Indonesian Journal of Data and Science

Volume 7 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v7i1.376

Research Article

Handling Imbalanced Data in Concept Map Proposition Quality Classification: A Comparative Study of Resampling and Class-Weighted SVM

Nurul Rismayanti [1,*]; Lukman Syafie [2]; A Sinra [3]; Rahmadani [4]
[1] Universitas Muslim Indonesia, Jl. Urip Sumoharjo No.km.5, Panaikang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231, Indonesia, nurulrismayanti.labfik.umi.ac.id
[2] Universiti Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia, lukman.syafie@s.unikl.edu.my
[3] UPTP. Wil. Makassar 1 Badan Pendapatan Daerah Prov. Sulsel, Indonesia, andisinra78@gmail.com
[4] Universitas Muslim Indonesia, Jl. Urip Sumoharjo No.km.5, Panaikang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231, Indonesia, rahmadani.iclabs@umi.ac.id

Correspondence should be addressed to Nurul Rismayanti; nurulrismayanti.labfik@umi.gmail

Received 03 Jan 2026; Accepted 15 June 2025; Published 31 March 2026

Abstract:
Concept maps are widely used in education to represent students' knowledge structures and conceptual understanding. At the proposition level, automatic quality classification can support faster and more consistent assessment than manual evaluation. However, proposition quality datasets are often imbalanced, causing classification models to favor the majority class and perform poorly on minority classes. This study aims to compare several imbalance handling techniques for concept map proposition quality classification using Support Vector Machine (SVM) and TF-IDF features. The dataset consisted of 691 propositions labeled into four quality classes, with class 3 dominating the distribution. The evaluated methods included baseline training without balancing, Random OverSampler, SMOTE, ADASYN, SMOTE-Tomek, SMOTE-ENN, and a cost-sensitive SVM with class_weight='balanced'. The dataset was divided into training and testing sets using an 80:20 stratified split, and balancing methods were applied only to the training data. Performance was evaluated using accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, and mean absolute error (MAE). The results show that the cost-sensitive SVM with class_weight='balanced' achieved the best overall performance, with an accuracy of 0.8417, macro precision of 0.7098, macro recall of 0.7320, macro F1-score of 0.7149, weighted F1-score of 0.8455, and MAE of 0.2158. Among the resampling methods, ADASYN and Random OverSampler were the most competitive, while SMOTE, SMOTE-Tomek, and SMOTE-ENN produced lower macro-level performance. These findings indicate that for sparse text-based concept map proposition data, class-weighted SVM is more effective than the evaluated resampling methods in improving balanced classification performance. The study highlights the importance of using both macro-level and ordinal-aware evaluation metrics when addressing imbalanced educational text classification tasks.

Keywords: Concept Map, Proposition Quality Classification, Imbalanced Data, Support Vector Machine, TF-IDF, Class-Weighted SVM.

1. Introduction

Concept maps have long been recognized as an effective educational tool for representing learners' knowledge structures and conceptual understanding [1], [2]. By explicitly showing the relationships among concepts, concept maps enable educators to assess not only whether students know particular concepts, but also how well they organize and connect those concepts meaningfully. Among the components of a concept map, propositions play a central role because they represent the basic semantic relationships between concepts. Consequently, proposition-level quality assessment is important for understanding the overall quality of a learner's concept map [3], [4]. Despite its pedagogical value, manual assessment of concept map propositions remains time-consuming, labor-intensive, and

potentially inconsistent, especially when applied to large numbers of student submissions. These limitations motivate the need for automated assessment approaches that are efficient, scalable, and reliable.

In recent years, machine learning approaches have increasingly been explored for educational assessment tasks, including text-based classification problems. For concept map proposition quality classification, machine learning offers the possibility of assigning propositions into predefined quality categories automatically, thereby supporting faster and more consistent evaluation [5]. In this setting, Support Vector Machine (SVM) is particularly relevant because it has demonstrated strong performance on sparse and high-dimensional textual representations such as TF-IDF. However, the effectiveness of such automated classification systems is often constrained by the class distribution of the available dataset. When the dataset is imbalanced, the trained model tends to favor the majority class, which may produce deceptively high overall accuracy while underperforming on minority classes. This issue is especially problematic in educational assessment, where minority classes may represent low-quality propositions that are crucial to detect for diagnostic and formative purposes.

Class imbalance has been widely acknowledged as a major challenge in supervised classification [6], [7]. A model trained on imbalanced data often learns biased decision boundaries because it is exposed more frequently to the dominant class than to minority classes. In practical terms, this means that the classifier may optimize its global performance by prioritizing majority-class predictions, while substantially reducing its sensitivity to underrepresented categories. In the context of concept map proposition assessment, such bias may weaken the fairness and usefulness of the automated evaluation process. A system that fails to identify lower-quality propositions accurately may overlook important learning deficiencies and thereby reduce its pedagogical value. Therefore, handling class imbalance is not only a technical concern but also an essential step toward building a fairer and more educationally meaningful automated assessment system.

