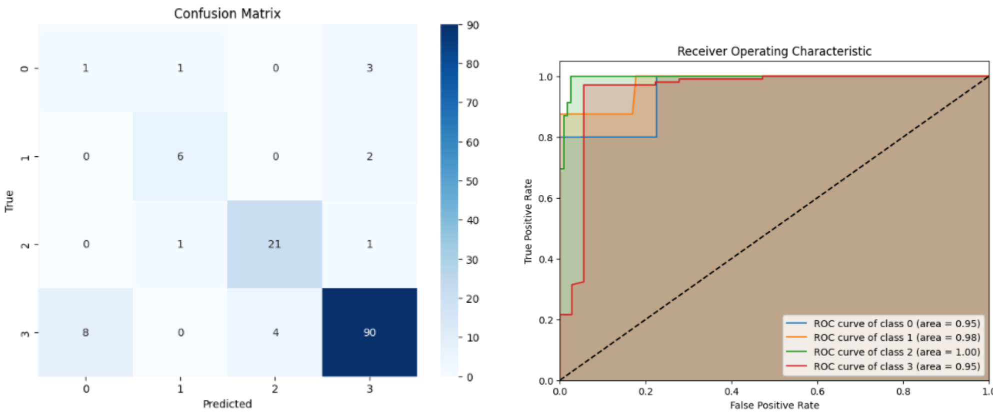
Table 2. Classification Performance Result

Classification	Result					Cross-Validation				
	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Accuracy	Precision	Recall	F1 Score	ROC-AUC
SVM	85.50%	**88.04%**	85.50%	86.64%	93.62%	85,07%	84,6%	85,07%	83,78%	94,05%
Decision Tree	82.60%	84.89%	82.60%	83,36%	85,29%	83,77%	85,45%	83,77%	83,48%	86,52%
Random Forest	85.51%	85.30%	85.51%	85.25%	93,48%	86,67%	87%	86,67%	85,76%	**95,81%**
KNN	**86.95%**	86.62%	**86.95%**	**86.73%**	**97,09%**	85,94%	85,92%	85,94%	84,73%	94,44%
Logistic Regression	84.78%	84,93%	84.78%	84.60%	93,73%	83,77%	82,49%	83,77%	81,82%	95,11%
FNN	84.78%	81.65%	84.78%	82.94%	93.69%	**92,32%**	**92,61%**	**92,32%**	**92,23%**	98,43%

As illustrated in Table 2, most models and classifiers have produced satisfactory predictions for the relevant outcomes. The overall performance of each model is primarily assessed using the accuracy metric, where KNN stands out as the highest accuracy rate at 86.95%. This indicates that KNN is effective in classifying instances within the dataset. Meanwhile, the highest precision metric is achieved by SVM with a score of 88.04%, enhancing its reliability in identifying relevant instances. Furthermore, KNN also leads in recall with an 86.95% score. KNN managed its strong performance across various metrics, reinforcing its position as a robust classifier. These findings underscore the effectiveness of these models in generating accurate forecasts and in handling the complexities inherent in the dataset. This version elaborates on each model's strengths while emphasizing the importance of the metrics in evaluating predictive performance.

Moreover, KNN achieves the highest F1 Score at 86.95%, representing a balance between recall and precision. This metric is valuable in scenarios with significant costs associated with false positives and false negatives. The ROC-AUC score evaluates the model's ability to distinguish between classes. KNN performs exceptionally well with 97.09%, demonstrating a strong capacity to differentiate between positive and negative classes. The outcomes of each evaluation are illustrated in the confusion matrix shown in Figure 3.

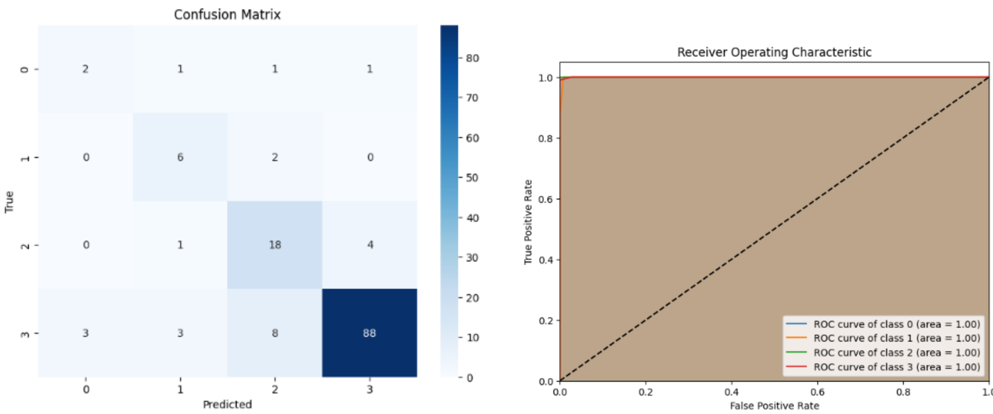
The results presented in Figure 3 (a) illustrate SVM classifier effectiveness in classifying our dataset, revealing distinct patterns within the data. An accuracy rate of 85.50% demonstrated the propositions' predictive capacity in the dataset. This suggests that SVM is reliable and effective in exploring underlying relationships among the data points. SVM demonstrated a high precision rate of 88.04%, highlighting its proficiency in minimizing false positive predictions. The precision value shown ensures that the instances are accurately classified. Figure 3 showcases the ROC-AUC value of 93.59%, underscoring its capability to handle the concept map's proposition dataset effectively. This indicates that the SVM model distinguishes between different classes across various thresholds, reinforcing its robustness and reliability in classification tasks.



