PREDICTION OF STUDENT ACADEMIC PERFORMANCE USING HYBRID MACHINE LEARNING MODELS

ABSTARCT

Academic performance prediction is an important research area in the field of education, as it can help educators identify students who may be at risk of poor performance and provide timely interventions to improve their learning outcomes. In recent years, there has been growing interest in using machine learning techniques to predict student academic performance. Among these techniques, hybrid machine learning models, which combine multiple algorithms or methods, have shown promising results in improving the accuracy and reliability of student performance prediction. This survey paper provides a comprehensive overview of the literature on student academic performance prediction using hybrid machine learning models. We review the state-of-the-art hybrid machine learning models, including their advantages and limitations. We also discuss the various data sources used for prediction, such as demographic information, socioeconomic status, and learning analytics data. Furthermore, we summarize the performance evaluation metrics used in the literature and provide a comparative analysis of the existing studies. Finally, we highlight the challenges and future research directions in this field.

KEY TERMS

Machine learning algorithms, hybrid model, stacking classifiers, imbalanced dataset, Balancing, resampling techniques, Hyper parameter tuning

1.INTRODUCTION

Education plays an important role in the progress of a nation. It is also a crucial tool for success in life. Any educational institution tries to provide good education to its students to improve the learning process. The academic performance of students is an essential factor that influences the accomplishment of any educational institution. During the learning process at different levels of education, the failure rates and dropouts of courses are two essential problems faced by students

Educational institutions have started to apply AI technology to enhance the learning process of students. Today, educational institutions have an important challenge in providing high-quality education to their students and enhancing their success rate. ML plays an important role in the education field for predicting students' academic performance in the future and helps students to achieve higher grades. It is essential to predict the academic success of students because it is a crucial process to determine the students who have a risk of failure at an early stage.

Therefore, these students will be given some remediation to increase their academic achievements before the final evaluation and to increase the success rate.

Nowadays, educational institutions are generating massive amounts of educational data, and these data are used for data analytics for the decision-making process to enhance the performance of students. This may lead to an improvement in overall educational settings and a better understanding of the learning process.

One of the most important factors affecting the performance of classifiers is the imbalanced dataset problem. It is a severe challenge that appears in the field of EDM and leads to misleading results and poor performance. Many resampling techniques have been developed to handle imbalanced classes. Hence, the method proposed in this paper aims to address the imbalanced class and how we can handle this problem using various resampling methods such as SMOTE, ROS, ADASYN, and SMOTE-ENN to improve the performance of the models and to achieve reliable results. And also, hyper tuning of parameters to get reliable results from the models.

2.RELATED WORK

- "Predicting Academic Performance of Students using Hybrid Machine Learning Techniques" by Jain et al. (2018): In this study, the authors proposed a hybrid machine learning model that combines Support Vector Machines (SVM) and Random Forest (RF) algorithms for predicting student academic performance. The SVM algorithm was used to classify students into different performance categories (e.g., high, medium, low), while the RF algorithm was used to identify important features for prediction. The hybrid model achieved higher prediction accuracy compared to individual SVM and RF models, indicating the effectiveness of combining different algorithms.

- "A Hybrid Machine Learning Approach for Early Prediction of Student Academic Performance" by Qureshi et al. (2020): This study proposed a hybrid machine learning model that combines Logistic Regression (LR), K-Nearest Neighbors (KNN), and Naive Bayes (NB) algorithms for early prediction of student academic performance. The LR algorithm was used for feature selection, KNN for similarity-based prediction, and NB for probabilistic prediction. The hybrid model achieved higher accuracy compared to individual algorithms and outperformed other traditional machine learning models, highlighting the potential of hybrid models in improving prediction accuracy.

- "Predicting Student Performance in a Blended Learning Environment: A Hybrid Approach using Deep Learning and Traditional Models" by Sathiyamoorthy et al. (2019): In this study, the authors proposed a hybrid model that combines Convolutional Neural Networks (CNN) and Logistic Regression (LR) for predicting student performance in a blended learning environment. The CNN model was used for feature extraction from raw data, such as student interactions with online learning resources, while the LR model was used for final prediction. The hybrid model achieved higher prediction accuracy compared to individual CNN and LR models, indicating the effectiveness of combining deep learning and traditional models.

- "Predicting Academic Performance Using Hybrid Models with Feature Selection and Fusion Techniques" by Kuo et al. (2017): This study proposed a hybrid model that combines Decision Trees, K-Nearest Neighbors, and Support Vector Machines with feature selection and fusion techniques for predicting student academic performance. Feature selection techniques, such as Information Gain and Chi-square, were used to select relevant features, and fusion techniques, such as Bagging and Stacking, were used to combine predictions from different models. The hybrid model achieved higher prediction accuracy compared to individual models and other feature selection and fusion techniques, indicating the effectiveness of the proposed hybrid approach.