To address imbalanced data, several strategies have been proposed in the literature. Resampling-based methods, such as Random OverSampler, SMOTE, and ADASYN, aim to improve the representation of minority classes by duplicating or synthesizing minority samples. Hybrid methods such as SMOTE-Tomek and SMOTE-ENN extend this idea by combining oversampling with cleaning mechanisms that attempt to remove noisy or overlapping instances. In contrast, cost-sensitive approaches modify the learning process itself by assigning greater importance to minority classes during model optimization. In SVM, this can be implemented through class weighting. Although these approaches are all designed to mitigate imbalance, their effectiveness may differ depending on the nature of the data, the classifier, and the feature representation used. For sparse text data represented through TF-IDF, generating synthetic samples may not always preserve meaningful semantic variation, making comparative evaluation particularly important.

Previous studies have shown that imbalance handling can improve minority-class recognition and produce better macro-level performance. However, in the context of concept map proposition quality classification, prior work has more often emphasized automatic scoring and educational interpretation than a systematic comparison of imbalance handling strategies. The earlier version of this study demonstrated that applying SMOTE-ENN improved recall, macro F1-score, and mean absolute error compared with a baseline SVM model, although overall accuracy decreased slightly. This result confirmed that class imbalance substantially affects proposition quality classification and that balancing strategies deserve further investigation. Nevertheless, relying on a single balancing method does not provide sufficient evidence regarding which imbalance handling strategy is most suitable for this task.

Accordingly, this study extends the previous work by conducting a comparative analysis of multiple imbalance handling techniques for concept map proposition quality classification using SVM [8], [9]. The evaluated approaches include training without balancing, Random OverSampler, SMOTE, ADASYN, SMOTE-Tomek, SMOTE-ENN, and a cost-sensitive SVM with class_weight='balanced'. The propositions are represented using TF-IDF features, and performance is assessed using accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, and mean absolute error (MAE). The inclusion of MAE is especially relevant because the target labels reflect ordered quality levels, meaning that prediction errors should also be interpreted according to their ordinal distance from the true class.

The contribution of this study is threefold. First, it provides a structured comparison between resampling-based and cost-sensitive imbalance handling methods in the specific domain of concept map proposition quality classification. Second, it offers empirical evidence regarding the suitability of these methods for sparse educational text data represented through TF-IDF. Third, it highlights the importance of evaluating imbalanced educational classification tasks using macro-level and ordinal-aware measures rather than relying solely on accuracy. Through these contributions, this study aims to identify a more effective and equitable strategy for automatic concept map proposition assessment and to support the development of more reliable educational evaluation systems.