(a) (b)

Figure 3. SVM results with confusion matrix (a) and ROC-AUC (b)

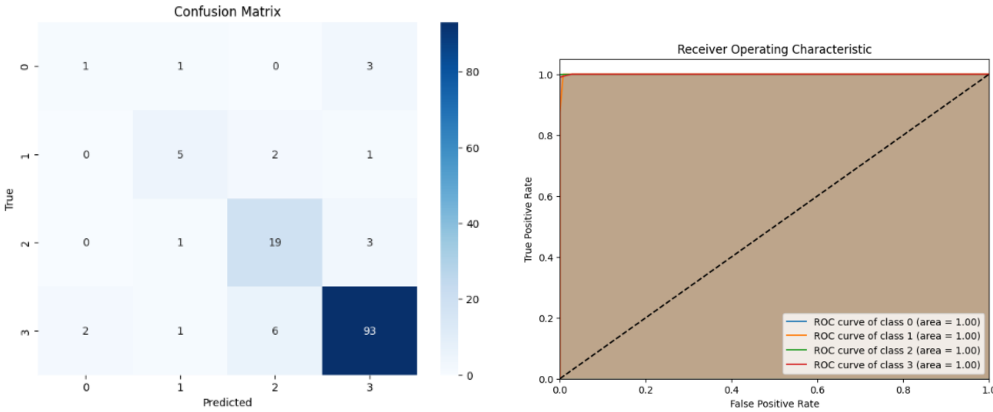
The DT in Figure 4 (a) achieved 82.60% accuracy, indicating its capability to identify the propositions, though slightly lower than that of SVM. This metric shows that while DT can capture many patterns, there may be instances where it struggles with certain classifications. The precision score of 84.89% further shows its ability to predict relevant instances, which is 82.60% accurate. This balance suggests that the model may generate many false negatives even while it is good at detecting actual positives. The F1 Score results of 83.36% highlight this model's overall performance in balancing precision and recall. In contrast to previous models used in this study, the ROC-AUC score of 85.29%, as shown in Figure 4 (b), suggests there is potential for improvement in class distinction.



(a) (b)

Figure 4. DT results with confusion matrix (a) and ROC-AUC (b)

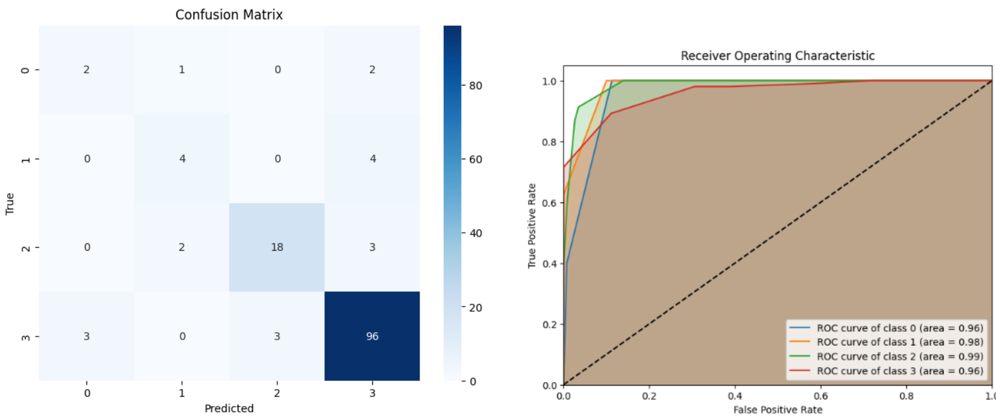
RF classifier performed better than the previous DT, achieving 85.51% accuracy in Figure 5 (a). This increased accuracy results from its ensemble technique, which aggregates the predictions to capture intricate patterns. RF also scored 85.30% in precision, indicating a stronger capability in minimizing false positives. Meanwhile, the recall rate achieved 85.51%, which better identifies true positive cases. The F1 Score is 85.25%, underscoring its balanced performance in maintaining precision and recall. With an ROC-AUC score of 93.48% in Figure 5 (b), RF significantly outperforms DT and showcases its ability to distinguish between classes.



