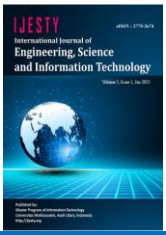
International Journal of Engineering, Science and Information Technology
Volume 5 No. 1 (2025) pp. 258-263
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i1.707
Research Paper, Short Communication, Review, Technical Paper



Optimization of Feature Selection on Student Complaint Data Using Recursive Feature Elimination to Improve Academic Service Quality

Hamzah Setiawan*, Arief Senja Fitrani, Yunianita Rahmawati, Cakra Wirabumi Putra, Firdausi Usqi Salsabila

Department of Informatics, Faculty of Science and Technology, Universitas Muhammadiyah Sidoarjo, Indonesia

**Corresponding author Email: hamzah@umsida.ac.id*

The manuscript was received on 1 March 2024, revised on 15 April 2024, and accepted on 1 January 2025, date of publication 9 January 2025

Abstract

This study aims to optimize the feature selection process on student complaint data regarding academic services in universities using the Recursive Feature Elimination (RFE) method. Student complaints' diverse and complex nature requires in-depth analysis to identify crucial features that affect service satisfaction. An accurate feature selection process can help universities understand the most frequently reported issues, enabling them to respond and improve services more effectively. The research utilizes complaint data from various aspects of academic services, such as administration, facilities, and faculty interactions. After preprocessing the data to remove noise and irrelevant entries, RFE is applied to select the most relevant features. Subsequently, a classification model is built using the selected features to identify the most significant complaint patterns. Model evaluation is conducted through cross-validation techniques to ensure accuracy and reliability, with metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that the RFE method significantly enhances model performance in selecting essential features, making the classification model more efficient and accurate in predicting student complaints. Thus, this study contributes significantly to assisting universities in enhancing the quality of academic services through a more targeted analysis of student complaints. Implementing this method will improve the complaint-handling process and increase overall student satisfaction.

***Keywords**: Academic Services, Classification Model, Feature Selection, Recursive Feature Elimination, Student Complaints.*

1. Introduction

In higher education, efficiently managing student complaints is crucial for maintaining and enhancing the quality of academic services [1]. These complaints serve as a vital feedback mechanism highlighting areas needing improvement within educational institutions. However, complaints' diverse and complex nature poses significant challenges in data management and analysis [2]. They encompass various issues, from administrative concerns to academic mismatches and infrastructure deficiencies. Thus, effectively addressing these concerns requires a sophisticated approach to understand and prioritize them, ensuring that the most urgent issues are addressed promptly [3]. Advanced data analysis techniques like Recursive Feature Elimination (RFE) offer a promising solution to these challenges. RFE is a feature selection method widely used in machine learning to enhance model performance by eliminating less essential features while retaining the most critical ones. By applying RFE to student complaint data, institutions can identify key factors often leading to dissatisfaction, allowing for more focused and efficient responses[4]. This study specifically focuses on optimizing the feature selection process using RFE on student complaint data to improve university academic service response mechanisms. By refining the data analysis process, this study aims to develop a robust model that predicts and effectively categorizes student complaints[5]. This enables educational institutions to tailor their problem-solving strategies more precisely, addressing the root causes of complaints and implementing strategic changes that enhance student satisfaction and academic outcomes [6].
Furthermore, applying the RFE method in managing student complaints necessitates integrating a comprehensive data system within universities [7]. This system must be capable of collecting, integrating, and analyzing data from various student complaint sources,

including online systems, direct communications, and satisfaction surveys. By consolidating all this data into a coherent platform, universities can more easily identify trends and patterns commonly emerging in complaints. This not only enhances the effectiveness of the analysis but also enables institutions to be proactive in addressing issues before they escalate into more serious problems[8]. This approach aims to create a more responsive academic environment that meets students' needs [9]. Integrating and using advanced technology in complaint management also opens opportunities for universities to innovate how they interact and communicate with students [10]. For example, developing a mobile application that allows students to submit complaints and monitor the status of their complaints in real time could be one innovative solution. This provides transparency and strengthens students' trust in a fair and open complaint-handling system. Thus, universities can build better relationships with students and gradually improve the quality of academic services and the institution's overall image. Ultimately, this study seeks to demonstrate how leveraging advanced data analysis techniques can transform student complaint management into an opportunity for quality improvement in academic services, thereby enhancing student satisfaction and the institution's overall reputation [11].

