E-ISSN : 2963-4946
Vol. 3 No. October 01, 2024
Asian Journal of Social and Humanities
https://ajosh.org/

Comparison of SVM, KNN, and Naïve Bayes Classification Methods in Predicting Student Transfers at BK Palu School

William Nugraha, Gerry Firmansyah, Agung Mulyo Widodo, Budi Tjahjono
Universitas Esa Unggul, Indonesia
E-mail: wiliamnugraha8@gmail.com, gerry@esaunggul.ac.id, agung.mulyo@esaunggul.ac.id, budi.tjahjono@esaunggul.ac.id

***Correspondence:** wiliamnugraha8@gmail.com

KEYWORDS
student transfers, classification methods, predictive modeling, BK palu school

ABSTRACT
Student transfers are a significant issue in schools and can affect the dynamics of education and student performance. This research aims to predict student transfers using a comparative analysis of three classification methods: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes. The study utilizes historical data from BK Palu School, covering the years 2022 to 2024, which includes demographic, academic, socio-economic, and student quality information. The methodology involves data collection, data preparation, algorithm selection, implementation, and evaluation of the three methods. The performance of the classification methods is assessed using metrics such as accuracy, precision, recall, and F1-score. The results indicate that SVM has the highest accuracy in predicting student transfers, followed by KNN and Naïve Bayes. This study contributes to identifying key factors influencing student transfers and offers schools a robust model to develop targeted strategies for reducing transfer rates. Ultimately, this research provides insights into optimizing student retention and improving the overall quality of education.

Introduction

Student transfers are a common issue in schools and can be caused by various factors such as academic problems, socio-economic conditions, and education quality. This phenomenon can affect the quality of education and school dynamics. According to the Central Statistics Agency, the student transfer rate in Indonesia increases every year. Data from the Department of Education shows that about 10% of students in Indonesia transfer schools annually, mainly due to academic and social environment issues (Wahyuni & Marbun, 2020).

To address this problem, an effective solution is needed to predict and tackle student transfer issues. This study aims to compare the performance of three classification methods: SVM, KNN, and Naïve Bayes in predicting student transfers at BK Palu School.

By understanding the strengths and weaknesses of each method, this study contributes to providing a more effective solution to reduce student transfer rates.
The research objectives are:

- Identifying the factors that influence student transfers.
- Comparing the performance of SVM, KNN, and Naïve Bayes methods in predicting student transfers.
- Providing recommendations based on the analysis results to reduce student transfer rates.

Previous research has discussed the use of classification methods in the educational context. According to (Kamila & Subastian, 2019) compared Naïve Bayes and KNN methods in detecting college dropouts with accuracies of 92% and 85%, respectively. According to (Budiyantara et al., 2020) examined KNN, Decision Tree, and Naïve Bayes methods to predict student graduation with the highest accuracy of 98.04% in Decision Tree. According to (Purba et al., 2022) used SVM to predict school dropouts with an accuracy of 97.27%. This study adopts a similar approach to predict student transfers at BK Palu School and compares the performance of SVM, KNN, and Naïve Bayes.

In predicting student transfers, it is crucial to identify the factors that contribute to this phenomenon. Various studies have pointed to academic performance, socio-economic background, school environment, and parental involvement as significant predictors of student mobility (Kabathova & Drlik, 2021). In this research, we focus on both academic and non-academic factors such as grades, attendance records, parental occupation, and peer relationships, which may contribute to a student's decision to transfer. By understanding these factors, schools can implement targeted interventions to retain students and prevent unnecessary transfers, thereby enhancing the overall educational experience (Monteverde-Suárez et al., 2024).

The three classification methods used in this study—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes—have been widely applied in various domains for predictive analysis. SVM, known for its ability to perform well in high-dimensional spaces, constructs a hyperplane that maximally separates different classes (Shilbayeh & Abonamah, 2021). KNN is a simple algorithm that classifies data points based on the majority class of their k-nearest neighbors. Naïve Bayes, on the other hand, is based on the application of Bayes' Theorem and assumes independence among features. Each of these methods has unique advantages and limitations that may influence their performance in predicting student transfers (Abu-Oda & El-Halees, 2015).

The selection of these methods is based on their differing strengths in handling classification problems with various data characteristics. For instance, SVM is particularly effective when dealing with non-linear relationships between variables, whereas KNN works well with smaller datasets and is easy to implement (Mariano et al., 2022). Naïve Bayes is known for its robustness in handling categorical data and its computational efficiency, making it an appealing choice when processing large datasets. In this study, these methods will be compared in terms of accuracy, precision, recall, and F1 score to determine which is most effective in predicting student transfers (Darmayanti et al., 2021).

The dataset used in this research includes records from BK Palu School, comprising both academic and non-academic variables. The data will be pre-processed by cleaning any missing values, normalizing numerical features, and converting categorical variables into a suitable format for machine learning algorithms (Janan & Ghosh, 2021). A portion of the data will be set aside as a testing set to evaluate the performance of the models.

