International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Predictive Analytics and Explainable Machine Learning for Early Identification of At-Risk Students Using an Early Warning Score

Snehal Ramesh Dedge,Dr. Kalpana Salunke

MCA Second Year JSPM University, Pune, India
Email: dedgesnehal3@gmail.com

----------**********----------

Abstract:

Institutions of higher education have been increasingly challenged by the need to identify students at risk of academic failure prior to significant declines occurring in performance. Traditional evaluation techniques based primarily on limited frequency of scheduled assessments have tended to be reactive rather than proactive, thus often missing the opportunity for timely intervention prior to experiencing negative consequences due to poor student performance. The rapid growth of online learning tools has produced large volumes of academic and behavioral data from thousands of students, creating a tremendous potential for leveraging this data to provide a proactive, data-informed solution. The present study outlines a predictive analytics framework that identifies at-risk students early using a weighted Early Warning Score (EWS) and machine learning models. The predictive model used to calculate the EWS leverages key indicators (e.g., student engagement, attendance patterns, workload stressors, and learning behaviours) to derive and evaluate potential risk levels associated with an at-risk student's academic performance. To enhance the transparency and establish trust in the predictions generated by the model, the framework incorporates Explainable Artificial Intelligence (XAI) techniques, in particular the use of SHAP (Shapley Additive explanations).

This allows educators to understand the degree to which each individual factor contributes to the derived prediction and provide data/information to enable educators to make informed decisions about implementing targeted interventions. The experimental results indicate that this predictive analytics framework demonstrates a high level of accuracy and high recall in identifying at-risk students. In addition, this predictive framework provides an adequate degree of interpretability for educators to use in determining the necessary performance-enhancing interventions and support services when implementing proactive academic monitoring. By providing a means of utilizing predictive analytics and incorporating an explanation of the factors provided in making the prediction, this framework represents a comprehensive solution for utilizing a proactive approach in supporting students and increasing their success through timely and relevant service strategies to improve student retention and overall student achievement.

Keywords — Predictive Analytics, Early Warning System, Explainable AI, SHAP, Student Performance Prediction.

----------**********----------

INTRODUCTION

Higher Education Institutions all over the world face very significant challenges regarding high levels of student attrition and poor academic performance. Many universities and colleges are always looking for ways to improve both student retention and student success. Unfortunately, many students experience problems that university personnel do not know about until the student has already experienced a significant decline in their academic performance. Traditional assessments of student learning such as

midterm and final exams typically only show that there is a problem with the student's academic performance after the student's performance has already declined, thereby producing difficulties for universities attempting to intervene early [29], [30]. Research in the area of educational data mining has found that students who ultimately do poorly academically or drop out of school demonstrate early warning signs in their academic and behavioural patterns of undesirable behaviour. Examples of academic and behavioural indicators that frequently precede the failure of students to achieve academic success include lower levels of engagement with course material, irregular attendance, and high levels of stress as a result of workload, as well as inappropriate study habits used while attempting to accomplish academic tasks. By identifying these types of early warning indicators, institutions can intervene in a more timely manner to provide academic supports and to provide more personalised support for students through a more tailored approach [5], [6].

Due to the rapid digital transformation of educational environments, the amount of data being generated related to students has increased significantly. Specifically, Learning Management Systems, Online Learning Systems, and Institutional databases capture massive amounts of data resulting from student attendance, student participation, student assignment submissions, and student learning activities. These data provide opportunities to utilise predictive models and machine learning techniques to help identify hidden behavioural patterns and potentially identify potential academic risks to students [2], [7]

Models of machine learning perform well when it comes to predicting how students do in school, including logistic regression, decision trees, and the various ensemble learning algorithms available today. A significant problem with most of these machine-learning models is that they are not interpretable. This means that complex models operate as a “black box” so that users/educators cannot tell why any particular prediction is made. Because of a lack of transparency, users/educators have less trust in an online decision-making system and will restrict their use in school settings (ref. 22, ref. 24). To address this challenge, Explainable Artificial Intelligence (XAI) techniques have been developed to provide interpretable insights into the predictions of machine learning algorithms.

