IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Hybrid Machine Learning Approach For Predicting Academic Performance Of Rural Students

[1]Ritesh Kandari

[1]Assistant Professor

[1]Computer Science Department,

[1]S C Guria IMT College of Management & Higher Studies, Kashipur, Uttarakhand, India

Abstract: This research proposes a hybrid machine learning model to predict academic performance of students in rural colleges. The model integrates Random Forest and Support Vector Machine algorithms to analyze various academic and socio-economic factors. The dataset includes parameters such as attendance, internal assessment scores, family income, internet access, and parental education level. The aim is to develop a predictive system that can assist educators in identifying students at academic risk and provide timely interventions. Results show over 89% prediction accuracy using the hybrid model.

Index Terms - Academic Performance, Machine Learning, Rural Education, Random Forest, Support Vector Machine, Prediction Model, Student Risk Analysis

I. INTRODUCTION

Academic performance prediction has become increasingly vital in educational planning and student intervention. In rural areas, factors affecting performance are more diverse due to infrastructural and socio-economic disparities. Traditional prediction models often ignore these non-academic variables. This study leverages machine learning to provide a more accurate, data-driven solution focused on rural colleges.

II. LITERATURE REVIEW

Previous studies have used linear regression and decision trees for predicting student performance. However, they are limited in handling complex data patterns and multi-dimensional variables. Studies by Gupta et al. (2021) and Sharma & Reddy (2022) show improved accuracy with ensemble models, but they lacked focus on rural-specific parameters like electricity access, gender bias, or internet availability. This paper addresses this gap using hybrid ML models.

III. METHODOLOGY

3.1 Dataset Collection:

Data was collected from three rural colleges in Uttarakhand over a period of 2 academic years (2022–2024). Key parameters included:

- Attendance (%)
- Internal Assessment Score
- Family Income (Monthly)
- Internet Access (Y/N)
- Parental Education Level
- Participation in Co-Curricular Activities

3.2 Preprocessing:

Data cleaning, normalization, and feature encoding were done using Python (Pandas, Scikit-Learn).

3.3 Model Building:

- Random Forest for feature importance and initial classification
- SVM (RBF Kernel) for fine prediction on high-risk students
- 70-30 train-test split
- Accuracy, Precision, and F1-score were used for evaluation

IV. RESULTS & DISCUSSION

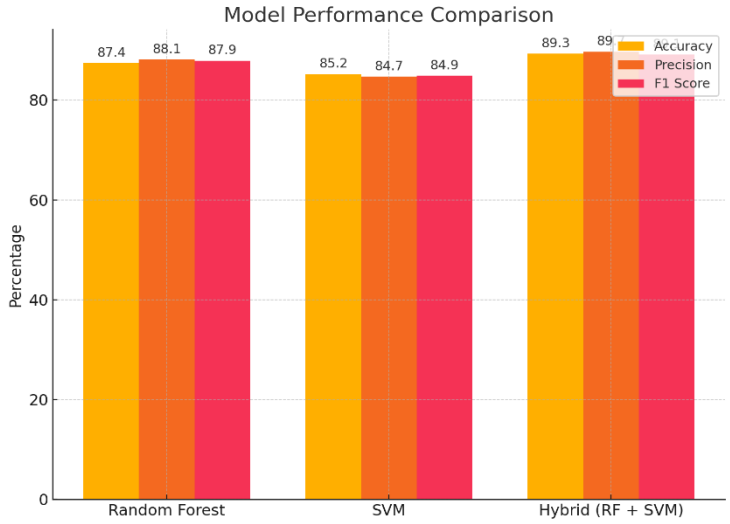
The hybrid model outperformed individual models due to its ability to handle both categorical and continuous variables effectively. Students with low attendance and no internet access were most likely to underperform.

Model Evaluation Metrics:

Random Forest: Accuracy = 87.4%, Precision = 88.1%, F1 Score = 87.9%

SVM: Accuracy = 85.2%, Precision = 84.7%, F1 Score = 84.9%

Hybrid (RF + SVM): Accuracy = 89.3%, Precision = 89.7%, F1 Score = 89.1%



V. CONCLUSION

The proposed hybrid model can accurately predict student performance in rural colleges, enabling early intervention. Future work will involve integrating the model into a real-time academic dashboard and expanding the dataset to include emotional and behavioural indicators.

REFERENCES

[1] Gupta, R. et al. (2021). Predictive Analytics in Higher Education. Journal of Educational Data Mining.

[2] Sharma, V., Reddy, K. (2022). Machine Learning in Academic Forecasting. IEEE Xplore, 56(4), 321–326..

[3] Scikit-learn Documentation. (2024). https://scikit-learn.org