DOI: 10.21554/hrr.042605

Predicting High School Academic Performance in Bosnia and Herzegovina Using Machine Learning Algorithms

Original research paper

Edisa Zilic and Mirdin Zilic

Faculty of Economics and Social Sciences, International Burch University, Bosnia and Herzegovina

Received: 2025/06/08
Accepted: 2025/10/18

Abstract

This study employs machine learning techniques to predict the academic success of high school students in Canton Sarajevo, aiming to identify key factors influencing performance and assess the accuracy of various predictive models. The results indicate that the Random Forest algorithm demonstrated significant improvements in precision compared to previous studies, highlighting its potential for use in educational decision-making. Key predictors of student performance included prior academic success, subject-specific grades, parental education, and engagement in academic activities such as homework and class participation. The study's practical implications suggest that educators and policymakers can use machine learning models to develop targeted interventions, improve student guidance, and implement early warning systems for at-risk students.

Keywords: *academic success, Canton Sarajevo, high school students, Machine Learning*

Academic performance plays a key role in shaping students' future opportunities. Schools are under steady pressure to identify students who may fall behind and provide timely support. As educational institutions collect more data, ranging from grades and attendance to behavioral and background information, many have turned to data-based decision-making to support student progress. Recent work from Bosnia and Herzegovina shows how school-sector data can power predictive models and guide early action, including studies on teacher outcomes that use survey-based features and machine learning (Zilic & Zilic, 2025). Machine learning models offer a practical way to use these data sources for forecasting, especially when early intervention can help prevent longer-term difficulties. This approach has already been applied locally to retention risks among teachers, with Random Forest used to flag at-risk staff and rank the strongest

Correspondence to: Mirdin Zilic, Faculty of Economics and Social Sciences, International Burch University
Email: mirdin.zilic@ibu.edu.ba
Phone: +38761707613
ORCID: https://orcid.org/0000-0002-2761-2096

predictors (Zilic & Zilic, 2025).

Several algorithms have been applied in this area, with Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes among the most frequently used. These models perform well in classification tasks and can handle mixed or incomplete datasets. Random Forest stands out for its ability to manage varied input types, maintain stable accuracy, and reduce overfitting. Research by Kabáthová and Drlík (2021) and Falát and Piscová (2022) confirmed its strong performance in predicting dropout and GPA at the university level. Still, most studies have focused on higher education, where larger datasets and stronger institutional infrastructure are available.

Random Forest builds its predictions from multiple decision trees trained on different subsets of data. This ensemble structure increases stability and improves accuracy, particularly when working with smaller or imbalanced datasets often found in education. Tomasevic et al. (2020) found that ensemble models performed better than individual classifiers when student interaction data were available. Similarly, Falát and Piscová (2022) reported that RF outperformed other models in GPA prediction, while also offering useful insights into which features influenced the outcome.

The usefulness of Random Forest also lies in its ability to incorporate a range of variables. Academic performance is shaped by more than test scores. Family obligations, employment during school, emotional well-being, commuting distance, and financial conditions all affect learning outcomes (Bin Mat et al., 2013; Madni et al., 2017). Studies have also shown the influence of emotional intelligence during academic transitions (Berger & Milem, 1999) and the importance of attendance, prior failures, extracurricular activities, and study habits (Alturki & Alturki, 2021; Arun et al., 2021). Broader social factors, such as gender, school environment, and neighborhood context, also play a role (Thiele et al., 2016). Models that can accommodate this diversity, like RF, offer a practical way to reflect the realities of school life in prediction.

RF has been used in a wide range of academic prediction tasks. Nachouki et al. (2023) used it to forecast final course grades and found that prior GPA and high school grades were the strongest indicators. In digital education settings, where data such as logins, assignment submissions, and forum activity are available, RF has shown strong results. Jawad et al. (2022) applied RF with SMOTE to account for data imbalance and improved the model's AUC to 0.96. Abubakar and Ahmad (2017) also demonstrated RF's effectiveness in identifying key engagement features in e-learning environments. Chen and Liu (2024) applied a genetic algorithm to enhance RF's tree structure, reaching over 93% classification accuracy, about 2% higher than standard RF. Balabied and Eid (2023) used RF to analyze data from the Open University Learning Analytics Dataset (OULAD), achieving an accuracy of 90% when flagging students likely to underperform in online learning environments. Such applications show how RF can support early warning systems, enabling institutions to respond more quickly to signs of disengagement or poor performance.