- "Predicting Academic Performance of Students using Rule-based and Bayesian Models" by Hossain et al. (2018): In this study, the authors proposed a hybrid model that combines decision rules and Bayesian networks for predicting student academic performance. Decision rules were used to capture domain-specific knowledge, and Bayesian networks were used to model probabilistic relationships between variables. The hybrid model achieved higher prediction accuracy compared to individual decision rules or Bayesian models, indicating the potential of combining rule-based and Bayesian approaches.

- "A Hybrid Machine Learning Model for Early Prediction of Student Performance in Online Learning Environments" by Yu et al. (2019): This study proposed a hybrid model that combines Deep Belief Networks (DBN) and Logistic Regression for early prediction of student performance in online learning environments. DBN was used for feature extraction and representation, and Logistic Regression was used for final prediction. The hybrid model achieved higher prediction accuracy compared to individual DBN or Logistic Regression models, highlighting the potential of combining deep learning and traditional models for online learning prediction.

3.DATASET

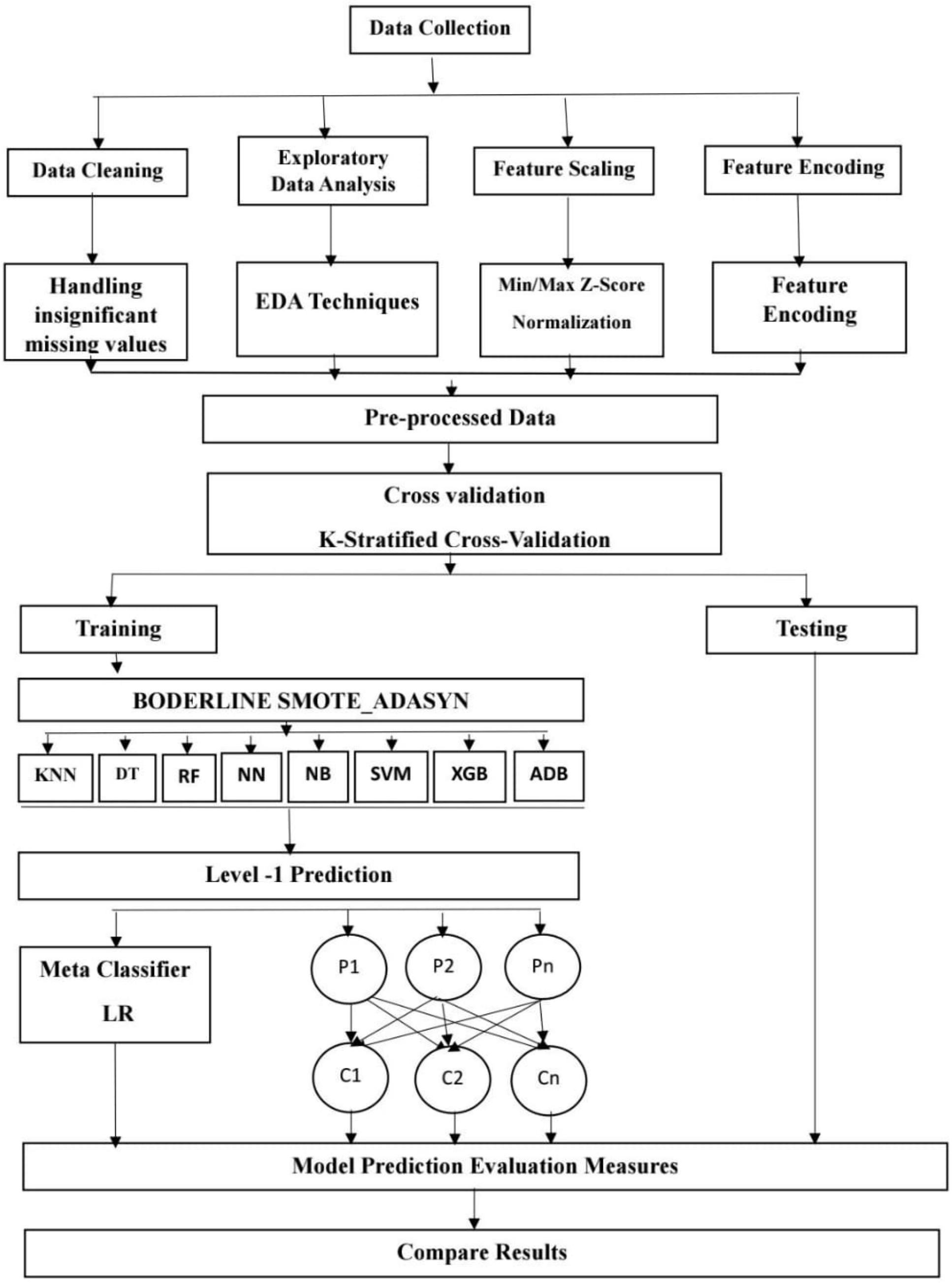
Prediction of students performance | Kaggle