(a) (b)

Figure 5. RF results with confusion matrix (a) and ROC-AUC (b)

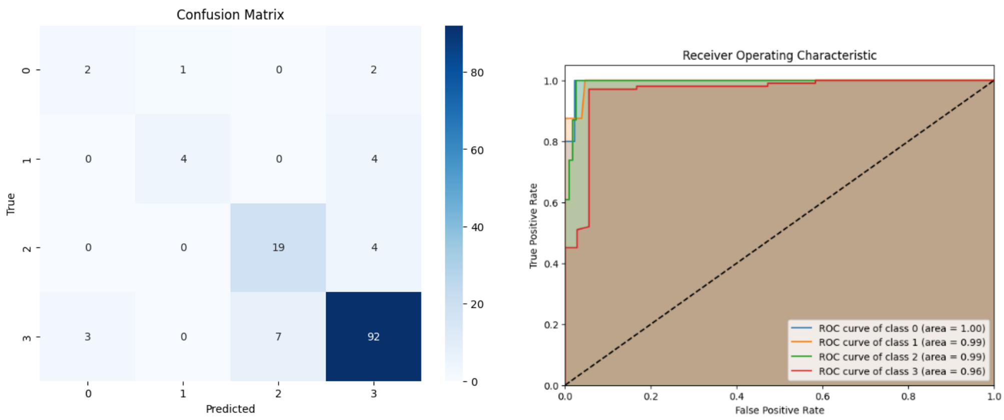
KNN classifier in Figure 6 (a) exhibited the best performance in our analysis amongst other algorithm scenarios, achieving 86.95% accuracy. This demonstrates how well KNN uses the proximity of data points to generate predictions and accurately classifies the propositions in the dataset. Furthermore, KNN also achieves an 86.62% precision score that emphasizes its effectiveness in minimizing false positives, while the recall rate matches its accuracy at 86.95%. A well-balanced trade-off between precision and recall is highlighted by the F1 Score of 86.73%, which makes KNN appropriate for applications where both metrics are essential. KNN achieved a high ROC-AUC score of 97.09%, indicated in Figure 6 (b), proving its ability to distinguish between classes across various thresholds. KNN has proven to outperform DT and RF, highlighting its predictive capabilities. KNNs also result in slightly higher precision than random forests, indicating better performance in reducing false positives. Both KNN and RF demonstrated identical recall rates, with KNN maintaining a strong ability to identify true positives. The F1 Score for KNN also surpasses Random Forest, reflecting its more favorable balance between precision and recall.



(a) (b)

Figure 6. KNN results with confusion matrix (a) and ROC-AUC (b)

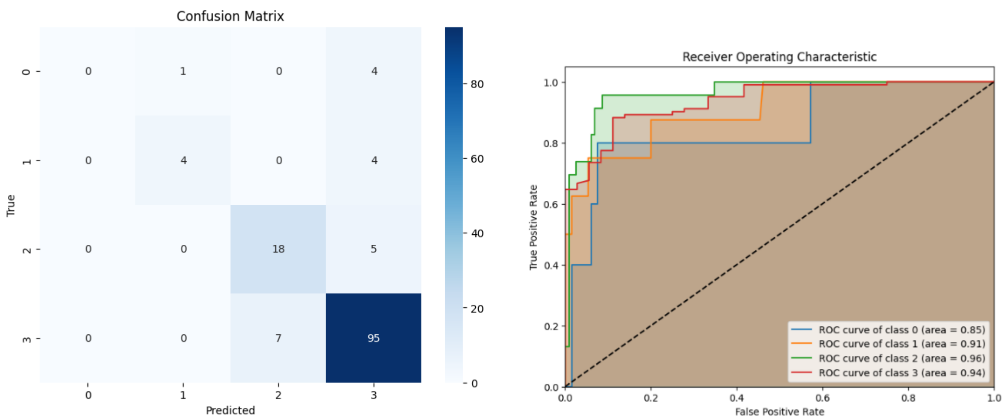
LR classifier in Figure 7 (a) achieves an accuracy of 84.78%. This outcome shows that, while the model performs better than other models, it successfully recognizes a decent percentage of the propositions in the dataset. It also shows a precision score of 84.93%, accurately reflecting their prediction of relevant instances. The recall rate matched its accuracy at 84.78%, suggesting that this model may miss some relevant cases similar to DT. The F1 Score of 84.60% indicates a respectable balance. Compared with other classifiers, LR performance metrics reveal strengths and limitations. While it performs slightly better than the DT, it does not quite reach other model levels. The ROC-AUC score shown in Figure 7 (b) for Logistic Regression is 93.73%, a lower result than KNN's impressive 97.09% and slightly higher than Random Forest's 93.48%. This suggests that while LR effectively distinguishes between classes, it may not be as robust as the ensemble methods in handling complex datasets.



