ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Student Progress Analysis and Educational Institutional Growth Prognosis Using Data Mining

S.Saranya[*1], R.Ayyappan[2], N.Kumar[3]

[*1,2] M.E Student, [3] Assistant Professor, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamil Nadu, India

sammanmeena.sharan@gmail.com

Abstract

Educational organization is one of the important parts of our society, playing a vital role for growth and development of any nation. Mining educational institution's information system can prove to be of great use to students as well as to the institution. The proposed system can aid in the betterment of student's performance by figuring educational, co-curricular, extra-curricular, behavioral and overall performance. Educational institutions can reap from the system- scope of different courses, best performing student, key areas to improve on, job placement issues. Various techniques in data mining such as data cleaning, data integration, and classification and regression analysis are used. The graphical output of the system is made by incorporating GoogleVis in the paper. The system is aimed to develop a faith on Data mining techniques so that present education and business system may adopt this as a strategic management tool.

Keywords: Data mining, Naive Bayes algorithm, GoogleVis, postgresql, EIG Prognosis

Introduction

Objective

So far, Education Institutional growth Prognosis has not been studded in BI **[3]** and has not been embedded in EDM. Using EDM to implement functionalities of BI **[3]** drives us to **'Educational Institutional Growth Prognosis '**- a bridge connecting EDM and BI. Also, EDM is aimed to understand students' learning strategies but not a **holistic student analysis**. But the proposed system regards the student's holistic performance. **Naive Bayes classification algorithm** [8] is used for classification, as it has quicker convergence. GoogleVis [7]. API is used to create interactive graphical output.

Overview

Student Progress Analysis and Educational Institutional Growth Prognosis help students and Education Institution's Management to improve their performance [1]. Introspection is a major factor towards improving performance, which is often neglected or forgotten. The paper introspects user's data and Institutional data and presents a graphical output for easy understanding. Traditionally, Educational Data Mining and Business Intelligence **[3]** remain isolated. The paper combines the two entities, mining student data and Institutional data. Students can utilize Education data mining [5]. detail in the paper to witness their performance and scrutinize them. Management personnel can utilize Business Intelligence [3] detail of the paper to improve the growth of the Institution as well as to improve the students. Periodic analysis of their growth of the users using the paper can be a great help in analyzing and predicting one's progress.

Various techniques in data mining [6] such as data cleaning, data integration, and classification and regression analysis are used. Data from the database is analyzed for discrepancy and missing entries. Identified instances are then cleaned and filled with a papered value. Missing instances are filled using 'similar cases' approach. Data mined from different tables are integrated together for effective visualization. Register Number of students is used as the common attribute in each datum. The paper implements Naïve Bayes algorithm to classify instances, instead of using SVM algorithm [10] which is used in the existing system. Naïve Bayes algorithm proves to be more worthy when analyzed using versatile training data. Classification is used to segregate students into three classes – Elite, Average and Poor. Student intake prediction is predicted using regression analysis techniques.

ISSN: 2277-9655
Impact Factor: 1.852

The graphical output of the system is made a treat to the eyes by incorporating GoogleVis in the paper. GoogleVis is a custom API from search engine giant Google. GoogleVis [7] Enables users to interact with the graphical output. Tables in the output provide options for the user to sort table with respect to a column. Pie charts and bar charts also provide dynamic interaction to the user.

Purpose

The purpose of the paper is to help students and management personnel to improve their performance. Students can use the system to analyze their academic performance, co-curricular activity analysis, extra-curricular activity analysis, job placement probability and improvement factor. On the other hand, management personnel can use the system to find bottlenecks and measure their performance. Management personnel can avail academic analysis, job placement probability, student intake prediction, student ranks and factors to improve on. Students and institution stakeholders of the institution can be highly benefited from this.

Scope

The paper reaches to the student community and ensures that they can improve tremendously in the realms of academics, extra and co-curricular activities, employability skills and behavioral skills. Management personnel can utilize the paper to analyze performance of students in academics, extra and co-curricular activities, employability skills and behavioral skills, view rank of students and can get to know the bottlenecks and hurdles in growth. Management personnel can analyze the performance of students batch-wise, department-wise and section-wise.

