



Comparative Study of Random Forest and Ordinal Regression in Concept Map Quality Assessment: The Role of TF-IDF, BERT, and SMOTE-based Balancing

Nurul Rismayanti^{a,1}; Didik Dwi Prasetya^{a,2,*}; Triyanna Widiyaningtyas^{a,3}; Tsukasa Hirashima^{b,4}

^a Universitas Negeri Malang, Jl. Semarang Gading Kasri, Kecamatan Klojen, Kota Malang, 65115, Indonesia

^b Hiroshima University, Chome-3-2 Kagamiyama, Higashihiroshima, Hiroshima 739-0046, Japan

¹ nurul.rismayanti.2305348@students.um.ac.id, ² didikdwi@um.ac.id, ³ triyannaw.ft@um.ac.id, ⁴ tsukasa-hirashimahiroshima-u.ac.jp

* Corresponding author

Article history: Received July 29, 2025; Revised July 30, 2025; Accepted September 19, 2025; Available online December 21, 2025

Abstract

Automatic assessment of concept map quality is an important challenge in the field of education, particularly in evaluating students' conceptual understanding objectively and efficiently. This study aims to compare the performance of two machine learning algorithms, namely Random Forest and Ordinal Regression, in classifying the quality of concept maps. The evaluation was conducted on three approaches to text feature representation: Term Frequency-Inverse Document Frequency (TF-IDF), Bidirectional Encoder Representations from Transformers (BERT), and a combination of both (TF-IDF + BERT). Additionally, this study compares the performance of the models under two dataset conditions: original data and data balanced using the Synthetic Minority Over-sampling Technique (SMOTE), to address the class imbalance that often occurs in educational data. The data used consists of a collection of propositions from students' concept maps that have been labeled with ordinal scores based on quality. Text representation is extracted using the TF-IDF and BERT approaches, and then used as input to build the classification model. Performance evaluation was conducted using the metrics of Accuracy, Precision, Recall, F1-score, Cohen's Kappa, and MAE. The results show that the Ordinal Regression model with TF-IDF representation combined with SMOTE achieved the best performance, with an accuracy of 0.8777, an F1-score of 0.8773, and a Cohen's Kappa of 0.7701. These results indicate that classical feature representations like TF-IDF remain effective in limited data scenarios, and that the SMOTE technique successfully improved the model's performance by reducing bias towards the majority class. This research contributes to the development of an automatic concept map assessment system and suggests optimal classification strategies for educational datasets with ordinal and imbalanced characteristics.

Keywords: Concept Map Classification, Random Forest, Ordinal Regression, SMOTE, TF-IDF.

Introduction

Concept maps are important tools for visualizing knowledge and representing semantic relationships between concepts, especially in the context of education and knowledge-based systems [1]. Its ability to externalize learners' understanding or the knowledge structure of a domain makes it widely used in cognitive assessments, curriculum planning, and even intelligent tutoring systems. However, the process of evaluating the quality of concept maps remains a challenge, especially due to the subjective nature of the assessment criteria and the variation in representation forms [2].

Several studies have attempted to automate this evaluation process by utilizing machine learning techniques. Graphical feature-based and textual content-based approaches have been developed to classify the quality of concept maps. Significant advancements in the field of Natural Language Processing (NLP), particularly with the introduction of the BERT (Bidirectional Encoder Representations from Transformers) model, have enhanced the contextual semantic representation of text [3]. This representation opens up new opportunities in evaluating the quality of concept map content with a more in-depth and standardized approach.

In the context of classification, two commonly used approaches are Random Forest (RF), an ensemble learning algorithm that is robust against noise and overfitting, and Ordinal Regression, which is naturally suited for modeling hierarchical target variables such as quality levels. However, challenges arise when these models are applied to imbalanced datasets, as commonly found in the education domain. To address this issue, resampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique) have been developed to enhance the representation of

minority classes [4], [5]. Various studies have shown that the combination of SMOTE and Random Forest provides significant performance improvements in various classification domains facing data imbalance issues. For example, [6] reported an increase in accuracy and AUC in air quality classification with RF and XGBoost after the application of SMOTE. [7] showed that sensitivity in lung cancer diagnosis increased from 55% to 94.5% with the application of SMOTE before RF. In the domain of psychology, [8] reported a 15% increase in accuracy in panic disorder classification, while [9] showed an improvement in precision and recall in sentiment analysis after combining RF and SMOTE. Although most of these studies focus on binary or multiclass classification, their approaches and findings are highly relevant to ordinal classification tasks, especially those involving hierarchical structures and imbalanced class distributions.