Other machine learning techniques, neural networks, decision trees, Naive Bayes, and SVM, have been widely studied. For example, Decision Trees reached up to 91% accuracy when using CGPA as the main feature (Jishan et al., 2015), while Neural Networks generally performed slightly lower, with results ranging from 71% to 75% depending on the predictors used (Ramesh et al., 2013; Jishan et al., 2015). Ensemble methods generally show better results than individual models (Romero & Ventura, 2010; Madni et al., 2017). However, outcomes can vary depending on the dataset and chosen features. For example, in Abubakar and Ahmad's (2017) study, Naive Bayes outperformed RF, but this was attributed to the structure of the dataset and specific variables used. SVM has also achieved strong performance when paired with oversampling methods, as shown by Thammasiri et al. (2014), who reached over 90% accuracy in student retention prediction.

Several studies in Southeast Europe have started to apply these methods to local education systems. Osmanbegovic and Suljic (2012) demonstrated the usefulness of combining Decision Trees and Neural Networks in predicting course success in Bosnia and Herzegovina. Pasic et al. (2023) examined engineering students in Canton Sarajevo and found that high school GPA and grades in math and physics were reliable predictors of university performance. Although they used a single decision tree due to limited data, the findings suggest potential for more advanced methods. In Serbia, Tomasevic et al. (2020) reported that Random Forest was among the top-performing models when applied to exam score prediction, especially when continuous assessment and engagement metrics were available.

In Bosnia and Herzegovina, academic performance research still relies mostly on descriptive analysis or official statistics. Differences between students across municipalities, school types, or socioeconomic groups are recognized but rarely examined using predictive models. Career guidance and academic support are typically generalized, often failing to reflect students' individual profiles or challenges. This disconnect limits the ability of schools to target support where it's most needed. Parallel projects in the same system have already demonstrated both the availability of rich, school-level data and the feasibility of ML workflows in this setting (Zilic & Zilic, 2025).

This study addresses that gap by applying the Random Forest algorithm to data collected from high school students in Canton Sarajevo. The dataset includes information on socio-demographic characteristics, academic records, school experiences, student habits, and future plans. All responses were anonymized and processed to detect patterns linked to academic success. Along with Random Forest, this research evaluates the performance of Support Vector Machine and Naive Bayes classifiers. It also assesses whether feature selection improves prediction accuracy. The study aims to support local schools in using available data more effectively and in building targeted academic support based on evidence.

Methods

Research Design and Methodological Justification

This study applied a quantitative research design supported by supervised machine learning techniques. The goal was to build predictive models that identify which factors most strongly influence high school students' academic performance. Given the dataset's mixed structure-numerical and categorical variables from survey responses, machine learning methods were well-suited for capturing patterns that traditional statistical techniques might overlook. Random Forest, Support Vector Machine (SVM), and Naive Bayes classifiers were selected for their consistent track record in handling classification tasks in educational research. Each model also supports various feature selection methods, allowing for refined analysis and improved accuracy.

Context of the Study

The research was conducted during the 2022/2023 academic year in Canton Sarajevo, Bosnia and Herzegovina. The canton includes a range of urban and suburban municipalities, reflecting a diverse student population. Secondary schools in the region face varying levels of infrastructure and resources, making predictive tools valuable for understanding student needs more precisely. The study context allowed for examining how machine learning can support academic planning in a setting that combines demographic variety and structural limitations.

Sample, Sampling, and Ethical Considerations

Participants were high school students enrolled in public and private schools across different municipalities, including Centar, Novo Sarajevo, Stari Grad, Ilidža, Ilijaš, Vogošća, Novi Grad, and Hadžići. The final dataset included 405 responses. Students were selected using a mix of purposive and convenience sampling. Schools were approached based on their geographic distribution and administrative willingness to take part.