SHAP (SHapley Additive exPlanations) is one of the most popular providing explanations for how models make predictions by calculating the contribution of every feature used in the model. The potential exists to create transparent predictions by using predictive models together with SHAP explained outputs so that users/educators can understand how to react (ref. 1, ref. 26). This research proposes a predictive analytics framework that combines the use of an Early Warning Score (EWS), supervised machine learning algorithms, and SHAP-based explainability. The EWS aggregates the behavioral and academic indicators into an overall risk measure used to assess a student’s risk of academic failure. To identify student risk levels, machine learning algorithms utilize prediction values, while the model's explainability layer specifies which attributes were used to arrive at those predictions. The purpose of this research project is to create an interpretable and dependable system that will detect at-risk students as early as possible. Through the combination of predictive analytics and explainable machine learning, this proposed framework is intended to enable educators to make data-driven decisions and provide timely interventions to improve student success and retention rates.

I. PROBLEM STATEMENT

Most current academic monitoring systems are primarily based on test scores and fixed cut-off scores. These methods are limited for a number of reasons, including: Late identification of academic risk Limited behavioral and engagement indicators Lack of interpretability in predictive systems Thus, there is a critical need for an intelligent system that will be able to detect at-risk students early on, explain why they have been identified as such, and assist in implementing timely strategies for intervention.

III. LITERATURE REVIEW

There has been considerable attention given to the use of machine learning techniques in educational

data mining to forecast student success and discover at-risk pupils. Initially, early investigations employed conventional statistical techniques (logistic regression and decision trees) as they were simple, efficient and interpretable. Although the models were useful to identify basic risk patterns in education, they were unable to accurately identify complex, high-dimensional data sets. To address these issues, researchers developed ensemble learning techniques such as Random Forest and Gradient Boosting to create predictions from a Combining many learners and capturing non-linear relationships with features for greater predictive accuracy than a single learner. Despite achieving greater predictive performance than other models, educators are concerned about the lack of transparency provided by these methods. Therefore, how do transactional, multi-level educational systems gain transparency when using these advanced models? Recent advancements in Explainable Artificial Intelligence have created new ways to improve transparency of predictive models through multiple techniques, such as the SHAP (SHapley Additive exPlanations method), which provides insights into how feature contributions are related to model predictions so that stakeholders can see how various attributes affect model outputs. Studies demonstrate that the increase of explainability has had a positive impact on user trust and adoption of these technologies by educators. The majority of research on Early Warning Systems (EWS) has also used an early warning system approach to monitor and assess employee performance using indicators (e.g., student attendance and purpose) for identifying students who may be at-risk. EWS was designed to identify students as at-risk by relying on rule-based methods or static scoring methods, which limits the predictive capabilities and adaptability of the WDS.

There continues to be a gap in knowledge regarding the connectivity between the traditional domain-based early warning score systems (EWSs) and explainable Machine Learning (ML) predictive models; this gap is addressed by using both EWSs and a Shapley value system of explainability by producing both accurate results and interpretable information for academic intervention purposes.

IV. RELATED WORK

Many researchers have evaluated predictive model applications for student performance assessments using logistic regression and decision trees as they provide a degree of practicality and interpretability. In addition, Ensemble modelling has also provided enhanced predictive accuracy (i.e. Random Forests and gradient boosting) but often produce less transparency than traditional models. Many methods of Explainable Artificial Intelligence (AI) have developed processes and frameworks to increase the understanding of AI and the associated interpretability; SHAP(SHapley Additive exPlanations) has become a mainstream framework for developing the associated contribution scores for each input or attribute that contributes to the prediction. Prior researchers have indicated that the use of SHAP within an Educational Prediction model has increased the perceived level of confidence in the student prediction models; however, the extent to which SHAP has been applied in conjunction with EWSs remains limited. This paper extends prior studies to include domain specific risk assessment scores used in association with Explainable ML predictive modelling techniques.