Despite the abundance of such research, there remains a gap in comparative studies on ordinal classification algorithms like Ordinal Regression and ensemble-based approaches like Random Forest in the context of evaluating the quality of concept maps, especially when using semantic representations from BERT and on imbalanced data. Furthermore, there have not been many studies that deeply explore the impact of applying SMOTE on the stability and accuracy of the model in such scenarios.

This research aims to fill that gap by comparing the effectiveness of the Random Forest algorithm and Ordinal Regression in classifying the quality of concept maps based on BERT embeddings, as well as analyzing the impact of applying SMOTE on the model's performance and stability. The dataset used consists of 32 concept maps in the domain of relational databases, which have an imbalanced class distribution. Model evaluation is conducted using metrics such as accuracy, precision, recall, F1-score, and Cohen's Kappa to measure the agreement between the model's predictions and the reference labels from human assessors.

With this approach, it is hoped that the research can make a significant contribution to the development of automatic classification systems for ordinal and imbalanced educational data, as well as demonstrate the potential integration of modern NLP techniques and classical classification approaches in evaluating the quality of concept maps.

Method

A. Research Design

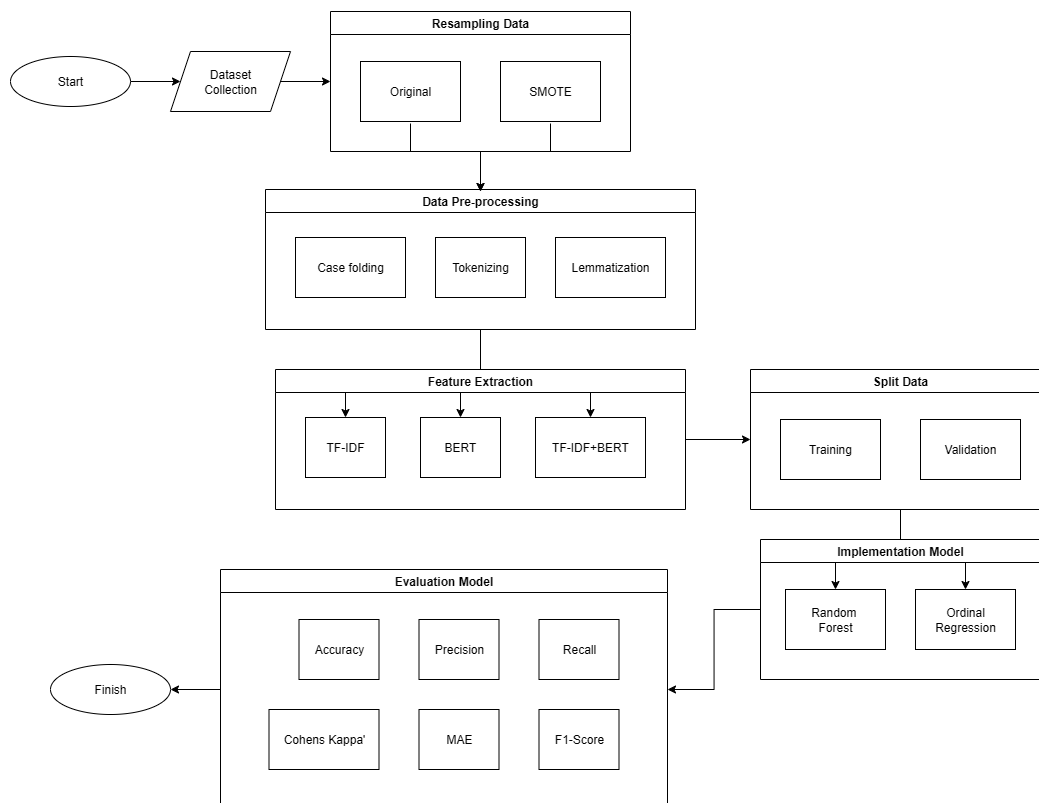


Figure 1. Research Design

Figure 1 shows the systematic flow of the research design applied to evaluate the performance of two classification algorithms Random Forest and Ordinal Regression in the task of classifying concept map quality. The research process begins with the stage of dataset collection, which is then divided into two scenarios: original data and data that has been resampled using SMOTE to address class imbalance. The data then undergoes a text preprocessing stage that includes case folding, tokenization, and lemmatization, before being extracted using three feature representation approaches: TF-IDF, BERT, and a combination of TF-IDF + BERT. After feature extraction, the data is divided into training and testing sets, and then used in the training and testing of two classification models. In the final stage, the model's performance is evaluated using seven key metrics, namely Accuracy, Precision, Recall, F1-Score, Cohen's Kappa, and MAE, to assess the accuracy and stability of predictions in the context of ordinal classification with an uneven class distribution.