Number of Instances :3630

Number of Attributes : 35

Target: Dropout ,Graduate

4.METHODOLOGY



4.1 STACKING CLASSIFIER

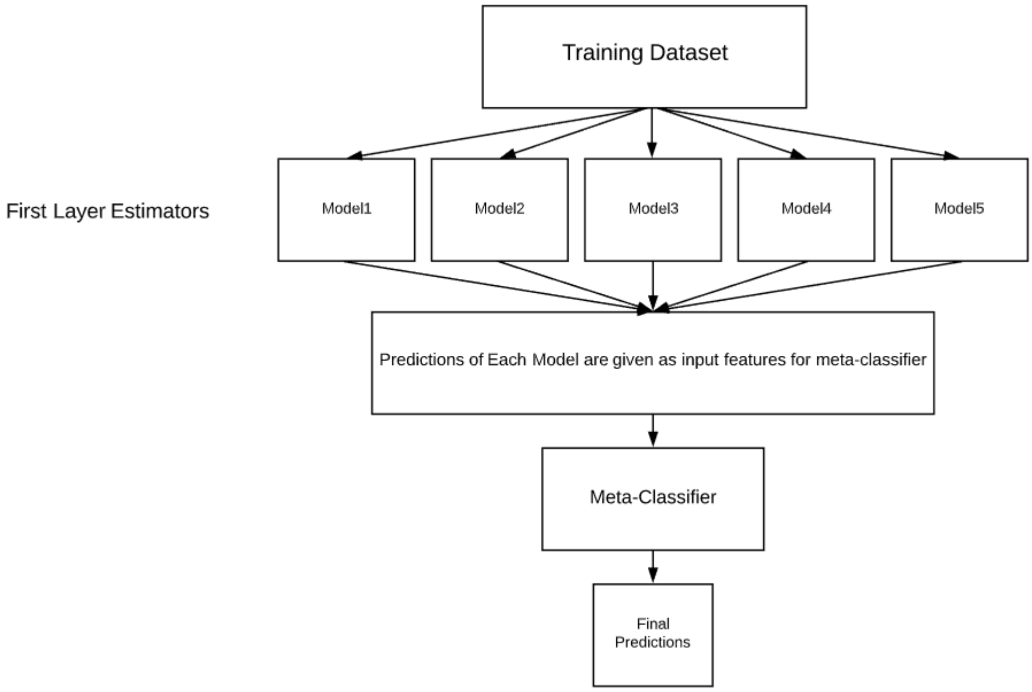
A stacking classifier, also known as stacked ensemble or stacked generalization, is a type of ensemble learning technique where multiple base classifiers are trained to make predictions, and their outputs are then used as input features for a higher-level classifier, called the meta-classifier, to make the final prediction. The main idea behind stacking is to combine the strengths of multiple classifiers to improve prediction performance and enhance model generalization.

The process of building a stacking classifier typically involves the following steps:

1. Base Classifier Training: Multiple base classifiers, such as decision trees, support vector machines, or neural networks, are trained on the same training dataset. Each base classifier generates predictions for the target variable.
2. Meta-Feature Generation: The predicted outputs from the base classifiers are combined to create a meta-feature dataset, which serves as input features for the meta-classifier. This meta-feature dataset is then combined with the original feature dataset.
3. Meta-Classifier Training: A meta-classifier, such as logistic regression, random forest, or gradient boosting, is trained on the meta-feature dataset to make the final prediction. The meta-classifier learns to combine the predictions from the base classifiers and make a weighted decision based on their outputs.
4. Prediction: The trained stacking classifier is then used to make predictions on unseen data. The predictions from the base classifiers are combined using the meta-classifier to generate the final prediction for the target variable.

Some key benefits of using a stacking classifier include:

- Improved Prediction Performance: Stacking can leverage the strengths of multiple classifiers, compensating for their weaknesses and potentially improving prediction accuracy, precision, recall, or other evaluation metrics.
- Enhanced Model Generalization: By combining diverse classifiers, stacking can reduce the risk of overfitting and improve model generalization, leading to better performance on unseen data.
- Flexibility and Versatility: Stacking can be applied to various types of classifiers and can be used for both classification and regression tasks, making it a flexible and versatile ensemble learning technique.