2. Method

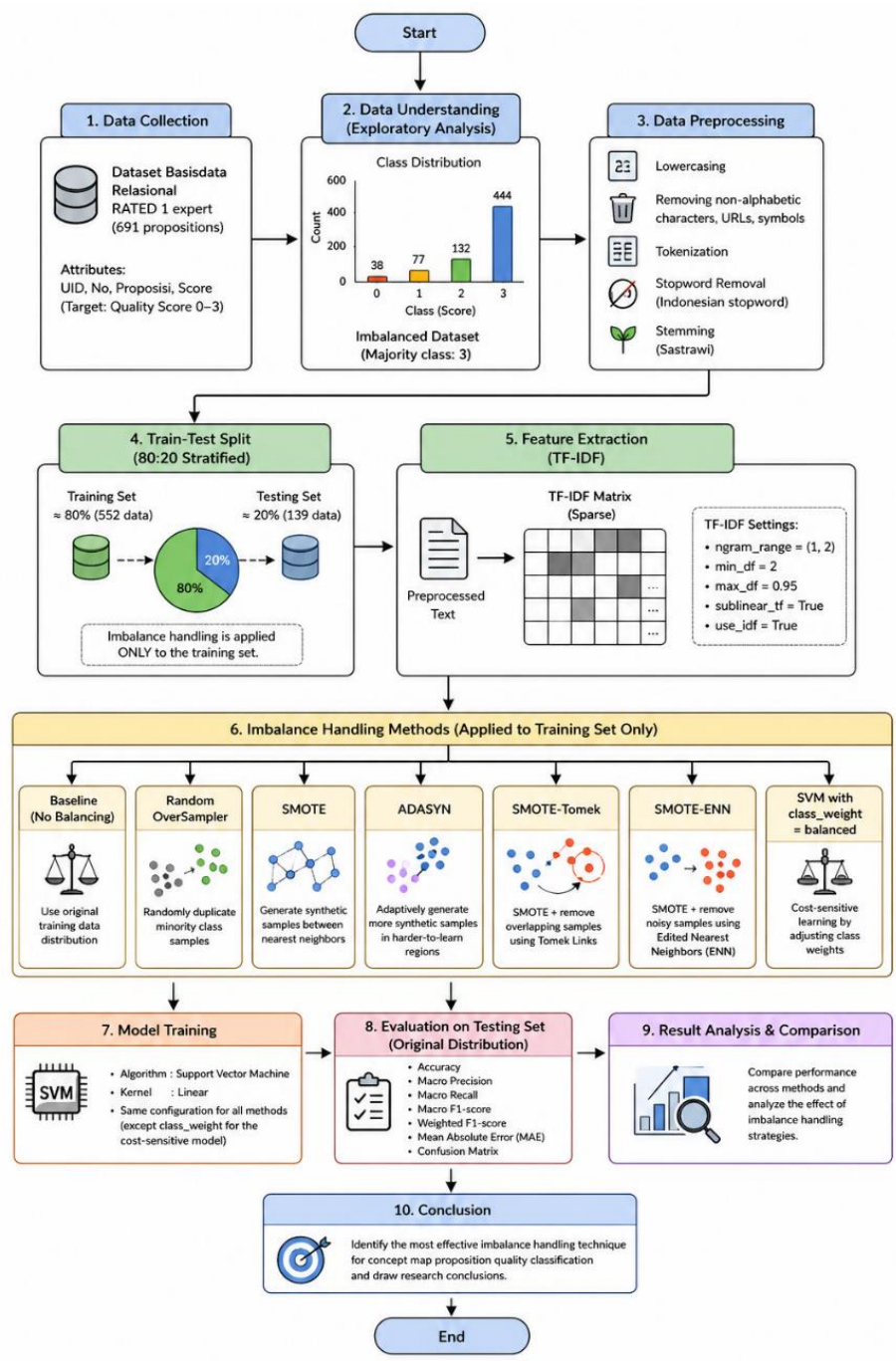

Figure 1. Research Workflow

Figure 1 illustrates the overall research workflow used in this study, starting from dataset preparation and text preprocessing to feature extraction, imbalance handling, model training, evaluation, and final result analysis.

Research Design

This study employed a comparative experimental design to evaluate the effectiveness of several imbalance handling techniques for concept map proposition quality classification. The experiment was conducted by using the same dataset, preprocessing steps, feature extraction method, classifier, and testing set across all scenarios, while varying only the imbalance handling strategy. This design was intended to ensure a fair comparison among methods and to isolate the contribution of each imbalance handling approach to classification performance.

The evaluated scenarios consisted of one baseline model without balancing and six imbalance handling strategies. The baseline model was trained using the original training data distribution. The imbalance handling strategies included Random OverSampler, SMOTE, ADASYN, SMOTE-Tomek, SMOTE-ENN, and a cost-sensitive Support Vector Machine with class_weight='balanced'. The overall workflow of the study included dataset preparation, text preprocessing, train-test splitting, TF-IDF feature extraction, imbalance handling, model training, and performance evaluation.

Dataset

The dataset used in this study was Dataset Basisdata Relasional RATED 1 expert, which contains proposition-level data from concept maps. Each record represents a single proposition in textual form and is associated with a quality score label. The dataset contains four main attributes: UID, No, Proposisi, and Score. The UID attribute identifies the source of the proposition, No indicates the proposition order within each source, Proposisi contains the proposition text, and Score represents the target quality label.

After removing incomplete records, the dataset consisted of 691 propositions. The quality labels were divided into four classes, ranging from 0 to 3. The class distribution was imbalanced, with class 3 dominating the dataset. Specifically, class 0 contained 38 samples, class 1 contained 77 samples, class 2 contained 132 samples, and class 3 contained 444 samples. This distribution indicates a substantial imbalance problem, which motivated the use of specialized imbalance handling techniques in this study.

Data Splitting

The dataset was divided into training and testing sets using an 80:20 ratio. A stratified split was applied in order to preserve the class proportion in both subsets. As a result, the training set contained approximately 80% of the data, while the remaining 20% were used as the testing set. The use of stratification was important to ensure that minority classes were still represented in the testing phase and that the evaluation reflected the original imbalance pattern of the dataset.

In this study, imbalance handling methods were applied only to the training data, while the testing data were kept unchanged. This decision was made to avoid data leakage and to ensure that model performance was evaluated on the original data distribution. Such a setup is essential in imbalanced classification studies because resampling the testing set would lead to overly optimistic results and would not reflect real-world performance.

Text Preprocessing

Since the input data were in textual form, preprocessing was applied to transform the raw propositions into a cleaner and more consistent representation before feature extraction [10], [11], [12]. The preprocessing pipeline consisted of several stages.

First, all proposition texts were converted to lowercase to reduce case variation. Second, non-alphabetic characters, URLs, and irrelevant symbols were removed. Third, the cleaned text was tokenized into individual words. Fourth, Indonesian stopwords were removed to eliminate commonly occurring words with limited discriminative value. Finally, stemming was applied using the Sastrawi stemmer in order to reduce inflected words to their root forms.

This preprocessing procedure was applied consistently to all proposition texts before splitting and feature extraction. The aim was to improve feature quality and reduce noise in the textual representation used by the classifier.