B. Data Collection

The data used in this study is a collection of concept map propositions from students in the Relational Database course. This dataset consists of 691 rows of data, which includes participant identity information (UID), proposition sequence numbers, proposition text, and proposition quality scores. This dataset was obtained from previous research by Prasetya et al. (2022) which examined the structure of knowledge and student perceptions in concept mapping activities [10]. All data have undergone an annotation process by experts using a standardized assessment rubric that considers clarity, accuracy, and the interconnection between concepts. Quality scores are given on an ordinal scale and used as target labels in classification. The "Proposition" column is used as the main feature in the text extraction process, while the "score" column becomes the target classification variable.

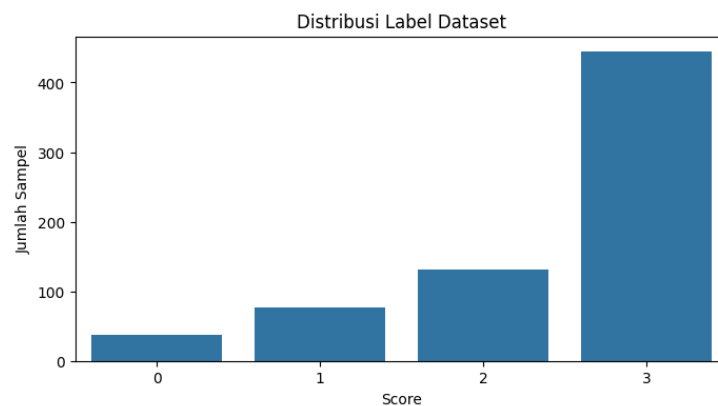


Figure 2. Distribution Label Score

The initial distribution of class labels in the dataset is shown in **Figure 2**, which indicates that the data has an imbalance between classes, with a dominance of the highest score class (score 3) and a significantly smaller number in the score 0 and 1 classes. This imbalance is a significant concern as it has the potential to cause model bias towards the majority class during the training process. Therefore, this initial analysis serves as the basis for applying the resampling technique using SMOTE, which is explained in the following section.

C. Resampling Data

One of the main challenges in this research is the class distribution imbalance in the concept map quality score labels. Based on the initial exploration of the dataset, it appears that there is a dominance of classes with the highest scores, while classes with low or medium scores are much fewer. This condition can cause the model to tend to predict the majority class and ignore patterns from the minority class. **Figure 2** shows the label distribution before the resampling process, where the class with score 3 significantly dominates compared to scores 0, 1, and 2.

To address this issue, this study employs a resampling technique known as Synthetic Minority Over-sampling Technique (SMOTE) [6], [8], [11]. SMOTE works by synthesizing new samples from the minority class based on interpolation between points in the feature space that are close to each other, thereby allowing for a more balanced class distribution without explicitly duplicating data [7], [12], [13]. This technique is specifically applied to the training data to avoid the risk of information leakage to the test data, which could affect the validity of the testing.

The resampling process results in a more balanced class distribution as shown in **Figure 3**, where each score class now has an equal number of samples. In the experiment, two versions of the data were used: the original data without resampling and the oversampled data using SMOTE, to comparatively assess the impact of SMOTE on the model's

performance and stability. This approach aims to evaluate the effectiveness of resampling in the context of ordinal classification, particularly in the education domain, which is often characterized by uneven score distributions.

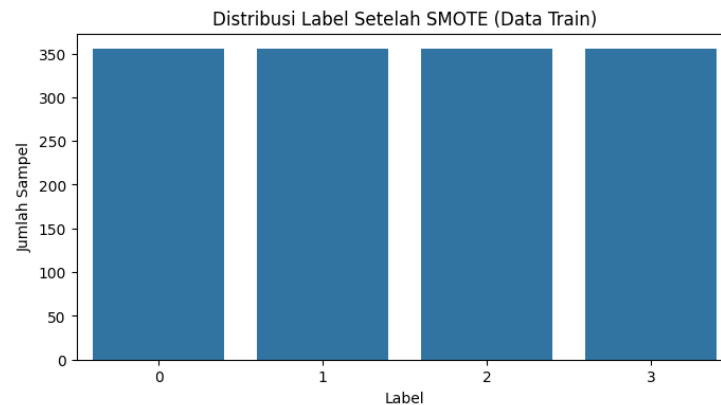


Figure 3. Distribution Label *Score* after SMOTE

D. Data Pre-processing

Before the feature extraction and model building process, the text data from the "Proposition" column undergoes a series of preprocessing stages to ensure the quality and consistency of the input. This process is important so that the text can be represented in a numerical form that is accurate and relevant for machine learning analysis. The three main steps used in this stage are case folding, tokenizing, and lemmatization [14]–[16].