(a) (b)

Figure 7. KNN results with confusion matrix (a) and ROC-AUC (b)

In Figure 8 (a), the FNN classifier achieved 84.78% accuracy, indicating performance in classifying the propositions. This accuracy matches with LR, suggesting both models are on par regarding overall capability. However, FNN hits a lower precision score at 81.65%, reflecting a reduced ability to minimize false positives compared to other classifiers. The recall rate for FNN is consistent with its accuracy at 84.78%, showing the ability to capture true positive instances. The F1 Score of 82.94% highlights a less favorable balance between precision and recall, suggesting that it might not be as effective as other models in maintaining this balance. FNN's performance metrics reveal both strengths and weaknesses. The ROC-AUC score shown by Figure 8 (b) for FNN achieved 93.69%, not explicitly detailed, but is generally expected to be lower than KNN's exceptional 97.09% and Random Forest's 93.48%, indicating its ability to distinguish between classes, although it still trails behind KNN's result. This model provides the foundation for classification tasks due to its ability to model complex relationships within the data. Its lower precision and F1 Score suggest that it requires further tuning to compete with other models in scenarios demanding high accuracy and robustness.



(a) (b)

Figure 8. KNN results with confusion matrix (a) and ROC-AUC (b)

An analysis of the classification performance of 6 distinct algorithms reveals their capabilities in handling the uniqueness of propositions within the dataset. KNN emerged as the most effective model from accuracy metrics up to the ROC score, indicating its proficiency in distinguishing between classes [63]. While SVM also performed well in accuracy and highest precision score, it demonstrated particular strengths in scenarios involving imbalanced data, effectively maximizing margins to produce precise results [64]. RF also provided an appropriate accuracy score and balanced precision-recall score, showcasing an ensemble learning approach that enhances reliability through multiple decision trees [65]. However, despite still producing decent results, LR and Decision Tree have lower accuracy and more limitations than earlier algorithms [66]. Both FNN and LR have similar accuracy to LR, yet their lower precision score suggests challenges in minimizing false positives.

KNN, SVM, and RF stand out for their accuracy and robustness in creating effective classification. The performance of DT and LR highlights their utility in simpler applications but suggests a refinement to enhance predictive capabilities in more complex scenarios [64]. FNN, while promising in modelling intricate relationships, may require optimization to improve its performance metrics relative to the higher-performing models.

4. CONCLUSION

The usefulness of several categorization algorithms in evaluating idea maps made by Bachelor of Informatics students has been shown in this study, underscoring the potential of AI to improve teaching methods through automated evaluation. Using an experimental approach, including data collection, preprocessing, representation generation using TF-IDF, and modelling with six distinct classification methods, we were able to derive valuable insights into student comprehension. The results indicate that KNN emerged as the most effective classifier across most evaluation metrics, proving its robustness in distinguishing between classes. SVM also performed well, especially in the precision metric, showcasing its ability to minimize false positives. RF provided a balanced performance, effectively capturing complex patterns through its ensemble learning approach. Meanwhile, even though DT and LR proved to have a good accuracy score, both classification algorithms revealed limitations in handling complex patterns. FNN can also model complex relationships, but this backpropagation model needs better optimization to enhance its predictive capabilities.

Our research explored the potential of the classification process in AI through the education sector, creating a path for better personalized learning experiences and improved learning strategies. As AI technologies continue to evolve, further research may focus on integrating sophisticated classification methods. More study can be done by refining these models and exploring additional algorithms to develop better accuracy and reliability of concept map assessments. This research can significantly enhance the assessment and understanding of student learning outcomes, fostering an effective educational environment.

5. ACKNOWLEDGEMENTS

The Acknowledgments section is optional. Research sources can be included in this section.

6. DECLARATIONS

AUTHOR CONTIBUTION

Conceptualization: Wahyu Styo Pratama, Didik Dwi Prasetya, Triyanna Widyaningtyas, and Muhammad Zaki Wiryawan.
Methodology: Wahyu Styo Pratama and Muhammad Zaki Wiryawan.
Investigation: Wahyu Styo Pratama.
Discussion of results: Wahyu Styo Pratama. Didik Dwi Prasetya, and Triyanna Widyaningtyas.
Writing – Original Draft: Wahyu Styo Pratama.
Writing – Review and Editing: Didik Dwi Prasetya, Triyanna Widyaningtyas, and Muhammad Zaki Wiryawan.
Resources: Didik Dwi Prasetya, Lalu Ganda Rady Putra, and Tsukasa Hirashima.
Supervision: Didik Dwi Prasetya, Triyanna Widyaningtyas, and Muhammad Zaki Wiryawan.
Approval of the final text: Wahyu Styo Pratama. Didik Dwi Prasetya, Triyanna Widyaningtyas, and Muhammad Zaki Wiryawan.

FUNDING STATEMENT

COMPETING INTEREST

REFERENCES

[1] M. Ghosh and A. Thirugnanam, “Introduction to Artificial Intelligence,” in *Artificial Intelligence for Information Management: A Healthcare Perspective*, K. G. Srinivasa, S. G. M., and S. R. M. Sekhar, Eds. Singapore: Springer Singapore, 2021, vol. 88, pp. 23–44, https://doi.org/10.1007/978-981-16-0415-7_2.