2. Method

The research methodology for this study comprises several carefully structured stages to optimize feature selection on student service complaint data using the Recursive Feature Elimination (RFE) method, thereby enhancing the quality of academic services. Here is an expanded description of each stage:

1. **Data Collection and Preprocessing**: Initially, the study involves systematically collecting student complaint data from diverse sources such as online complaint management systems, email communications, and student satisfaction surveys. Following collection, the data undergoes rigorous preprocessing to eliminate noise and incomplete entries, ensuring the data is clean and formatted uniformly for subsequent analysis[3].
2. **Exploratory Data Analysis**: Before applying the feature selection method, exploratory data analysis (EDA) is conducted to understand the data's underlying structure and characteristics thoroughly. This stage involves performing descriptive statistical analysis, visualizing data trends and distributions, and pinpointing preliminary features that appear to be relevant to the students' complaints[12].
3. **Application of Recursive Feature Elimination (RFE)**: This crucial stage involves implementing the RFE method to systematically eliminate less impactful features while retaining those most vital for model accuracy. The process iteratively evaluates each potential feature subset by considering metrics such as the accuracy and complexity of the resulting classification models [13].
4. **Model Evaluation**: After determining the optimal set of features, a classification model is constructed and rigorously evaluated using cross-validation techniques. This step is essential to verify that the model is robust, performs consistently across different subsets of the data, and avoids overfitting. Key performance metrics such as accuracy, precision, recall, and the F1-score are utilized to assess the model's efficacy [14].
5. **Validation and Implementation**: Upon successful evaluation, the model is implemented in a real-world setting to validate its effectiveness in addressing and predicting student complaints. This real-world testing provides critical feedback to refine the model further and ensure its practical applicability in enhancing academic service quality.
6. **Documentation and Dissemination**: Finally, all findings, methodologies, and insights gained from the study are meticulously documented. The comprehensive documentation covers detailed methodologies, analysis results, practical implications, and recommendations for future applications. This documentation is aimed at scholarly publication in academic journals and sharing with stakeholders who could benefit from the improved complaint management strategies [15].
7. These methodological stages are designed to achieve the research objectives and contribute significantly to the field by providing a robust model for effectively managing student complaints in educational settings.

These methodological stages are designed to achieve the research objectives and contribute significantly to the field by providing a robust model for effectively managing student complaints in educational settings.

0	SUGGESTION	SU	DPAL	DSTI	DA	DPMBKA	DK	LIBRARY
1	Facilities are less supportive	1	1	0	0	0	0	0
2	wifi is not smooth	0	0	1	0	0	0	0
3	if students forget or	1	1	0	0	0	0	0
4	please fix the wifi again	0	0	1	0	0	0	0
5	campus parking	1	1	0	0	0	0	0
6	he parking lot must be arranged properly and neatly	0	1	0	0	0	0	0
7	additional facilities and infrastructure for lifts	0	1	0	0	0	0	0
8	inadequate wifi quality	0	0	1	0	0	0	0
9	wifi can't connect	0	0	1	0	0	0	0
10	Please improve campus facilities, especially from	0	0	1	0	0	0	0
11	he wifi is slow	0	1	1	0	0	0	0
12	service	1	0	0	0	1	0	0
13	library	0	0	0	0	0	0	1
14	wifi is slow	0	0	1	0	0	0	0
15	the wifi is often unstable	0	0	1	0	0	0	0
16	the air conditioner is not cold	0	1	0	0	0	0	0
17	the book library can be expanded further	0	0	0	0	0	0	1
18	service improved	1	0	0	0	1	0	0
19	the wifi is not smooth enough to use	0	0	1	0	0	0	0
20	Pusdakom	0	0	1	0	0	0	0