1. Case folding: that is the process of converting all letters in the text to lowercase. This step aims to eliminate unnecessary differences in meaning between the same words but with different capitalizations (for example, "Data" and "data").
2. Tokenizing: which means breaking the text into word units or tokens. This stage is very important to prepare the text before feature extraction based on frequency or embedding. Tokenization is performed using a space and punctuation-based approach, adjusted to standard tokenizers in natural language processing libraries such as nltk or transformers
3. Lemmatization: which involves changing each token to its base form (lemma), considering grammatical contexts such as nouns, verbs, or adjectives. Unlike stemming, which only cuts words to their root forms heuristically, lemmatization retains the meaning of words with the help of linguistic dictionaries. This process helps simplify word variations and enhance the generalization of features extracted from the text.

All these preprocessing stages are consistently applied to the entire dataset before the feature extraction process is carried out, to ensure that the semantic information from each proposition can be effectively captured by the machine learning model.

E. Representasi Fitur

After going through the pre-processing stage, the proposition text from the dataset is represented in the form of numerical features so that it can be processed by machine learning algorithms. In this study, three feature representation approaches are used, namely Term Frequency-Inverse Document Frequency (TF-IDF), Bidirectional Encoder Representations from Transformers (BERT), and a combination of TF-IDF + BERT. These three approaches were chosen to examine the performance differences between conventional representation methods and context-based semantic representation.

The first representation is TF-IDF, a statistical technique that measures the importance of a word in a document relative to the entire corpus. TF-IDF works by calculating the term frequency, calibrated with the inverse document frequency [17]–[19]. This representation produces a high-dimensional and sparse feature matrix, but it has proven effective for many text classification tasks, especially in scenarios with limited data.

The second representation uses BERT, a transformer-based deep learning model that generates contextual embedding vectors for each token in the sentence. In this study, the pre-trained BERT model (bert-base-uncased) from the Hugging Face Transformers library [20], [21] was used. Each proposition is represented using the vector from the [CLS] token as the overall sentence representation. BERT allows the model to understand the meaning of words based

on the overall context of the sentence, but it requires greater computational capacity and generally works optimally on large-sized datasets [22].

The third approach is a combination of TF-IDF and BERT, which integrates the statistical representation power of TF-IDF with the semantic capabilities of BERT embeddings [23], [24]. The features from both representations are concatenated and used as input to the classification model. The goal of this combination is to merge the advantages of both: the stability of the TF-IDF representation and the contextual understanding of BERT. Each representation approach is applied separately in experiments and their performance is compared to assess the impact of the representation on the model's accuracy and prediction stability.

F. Modelling

This research uses two main classification algorithms, namely Random Forest and Ordinal Regression, each of which has different approaches and assumptions towards classification problems, especially with ordinal labeled data.

- Random Forest

Random Forest is a bagging-based ensemble learning algorithm that combines predictions from a number of decision trees to produce more accurate and stable classification results [25]–[27]. Each tree is trained using a random subset of the data and a random subset of features, thereby naturally reducing the risk of overfitting. Mathematically, the final prediction \hat{y} from the Random Forest on input x is defined as:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_N(x)\} \quad (7)$$

where $h_i(x)$ is the prediction from the i -th tree and N is the total number of trees in the ensemble.

- Ordinal Regression

Ordinal regression, or ordinal logistic regression (proportional odds model), is a statistical method used to model the relationship between one or more predictor variables and an ordinal (ranked) target variable [28]–[30]. This model assumes that there is a threshold between each class category. The basic function of this model is:

$$\log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \theta_j - \beta^T x, \quad j = 1, 2, \dots, J - 1 \quad (7)$$

where,

Y is the ordinal target variable with J category,

θ_j is the threshold or cutoff for the j -th category,

β is the vector of regression coefficients,

x is the vector of input features.

G. Evaluasi dan Metric

Model performance evaluation is conducted using a number of metrics designed to assess the accuracy, consistency, and capability of the model in handling imbalanced ordinal labeled data. In this study, primary metrics such as Accuracy, Precision, Recall, F1-Score, and Cohen's Kappa were used to evaluate classification performance. Additionally, Mean Absolute Error (MAE) is added as a more suitable metric to evaluate how far the model's predictions deviate from the ordinal labels in order. The use of MSE and RMSE is still included as a complement to quantitative evaluation, but they are not the primary metrics in the context of ordinal classification.