System Analysis

Existing System

Existing systems considers EDM and BI as distinct entities, as a result educational institutions may not be able to unleash Mining Intelligence under one roof. Holistic student performance is neglected in the existing system.SVM classifier[10] used in existing systems prove to be slow in earning contrasting Naive Bayes classifier[8].Interactive output is not available in any of the existing systems

Proposed System

The proposed system incorporates BI **[3]** in EDM, resulting in an integrated solution. Better performing Naïve Bayes algorithm is used. The paper aims to reach to the student community to ensure that they can improve tremendously in the realms of academics, extra and co-curricular activities, employability skills and behavioral skills. Management personnel can utilize the paper to analyze performance of students in academics, extra and co-curricular activities, employability skills and behavioral skills, view rank of students and can get to know the bottlenecks and hurdles in growth. Management personnel can analyze the performance of students batch-wise, department-wise and section-wise.

System Design And Methods

System Architecture

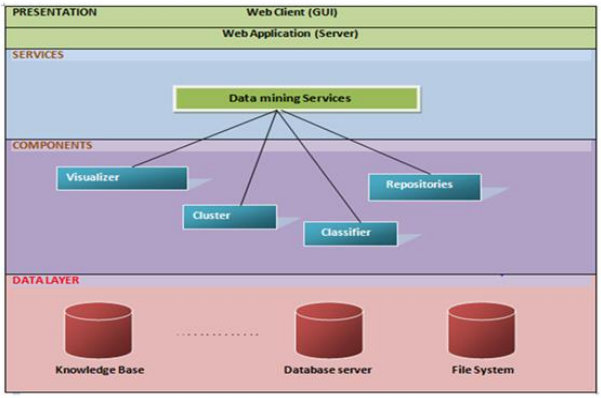


Fig.1: Architecture diagram

In this fig.1 we showed an overall architecture of the system. In presentation layer graphical user interface is a web client and server is a web application, it will get input from the user and process by executing queries in background and retrieve the information from the database by web application. Then service layer, the services have been provided by data mining [6] using various components like visualize for visualizing the result graphical form, cluster [9] to group the related fields for showing the result in simplified manner, classifier will classify the data by training the data using some datasets, In data layer data will be processed from the data server and file servers.

System Methodologies

Naïve Bayes Classifier:

Naïve Bayes classifier [8]. is used to classify instances in management personnel detail. This algorithm is used in academic analysis, co-curricular analysis, extracurricular analysis and job placement analysis and Training data**.**

Linear Regression:

Linear regression analysis is used to predict student intake in the university

ISSN: 2277-9655
Impact Factor: 1.852

[2]. Factors such as employability ratio and student count are taken into consideration for prediction.

Each of these factors is given a weight and a two variable equation is formed to apply single variable linear regression technique.

Data Cleaning:

The system uses the similarities between the rows (observations) to fill in missing values. Euclidean distance is used to obtain missing values. Package 'Cluster' [9] is used to achieve this.

System Implementation

Student detail

Job Placement Prognosis

Individual student job placement probability is measured. Parameters include arrear history, CGPA, radiating* and receptive** participation and extra-curricular activities

Description

Specified parameters are fetched from the database and are cleaned thoroughly. The fields, arrear history and standing arrears are normalized to a scale of 10. Employability score is calculated with the normalized values. Naïve Bayes classifier [8] is used to classify the student to any of three classes namely highly probable, fairly probable and least probable for a job.

* radiating participation- paper presentation, paper developed, seminars presented etc

** Receptive- workshops, seminars attended etc.

Input: CGPA, arrear history, extra-curricular and co-curricular activities and employability skills

Output: A bar chart of parameters used for the prediction and the class of job probability the student belongs to.

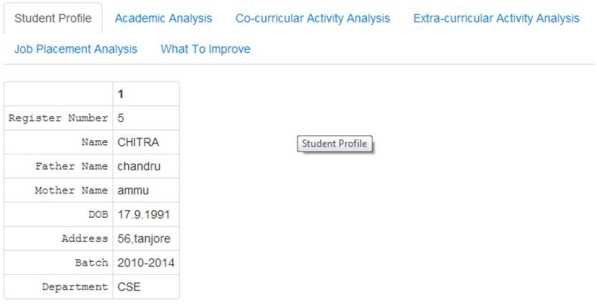


Fig.2

Academic Performance Analysis

Student's academic progress is analyzed to fall into elite performer, average performer or poor performer class.

