# PREDICTING STUDENT DROPOUT IN GOVERNMENT SCHOOLS

Anjana V Murthy, Harshitha Batta, Neha Ganesh Shastri, Trupti J and Prof. Suresh Jamadagni

Department of Computer Science and Engineering, PES University

## Problem Statement

There has been growth in the educational sector in India over the years, but there is still a major concern and challenge of students dropping out in the early stages of education.

By predicting the students that are likely dropout and understanding the various factors due to which they leave schools, we can help retain these students and ensure their education.

## Background

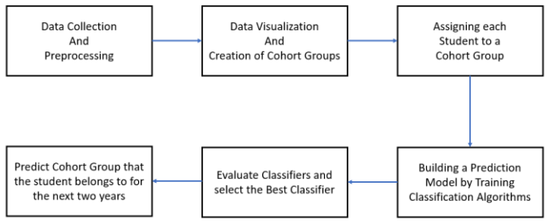
The basic idea of the project is to identify the socio-economic factors due to which students of government schools decide to discontinue their education and predict which students are likely to dropout.

## Dataset and Features / Project Requirements / Product Features

The Data has been collected from Urban Areas like North Bangalore and Rural areas like Tarikere, Chikkamagaluru.

The dataset consists of various socio-economic features like Parents Occupation, Income and Education along with Age and Siblings.

## Design Approach/Methods




A Cohort analysis has been performed on the data collected for the students who have dropped out and various classifiers have been compared to find the best model to predict the cohort group the student belongs to.

## Result and Discussion

On performing cohort analysis on students that have dropped out from Rural and Urban Areas a total of 28 and 35 cohort groups have been formed.

For predicting the cohort group and determining the dropout rate of students studying in schools, various classification algorithms: Naive Bayes, Decision Tree, Random Forest and Gradient Boost have been trained and based on the results Random Forest performed the best with 99.4% Accuracy.

## Summary of Project Outcome

Based on the Urban data 28 cohorts have been identified with corresponding reasons for dropping out and similarly for Rural 35 cohort groups have been formed.

For both Urban and Rural Areas, Random Forest yielded the best results.

A user interface is built to identify the risk and rate of dropout of students currently studying in schools along with a few graphs describing the distribution of results

## Conclusion and Future Work

Cohort analysis has been performed on the students who have dropped out and various prediction algorithms have been compared.

For both Urban and Rural Areas, Random Forest yielded the best results.

This application would be very useful to identify students with high risk of dropping out and suitable actions can be taken to keep them in school

## References

1. Malik Hassan, Tabasum Mirza, Prediction of School drop outs with the help of machine learning algorithms", GIS SCIENCE JOURNAL, July 2020.
2. Ekansh Maheshwari, Chandrima Roy, Manjusha Pandey and Siddharth Swarup Rautray, "Prediction of factors associated with the dropout rates of primary to high school students in India using Data Mining Tools", Frontiers in Intelligent Computing: Theory and Applications (pp.242-251), January 2020.
3. Syed Rooquiyya Tabassum," Study of School Dropouts in India: Sensitivity at Display" Journal of Applied Linguistics and Language Research Volume 6, Issue 3, 2019, pp. 1-10"
4. Sateesh Gouda, M and Dr.T.V Sekher. "Factors Leading to School Dropout in India: An Analysis of National Family Health


Anjana V Murthy


Neha Shastri


Trupti J


Harshitha Batta


Professor Suresh Jamadagni