- Accuracy measures the proportion of correct predictions to the total predictions, but does not account for class imbalance.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

- Precision measures the accuracy of positive predictions:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (5)$$

- Recall (Sensitivity) measures how well the model finds all positive data:

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

- F1-Score is the harmonization between precision and recall:

$$F - measure = \frac{2(presisi \times recall)}{(presisi + recall)} \quad (7)$$

- MAE (Mean Absolute Error) is used to evaluate how far the model's predictions deviate from the ordinal labels in order:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

- Cohen's Kappa

Cohen's Kappa is used to measure the level of agreement between two annotators or between the model's predictions and the actual labels, controlling for the possibility of agreement occurring by chance [31], [32].

$$\kappa = \frac{\rho_o - \rho_e}{1 - \rho_e} \quad (7)$$

where,

ρ_o is the proportion of actual agreement (observed agreement),

ρ_e is the proportion of agreement expected by chance (expected agreement).

The Kappa value ranges from -1 to 1, with:

$\kappa < 0$ no agreement,

$0.0 < \kappa \leq 0.4$ low agreement

$0.4 < \kappa \leq 0.6$ moderate

$0.6 < \kappa \leq 0.8$ good

$\kappa > 0.8$ very good

This metric is highly relevant in ordinal classification because it considers the level of error between categories, not just the number of incorrect predictions.

Results and Discussion

A. Results

This study compares the performance of two classification algorithms, namely Random Forest and Ordinal Regression (using Logistic Regression), on three feature representation approaches: TF-IDF, BERT, and a combination of TF-IDF + BERT, with two data scenarios: original data and data resulting from oversampling using SMOTE. Evaluation is conducted using seven metrics: Accuracy, Precision, Recall, F1-Score, Cohen's Kappa, and MAE. The following table summarizes the experimental results based on the three feature representations:

Table 1. Model Evaluation Results Based on Feature Representation and Resampling Technique

Representation	Model	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa	MAE
TF-IDF	Random Forest (Original)	0.848921	0.845553	0.848921	0.835516	0.690325	0.3381
TF-IDF	Ordinal Regression (Original)	0.848921	0.825741	0.848921	0.824811	0.679794	0.3381
TF-IDF	Random Forest (SMOTE)	0.870504	0.873919	0.870504	0.865223	0.748442	0.3887
TF-IDF	Ordinal Regression (SMOTE)	0.877698	0.878247	0.877698	0.877251	0.770069	0.2674
BERT	Random Forest (Original)	0.805755	0.822956	0.805755	0.779834	0.577698	0.21

Representation	Model	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa	MAE
BERT	Ordinal Regression (Original)	0.834532	0.818862	0.834532	0.819809	0.669697	0.4971
BERT	Random Forest (SMOTE)	0.834532	0.836994	0.834532	0.830387	0.6829	0.4234
BERT	Ordinal Regression (SMOTE)	0.827338	0.809726	0.827338	0.813731	0.662246	0.3892
TF-IDF + BERT	Random Forest (Original)	0.820144	0.839529	0.820144	0.796737	0.611688	0.4163
TF-IDF + BERT	Logistic Regression (Original)	0.841727	0.824714	0.841727	0.826074	0.681624	0.4582
TF-IDF + BERT	Random Forest (SMOTE)	0.848921	0.844124	0.848921	0.841555	0.705598	0.3976
TF-IDF + BERT	Logistic Regression (SMOTE)	0.848921	0.848125	0.848921	0.845522	0.709783	0.3402