[2] C. Chhabra and S. Choudhary, “Assessing Intelligence Text Classification Techniques,” in *Decision Analytics for Sustainable Development in Smart Society 5.0*, V. Bali, V. Bhatnagar, J. Lu, and K. Banerjee, Eds. Singapore: Springer Nature Singapore, 2022, pp. 55–63, https://doi.org/10.1007/978-981-19-1689-2_4.

[3] D. Kim, H.-G. Kang, K. Bae, and S. Jeon, “An artificial intelligence-enabled industry classification and its interpretation,” *Internet Research*, vol. 32, no. 2, pp. 406–424, Mar. 2022, https://doi.org/10.1108/INTR-05-2020-0299.

[4] I. Dawar, N. Kumar, S. Negi, S. Pathan, and S. Layek, “Text Categorization using Supervised Machine Learning Techniques,” in *2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*. Riyadh, Saudi Arabia: IEEE, Mar. 2023, pp. 185–190, https://doi.org/10.1109/WiDS-PSU57071.2023.00046.

[5] S. N. Bardab, T. M. Ahmed, and T. A. A. Mohammed, “Data mining classification algorithms: An overview,” *International Journal of Advanced and Applied Sciences*, vol. 8, no. 2, pp. 1–5, Feb. 2021, https://doi.org/10.21833/ijaas.2021.02.001.

[6] N. Dutta, U. Subramaniam, and S. Padmanaban, “Mathematical models of classification algorithm of Machine learning,” in *International Meeting on Advanced Technologies in Energy and Electrical Engineering*. Fez, Morocco,: Hamad bin Khalifa University Press (HBKU Press), 2020, https://doi.org/10.5339/qproc.2019.imat3e2018.3.

[7] V. V. Bobrova, O. I. Bantikova, and V. A. Novikova, “Modeling the academic performance of students based on intelligent analysis of educational data,” *Economic Analysis: Theory and Practice*, vol. 22, no. 2, pp. 235–253, Feb. 2023, https://doi.org/10.24891/ea.22.2.235.

[8] A. J. Martin, M. Maria, and F. Sagayaraj, “Learners Classification for Personalized Learning Experience in e-Learning Systems,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 690–697, 2021, https://doi.org/10.14569/IJACSA.2021.0120485.

[9] M. Savic, D. Gunter, E. Curtis, and A. Paz Pirela, “Productive Failures: From Class Requirement to Fostering a Support Group,” *International Journal of Educational Psychology*, vol. 10, no. 3, pp. 271–294, Oct. 2021, https://doi.org/10.17583/ijep.5994.

[10] H. Van Pham, L. H. T. Thuy, N. C. Hung, N. Q. Dich, S. L. Ngoc, and P. Moore, “Mentor and mentee matching: Using fuzzy logic with a maximal length matching algorithm, expressed preferences, and expert knowledge,” *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 3, pp. 4071–4087, Aug. 2023, https://doi.org/10.3233/JIFS-223820.

[11] L. Köbis and C. Mehner, “Ethical Questions Raised by AI-Supported Mentoring in Higher Education,” *Frontiers in Artificial Intelligence*, vol. 4, p. 624050, Apr. 2021, https://doi.org/10.3389/frai.2021.624050.

[12] S. Datta, “Role of Artificial Intelligence in Education,” *International Journal of English Learning & Teaching Skills*, vol. 4, no. 4, pp. 1–9, Jul. 2022, https://doi.org/10.15864/ijelts.4408.

[13] A. Harry and S. Sayudin, “Role of AI in Education,” *Interdiciplinary Journal and Hummanity (INJURITY)*, vol. 2, no. 3, pp. 260–268, Mar. 2023, https://doi.org/10.58631/injurity.v2i3.52.

[14] S. V. Tytenko, “Concept Maps, Their Application Types and Methods in Information and Learning Systems,” *KPI Science News*, vol. 4, no. 8, pp. 70–78, Mar. 2021, https://doi.org/10.20535/kpisn.2020.4.227090.

[15] P. Eachempati, K. Ramnarayan, K. K. Ks, and A. Mayya, “Concept Maps for Teaching, Training, Testing and Thinking,” *MedEdPublish*, vol. 9, p. 171, Aug. 2020, https://doi.org/10.15694/mep.2020.000171.1.

[16] D. D. Prasetya, T. Hirashima, and Y. Hayashi, “Study on Extended Scratch-Build Concept Map to Enhance Students’ Understanding and Promote Quality of Knowledge Structure,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 144–153, 2020, https://doi.org/10.14569/IJACSA.2020.0110420.

[17] D. D. Prasetya, T. Widiyaningtyas, and T. Hirashima, "Interrelatedness patterns of knowledge representation in extension concept mapping," *Research and Practice in Technology Enhanced Learning*, vol. 20, no. 9, pp. 1–18, May 2024, https://doi.org/10.58459/rptel.2025.20009.

[18] K. E. De Ries, H. Schaap, A.-M. M. J. A. P. Van Loon, M. M. H. Kral, and P. C. Meijer, "A literature review of open-ended concept maps as a research instrument to study knowledge and learning," *Quality & Quantity*, vol. 56, no. 1, pp. 73–107, Feb. 2022, https://doi.org/10.1007/s11135-021-01113-x.