Fig 1. Data Pre-Processing

3. Results and Discussion

3.1. Data Preprocessing

The first step in analyzing student complaint data involves collecting information from various sources, such as online complaint management systems, emails, and student satisfaction surveys, accumulating 3,323 entries. These sources provide diverse information related to the issues students face, making it crucial to ensure that the data collected is complete and relevant to the analysis topic. Once the data is gathered, the next stage is preprocessing [16]. This process aims to clean the data from noise, such as irrelevant or duplicate information, and to address incomplete entries [17]. For example, some entries may lack sufficient information for use, such as complaints without detailed explanations or from anonymous senders. Incomplete entries will be removed or supplemented with imputation methods if possible. Next, the available data will be normalized to ensure a consistent format [18]. This includes converting date formats to a uniform standard, rewriting text in lowercase, and removing unnecessary punctuation or symbols. This process facilitates further analysis and ensures that the model used adequately processes the data. Once the data is clean and structured, it is ready for the following stages of analysis, such as complaint classification, identifying common patterns, and modeling to find effective solutions. The preprocessing step is vital to ensure data quality and to enhance the accuracy of the analysis results.

```
# Download NLTK stop words (do this only once)
nltk.download('stopwords')
nltk.download('punkt')

# Load stop words in Indonesian
stop_words = set(stopwords.words('indonesian'))

# Function to remove stop words
def remove_stop_words(text):
    word_tokens = word_tokenize(text)
    filtered_text = [word for word in word_tokens if word not in stop_words]
    return ' '.join(filtered_text)

# Apply the function to remove stop words from the 'saran' column
df['saran'] = df['saran'].apply(remove_stop_words)

# Display the first few entries to verify the changes
print(df['saran'].head(22))
```

Fig 2. Code Pre-Processing

3.2. Exploratory Data Analysis

The descriptive analysis found that the most dominant label in this dataset is the 'dsti' category, with 967 occurrences. This indicates that most student complaints are related to the DSTI (Directorate of System and Information Technology) unit. Meanwhile, labels such as 'dpal', 'su', and 'library' also have significantly high frequencies. Subsequently, performance metrics of the classification, such as precision, recall, and F1-score for each label, were analyzed. The highest precision was found in the 'dsti' label with a value of 0.96, indicating that the model effectively identifies complaints in this category. However, the recall for some categories, such as 'SU' and 'DPAL', is lower, meaning that the model does not always identify all complaints within these categories. Labels like 'DPMBKA' and 'DK' show very low performance, with precision, recall, and F1-scores approaching zero, indicating difficulties in classifying infrequent complaints. The macro-average metrics suggest that, in general, the model performs exceptionally well in handling data imbalances with a weighted average F1-score of 0.65. However, there are significant challenges in identifying some categories with low complaint frequencies. These analysis results underscore the importance of conducting further feature selection and adjustments before implementing more complex models like RFE to improve classification performance [19].

```
# Menghitung jumlah kemunculan untuk setiap label dalam array y_train atau y_test
label_counts = np.sum(y_train, axis=0)  # Untuk dataset multi-label, misalnya

# Calculating dominant label
dominant_label_index = np.argmax(label_counts)
dominant_label_name = df.columns[dominant_label_index + 1]  # Adjust index based on your DataFrame structure
dominant_label_count = label_counts[dominant_label_index]  # Menyimpan jumlah kejadian label yang dominan

print(f"Label yang paling dominan adalah '{dominant_label_name}' dengan {dominant_label_count} kejadian.")
```

```
Index(['saran', 'su', 'dpal', 'dsti', 'da', 'dpmbka', 'dk', 'perpustakaan'], dtype='object')
              precision    recall  f1-score   support

          SU       0.72      0.26      0.38       113
        DPAL       0.86      0.52      0.65       194
        DSTI       0.96      0.81      0.88       238
          DA       1.00      0.30      0.46        20
      DPMBKA       0.00      0.00      0.00         6
          DK       0.00      0.00      0.00        11
Perpustakaan       1.00      0.29      0.45        51

   micro avg       0.91      0.54      0.68       633
   macro avg       0.65      0.31      0.40       633
weighted avg       0.87      0.54      0.65       633
 samples avg       0.48      0.47      0.47       633

Hamming Loss: 0.0696
Label yang paling dominan adalah 'dsti' dengan 967 kejadian.
```

Fig 3. Results of Exploratory Data Analysis

3.3. Implementation of Recursive Feature Elimination (RFE)

Before applying the Recursive Feature Elimination (RFE) method, an exploratory analysis was conducted to understand the distribution, patterns, and essential characteristics of the complaint data. The first step was descriptive statistical analysis, including features such as 'saran', 'su', 'dpal', 'dsti', 'da', 'dpmbka', 'dk', and 'library'. This analysis provided an overview of the occurrence and frequency of each complaint category and helped identify which features were relevant for further study [14]. At this stage, the Recursive Feature Elimination (RFE) method was implemented to perform automatic feature selection. RFE selects the most relevant and impactful subset of features on model performance. Each particle in the swarm represents a potential solution, a subset of the available features. Evaluation criteria such as classification accuracy and model complexity were used for assessment.