Feature Extraction Using TF-IDF

After preprocessing, the proposition texts were transformed into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a widely used text representation method that assigns higher weights to terms that occur frequently within a document but less frequently across the corpus [13], [14], [15], [16]. This weighting scheme helps emphasize informative words while reducing the influence of overly common terms.

In this study, TF-IDF vectorization was configured with a unigram and bigram range (ngram_range=(1,2)), a minimum document frequency of 2, a maximum document frequency of 0.95, and sublinear term frequency scaling. These settings were chosen to capture both single-word and two-word phrase patterns while filtering out extremely rare and extremely common terms. The TF-IDF model was fitted on the training data and then applied to transform both the training and testing sets into sparse numerical feature matrices [17], [18].

Imbalance Handling Techniques

To address the class imbalance problem, six imbalance handling strategies were compared with the baseline model.

a. Random OverSampler
   This method randomly duplicates existing minority class samples in the training data until a more balanced class distribution is achieved.
b. SMOTE (Synthetic Minority Oversampling Technique)
   SMOTE generates synthetic minority class samples by interpolating between existing minority instances and their nearest neighbors [19], [20].
c. ADASYN (Adaptive Synthetic Sampling)
   ADASYN extends the idea of synthetic oversampling by generating more synthetic samples in regions where the minority class is harder to learn.
d. SMOTE-Tomek
   This hybrid method combines SMOTE oversampling with Tomek Links cleaning in order to reduce class overlap after synthetic sample generation.
e. SMOTE-ENN
   This method combines SMOTE with Edited Nearest Neighbors (ENN), where synthetic minority samples are created first and then ambiguous or noisy samples are removed based on nearest neighbor rules [9], [21].
f. Cost-sensitive SVM with class_weight='balanced'
   Instead of modifying the training data distribution, this approach adjusts the classifier by assigning larger penalties to errors on minority classes. In this way, the classifier becomes more sensitive to underrepresented classes during training.

All resampling methods were applied only to the TF-IDF-transformed training data. The testing set was never resampled. This setup ensured that all methods were compared under the same evaluation conditions.

Classification Model

The classifier used in all experiments was Support Vector Machine (SVM) with a linear kernel [22], [23]. SVM was selected because it is known to perform well in text classification tasks involving high-dimensional sparse features such as TF-IDF. The linear kernel was chosen because it is computationally efficient and generally suitable for document classification problems.

For the baseline and all resampling-based scenarios, the same linear SVM configuration was used. In the cost-sensitive scenario, the SVM classifier was configured with class_weight='balanced', while all other settings remained unchanged. By keeping the classifier configuration consistent across scenarios, the comparison focused specifically on the effect of the imbalance handling strategy rather than on differences in classifier tuning.

Performance Evaluation

Model performance was evaluated on the testing set using several metrics: accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, and mean absolute error (MAE).

- Accuracy measures the proportion of correct predictions among all predictions.
- Macro precision calculates precision independently for each class and then averages the results equally across classes.
- Macro recall measures the average ability of the model to correctly identify samples from each class.
- Macro F1-score provides the harmonic mean of macro precision and macro recall, making it particularly useful for imbalanced datasets.
- Weighted F1-score accounts for class frequency when averaging F1 values.
- Mean Absolute Error (MAE) measures the average absolute difference between the predicted class and the true class.

The inclusion of MAE was important because the target labels represent ordered quality levels from 0 to 3. Therefore, classification errors should not only be judged as right or wrong, but also based on how far the predicted quality level deviates from the actual one. In addition to these numerical metrics, confusion matrices were used to analyze class-level prediction patterns and identify which classes were most frequently confused.

3. Result and Discussion

Experimental Results

This section presents the comparative results of seven experimental scenarios for concept map proposition quality classification using TF-IDF and Support Vector Machine (SVM). The compared methods consisted of a baseline model without balancing, four resampling methods (Random OverSampler, SMOTE, ADASYN, SMOTE-Tomek, and SMOTE-ENN), and one cost-sensitive approach using SVM with class_weight='balanced'. The evaluation was performed on the same testing set using accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, and mean absolute error (MAE).

Table 1. Data Crawling Results

Method	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted F1	MAE
SVM_ClassWeight_Balanced	**0.8417**	**0.7098**	**0.732**	**0.7149**	**0.8455**	0.2158
ADASYN	0.8345	0.7057	0.7285	0.7128	0.8369	0.223
RandomOverSampler	0.8345	0.7155	0.7081	0.7042	0.8359	0.223
Baseline_No_Balance	0.8345	0.7433	0.6164	0.6582	0.8162	0.295
SMOTE_ENN	0.8345	0.661	0.6577	0.6508	0.842	0.2374
SMOTE_Tomek	0.8345	0.6649	0.6513	0.6501	0.8333	**0.2158**
SMOTE	0.8273	0.656	0.6485	0.6422	0.8272	**0.2158**

Based on **Table 1**, the best-performing method was SVM with class_weight='balanced', which achieved the highest values in accuracy (0.8417), macro precision (0.7098), macro recall (0.7320), macro F1-score (0.7149), and weighted F1-score (0.8455). Although SMOTE and SMOTE-Tomek achieved the same MAE value as the best model, their macro-level performance remained lower. These results indicate that the cost-sensitive strategy was more effective and more stable than the evaluated resampling methods in this task.