[19] D. D. Prasetya and T. Hirashima, "Associated Patterns in Open-Ended Concept Maps within E-Learning," *Knowledge Engineering and Data Science*, vol. 5, no. 2, pp. 179–187, Dec. 2022, https://doi.org/10.17977/um018v5i22022p179-187.

[20] V. Maphosa and M. Maphosa, "Artificial intelligence in higher education: A bibliometric analysis and topic modeling approach," *Applied Artificial Intelligence*, vol. 37, no. 1, pp. 1–23, Dec. 2023, https://doi.org/10.1080/08839514.2023.2261730.

[21] T. Bleckmann and G. Friege, "Concept maps for formative assessment: Creation and implementation of an automatic and intelligent evaluation method," *Knowledge Management & E-Learning: An International Journal*, vol. 15, no. 3, pp. 433–447, Jul. 2023.

[22] A. W. Fazil, M. Hakimi, R. Akbari, M. M. Quchi, and K. Q. Khaliqyar, "Comparative Analysis of Machine Learning Models for Data Classification: An In-Depth Exploration," *Journal of Computer Science and Technology Studies*, vol. 5, no. 4, pp. 160–168, Dec. 2023, https://doi.org/10.32996/jcsts.2023.5.4.16.

[23] M. S. Chowdhury, M. N. Rahman, M. S. Sheikh, M. A. Sayeid, K. H. Mahmud, and B. Hafsa, "GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh," *Heliyon*, vol. 10, no. 1, p. e23424, Jan. 2024, https://doi.org/10.1016/j.heliyon.2023.e23424.

[24] R. Rofik, R. A. Hakim, J. Unjung, B. Prasetiyo, and M. A. Muslim, "Optimization of SVM and Gradient Boosting Models Using GridSearchCV in Detecting Fake Job Postings," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 419–430, Mar. 2024, https://doi.org/10.30812/matrik.v23i2.3566.

[25] R. Mondal, M. D. Do, N. U. Ahmed, D. Walke, D. Micheel, D. Broneske, G. Saake, and R. Heyer, "Decision tree learning in Neo4j on homogeneous and unconnected graph nodes from biological and clinical datasets," *BMC Medical Informatics and Decision Making*, vol. 22, no. S6, pp. 1–12, Mar. 2023, https://doi.org/10.1186/s12911-023-02112-8.

[26] K. J. Grimm and R. Jacobucci, "Reliable Trees: Reliability Informed Recursive Partitioning for Psychological Data," *Multivariate Behavioral Research*, vol. 56, no. 4, pp. 595–607, Jul. 2021, https://doi.org/10.1080/00273171.2020.1751028.

[27] Aqib Fawwaz Mohd Amidon, Zakiah Mohd Yusoff, Nurlaila Ismail, and Mohd Nasir Taib, "KNN Euclidean Distance Model Performance on Aquilaria Malaccensis Oil Qualities," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 48, no. 2, pp. 16–28, Jul. 2024, https://doi.org/10.37934/araset.48.2.1628.

[28] K. Urkalan and G. T. V., "Concept Map Information Content Enhancement Using Joint Word Embedding and Latent Document Structure:," *International Journal on Semantic Web and Information Systems*, vol. 16, no. 4, pp. 45–60, Oct. 2020, https://doi.org/10.4018/IJSWIS.2020100103.

[29] W. Jia, Y. Qin, and C. Zhao, "Rapid detection of adulterated lamb meat using near infrared and electronic nose: A F1-score-MRE data fusion approach," *Food Chemistry*, vol. 439, pp. 1–15, May 2024, https://doi.org/10.1016/j.foodchem.2023.138123.

[30] G. Phillips, H. Teixeira, M. G. Kelly, F. Salas Herrero, G. Várbíró, A. Lyche Solheim, A. Kolada, G. Free, and S. Poikane, "Setting nutrient boundaries to protect aquatic communities: The importance of comparing observed and predicted classifications using measures derived from a confusion matrix," *Science of The Total Environment*, vol. 912, pp. 1–16, Feb. 2024, https://doi.org/10.1016/j.scitotenv.2023.168872.

[31] Y. Liu, Y. Li, and D. Xie, "Implications of imbalanced datasets for empirical ROC-AUC estimation in binary classification tasks," *Journal of Statistical Computation and Simulation*, vol. 94, no. 1, pp. 183–203, Jan. 2024, https://doi.org/10.1080/00949655.2023.2238235.

[32] D. D. Prasetya, A. Pinandito, Y. Hayashi, and T. Hirashima, "Analysis of quality of knowledge structure and students' perceptions in extension concept mapping," *Research and Practice in Technology Enhanced Learning*, vol. 17, no. 1, pp. 1–25, Dec. 2022, https://doi.org/10.1186/s41039-022-00189-9.