All participants were informed about the purpose of the research through a brief statement at the beginning of the survey. Participation was voluntary and anonymous. The instrument contained no questions that could lead to identification. Prior to the full launch, the questionnaire was reviewed by educators and tested with a small pilot group of 50 students to confirm clarity and appropriateness. No further revisions were necessary. Data were collected using Google Forms and stored in a password-protected database accessible only to the researchers.

Research Tools and Procedures

The main instrument for data collection was a structured questionnaire consisting of 57 items, organized into four sections: socio-demographic characteristics, academic history and school experience, behavior and personal habits, and expectations and plans after graduation. The questionnaire design drew on prior studies and was adapted to the educational context of Canton Sarajevo (Benford & Gess-Newsome, 2006; Cortez & Silva, 2008). After data collection, all responses were exported into Python for further processing and analysis. The analysis was carried out using several key libraries: Pandas for data manipulation, NumPy for numerical operations, Scikit-learn for training and evaluating machine learning models, and Matplotlib and Seaborn for generating visualizations.

To prepare the data for modeling, categorical variables were converted into numerical formats using two mapping strategies. Numeric range mapping was applied to ordered categories such as parental education levels, study time, and extracurricular participation. For binary variables like relationship status and scholarship status, dummy variable mapping was used. The dataset was reviewed for quality, and no missing, inconsistent, or duplicate entries were identified, making the data ready for analysis without further cleaning.

Data Analysis Techniques

The analysis involved training and evaluating three machine learning models: Random Forest, Support Vector Machine (SVM), and Naive Bayes. Each algorithm was optimized through specific feature selection techniques tailored to its structure:

- Random Forest: Used internal feature importance rankings to identify the top ten predictors of student performance. Model tuning included adjusting parameters such as tree depth, number of trees, and minimum split size.
- Naive Bayes: Applied Chi-square testing, mutual information scoring, and L1 regularization (Lasso) to determine relevant features. Laplace smoothing was used to prevent zero probabilities in categorical data.
- SVM: Employed Recursive Feature Elimination (RFE), SelectKBest with mutual information, and L1 regularization to enhance model performance. Hyperparameters such as kernel type, regularization (C), and gamma were fine-tuned.

Model performance was evaluated using accuracy, precision, recall, and F1-score. Cross-validation helped assess consistency and avoid overfitting. These techniques were chosen for their suitability in handling structured survey data and for their ability to highlight which features most affect academic outcomes.

Results

This section presents the outcomes of applying three machine learning models, Random Forest, Naive Bayes, and Support Vector Machine (SVM), to a dataset collected from high school students in Canton Sarajevo. The models were evaluated using standard classification metrics to determine their effectiveness in predicting student success based on survey responses.

Model Evaluation Criteria

To assess the quality of each model, the following metrics were applied:

- Training Accuracy – how well the model fit the training data.
- Testing Accuracy – how well the model generalized to new data.
- Precision – proportion of predicted successful students that were correct.
- Recall – proportion of actual successful students that the model identified correctly.
- F1-Score – a measure that balances precision and recall, particularly relevant in educational datasets where class imbalance may be present.

These metrics provide a practical overview of how well each model performed in identifying students likely to achieve higher academic outcomes.

Random Forest Results

Random Forest was first trained using all available features. The initial testing accuracy reached 75.41%. After selecting the ten most informative variables based on feature importance, the model's performance improved. The selected features included: last overall success, math grade, mother tongue language, foreign language, scholarship, homework, class, the distance of the school, mother's education, and the number of siblings.

Using these variables alone, the testing accuracy increased to 82.72%. The precision score was 0.72, with a recall of 0.75 and an F1-score of 0.73 (Table 1). Training accuracy remained high, indicating consistent learning.