The first step was to determine the feature names used. The code feature_names = df.columns.tolist() ensures that all column names from the dataset are stored in the feature_names variable for easy use in feature selection [20]. Next, the RandomForestClassifier was used as the base model for feature selection. RFE with the Random Forest model was configured to select the top 10 features based on classification performance iteratively. A pipeline was used to integrate the feature selection and classification processes into a single unit for the evaluation process. The data was then split into training and testing sets using train_test_split with an 80% training and 20% testing ratio[21]. After the model was trained with the training data, predictions were made on the test data, and the evaluation results were displayed using metrics such as precision, recall, F1-score, and accuracy. From the evaluation results, it was evident that the model had an accuracy of 85%. However, there was a performance discrepancy between the two tested classes. Class 0 had a precision of 0.87 and a recall of 0.96, indicating that the model was very good at detecting this class.

On the other hand, Class 1 had a lower precision of 0.61 and a recall of 0.29, indicating that the model struggled to identify complaints in this class. RFE successfully selected the top 10 features, but only eight were valid upon extraction, one of which was the 'library' feature. This process ensures that only the most relevant features are used in further modelling, enhancing performance and reducing model complexity.

3.4. Evaluation Model

After performing feature selection using the Recursive Feature Elimination (RFE) method, the next step is to build a classification model using the selected features. To ensure that the model built has stable performance and does not suffer from overfitting the cross-validation (CV) technique is employed for evaluation. This technique divides the data into several subsets (folds), then trains the model on each subset and evaluates it on another subset, thus producing more comprehensive evaluation results. The results from the evaluation using 5-fold cross-validation show that the model has an average accuracy of 86.34% with a slight standard deviation of about +/- 2%. This indicates that the model performs quite well overall and can handle data variations without being overly affected by overfitting. Next, parameter optimization is conducted using the GridSearchCV technique to find the best parameter combination for the Random Forest model. This process involves searching for optimal parameters from several aspects, such as the number of estimators (n_estimators), maximum depth of the decision tree (max_depth), and node splitting criteria (criterion). The tested parameters include n_estimators (100, 200, 300), max_features ('auto', 'sqrt', 'log2'), and max_depth (4 to 8). This helps enhance model accuracy by determining the best configuration based on training data. After the GridSearchCV process is completed, it is found that the best parameter combination for the Random Forest model is using the 'entropy' criterion, max_depth of 5, max_features 'auto', and n_estimators of 100. This combination produces the highest accuracy of 86.34% on the cross-validation data, indicating improved performance compared to the initial model. This model demonstrates good and stable performance, with consistent evaluation results. Cross-validation ensures that the model is not overfitting and can handle data variations. At the same time, parameter optimization through GridSearchCV helps find the most optimal parameter combination to enhance model performance.