[33] S. S. Bisarya, A. Shukla, and S. Kumar, "Data Preparation in Context of Social Sciences Research," *International Journal For Multidisciplinary Research*, vol. 5, no. 3, pp. 1–10, Jun. 2023, https://doi.org/10.36948/ijfmr.2023.v05i03.3937.

[34] M. Hameed and F. Naumann, "Data Preparation: A Survey of Commercial Tools," *ACM SIGMOD Record*, vol. 49, no. 3, pp. 18–29, Dec. 2020, https://doi.org/10.1145/3444831.3444835.

[35] D. Varma, A. Nehansh, and P. Swathy, "Data Preprocessing Toolkit : An Approach to Automate Data Preprocessing," *Interantional Journal of Scientific Research in Engineering and Management*, vol. 07, no. 03, Mar. 2023, https://doi.org/10.55041/IJSREM18270.

[36] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *Journal of Physics: Conference Series*, vol. 1567, no. 3, p. 032024, Jun. 2020, https://doi.org/10.1088/1742-6596/1567/3/032024.

[37] J. Rejito, A. Atthariq, and A. S. Abdullah, "Application of text mining employing k-means algorithms for clustering tweets of Tokopedia," *Journal of Physics: Conference Series*, vol. 1722, no. 1, pp. 1–10, Jan. 2021, https://doi.org/10.1088/1742-6596/1722/1/012019.

[38] A. K. Ohm and K. K. Singh, "Study of Tokenization Strategies for the Santhali Language," *SN Computer Science*, vol. 5, no. 7, pp. 1–8, Aug. 2024, https://doi.org/10.1007/s42979-024-03083-x.

[39] E. Dotan, G. Jaschek, T. Pupko, and Y. Belinkov, "Effect of tokenization on transformers for biological sequences," *Bioinformatics*, vol. 40, no. 4, pp. 1–15, Mar. 2024, https://doi.org/10.1093/bioinformatics/btae196.

[40] Z. Abidin, A. Junaidi, and Wamiliana, "Text Stemming and Lemmatization of Regional Languages in Indonesia: A Systematic Literature Review," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 2, pp. 217–231, Jun. 2024, https://doi.org/10.20473/jisebi.10.2.217-231.

[41] Z. A. Merrouni, B. Frikh, and B. Ouhbi, "EXABSUM: A new text summarization approach for generating extractive and abstractive summaries," *Journal of Big Data*, vol. 10, no. 1, pp. 1–34, Oct. 2023, https://doi.org/10.1186/s40537-023-00836-y.

[42] I. Akhmetov, A. Pak, I. Ualiyeva, and A. Gelbukh, "Highly Language-Independent Word Lemmatization Using a Machine-Learning Classifier," *Computación y Sistemas*, vol. 24, no. 3, Sep. 2020, https://doi.org/10.13053/cys-24-3-3775.

[43] J. Zhou, Z. Ye, S. Zhang, Z. Geng, N. Han, and T. Yang, "Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data," *Heliyon*, vol. 10, no. 16, pp. 1–22, Aug. 2024, https://doi.org/10.1016/j.heliyon.2024.e35945.

[44] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *Journal of Physics: Conference Series*, vol. 2171, no. 1, pp. 1–8, Jan. 2022, https://doi.org/10.1088/1742-6596/2171/1/012021.

[45] U. Rani and K. Bidhan, "Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA," *Journal of scientific research*, vol. 65, no. 01, pp. 304–311, 2021, https://doi.org/10.37398/JSR.2021.650140.

[46] Bogdan Khmelnitsky Melitopol State Pedagogical University (Ukraine), E. Murtaziiev, Y. Syusyukan, and Bogdan Khmelnitsky Melitopol State Pedagogical University (Ukraine), "Mathematical Modeling: Main Stages and Classification of Models," *Modern problems of modeling*, vol. 24, pp. 140–146, Dec. 2022, https://doi.org/10.33842/2313125X-2022-24-140-146.

[47] L. Sundberg and J. Holmström, "Innovating by prompting: How to facilitate innovation in the age of generative AI," *Business Horizons*, vol. 67, no. 5, pp. 561–570, Sep. 2024, https://doi.org/10.1016/j.bushor.2024.04.014.

[48] N. Amaya-Tejera, M. Gamarra, J. I. Vélez, and E. Zurek, "A distance-based kernel for classification via Support Vector Machines," *Frontiers in Artificial Intelligence*, vol. 7, pp. 1–15, Feb. 2024, https://doi.org/10.3389/frai.2024.1287875.

[49] K. Wang, J. Liu, and X. Sun, "Support vector machine in big data: Smoothing strategy and adaptive distributed inference," *Statistics and Computing*, vol. 34, no. 6, pp. 1–19, Dec. 2024, https://doi.org/10.1007/s11222-024-10506-5.

[50] J. Wekalao and N. Mandela, "Terahertz metasurface biosensor for high-sensitivity salinity detection and data encoding with machine learning optimization based on random forest regression," *Optical and Quantum Electronics*, vol. 56, no. 11, pp. 1–39, Nov. 2024, https://doi.org/10.1007/s11082-024-07777-7.