Table 1.
Random Forest Model Performance

Metric	Value
Training Accuracy	0.9964
Testing Accuracy	0.7540
Accuracy	0.8272
Precision Score	0.7154
Sensitivity/Recall Score	0.7540
F1 Score	0.7337

Naive Bayes Results

Naive Bayes was tested with three feature selection strategies: Chi-square test, Lasso regularization, and Mutual Information. The best outcome was achieved with Mutual Information, resulting in 76.54% accuracy. The key variables identified with this method were: the student's previous overall success, math grade, mother tongue grade, foreign language mark, father's education, scholarship status, homework completion, engagement in post-school discussions, participation in class, and gender (Table 2).

Table 2.
Top Predictors for Naive Bayes (Mutual Information)

Feature
Last overall score
Math grade
Native language grade
Foreign language grade
Father's education
Scholarship
Homework time
Discussions outside class
Class participation
Gender

The classification metrics for this model are shown below (Table 3).

Table 3.
Naive Bayes Performance

Metric	Precision	Recall	F1-Score
Accuracy			0.77
Macro avg	0.60	0.59	0.59
Weighted avg	0.76	0.77	0.76

Support Vector Machine (SVM) Results

Support Vector Machine achieved the highest testing accuracy among the three models, reaching 83.95%. Feature selection techniques, including Lasso, Recursive Feature Elimination (RFE), and SelectKBest with Mutual Information, were applied to improve the model. However, none of these methods led to further improvements in accuracy.

The best results were obtained using the full feature set. The model showed: Training accuracy of 86.42%, Precision of 0.56, Recall of 0.58, and F1-score of 0.57 (Table 4).

Table 4.
Support Vector Machine Performance

Metric	Value
Training Accuracy	0.8642
Testing Accuracy	0.8395
Precision Score	0.5556
Sensitivity/Recall Score	0.5792
F1 Score	0.5655
Accuracy	0.8395

Discussion

This study compared three machine learning algorithms, Random Forest, Support Vector Machine (SVM), and Naive Bayes, using a dataset collected from high school students in Canton Sarajevo. SVM achieved the highest testing accuracy (83.95%), followed closely by Random Forest (82.72%) after applying feature selection. Naive Bayes, which initially performed at 68.85%, improved to 76.54% when mutual information was used to refine the input variables. These results demonstrate that carefully developed models can classify academic success with considerable precision when supplied with well-organized data reflecting both academic and behavioral dimensions.

The outcomes correspond with several earlier studies in educational data mining that reported favorable performance for Random Forest and SVM in similar classification tasks. Falát and Piscová (2022) confirmed Random Forest's capability in GPA prediction, particularly in situations where datasets included a wide variety of features. The improved accuracy in this study compared to Abubakar and Ahmad (2017), who reported 76.9% for Random Forest, likely reflects differences in variable selection. The inclusion of behavioral, socio-demographic, and extracurricular factors contributed to the stronger results here. Similar observations were made by Tomasevic et al. (2020), who found that ensemble models, including Random Forest, produced stronger results when data included information about student engagement. SVM outperformed the other models in this study, reflecting findings from Thammasiri et al. (2014), where SVM surpassed 90% accuracy when used with SMOTE. Although no oversampling method was used here, the strong performance of SVM shows that it remains competitive even without additional preprocessing. However, this study also revealed that SVM's performance was more sensitive to feature selection methods. Techniques such as Lasso and Recursive Feature Elimination offered no improvements and in some cases reduced accuracy. This aligns with other work indicating that SVM performs best with well-balanced and clearly defined inputs. In contrast, Naive Bayes improved only when mutual information was applied, indicating that its assumptions of feature independence do not align easily with educational datasets where variables often correlate, for example, parental education and academic success.

Compared to the higher education settings frequently covered in existing literature, this study focused on high school students, incorporating a broader array of features including class behavior, homework habits, extracurricular involvement, and future aspirations. Findings align with local ML studies in the same school system that also emphasized the value of behavioral and support variables captured through surveys (Zilic & Zilic, 2025). Past academic performance in math, native language, and foreign language consistently ranked among the most useful predictors, reinforcing conclusions from Nachouki et al. (2023) and Pasic et al. (2023) that prior achievement is a reliable basis for estimating future success. The presence of these same predictors across different model types also lends weight to their relevance for school-level interventions.