TABLE I
RELATED WORK ON STUDENT PERFORMANCE PREDICTION

Study/Author	Method Used	Key Contribution	Limitation
Traditional Models	Logistic Regression, Decision Trees	Simple, interpretable models for student performance prediction	Lower accuracy On complex datasets
Ensemble Models	Random Forest, Gradient Boost- ing	Improved prediction accuracy using multiple learners	Lack of interpretability (black-box behavior)

XAI-Based Student	SHAP (Explain-able AI)	Provides feature-level contribution and improves trust in predictions	Limited integration with structured scoring sys-tems
Proposed Approach	EWS + ML + SHAP	Combines domain-based risk scoring with explainable machine learning for early detection	Requires proper feature engineer- ing and domain knowledge

V. PROPOSED SYSTEM ARCHITECTURE

The system proposed is layered and modular, which will allow for reproducibility and scalability in composition. The components of the architecture are:

• The Data Ingestion Layer captures both Academic Data, (Academic, Behavioural), and Engagement Data.

• The Pre-processing Layer provides functionality for managing missing values, normalising the dataset and encoding the required feature set.

• The Early Warning Score Computation Layer estimates and calculates Early Warning Scores.

• The Machine Learning Layer will predict Academic Risk Categories.

• The Explainability Layer generates SHAP explanations for the machine learning predictions. Refer to Fig. 1 for the visual representation of the Proposed System Architecture to predict Student Risk.

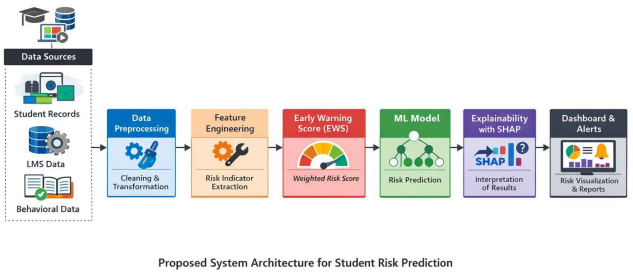

Fig. 1. Proposed System Architecture for Student Risk Prediction

VI. DATASET INFORMATION

The dataset for this study consists of attributes of student records related to academic metrics such as Attendance and Engagement, Learning Behaviours and Amount of Workload. Continuous variables were normalised for value scale consistency, while categorical variables were transformed using label encoding approaches. The dataset has been divided into training and testing datasets so that model performance could be evaluated consistently using a random 80/20 distribution.

VII. EARLY WARNING SCORE FORMULATION

The Early Warning Score is a composite risk metric that combines Academic and Behavioural Areas (Risk) into one score.

$$EWS = 0.35\ Ed + 0.30\ Ar + 0.20\ Ws + 0.15\ Lh \quad (1)$$

Where Ed = Engagement Deficit, Ar = Attendance Risk, Ws = Workload Stress, Lh = Learning Habit Weakness, and corresponding weights were selected based on validity and relevance to education literature. A higher EWS score indicates higher Academic Risk.

VIII. MACHINE LEARNING MODEL

A. Model Selection and Evaluation

The models used for the prediction of at-risk students (students who will drop out of school)

were based on supervised learning algorithms. Logistic regression and Random Forest were the two models given the most attention for this application because of their ability to deliver relatively similar performance, as well as their ability to provide equal interpretability of the learning data. Logistic regression was used as an initial model because it allows for the clear interpretation of feature importance. Random Forest was selected as the second model, which enables the identification of complex, non-linear relationships between features through a collection of decision trees, while providing a higher level of predictive accuracy for the number of values to be predicted. The predictive power of the models was evaluated using a variety of key performance indicators (KPIs), such as accuracy, precision, recall, and F1 score. The metric given the most consideration in the evaluation of the models was the recall metric, which is critical to the overall success of the system. The ultimate goal of the project is to accurately identify at-risk students. In other words, the quality of the recall metric is critical because a large volume of false negatives (i.e. at-risk students who are classified as "not at-risk") has serious consequences in education settings where missing an at-risk student would result in severe academic effects. In addition to the KPIs used to evaluate the performance of the models, the level of interpretability of both models with the associated data was very important in the decision on which model to use to predict at-risk students. By ensuring that the educators can understand and trust the results of the models, a higher level of interest in the system will likely occur.