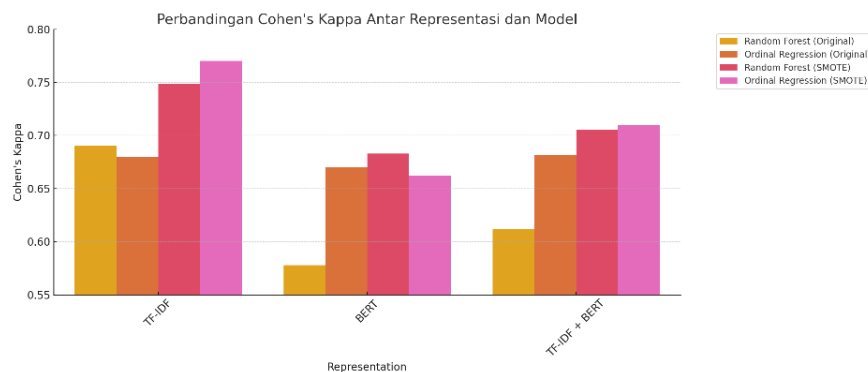


Figure 4. Comparison of Cohen's Kappa between Representations and Models

Table 1 shows that feature representation with TF-IDF combined with SMOTE and the Ordinal Regression algorithm yields the best overall performance, with an accuracy of 0.8777, F1-Score of 0.8773, and Cohen's Kappa of 0.7701. The Random Forest model with SMOTE on TF-IDF also shows very high performance, with an F1-Score of 0.8652 and RMSE of 0.5876. Meanwhile, the BERT-based approach generally produces lower metric values, especially in the Random Forest model, with the lowest recorded accuracy of 0.8058. In the combined TF-IDF + BERT approach, the model's performance is slightly better than BERT alone but has not yet surpassed the results of pure TF-IDF, although it shows fairly stable Cohen's Kappa values.

B. Discussion

The research results show that TF-IDF representation is still very effective for text classification tasks on small to medium-sized datasets, especially when combined with the SMOTE resampling technique that can address class imbalance. The high performance of the Ordinal Regression model with TF-IDF + SMOTE can be explained by the model's ability to understand hierarchical class relationships and by the more stable nature of the TF-IDF representation in capturing word frequencies compared to complex semantic embeddings. On the contrary, BERT, which is an advanced context-based representation, failed to demonstrate its superiority in this experiment. This is most likely due to the limited amount of data, where transformer-based models like BERT require much larger datasets to learn optimally. The combination of TF-IDF and BERT yields fairly competitive results, but still falls short of the performance of pure TF-IDF, indicating that the integration of two types of representations does not always guarantee performance improvement, especially when the data is insufficient to leverage the advantages of contextual representation. The results of this study are in line with several previous studies that demonstrate the effectiveness of combining balancing techniques such as SMOTE with ensemble-based algorithms or regression on imbalanced educational data. For example, studies [5], [33] show that SMOTE significantly improves classification performance in the text-based education domain. Furthermore, the TF-IDF approach used in this study has proven to be still relevant for small to medium-sized datasets, as demonstrated in study [34] in frequency-based spam classification. Thus, the findings in this study reinforce the evidence that classical methods such as TF-IDF remain competitive when combined with balancing strategies and classification models that align with the nature of ordinal data.

Conclusion

This research aims to compare the performance of the Random Forest and Ordinal Regression algorithms in classifying the quality of concept maps using three feature representation approaches, namely TF-IDF, BERT, and a combination of TF-IDF + BERT, as well as two data scenarios: original data and data that has been resampled with SMOTE. Based on the evaluation results using accuracy, F1-score, and Cohen's Kappa metrics, it was found that the TF-IDF representation approach with SMOTE consistently produced the best performance, particularly in the Ordinal Regression model, which recorded an accuracy of 0.8777, an F1-score of 0.8773, and a Cohen's Kappa of 0.7701. These findings indicate that the use of resampling techniques such as SMOTE significantly improves the performance and stability of the model in handling imbalanced class distributions. On the other hand, the BERT approach, which theoretically excels in capturing contextual meaning, has not yet shown optimal performance in this study, likely due to the limited amount of available data. Meanwhile, the combined approach of TF-IDF and BERT yielded competitive results but did not surpass the effectiveness of TF-IDF independently. Overall, this study concludes that the combination of simple representations such as TF-IDF with the SMOTE data balancing technique and the use of the Ordinal Regression model is an effective strategy for classifying the quality of concept maps on datasets with ordinal characteristics and uneven class distribution.

