

Paper ID: MISS21\_98

# Prediction and Analysis of Dropout Secondary Students in India Using Machine Learning Classification Algorithms

Sagar Chakraborty\*, Arpan Ghoshal

Seacom Engineering College, Jaladhulagori, Howrah, West Bengal, India

\*Corresponding author

## Abstract

**Background:** As per the UNICEF SDG-4, "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" should be achieved by 2030. To build a healthy society, education should be reached to all. According to the MHRD educational statistical report [1], the dropout rate at the secondary school level in India is more than 17%, while the dropout rate at upper-primary and primary levels is 1.8% and 1.5% which itself is a very alarming statistic. Secondary level education is of utmost necessity for acquiring skills and knowledge without which socio-economic development of individual being will be affected which will create a barrier to the overall development of a country. So, early prediction for Secondary school dropout in India is very essential to achieve SDG-4 by 2030. Some of the early works regarding the prediction of Secondary school dropouts are given here. "IoT System for School Dropout Prediction Using Machine Learning Techniques Based on Socioeconomic Data" proposed by Francisco A. da S. Freitas [2]. "PREDICTION OF SCHOOL DROP OUTS WITH THE HELP OF MACHINE LEARNING ALGORITHMS" was proposed by Hassan [3].

A Machine Learning Approach to Identify the Students at the Risk of Dropping Out of Secondary Education in India" was proposed by Nangia [4].

**Objective:** To predict and analyze the secondary school dropout students in India using Machine Learning classification algorithms like Logistics Regression, Naive Bayes, K-Nearest Neighbours (KNN), Decision Tree, Random Forest algorithms. After analyzing the performance of all the above-mentioned models, propose the best possible model for accurately predicting secondary school dropout students of India.

**Methodology:** The student dropout data is collected from the Ministry of Education-Govt. of India (<https://www.education.gov.in/en>). This dataset has been applied to the Logistic Regression, KNN, Naive Bayes, Decision Tree, and Random forest machine learning models. We use the Orange Data Miner tool for predicting the Dropout students of India using the above-mentioned algorithms and also measure the performance of each algorithm. We also measure the RMSE values for each algorithm.

**Result and discussion:** After measuring the accuracy and performance of the Logistic Regression, Decision Tree, Random Forest, and KNN models, it is found that the Random Forest model is the most accurate and robust model compared to other classification models.

**Future Work:** Prediction will be more accurate if we consider additional socio-economical and geographical parameters in this model. Deep learning models can be used to find more accurate predictions.



## References

- [1] Ministry of Education - Government of India, <https://www.education.gov.in/en/statistics-new>
- [2] Freitas, Francisco A. da S. and Vasconcelos, Francisco F. X. and Peixoto, Solon A. and Hassan, Mohammad Mehedi and Dewan, M. Ali Akber and Albuquerque, Victor Hugo C. de and Filho, Pedro P. Rebouças , "IoT System for School Dropout Prediction Using Machine Learning Techniques Based on Socioeconomic Data"
- [3] Hassan, Malik & Mirza, Tabasum, "PREDICTION OF SCHOOL DROP OUTS WITH THE HELP OF MACHINE LEARNINGALGORITHMS"
- [4] Nangia, Sagarika & Anurag, Jhilmil & Gambhir, Ishani, "A Machine Learning Approach to Identify the Students at the Risk of Dropping Out of Secondary Education in India." 10.1007/978-981-15-2475-2\_51.