The three algorithms will be trained on the remaining data and subsequently tested to assess their prediction accuracy in detecting which students are likely to transfer (Sari et al., 2022).

After applying the classification methods, the study will analyze the performance of each model. The results will include accuracy metrics for each algorithm, as well as a detailed comparison of their strengths and weaknesses. For example, while SVM might show better performance in complex datasets, it might also require more computational power compared to KNN and Naïve Bayes (Lottering et al., 2020). On the other hand, Naïve Bayes may outperform the others when the assumption of independence between features holds, but it could struggle with highly correlated data. Understanding these nuances will help in determining which algorithm is most suitable for addressing student transfer prediction at BK Palu School (Yujiao et al., 2023).

Based on the analysis results, recommendations will be provided for BK Palu School to reduce student transfer rates. If one of the methods shows significant accuracy, it can be used as part of an early warning system for identifying students at risk of transferring. Additionally, insights gained from the influential factors that predict student transfers can be used to create intervention programs, such as academic support, counseling services, and community-building activities, aimed at improving student retention (Alban & Mauricio, 2019).

In conclusion, this study contributes to the existing body of knowledge by comparing SVM, KNN, and Naïve Bayes in predicting student transfers, a relatively under-explored area in the educational context. By identifying key factors and recommending suitable strategies for schools to reduce transfer rates, this research has the potential to provide a practical solution to the ongoing issue of student mobility. Through the implementation of these classification methods, schools can take proactive steps to foster a stable and supportive learning environment for students.

The urgency of this research arises from the growing concern over student transfers, which significantly affect school dynamics, student performance, and educational outcomes. In particular, student transfers can disrupt learning, create administrative challenges, and lower overall school retention rates. At BK Palu School, addressing the root causes and predicting student transfers can help prevent unnecessary disruptions and ensure a more stable learning environment. With a rising trend in student mobility, it is essential to develop predictive tools that can assist schools in identifying students at risk of transferring and implementing timely interventions to reduce these rates.

The novelty of this research lies in its comparative approach to utilizing three different machine learning methods—SVM, KNN, and Naïve Bayes—to predict student transfers. While previous studies have focused on using individual algorithms to predict academic outcomes or dropouts, this study uniquely compares the strengths and weaknesses of these methods specifically in the context of student transfers. The research not only highlights the effectiveness of each algorithm in this particular application but also introduces a new perspective by applying these methods to real-world data from a specific school, thereby providing a more localized and actionable analysis.

The research aims to identify the most effective classification method for predicting student transfers at BK Palu School, comparing the performance of SVM, KNN, and Naïve Bayes. Through this comparison, the study seeks to uncover key factors that influence student transfers and offer insights into how schools can better anticipate and manage this issue. The ultimate goal is to provide a robust predictive model that can be

integrated into school management systems to flag potential transfer risks early, enabling schools to take proactive steps in retaining students.

The benefits of this research extend beyond the immediate findings. By understanding the predictive power of various algorithms, schools can develop more efficient strategies to mitigate student transfers, ultimately improving retention rates and the overall quality of education. Furthermore, the insights into the contributing factors behind student transfers can help educators tailor interventions, such as academic support programs, counseling services, and community-building activities, to better meet the needs of at-risk students. This research provides a framework for other schools to adopt similar predictive models, promoting a data-driven approach to reducing student mobility and improving educational stability.

Research Methods

The method used in this study is a comparative classification method utilizing three different algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes. The research focuses on comparing the performance of these three algorithms in predicting student transfers at BK Palu School. The evaluation process compares performance metrics such as accuracy, precision, recall, and F1-score to determine which algorithm is most effective in predicting student transfers.

The data used in this study consists of historical information on student transfers from BK Palu School, spanning the period from 2022 to 2024. The data includes various categories such as demographic information (age, gender, and family background), academic data (exam scores, academic achievements, and attendance), socio-economic data (family income and parents' occupation), and student quality indicators (participation in extracurricular activities and relationships with classmates and teachers). The method for this study follows several steps. First, student data is collected from school records, followed by a data preparation phase where the data is cleaned and processed to ensure consistency and quality. Afterward, the classification algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes, are selected for analysis. These algorithms are then implemented on the dataset to predict student transfers, and their performance is evaluated based on accuracy, precision, recall, and F1-score.

The Support Vector Machine (SVM) is a classification method that finds a hyperplane to separate data into two different classes with maximum margin, making it highly effective for binary classification, especially in high-dimensional spaces. K-Nearest Neighbors (KNN), on the other hand, is an instance-based method that classifies samples based on the majority of their nearest neighbors in the feature space. While simple and intuitive, KNN can be slower when dealing with large datasets. The Naïve Bayes algorithm operates as a probabilistic classifier using Bayes' Theorem, assuming independence among the features, and is known for its efficiency and effectiveness when working with large datasets. Finally, the performance of these three classification methods is evaluated and compared using key metrics such as accuracy, precision, recall, and F1-score to determine which method provides the most effective predictions for student transfers at BK Palu School.