Model behavior followed expected patterns. Random Forest performed consistently and provided interpretability through its ability to rank variable importance. Its structure allows it to manage both categorical and numerical inputs, which made it well-suited to the variety of features in the dataset. SVM, while performing strongly, required closer attention to input quality and feature selection. Naive Bayes, although simpler, showed improvement when the right variables were included, though its lower accuracy suggests that its assumptions make it less suitable for this type of data without adjustment.

Beyond academic indicators, the findings revealed a group of behavioral and contextual factors that also contributed to student success. Variables such as class participation, homework completion, volunteering, and engagement in academic discussions outside the classroom emerged as relevant in multiple models. These patterns suggest that school performance is influenced not only by prior knowledge but also by a student's willingness to engage in their educational environment. A similar pattern, non-academic conditions shaping outcomes, appeared in exploratory analyses of teacher attrition across regions (Zilic & Zilic, 2024). In particular, Random Forest captured these patterns effectively, pointing to its value in identifying subtle connections between variables that may not be immediately apparent.

Taken together, the models offer a way to identify students who may need additional academic guidance, even before their performance visibly declines. Prior success, classroom behavior, and background information such as parental education or commuting distance, can be used to understand risk factors and provide timely support. These insights can serve teachers and school leaders who are looking for ways to move beyond traditional assessment tools and offer more proactive, student-centered support. Rather than classifying students in a rigid manner, these models can guide counseling, enrichment, or mentoring by helping to clarify where and how a student may need assistance.

This research highlights the benefit of applying predictive models to high school education in Bosnia and Herzegovina. The results suggest that institutions with access to student-level data can use machine learning to identify patterns of success and challenge, and to build interventions that better reflect the diversity of students' experiences. As the use of data grows across education systems, these findings may serve as a reference point for schools considering how to connect data science with day-to-day teaching and student support.

Conclusion

This research examined how machine learning models can be used to predict academic success in secondary schools and identify the factors that most influence student performance. By applying Random Forest, Naive Bayes, and Support Vector Machine to survey data from students in Canton Sarajevo, the study offered a comparison of model effectiveness and generated practical insights for educators.

SVM produced the highest classification accuracy, followed closely by Random Forest. Naive Bayes showed improvement with the use of feature selection. Across all models, the most relevant predictors included grades in core subjects, past academic success, parental education, homework habits, and participation in school-related activities. These findings suggest that prediction models can support earlier identification of students who may need additional support and guide more targeted educational planning.

Beyond confirming earlier international findings, this study adds value by examining secondary education in Bosnia and Herzegovina, a level that has rarely been included in data-driven prediction research. By showing that models such as Random Forest and SVM deliver strong results with local student data, the research offers a foundation for further studies in education systems with similar characteristics and challenges.

For schools and policymakers, the results offer several possible applications. Models like the ones tested here can help educators detect risk factors early, allocate resources more effectively, and refine interventions based on actual data. Guidance counselors can use these insights when advising students, while classroom teachers may adapt their approaches based on student profiles. At a broader level, the use of data-driven strategies can inform system-wide decisions and shape policy that reflects student needs more accurately.

This study was limited to high schools within one administrative region. While the sample was diverse in school type and location, the geographic scope restricts how broadly the findings can be applied. The reliance on self-reported data introduces the possibility of bias, and the absence of longitudinal tracking means that changes in performance over time were not captured.

Future work should aim to include a wider range of schools across Bosnia and Herzegovina and extend the analysis across multiple school years. Adding longitudinal data would help understand how learning patterns develop over time. Researchers may also consider experimenting with hybrid algorithms or deep learning techniques, as well as refining data collection to include standardized assessments and teacher evaluations. By broadening the dataset and improving model inputs, future research can build on these findings to further support data-informed educational practices.

References

Abubakar, Y., & Ahmad, N.B. (2017). Prediction of Students' Performance in E-Learning Environment Using Random Forest. International Journal of Innovative Computing, 7(2). https://doi.org/10.11113/ijic.v7n2.143.