Description

Given the register number of a student, the student can view a bar plot of his academic performance. The category of his academic class- elite or average or poor performer is also displayed. The class of the student is obtained using a naïve Bayer classifier.

Input: Arrear history, CGPA, Radiating, Receptive, Extracurricular activities

Output: pie chart and an academic class

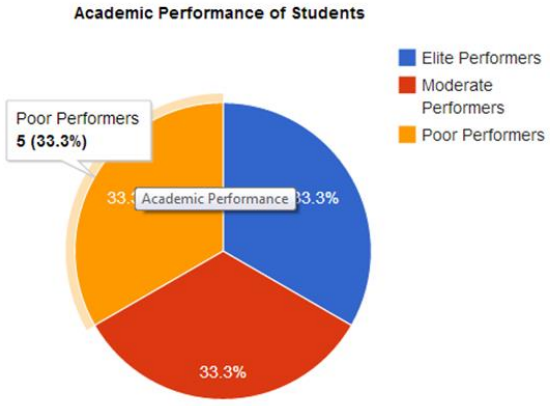


Fig.3

Co-Curricular Activity Analysis

Student's Co-curricular activity is analyzed to fall into elite performer, average performer or poor performer class.

Description

Given the register number of a student, the student can view a bar plot of his co-curricular activity performance. The category of his activity class- elite or average or poor performer is also displayed. The class of the student is obtained using a naïve Bayer classifier.

Input: Poster presentation, paper presentation, workshop and register number.

Output: Bar plot and a activity class

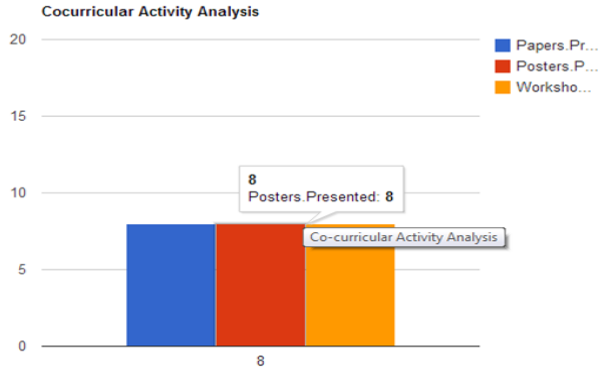


Fig.4

ISSN: 2277-9655
Impact Factor: 1.852

Exrtacurricular Activity Analysis

Student's extracurricular activity is analyzed to fall into elite performer, average performer or poor performer class.

Description

Given the register number of a student, the student can view a bar plot of his co-curricular activity performance. The category of his activity class- elite or average or poor performer is also displayed. The class of the student is obtained using a naïve Bayer classifier.

Input: NSS, NCC, Sports and register number.

Output: Bar plot and a activity class

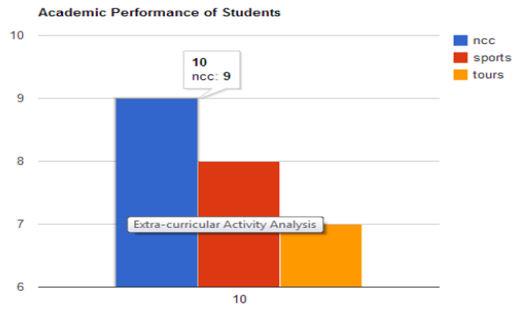


Fig.5

Identifying Bottlenecks

A realm in which the student has to improve on is mined here.

Description

Bottleneck in a student's progress is analyzed and is displayed as a sliced chart. The best and worst areas are explicitly highlighted.

Input: Academics, Seminar Attended, Paper Presented, Poster Presented, Paper Developed, IPT Attended, NSS, NCC, Sports, Workshop, Aptitude Skills, Communication, Body Language and Behavioral Skills.

Output: pie chart and an activity class

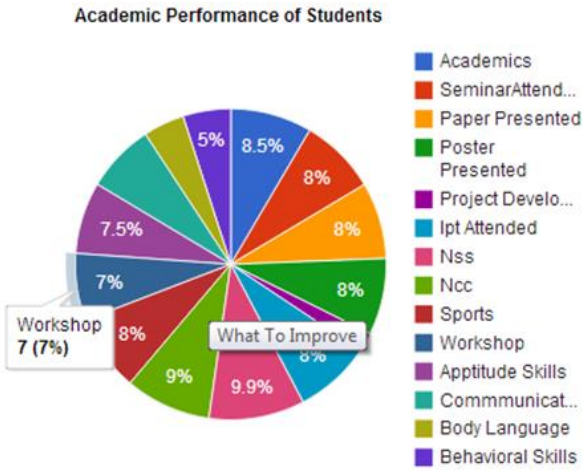


Fig.6

Management Personnel Detail

Job Placement Analysis

Analyzes what percentage of the student community is likely to be placed in recruiting companies.