Among the resampling methods, ADASYN produced the second-best performance, followed by Random OverSampler. In contrast, SMOTE, SMOTE-Tomek, and SMOTE-ENN did not outperform the cost-sensitive

approach, suggesting that synthetic sample generation and hybrid cleaning methods were less effective for this sparse text classification setting.

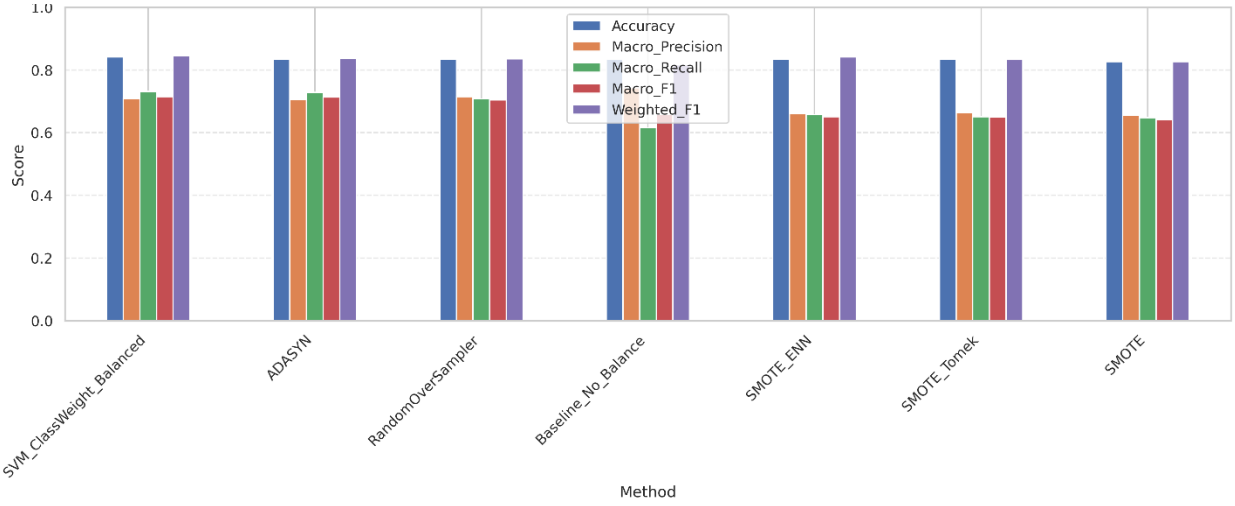

Figure 2. Comparison of evaluation metrics across methods

Figure 2 shows that the performance differences among methods are more clearly visible when macro-level metrics are considered. Although several methods produced similar accuracy values, their macro recall and macro F1-score differed considerably. This confirms that accuracy alone is not sufficient for evaluating imbalanced classification problems. A model may achieve competitive accuracy by correctly predicting the majority class, while still performing poorly on minority classes.

The figure also shows that class_weight='balanced' consistently remained among the top values across all major metrics. ADASYN and Random OverSampler also improved macro-level performance compared with the baseline, indicating that handling imbalance is indeed beneficial. However, the most stable overall profile was achieved by the class-weighted SVM.

Analysis of Class-Level Performance

To better understand the behavior of the models, class-level metrics were analyzed. Since the main issue in this study is class imbalance, special attention should be given to the minority classes, especially class 0 and class 1.

Table 2. Sentiment

Method	Class	Recall	F1-score
Baseline_No_Balance	0	0.125	0.1818
Baseline_No_Balance	1	0.7333	0.8148
Baseline_No_Balance	2	0.6296	0.7391
Baseline_No_Balance	3	**0.9775**	0.8969
RandomOverSampler	0	0.5	0.5
RandomOverSampler	1	0.7333	0.6286
RandomOverSampler	2	0.6667	0.766
RandomOverSampler	3	0.9326	0.9222
ADASYN	0	0.5	0.5
ADASYN	1	**0.8**	**0.6857**
ADASYN	2	**0.7037**	0.7451
ADASYN	3	0.9101	0.9205
SVM_ClassWeight_Balanced	0	**0.625**	**0.5263**

Method	Class	Recall	F1-score
SVM_ClassWeight_Balanced	1	0.6667	0.625
SVM_ClassWeight_Balanced	2	0.7037	**0.7755**
SVM_ClassWeight_Balanced	3	0.9326	**0.9326**

Table 2 demonstrates that the baseline model strongly favored the majority class. This is evident from the very high recall for class 3 (0.9775), while recall for class 0 was only 0.1250. In other words, the baseline model correctly recognized almost all majority-class samples but failed to identify most minority-class samples.