In order to provide greater transparency in predicting models, we incorporated explainability techniques (SHAP). The resultant predictive model was integrated into a complete predictive pipeline, which consists of 3 key stages: Data Preprocessing, Model Prediction and Performance Evaluation. The data preprocessing stage includes dataset cleaning, normalising and feature engineering (i.e. creating new features from existing variables). The prediction stage applies the machine learning model on the cleaned dataset to generate risk classifications for each policyholder. Finally, during the evaluation stage of the predictive pipeline, we continually monitor the effectiveness of the predictive model based on real-world performance metrics. This modular predictive pipeline allows for scalability, consistency and ease of application in academic settings when performing modelling tasks.

IX. EXPLAINABILITY USING SHAP

To ensure the predictive model is transparent and interpretable, SHAP (SHapley Additive exPlanations) was used to explain global and local predictions. SHAP is based on concepts from cooperative game theory. With this concept, we assign each variable a "shap" or contribution value that indicates how much associated input contributed to the final predictive score for the observational unit. By utilising SHAP to support our model, we will gain a clear picture of how the various independent variables impact the predictions that are generated for the model.

Using analysis called SHAP, we can see which factors in a dataset have the most impact consistently on predicting the risk of being an at-risk student,(by feature) and also how each specific feature value of a student increases or decreases the individual risk of being at-risk, i.e., each student is classified either as at-risk of becoming a drop-out or not. The results of the analysis show us that the two main driving factors that affect students' predictions of being classified as at-risk are their Engagement Deficit and Attendance Risks. Students who have low levels of engagement and have irregular attendance patterns will most likely be detected as At-Risk. There are other factors that impact the prediction of being an at-risk student such as student workload stress or learning behaviors, but have minimal effects impacting their at-risk prediction. The analysis performed with SHAP provides both a global (overall) and an instance-level (specific) explanation of why students have an at-risk classification, which will help build educators' trust in the prediction model, enabling them to view each student's individual prediction, and

thus allow educators to implement Data Driven Decisions based upon the fact that educator implementations will be appropriately targeted to meet the individual needs of each student.

X. RESULTS AND EVALUATION

The predictive performance of the model was well documented, especially when classifying the high-risk students. A Confusion Matrix illustrates the model's classification accuracy when identifying high-risk students. The Recall Score was high showing that the majority of at-risk students have been correctly predicted. This is critical for most of the early-stage identification systems of at-risk students.

