University Admit Eligibility predictor

Literature survey

Every year the number of students seeking admission for the graduate studies is constantly increasing. As a result the competition gets tougher and the chances of admission becomes unpredictable. Given the growth of new programs and number of admission, a student is often unaware of the existence of such programs. As university education has become a basic part of most people's preparation for working life, admission to university is henceforth a topic of importance. How a student chooses a university, and conversely how a university chooses a student, determines the success of both sides in carrying through the education.

Recommendation Problem

In today's fast-paced world, every technological innovation influences the importance of higher education. Being the top destination for higher education universities of United States enroll thousands of international students every year. For students across the world who wish to pursue graduate studies in the USA, choosing a suitable college and universities and earning an admit is a challenge. Although, many Internet resources and forums are available, they do not offer satisfactory suggestions, as most of them are based on assumptions from college rankings and not the actual statistical relations. From a student's perspective cost of application and amount of dedication and determination to the process is also high. Hence, to guide the students in an efficient manner, the University Recommendation and Admission Prediction System has to be developed, based on the input of the students' academic data. The choice of a university that is suitable for a any undergraduate student can be difficult decision to make. Reputation of the university, perceived difficulty of the degree program, distance from home, tuition and living cost, student's areas of academic strength as well as actual scores achieved are just some of the factors that may be considered by a undergraduate student. Likewise, the university has its own set of admission

criteria, mainly based on academic standard of the student to be admitted, admission based onminority and gender representation, local vs. domestic vs. overseas student proportion. Choosing the most suitable among the many thousands of candidates that apply to a university every year is not a trivial matte.

Related Work

In the past, a lot of work on employing data mining techniques in the field of higher education were done. Few recommender systems to suggest course and university based on a student's academic record were developed. Those systems employed decision tree classifier and fuzzy c-means clustering techniques using WEKA tool kit and it was aimed to help the students choose a stream which will suit their skill sets. Another different recommender system was built to help students with their academic stats. They help in making decisions about what course to select based on a student's schedule, stream and professors. Here the model was trained based on past 7 years data for particular university and classifiers for every subject was modeled based on cumulative GPA. On the other hand, some recommendation system were modeled to help university to know about their students by keeping track of their time, extra-curricular activities and achievements, in addition to their academic potential. This helps them to identify and categorize the students depending on the need using two step algorithm and K-means. However, there was no access to any of the data-set used in the above mentioned works.

Data Mining Techniques

Finding patterns in data dates back to 6th century BC succeeding the invention if the abacus made of bamboo rod in ancient China. Ancient China and Greece used statistics to help administrators govern monetary and military matters. In the eighteenth century, two branches of statistics evolved. The two branches were classical statistics and Bayesian statistics. Classical statistics was inspired from mathematical works of Laplace and Gauss, which considered the joint probability. Bayesian statistics, on the other hand, considered the probability of an event occurring will be equal to the probability of its past occurrence multiplied by the likelihood of its future occurrence. Data mining techniques use either approach. Data mining is a comparatively new field in statistics. One

of the early definitions of data mining from Frawley et al.(1991) defined data mining as the non trivial extraction of implicit, previously unknown, and potentially useful information from data. John(1997) explained data mining as a new name for a previous process of finding patterns in data. John clarifies that the search for patterns in data has neen going in in different fields for a long time, but a common name like data mining brought all these different fields together to focus on universal resolutions. Knowledge discovery and data mining are very closely related. “Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potential useful and ultimately understandable patterns in data”(Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996).

Most definitions of data mining focus on finding patters in the data and model building. Initial definitions of data mining was confined to the process of model building but was later extended to model evaluation. Finding patterns in data is a complicated process that can be accomplished by statistical algorithms that can also evaluate variable relationships. The modern knowledge discovery process combines the algorithms to find patterns and also evaluates the entire process of building and evaluating models

Hand et al.(200) summarized some of the major data mining activities as follows:

1. Exploratory data analysis: includes techniques that help in inspecting a data set using graphical charts and descriptive statistics.

2. Descriptive modelling: includes finding probability distributions, finding relationships between models and partitioning the data into groups.

3. Predictive modelling: includes building a statistical model to predict one variable using another variable.

4. Discovering patterns and rules: includes activities from finding combinations of items that occur frequently in transaction databases.

5. Retrieval by content: includes activities of finding patterns in a new data set similar to some know pattern of interest.

Data Mining Applications in Higher Education

Data mining techniques are extensively used in business applications. Nearly all data mining techniques used in business applications can be applied in solving higher education problems. Luan (2002) clarifies some data mining questions in the business sector and their equivalent in higher education:

1. Business question: Who are my most profitable customers? (Equivalent in higher education: Who are the students taking more credit hours?)

2. Business question: Who are my repeat website visitors? (Equivalent in higher education: Which students are likely to return for more classes?)

3. Business question: who are my loyal customers? (Equivalent question in higher education: Which students persist and graduate?)

4. Business question: who is likely to increase their purchase? (Equivalent in higher education: Which alumni are likely to donate?)

5. What clients are likely to defect to my rivals? (Equivalent in higher education: what type of courses can we offer to keep our students?)