After applying imbalance handling, minority-class performance improved substantially. Both Random OverSampler and ADASYN increased class 0 recall from 0.1250 to 0.5000. However, the best result for class 0 was obtained by SVM with class_weight='balanced', which achieved a recall of 0.6250 and an F1-score of 0.5263. This indicates that the cost-sensitive approach was more effective in improving minority-class recognition without sacrificing the performance of other classes too severely.

For class 1 and class 2, ADASYN and class-weighted SVM also showed better balance than the baseline. Class 3 remained the easiest class to classify across all methods, which is expected given its dominance in the dataset. Nevertheless, the best approach was not the one with the highest class 3 recall, but rather the one that achieved a more balanced performance across all classes.

Confusion Matrix

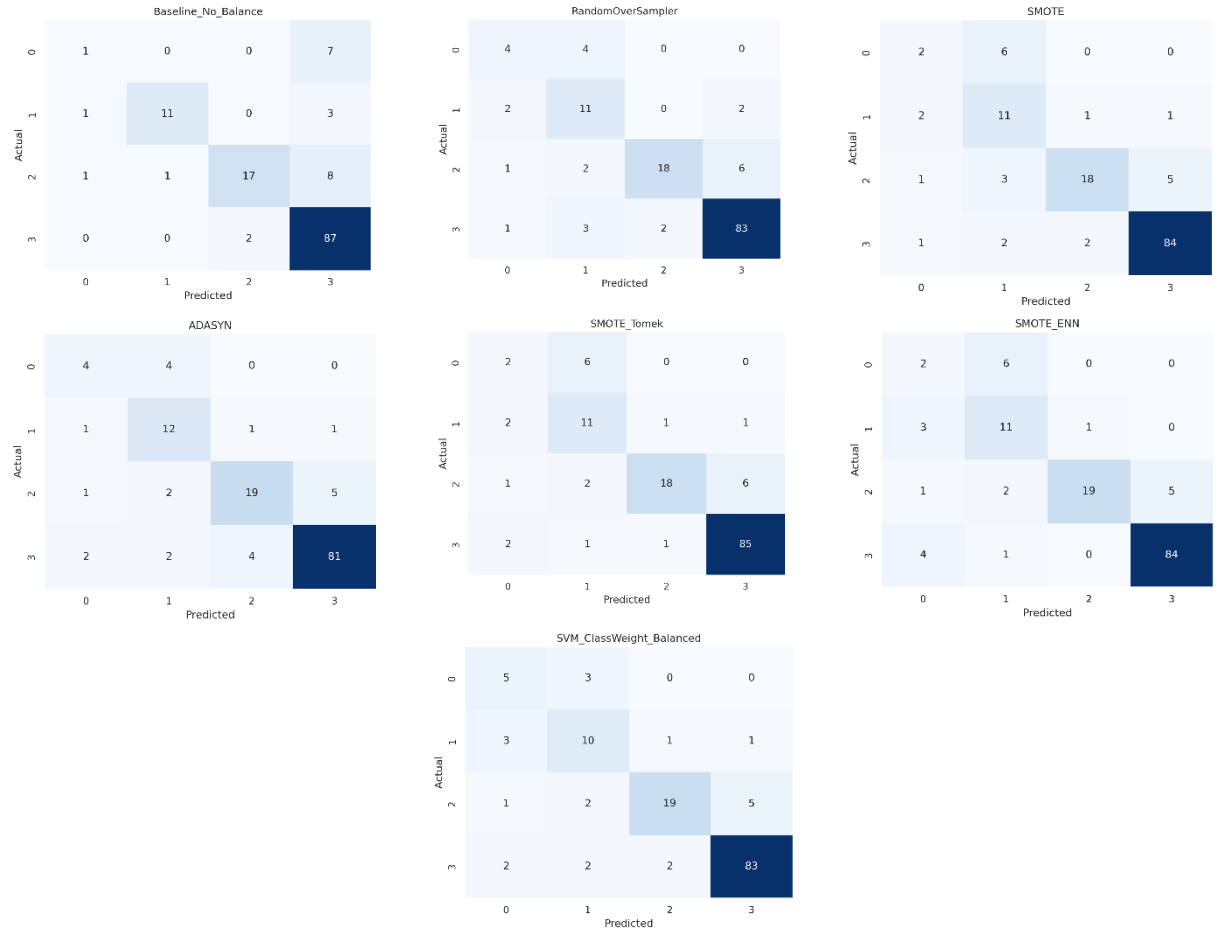

Figure 3. Confusion Matrices of All Compared Methods

Figure 3 provides a clearer view of prediction patterns for each method. The baseline confusion matrix reveals that many samples from class 0 were misclassified as class 3. This finding confirms that the baseline model was highly biased toward the majority class. Such a pattern is common in imbalanced datasets, where the classifier learns a decision boundary that is dominated by the most frequent class.

The confusion matrices for Random OverSampler and ADASYN show a noticeable improvement in class 0 recognition. More minority-class samples were predicted correctly, indicating that these methods helped the model become more sensitive to underrepresented classes. However, some trade-off remained, as a small reduction in majority-class dominance could also be observed.