Description

Specified parameters are fetched from the database and are cleaned thoroughly. The fields, arrear history and standing arrears are normalized to a scale of 10.Employability score is calculated with the normalized values. Naïve Bayes classifier is used to classify the instances to any of three classes namely highly probable, fairly probable and least probable for a job.

* radiating participation- paper presentation, paper developed, seminars presented etc

** Receptive- workshops, seminars attended etc.

Input: CGPA, arrear history, extra-curricular and co-curricular activities and employability skills

Output: A pie chart of the three different classes of job probabilities.

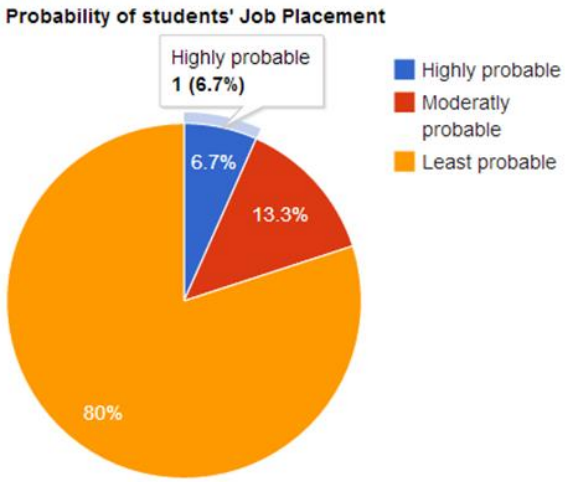


Fig.7

Student Intake Prognosis

A predictive analysis of student intake **[4]** is visualized here.

Description

A toll on the student count is taken and linear regression is used to predict the intake of student for the forthcoming years.

Input: ***count*** *of students in each year and unemployment rate.*

Output: line chart of student count that can be expected.

ISSN: 2277-9655
Impact Factor: 1.852

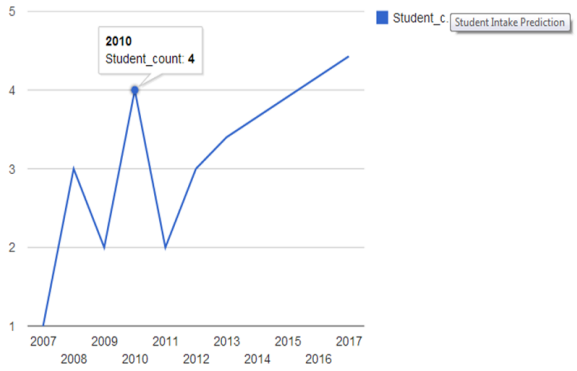


Fig.8

STUDENT RANKS

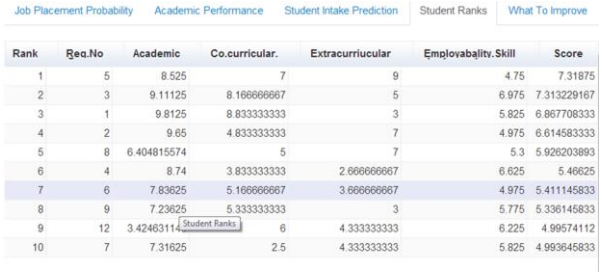
A collective list of student is analyzed and listed

Description

Based on a holistic approach, students are listed in a table. The table can be sorted with respect to any column.

Input: academics, extracurricular, co-curricular activity, employability skill and behavioral skill.

Output: Table with evaluation parameter and consolidated score.