References

- [1] C. Yang, "Neural Concept Map Generation for Effective Document Classification with Interpretable Structured Summarization," *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1629–1632, 2020, doi: [10.1145/3397271.3401312](https://doi.org/10.1145/3397271.3401312).
- [2] H. Choi, "Enhancement of Knowledge Concept Maps Using Deductive Reasoning with Educational Data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14798. pp. 104–116, 2024, doi: [10.1007/978-3-031-63028-6_9](https://doi.org/10.1007/978-3-031-63028-6_9).
- [3] S. Gao, "Research on Intention Recognition of Educational Counseling Combining BERT and Concept Map," *Proceedings - 2023 12th International Conference of Information and Communication Technology, ICTech 2023*. pp. 236–240, 2023, doi: [10.1109/ICTech58362.2023.00054](https://doi.org/10.1109/ICTech58362.2023.00054).
- [4] W. Satria and M. Riassetiawan, "Essay answer classification with smote random forest and adaboost in automated essay scoring," *IJCCS*, vol. 17, no. 4, p. 359, 2023, doi: <https://doi.org/10.22146/ijccs.82548>.
- [5] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 1, p. 54, Jan. 2023, doi: [10.3390/info14010054](https://doi.org/10.3390/info14010054).
- [6] F. P. Arifianti and A. Salam, "XGBoost and Random Forest Optimization using SMOTE to Classify Air Quality," *Adv. Sustain. Sci. Eng. Technol.*, 2024, doi: <https://doi.org/10.26877/asset.v6i1.18136>.
- [7] C. Michael Lauw, H. Hairani, I. Saifuddin, J. Ximenes Guterres, M. Maariful Huda, and M. Mayadi, "Combination of Smote and Random Forest Methods for Lung Cancer Classification," *Int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, pp. 59–64, Sep. 2023, doi: [10.30812/ijecsa.v2i2.3333](https://doi.org/10.30812/ijecsa.v2i2.3333).
- [8] D. Nurmalasari, H. R. Yuliantoro, and D. H. Qudsi, "Improving Panic Disorder Classification Using SMOTE and Random Forest," *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 272–279, 2024, doi: <https://doi.org/10.30871/jaic.v8i2.8315>.
- [9] K. Karfindo, R. Turaina, and R. Saputra, "Optimalisasi Klasifikasi Umpan Balik Mahasiswa Terhadap Layanan Kampus dengan Sinergi Random Forest dan Smote," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 6, no. 6, pp. 820–827, Jan. 2024, doi: [10.32672/jnkti.v6i6.7269](https://doi.org/10.32672/jnkti.v6i6.7269).
- [10] D. D. Prasetya, A. Pinandito, Y. Hayashi, and T. Hirashima, "Analysis of quality of knowledge structure and students' perceptions in extension concept mapping," *Res. Pract. Technol. Enhanc. Learn.*, vol. 17, no. 1, p. 14, Dec. 2022, doi: [10.1186/s41039-022-00189-9](https://doi.org/10.1186/s41039-022-00189-9).
- [11] F. Kamalov, S. E. Choutri, and A. F. Atiya, "Analytical formulation of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Gulf J. Math.*, vol. 19, no. 1, pp. 400–415, Jan. 2025, doi: [10.56947/gjom.v19i1.2639](https://doi.org/10.56947/gjom.v19i1.2639).

