

# **Analysis of School Performance in Delhi, India using Decision Tree Algorithm**

**Abhishek Bhatnagar and Nomes B. Bolia**

Department of Mechanical Engineering,

Indian Institute of Technology Delhi

New Delhi 110016, India

[abhishekjmi11@gmail.com](mailto:abhishekjmi11@gmail.com), [mez158446@iitd.ac.in](mailto:mez158446@iitd.ac.in); [nomesh@iitd.ac.in](mailto:nomesh@iitd.ac.in)

## **Abstract**

Education data mining is data mining on an education dataset with an emphasis to understand the underlying patterns in student learning and its performance. The present study pertains to an analysis of performance of schools in Delhi, India using a decision tree approach. This helps to classify the schools that according to their performance. Data for government schools in Delhi, India for grade XII has been procured from the Delhi government education portal and the Unified District Information System for Education (UDISE) portal.

There are three measures of the school performance, namely, score in the individual subjects, quality index of a school and the pass percentage. Scores in a subject is the average score obtained by the students who have appeared in the given subject. Quality index of a student is the total score across the five subjects in which the student has appeared. Quality index of a school is the average value of quality index of the individual students of a school. Pass percentage is the percentage of students that have obtained the minimum passing score in all the subjects. The aim of schools is to maximize their pass percentage (typically, 100%) as well as the scores, however, first it is desired to improve the pass percentage to have minimum levels of attainment for all the students.

The school input data used in the analysis pertains to its infrastructure that includes classrooms, playground, library, etc; its operations that comprises the medium of instruction, implementation of continuous and comprehensive evaluation, number of working days, working hours per day, number of inspections by external officials and other concerned authorities; school enrolment, and number and distribution of teachers; type of school (boys', girls' or co-educational) and school category (based on grades offered). The variables that are correlated have been removed to yield a reliable rule set.

The schools are classified based on whether they achieve the minimum acceptable level of performance. The schools are classified with two target labels: (i) those that achieve minimum acceptable performance, (ii) those that do not have minimum acceptable performance. The decision rule set can be used by the policy makers to identify and take appropriate corrective actions in the schools to help them with minimum acceptable performance. The work can be extended to other classification algorithms as well.

## **Keywords**

School Performance, Decision Tree, Classification and Educational Attainment.

## **Acknowledgements**

The authors would like to thank Animesh Jaiswal, Ansh Jain and Utkarsh N. Singh (Alumni of IIT Delhi) for their help in downloading and processing the data. The authors would also like to thank Maansi Gupta (PhD student, IIT Delhi), and, Prem Chand and Rama Shankar for carrying out related work.

## **Biographies**

**Abhishek Bhatnagar** is a Doctoral Candidate in the Department of Mechanical Engineering, IIT Delhi. He completed his B.Tech in Mechanical Engineering from Jamia Millia Islamia followed by M.Tech in Industrial Engineering from IIT Delhi. He is the recipient of the Ministry of Urban Development's fellowship for the academic year 2015-16 & 2016-17. His areas of interest are Operations Research, Statistics, Operations Planning and Control and Public Policy. He currently works on efficient planning and management in school education.

**Nomesh B. Bolia** is a Professor in the Department of Mechanical Engineering, IIT Delhi. He also has two years of work experience at Infotech Financials Pvt Ltd, Mumbai and TIFR, Mumbai. He received his B.Tech in Mechanical Engineering from IIT Bombay and PhD from University of North Carolina at Chapel Hill. His areas of research are Operations Research, Stochastic Modeling; their applications to Health Systems, Transportation and Governance. His past works include projects with Container Corporation of India (CONCOR), Philips, Department of Science and Technology, Government of India, HUDCO, JCB, International Water Management Institute, Central Square Foundation etc.