Rank	Reg.No	Academic	Co.curricular.	Extracurriucular	Employabality.Skill	Score
1	5	8.525	7	9	4.75	7.31875
2	3	9.11125	8.166666667	5	6.975	7.313229167
3	1	9.8125	8.833333333	3	5.825	6.867708333
4	2	9.65	4.833333333	7	4.975	6.614583333
5	8	6.404815574	5	7	5.3	5.926203893
6	4	8.74	3.833333333	2.666666667	6.625	5.46625
7	6	7.83625	5.166666667	3.666666667	4.975	5.411145833
8	9	7.23625	5.333333333	3	5.775	5.336145833
9	12	3.4246311	6	4.333333333	6.225	4.99574112
10	7	7.31625	2.5	4.333333333	5.825	4.993645833

Fig.9

Identifying Bottlenecks

Various parameters are analyzed to identify bottlenecks of growth.

Description: A holistic approach is made to identify factors to improve in. Such factors include student and staff details.

Input: academics, extracurricular, co-curricular activity, employability skill and behavioral skill.

Output: Column chart of each factor against the maximum scope is plotted

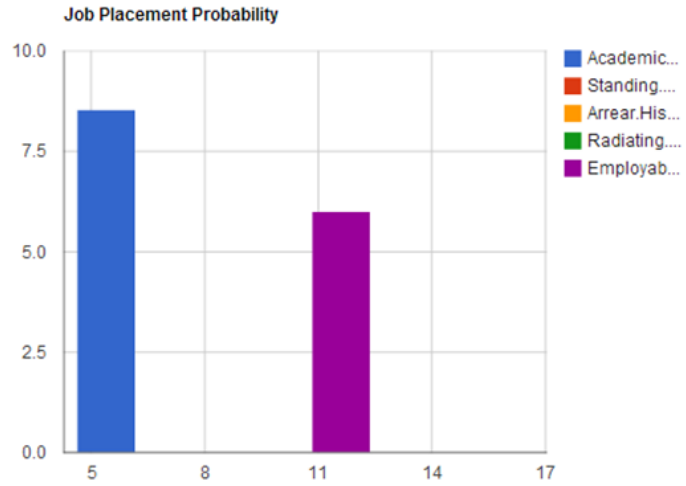


Fig.10

Results And Discussions

Naïve Bayes classifier algorithm [8] was used as a classifier, which exhibits better performance than traditionally used Support Vector Machine classifier [10]. All missing data are cleaned and any missing values are replaced by most probable substitution using 'Similar Cases' approach of filling missing values. Linear Regression analysis is used for forecast of values. To make the output visually pleasing and interactive, GoogleVis API [7] is used. The graphs animate on mouse hover.

Conclusion

In this paper, we implemented to serve students and management personnel to review their performance and thus find the realms to improve on. Data mining techniques data cleaning, data integration, and classification and regression analysis were effectively utilized. By using GoogleVis. we showed all the output of our concept in graphs.

References

[1] Sajadin Sembiring, M. Zarlis, Dedy Hartama, Ramliana S..," Prediction of student academic performance by an application of data mining techniques", Elvi Wani2011 International Conference on Management and Artificial Intelligence, IPEDR vol.6 (2011) © (2011) IACSIT Press, Bali, Indonesia

[2] WILAIRAT YATHONGCHAI, CHUSAK YATHONGCHAI, KITTISAK KERDPRASOP, NITTAYA KERDPRASOP..,

ISSN: 2277-9655
Impact Factor: 1.852

"2.Factor Analysis With Data Mining Technique In Higher Educational Student Drop Out" , Latest Advances in Educational Technologies.

[3] Yang Hang, Fong, S." A Framework of Business Intelligence-Driven Data Mining for E-business" INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference on 25-27 Aug. 2009

[4] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M.Inayat Khan..," Data Mining Model for Higher Education System".

[5] Sonali Agarwal, G. N. Pandey, and M. D. Tiwari.., "Data Mining in Education: Data Classification and Decision Tree Approach"

[6] Adriaans, P. and Zantinge, D. (1996)." Data Mining". Addison-Wesley.

[7] John Verzani CUNY/College of Staten Island John Verzani, CUNY/College of Staten Island.," gWidgetsWWW: Creating Interactive Web Pages within R".

[8] Bouckaert, R. (2004), "Naive Bayes Classifiers That Perform Well with Continuous Variables", Lecture Notes in Computer Science, Volume 3339, Pages 1089 – 1094.

[9] Amandeep Kaur Mann, Navneet Kaur.," Survey Paper on Clustering Techniques", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013.

[10] Danny Roobaert. DirectSVM: A fast and simple support vector machine perceptron. In Proceedings of IEEE International Workshop on Neural Networks for Signal Processing, Sydney, Australia, December 2000.