-
- [12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [13] I. K. Dharmendra, I. M. A. W. Putra, and Y. P. Atmojo, "Evaluation of the Effectiveness of SMOTE and Random Under Sampling in Emotion Classification of Tweets," *INFORMATICS Educ. Prof. J. Informatics*, vol. 9, no. 2, p. 182, Dec. 2024, doi: [10.51211/itbi.v9i2.3183](https://doi.org/10.51211/itbi.v9i2.3183).
- [14] M. Sergii V. and N. Oleksandr V., "Data preprocessing and tokenization techniques for technical Ukrainian texts," *Appl. Asp. Inf. Technol.*, vol. 6, no. 3, pp. 318–326, Sep. 2023, doi: [10.15276/aait.06.2023.22](https://doi.org/10.15276/aait.06.2023.22).
- [15] M. V. J. da Silva, E. E. Santana, F. M. F. Lobato, and A. F. L. Jacob Jr., "Preprocessing Applied to Legal Text Mining: analysis and evaluation of the main techniques used," in *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2023)*, Sep. 2023, pp. 1010–1021, doi: [10.5753/eniac.2023.234555](https://doi.org/10.5753/eniac.2023.234555).
- [16] A. Kathuria, A. Gupta, and R. K. Singla, "A Review of Tools and Techniques for Preprocessing of Textual Data," 2021, pp. 407–422.
- [17] K. Yusupov, "Comparative Analysis of Machine Learning and Deep Learning Models for Email Spam Classification Using TF-IDF and Word Embedding Techniques," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 231, pp. 114–122, 2025, doi: [10.1007/978-3-031-76452-3_11](https://doi.org/10.1007/978-3-031-76452-3_11).
- [18] G. Popoola, "Sentiment Analysis of Financial News Data using TF-IDF and Machine Learning Algorithms," *2024 IEEE 3rd International Conference on AI in Cybersecurity, ICAIC 2024*. 2024, doi: [10.1109/ICAIC60265.2024.10433843](https://doi.org/10.1109/ICAIC60265.2024.10433843).
- [19] S. M. M. Hossain, "TF-IDF feature-based spam filtering of mobile SMS using a machine learning approach," *Applied Intelligence for Industry 4.0*. pp. 162–175, 2023.
- [20] Y. Qiu, "ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems," *Intell. Syst. with Appl.*, vol. 21, 2024, doi: [10.1016/j.iswa.2023.200308](https://doi.org/10.1016/j.iswa.2023.200308).
- [21] A. S. Khan, "Integrating BERT Embeddings with SVM for Prostate Cancer Prediction," *Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024*. pp. 574–579, 2024, doi: [10.1109/ICEEICT62016.2024.10534547](https://doi.org/10.1109/ICEEICT62016.2024.10534547).
- [22] A. C. Mazari, "BERT-based ensemble learning for multi-aspect hate speech detection," *Cluster Comput.*, vol. 27, no. 1, pp. 325–339, 2024, doi: [10.1007/s10586-022-03956-x](https://doi.org/10.1007/s10586-022-03956-x).
- [23] J. W. Sun, "Text Classification Algorithm Based on TF-IDF and BERT," *Proceedings - 2022 11th International Conference of Information and Communication Technology, ICTech 2022*. pp. 533–536, 2022, doi: [10.1109/ICTech55460.2022.00112](https://doi.org/10.1109/ICTech55460.2022.00112).
- [24] L. Gomes, "BERT- and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: A comparative study," *Inf. Softw. Technol.*, vol. 160, 2023, doi: [10.1016/j.infsof.2023.107217](https://doi.org/10.1016/j.infsof.2023.107217).
- [25] A. Gupta and S. Gupta, "Enhanced Classification of Imbalanced Medical Datasets using Hybrid Data-Level, Cost-Sensitive and Ensemble Methods," *Int. Res. J. Multidiscip. Technovation*, pp. 58–76, Apr. 2024, doi: [10.54392/irjmt2435](https://doi.org/10.54392/irjmt2435).
- [26] K. V. Ramana, Y. B. S. S. P. Ponsudha, S. Pd, and A. V. Sangeetha, "Applying Cost-Sensitive Learning Methods to Improve Extremely Unbalanced Big Data Problems Using Random Forest," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, May 2023, pp. 1–7, doi: [10.1109/ACCAI58221.2023.10199250](https://doi.org/10.1109/ACCAI58221.2023.10199250).
- [27] H. Cui, H. Xu, and J. Li, "Optimization of random forest algorithm based on mixed sampling additional feature selection," in *2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Jan. 2023, pp. 461–467, doi: [10.1109/ICCECE58074.2023.10135433](https://doi.org/10.1109/ICCECE58074.2023.10135433).
-

-
- [28] J. Simićević, "Ordinal Regression Model of Parking Search Time," *Promet - Traffic Transp.*, vol. 35, no. 6, pp. 904–916, 2023, doi: [10.7307/ptt.v35i6.291](https://doi.org/10.7307/ptt.v35i6.291).
- [29] Y. Tajima, "Ordinal Regression Based on the Distributional Distance for Tabular Data," *IEICE Trans. Inf. Syst.*, no. 3, pp. 357–364, 2023, doi: [10.1587/transinf.2022EDP7071](https://doi.org/10.1587/transinf.2022EDP7071).
- [30] C. Lee, "Ordinal Regression for Beef Grade Classification," *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, vol. 2023. 2023, doi: [10.1109/ICCE56470.2023.10043530](https://doi.org/10.1109/ICCE56470.2023.10043530).
- [31] N. Habbat, "Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning," *Eng. Appl. Artif. Intell.*, vol. 126, 2023, doi: [10.1016/j.engappai.2023.106999](https://doi.org/10.1016/j.engappai.2023.106999).
- [32] A. Zaboli, "Human intelligence versus Chat-GPT: who performs better in correctly classifying patients in triage?," *Am. J. Emerg. Med.*, vol. 79, pp. 44–47, 2024, doi: [10.1016/j.ajem.2024.02.008](https://doi.org/10.1016/j.ajem.2024.02.008).
- [33] W. Satria and M. Riasetiawan, "Essay Answer Classification with SMOTE Random Forest and AdaBoostin Automated Essay Scoring," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 17, no. 4, p. 359, Oct. 2023, doi: [10.22146/ijccs.82548](https://doi.org/10.22146/ijccs.82548).
- [34] S. M. M. Hossain, K. M. A. Kamal, A. Sen, and I. H. Sarker, *TF-IDF Feature-Based Spam Filtering of Mobile SMS Using a Machine Learning Approach*. 2023.