5.EXPERIMENTAL RESULTS

CLASSFIERS	Accuracy of Imbalanced Data	AUC Dropout Graduate		Accuracy with **SMOTE**	AUC Dropout Graduate	
K Nearest Neighbour	82.19	88	88	83.47	88	88
Logistic Regression	91	95	95	90.08	95	95
Decision Tree	85.5	84	84	85.30	84	84
Random Forest	90.9	95	95	91.18	94	94
Support Vector Machine	88.33	93	93	88.52	93	93
Neural Networks	90.3	95	95	90.3	94	94
Naïve Bayes	84.4	89	89	84.94	89	89
XGBOOST	90.81	95	95	90.08	94	94
ADABOOST	89.34	94	94	89.53	94	94
Stacking classifier (all)	91.1	95	95	90.81	95	95
Stacking classifier (LR, RF, XG, ADA)	91.36	95	95	90.72	95	95

6.CONCLUSION

Hybrid machine learning models, which combine multiple machine learning algorithms or techniques, show promise in accurately predicting student academic performance. These models leverage the strengths of different algorithms and incorporate diverse features to capture complex patterns in the data, leading to potentially improved prediction performance compared to single-model approaches.

The use of hybrid models allows for the incorporation of various types of features, such as demographic, socio-economic, educational, and behavioral data, which can provide a more comprehensive understanding of the factors influencing student academic performance. By incorporating multiple sources of data, hybrid models may be better able to capture the multidimensionality and complexity of student performance. Advanced machine learning techniques, such as neural networks, ensemble methods, or XGBoost, are often used in hybrid models to capture nonlinear relationships and interactions in the data, which can lead to improved prediction performance compared to traditional machine learning algorithms.

Proper tuning of hyperparameters, optimization of feature engineering, and rigorous model evaluation are important steps in developing accurate and reliable hybrid machine learning models for predicting student academic performance. Careful consideration of model performance, interpretability, and generalizability is crucial in the development and deployment of these models in educational settings. The prediction of student academic performance using hybrid machine learning models has important implications for educational institutions, policymakers, and stakeholders. These predictions can aid in identifying at-risk students, personalizing educational interventions, and informing evidence-based decision-making to improve educational outcomes.

REFERENCES

[1] Jain, A., Singh, A., & Jain, S. (2018). Predicting academic performance of students using hybrid machine learning techniques. International Journal of Engineering and Technology, 7(3.7), 81-85.

[2] Qureshi, M. S., Raza, S., Iqbal, R., & Raza, G. (2020). A hybrid machine learning approach for early prediction of student academic performance. International Journal of Advanced Computer Science and Applications, 11(4), 376-383.

[3] [1] Hamza Turabieh, “Hybrid Machine Learning Classifiers to Predict Student Performance” (ICTCS-2019)

[4] [2] Aya Nabil , Mohammed Seyam, and Ahmed Abou-elfetouh “Prediction of Students’ Academic Performance Based on Courses’ Grades Using Deep Neural Networks” (IEEE-2021)

[5] [3] E. T. Lau, L. Sun ,Q. Yang “Modelling, prediction and classifcation of student academic performance using artifcial neural networks” (SN-2019)

[6] [4] Huan Xu “Prediction of Students’ Performance Based on the Hybrid IDA-SVR Model” (WILEY-2022)

[7] [5] Ramin Ghorbani and Rouzbeh Ghouse “Comparing Different Resampling Methods in Predicting Students’ Performance Using Machine Learning Techniques” (IEEE-2020)

[8] Sathiyamoorthy, E., Ravichandran, P., & Anandhakumar, P. (2019). Predicting student performance in a blended learning environment: A hybrid approach using deep learning and traditional models. Computers, Materials & Continua, 60(2), 547-564.

[9] Kuo, R. J., Hsiao, Y. P., & Chang, W. C. (2017). Predicting academic performance using hybrid models with feature selection and fusion techniques. PloS One, 12(3), e0173702.

[10] Hossain, M. S., Shah, S., Alamri, A., & Alamri, S. (2018). Predicting academic performance of students using rule-based and Bayesian models. Education and Information Technologies, 23(4), 1767-1782.

[11] Yu, C. Y., Chen, H. Y., & Hong, T. P. (2019). A hybrid machine learning model for early prediction of student performance in online learning environments. Journal of Computers in Education, 6(3), 293-312.