[51] N. Brugnone, N. Benkler, P. Revay, R. Myhre, S. Friedman, S. Schmer-Galunder, S. Gray, and J. Gentile, "Is from ought? A comparison of unsupervised methods for structuring values-based wisdom-of-crowds estimates," *Journal of Computational Social Science*, vol. 7, no. 2, pp. 1327–1377, Oct. 2024, https://doi.org/10.1007/s42001-024-00273-8.

[52] R. Oktafiani, A. Hermawan, and D. Avianto, "Max Depth Impact on Heart Disease Classification: Decision Tree and Random Forest," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 160–168, Feb. 2024, https://doi.org/10.29207/resti.v8i1.5574.

[53] U. G. Inyang, F. F. Ijebu, F. B. Osang, A. A. Afoluronsho, S. S. Udoh, and I. J. Eyoh, "A Dataset-Driven Parameter Tuning Approach for Enhanced K-Nearest Neighbour Algorithm Performance," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, no. 1, pp. 380–391, Jan. 2023, https://doi.org/10.18517/ijaseit.13.1.16706.

[54] S. Novo, G. Aneiros, and P. Vieu, "A k NN procedure in semiparametric functional data analysis," *Statistics & Probability Letters*, vol. 171, pp. 1–15, Apr. 2021, https://doi.org/10.1016/j.spl.2020.109028.

[55] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic Regression in Clinical Studies," *International Journal of Radiation Oncology*Biology*Physics*, vol. 112, no. 2, pp. 271–277, Feb. 2022, https://doi.org/10.1016/j.ijrobp.2021.08.007.

[56] A. Zaidi and A. S. M. Al Luhayb, "Two Statistical Approaches to Justify the Use of the Logistic Function in Binary Logistic Regression," *Mathematical Problems in Engineering*, vol. 2023, no. 1, pp. 1–11, Jan. 2023, https://doi.org/10.1155/2023/5525675.

[57] A. Vella and M. A. Alonso, "Maximum likelihood estimation in the context of an optical measurement," in *Progress in Optics*. Elsevier, 2020, vol. 65, pp. 231–311, https://doi.org/10.1016/bs.po.2019.11.007.

[58] D. Chen, J. Ye, and W. Ye, "Interpretable selective learning in credit risk," *Research in International Business and Finance*, vol. 65, pp. 1–13, Apr. 2023, https://doi.org/10.1016/j.ribaf.2023.101940.

[59] E. H. Houssein, M. E. Hosney, M. M. Emam, D. Oliva, E. M. Younis, A. A. Ali, and W. M. Mohamed, "Optimizing feed-forward neural networks using a modified weighted mean of vectors: Case study chemical datasets," *Swarm and Evolutionary Computation*, vol. 89, pp. 1–29, Aug. 2024, https://doi.org/10.1016/j.swevo.2024.101656.

[60] M. F. Nia, E. Hu, and M. H. Ghayesh, "Direct and inverse simulations of hydrodynamic and thermal characteristics in a room with random boundary conditions by feedforward neural network modelling," *Energy and Built Environment*, p. S2666123324000813, Aug. 2024, https://doi.org/10.1016/j.enbenv.2024.08.005.

[61] Z. Zhang, F. Feng, and T. Huang, "FNNS: An Effective Feedforward Neural Network Scheme with Random Weights for Processing Large-Scale Datasets," *Applied Sciences*, vol. 12, no. 23, pp. 1–14, Dec. 2022, https://doi.org/10.3390/app122312478.

[62] A. Kar, N. Nath, U. Kemprai, and Aman, "Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection," *International Journal of Communications, Network and System Sciences*, vol. 17, no. 02, pp. 11–29, 2024, https://doi.org/10.4236/ijcns.2024.172002.

[63] F. Aldi, I. Nozomi, and S. Soeheri, "Comparison of Drug Type Classification Performance Using KNN Algorithm," *SinkrOn*, vol. 7, no. 3, pp. 1028–1034, Jul. 2022, https://doi.org/10.33395/sinkron.v7i3.11487.

[64] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," *Journal of Data Analysis and Information Processing*, vol. 08, no. 04, pp. 341–357, 2020, https://doi.org/10.4236/jdaip.2020.84020.

[65] H. A. Elzeheiry, S. Barakat, and A. Rezk, "An Efficient Ensemble Model for Various Scale Medical Data," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 1283–1305, 2022, https://doi.org/10.32604/cmc.2022.027345.

[66] P. A. Sunarya, U. Rahardja, S. C. Chen, and Y.-M. Lic, “Deciphering Digital Social Dynamics: A Comparative Study of Logistic Regression and Random Forest in Predicting E-Commerce Customer Behavior,” *Journal of Applied Data Sciences*, vol. 5, no. 1, pp. 100–113, Jan. 2024, https://doi.org/10.47738/jads.v5i1.155.