The confusion matrix of SVM with class_weight='balanced' shows the best balance overall. In this method, class 0 was predicted correctly more often than in other methods, while classes 2 and 3 still maintained strong recognition performance. This supports the quantitative results in Table 1 and Table 2, where the class-weighted SVM achieved the best macro-level scores.

Discussion

The results demonstrate that imbalance handling has a substantial effect on concept map proposition quality classification. The baseline model produced a relatively competitive accuracy value, but this was largely due to its strong bias toward class 3, the majority class. Such a model may appear acceptable when evaluated only by accuracy, but it is not suitable for educational assessment tasks where minority classes are pedagogically important. In this study, minority classes represent lower-frequency proposition quality levels that should be identified reliably to support fairer and more meaningful evaluation.

Among the evaluated methods, the best results were obtained by SVM with class_weight='balanced'. This finding is important because it shows that directly incorporating class sensitivity into the learning process can be more effective than modifying the training data distribution. In text classification tasks using TF-IDF, the feature space is sparse and high-dimensional. Under such conditions, synthetic sample generation methods such as SMOTE and ADASYN may not always produce semantically meaningful new samples. As a result, the benefits of resampling may be limited, especially when the original text instances are short and highly varied.

Although ADASYN and Random OverSampler also improved minority-class performance, they remained slightly below the class-weighted SVM in macro-level metrics. ADASYN was particularly competitive, suggesting that adaptive oversampling can still be useful when the goal is to improve minority-class recognition. However, the results indicate that a cost-sensitive approach is more reliable in this dataset.

Interestingly, hybrid methods such as SMOTE-Tomek and SMOTE-ENN did not achieve the best results, even though these methods are often expected to perform well by combining oversampling and data cleaning. One possible explanation is that the generated synthetic vectors and cleaning steps may not align well with the nature of sparse textual data. In this setting, class weighting appears to preserve the original data structure more effectively while still improving the sensitivity of the classifier to minority classes.

Another important observation is the role of MAE. Since the labels represent ordered quality levels from 0 to 3, MAE provides additional insight beyond standard classification metrics. The best method achieved a relatively low MAE while also producing the highest macro recall and macro F1-score. This suggests that the predicted labels were not only more balanced across classes, but also closer to the true quality levels in ordinal terms.

Overall, the findings suggest that imbalance handling should be considered an essential part of automated concept map proposition assessment. More specifically, for sparse text data represented with TF-IDF and classified using SVM, class-weighted learning may be a more suitable strategy than many commonly used resampling techniques. This result provides a practical recommendation for future research and system development in educational text classification.

4. Conclusion

This study investigated the effect of several imbalance handling techniques on concept map proposition quality classification using TF-IDF features and Support Vector Machine (SVM). The comparative results confirmed that class imbalance significantly affects classification performance, particularly for minority classes. Although the baseline model achieved competitive accuracy, its performance was heavily biased toward the majority class, leading to weak recognition of minority-class propositions.

Among the evaluated methods, the best results were obtained by the cost-sensitive SVM with class_weight='balanced'. This method achieved the highest accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score, while also maintaining a low MAE. These findings indicate that directly incorporating class sensitivity into the learning process was more effective than the evaluated resampling methods for this dataset. In particular, the class-weighted approach provided the best balance between overall performance and fairness across classes.

The results also showed that ADASYN and Random OverSampler were the most competitive resampling-based methods, as both improved minority-class recognition compared with the baseline model. However, SMOTE, SMOTE-Tomek, and SMOTE-ENN did not outperform the class-weighted SVM. This suggests that for sparse text data represented with TF-IDF, synthetic resampling does not always provide the most suitable solution, likely because generated samples may not fully preserve meaningful textual variation.

Overall, this study demonstrates that imbalance handling is essential in automated concept map proposition assessment, and that cost-sensitive learning can be a more effective alternative than commonly used resampling techniques in this setting. The findings contribute practical guidance for developing fairer and more reliable educational text classification systems.

For future work, several directions can be explored. First, the experiments can be extended to other classifiers in order to examine whether the same pattern holds beyond SVM. Second, more advanced text representations such as word embeddings or contextual embeddings can be investigated. Third, cross-validation and statistical significance testing can be applied to strengthen the robustness of the results. These extensions may provide a deeper understanding of imbalance handling strategies in concept map quality classification and educational text mining more broadly.

References:

[1] 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: https://doi.org/10.1186/s41039-022-00189-9.

[2] D. D. Prasetya, T. Widiyaningtyas, and T. Hirashima, “Interrelatedness patterns of knowledge representation in extension concept mapping,” *Res. Pract. Technol. Enhanc. Learn.*, vol. 20, p. 009, May 2024, doi: https://doi.org/10.58459/rptel.2025.20009.

[3] T. Evans and I. Jeong, “Concept maps as assessment for learning in university mathematics,” *Educ. Stud. Math.*, vol. 113, no. 3, pp. 475–498, Jul. 2023, doi: https://doi.org/10.1007/s10649-023-10209-0.