Results and Discussions

The implementation results show that the SVM method has the highest accuracy in predicting student transfers with an accuracy of 95%, followed by Naïve Bayes with an accuracy of 92%, and KNN with an accuracy of 88%. The following table shows the performance comparison of the three methods:

Table 1. Prediction Result

Method	Accuracy	Precision	Recall	F1-score
SVM	95%	94%	96%	95%
Naïve Bayes	92%	91%	93%	92%
KNN	88%	87%	89%	88%

The SVM method shows the best performance in terms of accuracy and prediction stability. Naïve Bayes also shows good performance, especially in terms of precision. KNN has lower performance compared to the other two methods but is still quite effective.

Strengths and weaknesses of each method are analyzed to provide better insights in the context of student transfer data. SVM excels in handling high-dimensional data and can produce a clear separating margin between classes. Naïve Bayes is very effective when features are independent, although this assumption is rarely fully met in practice. KNN, while simple, requires optimal K value selection and can become inefficient for large datasets.

Conclusion

This study demonstrates that the SVM method outperforms KNN and Naïve Bayes in predicting student transfers at BK Palu School, exhibiting higher accuracy and prediction stability. The findings contribute to the development of more effective strategies aimed at reducing student transfer rates. Suggested future developments include the exploration of ensemble methods and testing on larger, more diverse datasets. Additionally, the study proposes several strategies to mitigate student transfers, such as providing school shuttle services to assist students with transportation difficulties, offering school fee discounts for high-achieving students as a motivational incentive, and providing discounts for families with multiple children enrolled at the same school to alleviate financial burdens. Implementing these strategies is expected to significantly reduce student transfer rates, thereby enhancing the stability and quality of education at BK Palu School.

References

Abu-Oda, G. S., & El-Halees, A. M. (2015). Data mining in higher education: university student dropout case study. *International Journal of Data Mining & Knowledge Management Process*, *5*(1), 15.

Alban, M., & Mauricio, D. (2019). Neural networks to predict dropout at the universities. *International Journal of Machine Learning and Computing*, *9*(2), 149–153.

Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *5*(2), 265–270.

Darmayanti, I., Subarkah, P., Anunggilarso, L. R., & Suhaman, J. (2021). Prediksi potensi siswa putus sekolah akibat pandemi covid-19 menggunakan algoritme k-nearest neighbor. *JST (Jurnal Sains dan Teknologi)*, *10*(2), 230–238.

Janan, F., & Ghosh, S. K. (2021). Prediction of student's performance using support vector machine classifier. *Proc. Int. Conf. Ind. Eng. Oper. Manag*, *11*(1), 7078–7088.

Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences*, *11*(7), 3130.

Kamila, V. Z., & Subastian, E. (2019). KNN vs Naive Bayes Untuk Deteksi Dini Putus Kuliah Pada Profil Akademik Mahasiswa. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *3*(2), 116–122.

Lottering, R., Hans, R., & Lall, M. (2020). A machine learning approach to identifying students at risk of dropout: A case study. *International Journal of Advanced Computer Science and Applications*, *11*(10), 417–422.

Mariano, A. M., Ferreira, A. B. de M. L., Santos, M. R., Castilho, M. L., & Bastos, A. C. F. L. C. (2022). Decision trees for predicting dropout in Engineering Course students in Brazil. *Procedia Computer Science*, *214*, 1113–1120.

Monteverde-Suárez, D., González-Flores, P., Santos-Solórzano, R., García-Minjares, M., Zavala-Sierra, I., de la Luz, V. L., & Sánchez-Mendiola, M. (2024). Predicting students' academic progress and related attributes in first-year medical students: an analysis with artificial neural networks and Naïve Bayes. *BMC Medical Education*, *24*(1), 74.

Purba, S. D., Harahap, L., & PANGGABEAN, J. F. R. (2022). Prediction of Students Drop Out With Support Vector Machine Algorithm. *Jurnal Mantik*, *6*(1), 582–586.

Sari, N. W. W., Mirwansyah, D., Fahrullah, F., Tandi, I. L., & Sopyani, A. H. H. (2022). Prediction of Students Drop Out With Naïve Bayes. *Jurnal Mantik*, *6*(3), 3858–3863.

Shilbayeh, S., & Abonamah, A. (2021). Predicting student enrollments and attrition patterns in higher educational institutions using machine learning. *Int. Arab J. Inf. Technol.*, *18*(4), 562–567.

Wahyuni, S., & Marbun, M. (2020). Implementation of data mining in predicting the study period of student using the naïve bayes algorithm. *IOP Conference Series: Materials Science and Engineering*, *769*(1), 12039.

Yujiao, Z., Weay, A. L., Shaomin, S., & Palaniappan, S. (2023). Dropout prediction model for college students in moocs based on weighted multi-feature and svm. *Journal of Informatics and Web Engineering*, *2*(2), 29–42.