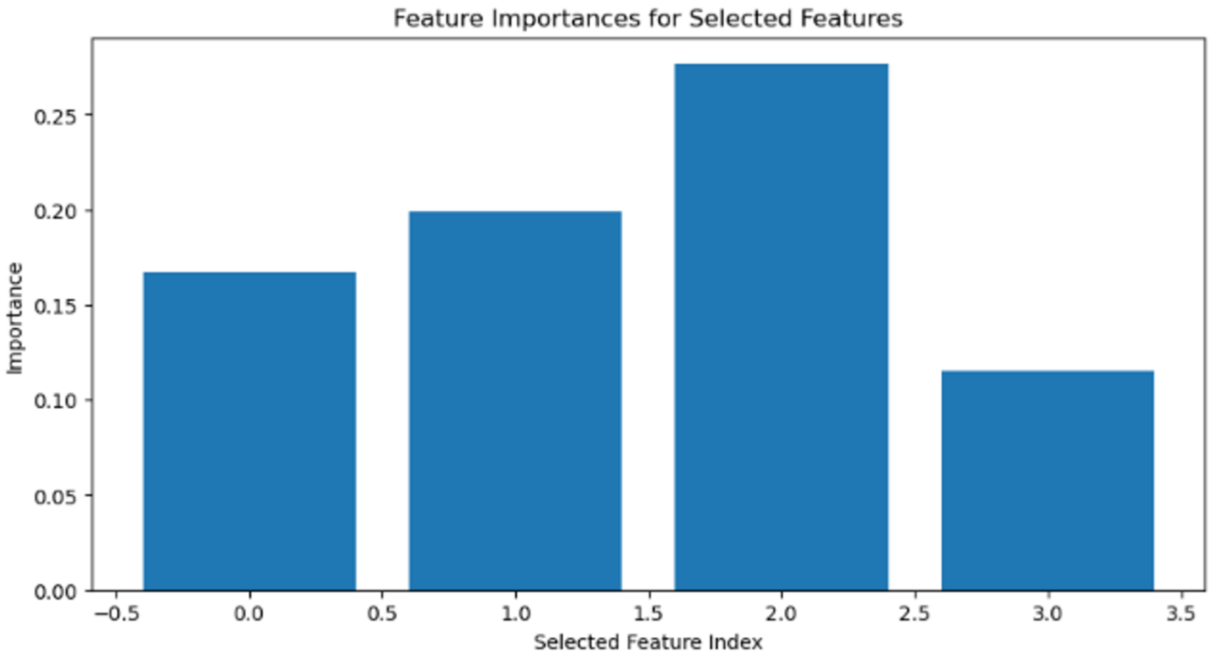


Fig 4. Selected Feature index

3.5. Validation and Implementation

After the model was developed using the optimal parameters from GridSearchCV, the next step was to test the model's performance on a test set that the model had not seen before. This optimized model uses RandomForestClassifier with the best parameters: 'entropy' criterion, max_depth of 5, max_features 'auto', and n_estimators of 100. The model is trained using a subset of previously selected training data, with features selected through the Recursive Feature Elimination (RFE) process. After that, predictions are made on the test data to measure the performance of the optimized model. The model evaluation based on the classification report shows that the model has an overall accuracy of 85% on the test data, meaning that 85% of the predictions are correct. The model's performance is further analyzed based on precision, recall, and F1-score metrics for each class (0 and 1). For class 0, the model shows excellent performance, with a precision of 0.86, meaning that 86% of the class 0 predictions are correct.

Additionally, the model has a recall of 0.98, meaning that the model correctly identified 98% of all actual class 0 examples. The F1-score for class 0 is 0.92, indicating a high balance between precision and recall. However, the performance for class 1 is lower. The precision for class 1 is 0.69, meaning that 69% of the class 1 predictions are correct, but the recall for this class is relatively low, only 0.22. This means the model only successfully identified 22% of all actual class 1 examples. The F1-score for class 1 is also relatively low, at 0.34, indicating that the model still struggles to detect and correctly predict for this class. Overall, the optimized model performs well for class 0 but shows weaknesses in detecting class 1. To address this, further adjustments to the model, such as handling imbalanced data or using sampling techniques, may be needed to provide more balanced results across both classes.