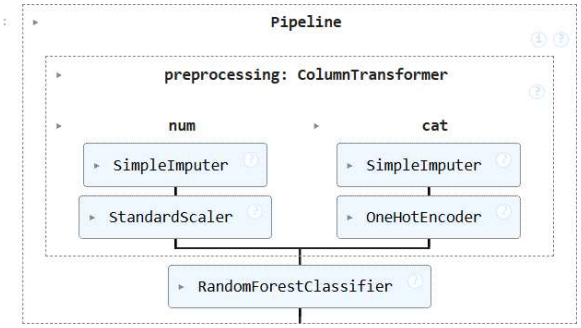

Fig. 2. Proposed Predictive Analytics Pipeline for Early Warning System

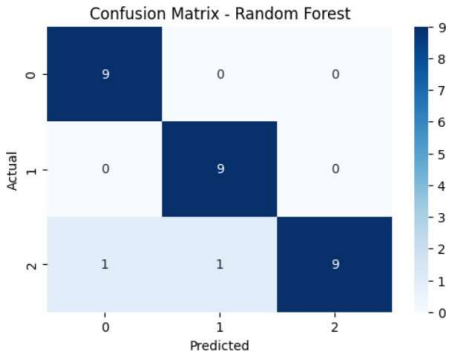

Fig. 3. Confusion Matrix Showing Classification Performance of the Model

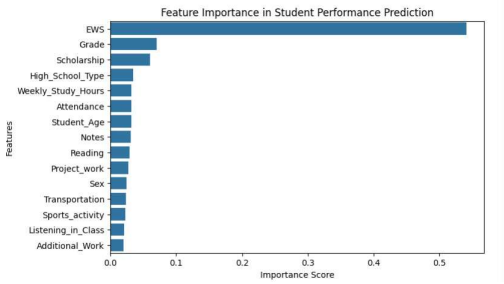

Fig. 4. Feature Importance Ranking for Academic Risk Prediction

Fig. 5. SHAP Summary Plot Showing Global Feature Contributions

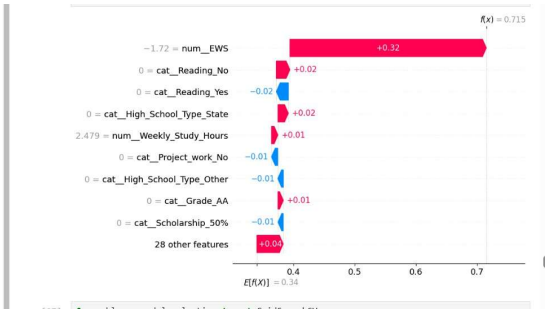

Figure 6. SHAP Waterfall

The explaining analysis of SHAP provides a way of interpreting both globally and locally the predicted outcomes produced by the model. The SHAP Summary Plot is presented in Figure 5 to show the global importance of the features across all predictions made by the model. The most important features (i.e. most important in regard to student academic risk classification) were engagement deficit and attendance risk. Additionally, Figure 6 shows an example of how to use a SHAP Waterfall plot to provide local explanations of how the features combine to contribute to a prediction for a particular student. This allows educators to gain insight into why students are predicted to be at risk for academic failure, therefore aiding educators who want to

provide targeted academic intervention for their students.

XI. DISCUSSION

By combining Early Warning Scores with explainable machine-learning methods we can provide a balance between accurate predictions of at-risk students and providing insight into what factors were used to make those predictions. Ultimately this combination helps to provide educators with the ability to identify at-risk students and understand why those students are predicted to be at risk, thus allowing them to provide targeted academic support to those students.

XII. LIMITATIONS AND FUTURE WORK

This system is limited in terms of historical data - academic record and behavioural data (which are a reflection of student's performance). It is assumed that the historical data is both truthful and relevant. However, in practice, this assumption cannot be held to be true, and, therefore, the overall performance of the model, as well as accuracy of predictions made by the model will vary as a result of the quality defects in the historical data. Another limitation exists because the model only utilizes 'static' data from pre-established datasets. Because the data utilized by the model does not include data from real-time data streams, it is possible that the model does not successfully incorporate the changes experienced by students over an extended period of time (i.e., as time progresses). Therefore, the applicability of the model will depend on the differences between various learning institutions and/or learning environments, particularly as they relate to their respective data distributions and the academic policies they have in place.

A further drawback is that the features and indicators of risk that were predefined and manual created have the potential to overlook some of the underlying characteristics associated with the students being modelled. The use of explainability techniques like SHAP may aid in determining how easy it is to interpret a model's predictions; however, they will not assist in improving the model's predictive accuracy In future iterations of this system, real-time data from learning management systems and student engagement/extracurricular activity sources will be incorporated, and a longitudinal analysis of student performance, can also be tracked and, therefore, provide a more accurate source of evidence, over an extended period of time than was possible in the initial implementation of the model. Continuous assessment of past/string-time analysis, as well as identifying correct predictive accuracy. Other areas for potential improvement might involve using more advanced or sophisticated deep learning models such as recurrent neural networks (RNNs) and/or transformer networks to more accurately model the temporal patterns associated with a student's behaviour. Exploring the possibility of using automated feature selection and/or feature learning techniques vs. reliance on manually selected features may also enhance this system. The ability to extend this system into a personalized recommendation system where students receive tailored recommendations based on their corresponding individual risk factors will also be of value. For instance, providing access to institutional dashboards and mobile applications should further assist educators and administrators with their ability to assist at-risk students. Future research will likely focus on enhancing model generalisation across different educational environments along with using fairness-aware machine learning approaches that will provide non-discriminatory and equitable prediction outcomes for each respective student group.

XIII. CONCLUSION

In summary, this research has provided a predictive analytics framework for identifying students who may be academically at risk through an integrated process that brings together Early Warning Scores; machine learning algorithms; and explainable artificial intelligence methodologies. The predictive analytics system will use multidimensional data collected about the students, such as engagement deficits, risk of attendance, stress from workload associated with schoolwork;

along with weaknesses due to learning self-regulation, to develop a thorough risk profile for each respective student.