[4] C. Cischke and S. T. Mueller, “Concept Mapping Assessments as a Tool for Judgment of Learning.” Jul. 29, 2022, doi: https://doi.org/10.31234/osf.io/69bjx.

[5] N. Rismayanti, D. D. Prasetya, T. Widiyaningtyas, and T. Hirashima, “Comparative Study of Random Forest and Ordinal Regression in Concept Map Quality Assessment: The Role of TF-IDF, BERT, and SMOTE-based Balancing,” *Ilk. J. Ilm.*, vol. 17, no. 3, pp. 336–345, 2025, doi: http://dx.doi.org/10.33096/ilkom.v17i3.2906.336-345.

[6] M. S. Mohosheu, M. A. al Noman, A. Newaz, Al-Amin, and T. Jabid, “A Comprehensive Evaluation of Sampling Techniques in Addressing Class Imbalance Across Diverse Datasets,” *2024 6th Int. Conf. Electr. Eng. Inf. Commun. Technol.*, 2024, doi: https://doi.org/10.1109/iceeict62016.2024.10534464.

[7] N. Habbat, “Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning,” *Eng. Appl. Artif. Intell.*, vol. 126, 2023, doi: https://doi.org/10.1016/j.engappai.2023.106999.

[8] F. De Engenharia, “Measuring the performance of ordinal classification,” *Int. J. Pattern Recognit. Arti¯cial*

Intell., vol. 25, no. 8, pp. 1173–1195, 2011, doi: https://doi.org/10.1142/S0218001411009093.

[9] D. Yilmaz Eroglu and M. S. Pir, "Hybrid Oversampling and Undersampling Method (HOUM) via Safe-Level SMOTE and Support Vector Machine," *Appl. Sci.*, vol. 14, no. 22, p. 10438, Nov. 2024, doi: https://doi.org/10.3390/app142210438.

[10] A. Tuppad and S. D. Patil, "Data Pre-processing Issues in Medical Data Classification," *2023 Int. Conf. ...*, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10275855/.

[11] J. Zhao, K. S. Chong, W. Shu, and ..., "A Data Pre-Processing Module for Improved-Accuracy Machine-Learning-based Micro-Single-Event-Latchup Detection," *2023 IEEE 9th Int. ...*, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10207447/.

[12] G. Ketepalli and P. Bulla, "Data Preparation and Pre-processing of Intrusion Detection Datasets using Machine Learning," *2023 Int. Conf. ...*, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10134025/.

[13] K. K. Agustiningsih, E. Utami, and M. A. Alsyaibani, "Sentiment Analysis of COVID-19 Vaccines in Indonesia on Twitter Using Pre-Trained and Self-Training Word Embeddings," *J. Ilmu Komput. dan Inf.*, vol. 15, no. 1, pp. 39–46, 2022, doi: https://doi.org/10.21609/jiki.v15i1.1044.

[14] A. Al Tawil, "Comparative Analysis of Machine Learning Algorithms for Email Phishing Detection Using TF-IDF, Word2Vec, and BERT," *Comput. Mater. Contin.*, vol. 81, no. 2, pp. 3395–3412, 2024, doi: https://doi.org/10.32604/cmc.2024.057279.

[15] 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, [Online]. Available: https://api.elsevier.com/content/abstract/scopus_id/85161154224.

[16] C. A. N. Agustina, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Computer Science*, vol. 234. pp. 156–163, 2024, doi: https://doi.org/10.1016/j.procs.2024.02.162.

[17] 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.

[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: https://doi.org/10.1109/ICAIC60265.2024.10433843.

[19] E. A. Aldhahri, A. A. Almazroi, M. H. Alkinani, N. Ayub, E. A. Alghamdi, and N. F. Janbi, "Smart Farming: Enhancing Urban Agriculture through Predictive Analytics and Resource Optimization," *IEEE Access*, vol. 13, no. October 2024, pp. 72375–72388, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3530006.

[20] N. T. Singh, "An Innovative URL-Based System Approach with ML Based Prevention," *Proceedings - International Conference on Computing, Power, and Communication Technologies, IC2PCT 2024*. pp. 1498–1503, 2024, doi: https://doi.org/10.1109/IC2PCT60090.2024.10486712.

[21] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, p. 37, Jan. 2025, doi: https://doi.org/10.3390/a18010037.

[22] X. Ning, "Classification of Sulfadimidine and Sulfapyridine in Duck Meat by Surface Enhanced Raman Spectroscopy Combined with Principal Component Analysis and Support Vector Machine," *Anal. Lett.*, vol. 53, no. 10, pp. 1513–1524, 2020, doi: https://doi.org/10.1080/00032719.2019.1710524.

[23] O. Hamidi, "Analysis of the Response of Urban Water Consumption to Climatic Variables: Case Study of Khorramabad City in Iran," *Adv. Meteorol.*, vol. 2021, 2021, doi: https://doi.org/10.1155/2021/6615152.