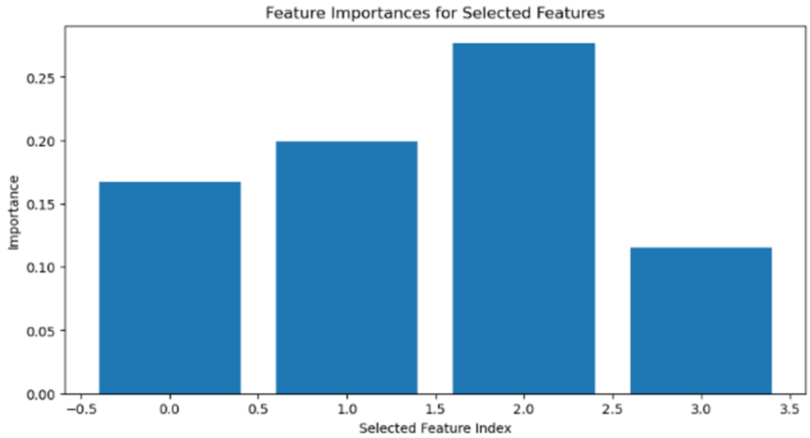


Fig 4. Selected Feature index

4. Conclusion

This research successfully optimized feature selection on student complaint data using the Recursive Feature Elimination (RFE) method to enhance the quality of academic services in higher education institutions. By implementing RFE, this study identified critical features that significantly influence the classification model of student complaints. The evaluation results show that the optimized model performs well, with an accuracy of approximately 86.34%, indicating the method's effectiveness in managing and accurately predicting student complaints. The optimized model demonstrated high accuracy in the dominant class but still required adjustments for courses with low frequency, highlighting the importance of handling imbalanced data. This study provides significant insights that proper feature selection and meticulous model validation can substantially improve prediction accuracy in academic services. Implementing this model is expected to enable higher education institutions to respond to and manage student complaints more effectively, thereby enhancing student satisfaction and the quality of educational services. Further exploration into the study reveals that the successful application of RFE refined the feature set and reduced the model's complexity, making it more efficient in real-time environments where quick decision-making is essential.

The precision of the model in identifying and categorizing complaints could aid institutions in pinpointing specific areas of concern, such as administrative processes, faculty interaction, or facility issues, which are common categories of student grievances. Moreover, the model's ability to distinguish between various types of complaints with high accuracy can lead to more personalized responses from the academic institution, fostering a more supportive and responsive educational environment. For instance, accurately identifying complaints related to IT services or library resources could prompt targeted improvements in these areas, directly addressing the students' issues and potentially preventing future complaints. The research also underscores the necessity for continuous model training and updates to adapt to new patterns in student behaviour and evolving academic environments. As student demographics and educational technologies change, so should the analytical models institutions rely on to monitor and respond to student needs. The study further suggests the potential for integrating these predictive models with student information systems and academic service platforms to create a cohesive framework for complaint management. This integration could streamline processes, reduce response times, and increase the efficiency of handling student inquiries and complaints, leading to a more agile and student-centred approach in higher education institutions. In addition to improving complaint management, the findings from this research could be applied to other areas of academic services, such as course evaluation and feedback on teaching methods. Using similar feature selection and machine learning techniques, educational institutions can enhance different aspects of academic quality, leading to a more comprehensive and practical educational experience for students.

In conclusion, implementing advanced data analysis techniques like RFE in educational settings represents a significant step forward in utilizing big data and machine learning to enhance the quality of academic services. This study demonstrates the practical applications of such technologies in real-world settings and sets a precedent for future research in educational quality improvement. It serves as a call to action for academic institutions to embrace these technologies to serve their student populations better and foster continuous improvement in the educational sector.

Acknowledgement

I want to express my deepest gratitude and thanks to Universitas Muhammadiyah Sidoarjo for the support provided through the internal grant that has been awarded. This grant has been a key factor in the success of my research. With this grant's support, I have completed a study that not only advances scientific knowledge but also provides direct benefits to the academic community and the wider society. This success is inseparable from the vision of Universitas Muhammadiyah Sidoarjo, which is always committed to supporting innovative research and development. This research represents a concrete step in implementing the scholarly values that the university consistently instils, and I am incredibly honoured to be a part of this effort. I hope that the results of this research can make a significant contribution and open new opportunities for further studies that will continue to support the university's achievements and the development of science.

References

[1] H. Garg, "A Novel Approach for Feature Selection by Harmony Search Optimization: Applications in Big Data. Future Generation Computer Systems," vol. 107, pp. 941–955, 2020.

[2] P. Patel, H., & Prajapati, "A Comprehensive Review on Feature Selection Techniques and Their Applications," *IEEE Access*, vol. 9, pp. 37731-37757., 2021.

[3] M. A. Alomari, O. A., Khader, A. T., & Al-Betar, "Feature Selection Using Particle Swarm Optimization in Big Data: A Review of State-of-the-Art. Applied Soft Computing," vol. 106907, p. 99, 2021.

[4] P. Arora, S., & Anand, "Binary Particle Swarm Optimization for Feature Selection in High-Dimensional Datasets. Swarm and Evolutionary Computation," vol. 54, p. 100663, 2020.

