Design and Development o Data Mining Models for The Prediction of Manpower Placement in the Technical Domain

Data mining is one of the hottest research areas nowadays as it has got wide variety of applications in common man's life to make the world a better place to live. It is all about finding interesting hidden patterns in a huge history data base. As an example, from a sales data base, one can find an interesting pattern like "people who buy magazines tend to buy news papers also" using data mining. Now in the sales point of view the advantage is that one can place these things together in the shop to increase sales. In this research work, data mining is effectively applied to a domain called placement chance prediction, since taking wise career decision is so crucial for anybody for sure. In India technical manpower analysis is carried out by an organization named National Technical Manpower Information System (NTMIS), established in 1983-84 by India's Ministry of Education & Culture. The NTMIS comprises of a lead centre in the IAMR, New Delhi, and 21 nodal centres located at different parts of the country. The Kerala State Nodal Centre is located at Cochin University of Science and Technology. In Nodal Centre, they collect placement information by sending postal questionnaire to passed out students on a regular basis. From this raw data available in the nodal centre, a history data base was prepared. Each record in this data base includes entrance rank ranges, reservation, Sector, Sex, and a particular engineering.

From each such combination of attributes from the history data base of student records, corresponding placement chances is computed and stored in the history data base. From this data, various popular data mining models are built and tested. These models can be used to predict the most suitable branch for a particular new student with one of the above combination of criteria. Also a detailed performance comparison of the various data mining models is done.

This research work proposes to use a combination of data mining models namely a hybrid stacking ensemble for better predictions. A strategy to predict the overall absorption rate for various branches as well as the time it takes for all the students of a particular branch to get placed etc are also proposed.

Finally, this research work puts forward a new data mining algorithm namely C 4.5 * stat for numeric data sets which has been proved to have competent accuracy over standard benchmarking data sets called UCI data sets. It also proposes an optimization strategy called parameter tuning to improve the standard C 4.5 algorithm.

As a summary this research work passes through all four dimensions for a typical data mining research work, namely application to a domain, development of classifier models, optimization and ensemble methods.