The formulation of the Early Warning Score helps aggregate the behaviours and academics into one composite score reflecting the entire at-risk academic level. With the use of machine learning algorithms, such as logistic regression and random forest classifiers, the system is able to detect very complex relationships with the student data and output highly accurate predictions of academic difficulties before they occur. A key contribution of this work is the incorporation of SHAP-based explainability into the predictive framework. Many traditional machine learning models operate as black box systems that leave educators with little faith when it comes to trusting the automated outputs of these methods. The use of SHAP-based explanations gives educators the ability to transparently interpret model output, and this will provide a quantification of how much each feature in the model is contributing to an individual prediction. This ability to interpret the models will allow educators to determine the underlying factors that are influencing a student's academic risk and design interventions appropriately for those at-risk students. Results from the experiments indicate that the proposed framework is achieving predictive performance, while also preserving the interpretability of the model.

The models will have high recall values, which means that the vast majority of at-risk students will be identified as such. This is important for early intervention systems. The explainability portion of the framework further enhances the usefulness of the framework and enables actionable insights to be gained about students' behavior and engagement patterns. Academic institutions can utilize predictive analytics and explainable machine learning combined to create a very effective tool for improving student retention and success rates. The proposed system identifies students who need additional support early on and provides explanatory information about predicted outcomes. This will help educators develop proactive and data-driven academic support plans for their institution.

REFERENCES

[1] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), 2017.

[2] R. S. Baker and K. Yacef, "The State of Educational Data Mining in 2009," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

[3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2016.

[4] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[5] P. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 601–618, 2010.

[6] C. Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005," Expert Systems with Applications, vol. 33, no. 1, pp. 135–146, 2007.

[7] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2012.

[8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.

[10] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, 2017.

[11] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Opti mization," Journal of Machine Learning Research, vol. 13, pp. 281–305, 2012.

[12] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.

[13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the ACM SIGKDD Conference, 2016.

[14] J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[15] M. Kuhn and K. Johnson, Applied Predictive Modeling. Springer, 2013.

[16] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation," Proceedings of

the International Joint Conference on Arti f icial Intelligence, 1995.
[17] S. Raschka and V. Mirjalili, Python Machine Learning. Packt Publishing, 2017.
[18] D. Berrar, "Cross-Validation," Encyclopedia of Bioinformatics and Com putational Biology, 2019.
[19] J. L. Herlocker et al., "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 5–53, 2004.
[20] A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," JAMA, vol. 319, no. 13, pp. 1317–1318, 2018.
[21] B. M. Marlin, "Collaborative Filtering: A Machine Learning Perspec tive," University of Toronto, 2004.
[22] P. Domingos, "A Few Useful Things to Know About Machine Learning," Communications of the ACM, vol. 55, no. 10, pp. 78–87, 2012.
[23] A. G´eron, Hands-On Machine Learning with Scikit-Learn and Tensor Flow. O'Reilly, 2019.
[24] S. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica, vol. 31, pp. 249–268, 2007.
[25] Z. C. Lipton, "The Mythos of Model Interpretability," Communications of the ACM, vol. 61, no. 10, pp. 36–43, 2018.
[26] D. Molnar, Interpretable Machine Learning. Lulu Press, 2020.
[27] C. Molnar, "Interpretable Machine Learning– A Guide for Making Black Box Models Explainable," 2020.
[28] E. M. B. Ferreira and M. H. A. C. Monteiro, "Early Warning Sys tems for Student Performance Prediction," IEEE Access, vol. 8, pp. 123456–123465, 2020.
[29] G. Siemens and R. S. J. D. Baker, "Learning Analytics and Educational Data Mining," Proceedings of the International Conference on Learning Analytics, 2012.
[30] K. Verbert et al., "Learning Analytics Dashboard Applications," Amer ican Behavioral Scientist, vol. 57, no. 10, pp. 1500–1509, 2013.
[31] B. Kitchenham, "Procedures for Performing Systematic Reviews," Keele University Technical Report, 2004.
[32] T. Mitchell, Machine Learning. McGraw Hill, 1997