[5] I. Osakwe, G. Chen, A. Whitelock-Wainwright, D. Gašević, A. Pinheiro Cavalcanti, and R. Ferreira Mello, "Towards automated content analysis of educational feedback: A multi-language study," *Comput. Educ. Artif. Intell.*, vol. 3, no. May 2021, 2022, doi: 10.1016/j.caeai.2022.100059.

[6] A. K. Sharma, A. Dhaka, A. Nandal, K. Swastik, and ..., "Big Data Analysis: Basic Review on Techniques,"*… Anal. Big Data …*, 2021, [Online]. Available: https://www.igi-global.com/chapter/big-data-analysis/272956

[7] W. Kim, "A Study on the Crime Prevention Smart System Based on Big Data Processing," *J. Korea Converg. Soc.*, 2020, [Online]. Available: https://www.koreascience.or.kr/article/JAKO202034352378871.page

[8] R. Goyena, "済無No Title No Title," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019, doi: 10.1017/CBO9781107415324.004.

[9] D. Haynes, P. Mitchell, and E. Shook, *Developing the Geospatial Big Data Benchmark: A Comparative Framework for Evaluating Raster Analysis on Big Data Platforms*. preprints.org, 2020. [Online]. Available: https://www.preprints.org/manuscript/202008.0504

[10] W. Alorainy, P. Burnap, H. Liu, M. Williams, and L. Giommoni, "Disrupting networks of hate: characterizing hateful networks and removing critical nodes," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, pp. 1–22, 2022, doi: 10.1007/s13278-021-00818-z.

[11] W. K. Chen, X., Xu, Y., & Wong, "Particle Swarm Optimization with XGBoost for Feature Selection and Classification. Pattern Recognition Letters," vol. 144, pp. 77–85, 2021.

[12] J. Lee, I. Yu, J. Park, and D. W. Kim, "Memetic feature selection for multilabel text categorization using label frequency difference," *Inf. Sci. (Ny).*, vol. 485, pp. 263–280, 2019, doi: 10.1016/j.ins.2019.02.021.

[13] C. Xu, C. Sun, G. Jiang, X. Chen, Q. He, and P. Xie, "Two-level multi-domain feature extraction on sparse representation for motor imagery classification," *Biomed. Signal Process. Control*, vol. 62, no. August 2019, p. 102160, 2020, doi: 10.1016/j.bspc.2020.102160.

[14] X. Wu, Z., Cai, Z., & Shao, "Feature Selection Using Particle Swarm Optimization with Phenotype-Based Local Search," *Inf. Sci. (Ny).*, vol. 507, pp. 16–32, 2020.

[15] S. F. Wamba, R. Dubey, A. Gunasekaran, and ..., "The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism," *Int. J. …*, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527319303184

[16] A. Sever, "Modeling Distributed Agile Software Development for Big Data Projects: Evolution in Process," *Curr. J. Appl. Sci. Technol.*, 2020, [Online]. Available: https://journalcjast.com/index.php/CJAST/article/view/30710

[17] E. Rendon, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and ..., "Data sampling methods to deal with the big data multi-class imbalance problem," *Appl. Sci.*, 2020, [Online]. Available: https://www.mdpi.com/2076-3417/10/4/1276

[18] W. C. S. IV and B. Krawczyk, "Multi-class imbalanced big data classification on Spark," *Knowledge-Based Syst.*, 2021, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705120307279

[19] P. Mansingh, B. K. Pattanayak, and B. Pati, "Bigdata Analytics and Deep Learning in Early Detection of Neurogenerative Diseases," *Intell. Cloud Comput.*, 2021, [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-6202-0_52

[20] N. Kumar, G. Akangire, B. Sullivan, K. Fairchild, and ..., "Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront," *Pediatric …*. nature.com, 2020. [Online]. Available: https://www.nature.com/articles/s41390-019-0527-0

[21] M. Abd Elaziz, H. M. Abu-Donia, R. A. Hosny, S. L. Hazae, and R. A. Ibrahim, "Improved evolutionary-based feature selection technique using extension of knowledge based on the rough approximations," *Inf. Sci. (Ny).*, vol. 594, pp. 76–94, 2022, doi: 10.1016/j.ins.2022.01.026.