Alturki, S., & Alturki, N. (2021). Using educational data mining to predict students' academic performance for applying early interventions. Journal of Information Technology Education: Innovations in Practice, 20, 121-137. https://doi.org/10.28945/4835.

Arun, D. K., Namratha, V., Ramyashree, B. V., Jain, Y. P., & Choudhury, A. R. (2021, January). Student Academic Performance Prediction using Educational Data Mining. In 2021 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-9). doi: 10.1109/ICCCI50826.2021.9457021.

Balabied, S. A. A., & Eid, H. F. (2023). Utilizing random forest algorithm for early detection of academic underperformance in open learning environments. PeerJ Computer Science, 9, e1708. https://doi.org/10.7717/peerj-cs.1708.

Benford, R., & Gess-Newsome, J. (2006). Factors Affecting Student Academic Success in Gateway Courses at Northern Arizona University. Online Submission.

Berger, J. B., & Milem, J. F. (1999). The Role of Student Involvement and Perceptions of Integration in a Causal Model of Student Persistence. Research in Higher Education, 40(6), 641–664. http://www.jstor.org/stable/40196897

Bin Mat, U., Buniyamin, N., Arsad, P. M., & Kassim, R. (2013, December). An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. In 2013 IEEE 5th conference on engineering education (ICEED) (pp. 126-130). IEEE.

Chen, M., & Liu, Z. (2024). Predicting performance of students by optimizing tree components of random forest using genetic algorithm. Heliyon, 10(12). https://doi.org/10.1016/j.heliyon.2024.e32570.

Cortez, P., & Silva, A. M. G. (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito, & J. Teixeira (Eds.), Proceedings of 5th Annual Future Business Technology Conference, Porto, 5-12.

Falát, L., & Piscová, T. (2022). Predicting GPA of University Students with Supervised Regression Machine Learning Models. Applied Sciences, 12(17), 8403. https://doi.org/10.3390/app12178403

Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. Sustainability, 14(22), 14795. https://doi.org/10.3390/su142214795

Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. Decision Analytics, 2, 1–25. https://doi.org/10.1186/s40165-014-0010-2

Kabáthová, J., & Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. Applied Sciences, 11(7), 3130. https://doi.org/10.3390/app11073130

Madni, H. A., Anwar, Z., & Shah, M. A. (2017). Data mining techniques and applications - A decade review. 23rd International Conference on Automation and Computing (ICAC), Huddersfield, UK, 2017, pp. 1-7, doi: 10.23919/IConAC.2017.8082090.

Nachouki, M., Mohamed, E. A., Mehdi, R., & Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. Trends in Neuroscience and Education, 33, 100214. https://doi.org/10.1016/j.tine.2023.100214

Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business, 10(1), 3–12.

Pasic, M., Vatres, A., Ferizbegovic, F., Bajric, H., & Pasic, M.. (2023). Machine learning prediction and analysis of students' academic performance. Periodicals of Engineering and Natural Sciences (PEN), 11(5), 27–46. https://doi.org/10.21533/pen.v11.i5.183

Ramesh, V., Parkavi, P., & Ramar, K. (2013). Predicting student performance: A statistical and data mining approach. International Journal of Computer Applications, 63(8). doi:10.5120/10489-5242.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. Expert Systems with Applications, 41(2), 321–330. https://doi.org/10.1016/j.eswa.2013.07.046

Thiele, T., Singleton, A., Pope, D., & Stanistreet, D. (2016). Predicting students' academic performance based on school and socio-demographic characteristics. Studies in Higher Education, 41(8), 1424–1446. https://doi.org/10.1080/03075079.2014.974528

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers & Education, 143, 103676. https://doi.org/10.1016/j.compedu.2019.103676

Zilic, E., & Zilic, M. (2024). Analysis of Factors Affecting Teacher Attrition in Bosnia and Herzegovina Using Exploratory Data Analysis. Zbornik Radova MES, 10, 277-288. https://doi.org/10.7251/BLCZR0723277Z.

Zilic, E., & Zilic, M. (2025). Using Machine Learning for Teacher Retention in Schools. In International Conference New Technologies, Development and Applications (pp. 110-118). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-95200-5_13.