

Estimating profile of successful IT & Computer Science student using data mining approach

^{#1}Krishna Nichal, ^{#2}Akshay Patil, ^{#3}Tejaswini Walbe, ^{#4}Krishna Mane, ^{#5}Tejas Nigade.



¹knichal25@gmail.com,
²patilakshay497@gmail.com,
³tejaswiniwalbe@gmail.com,
⁴manekrishna1995@gmail.com,
⁵nigadetejas@gmail.com

PDEA's College of Engineering, Manjari (bk), Pune

ABSTRACT

The study presented in this paper aims to explore student's characteristics to determine performance of student or group of students based on their previous education and socio- demographic characteristics. Descriptive data mining method, cluster analysis is applied in the analysis process. Data used in the research is collected from the Engineering first year, second year, third year, final year IT and Computer Science students. On that student data, we apply the classification K-Nearest Neighbor (KNN) Algorithm for the analysis purpose. Research result indicates performance of student which is helpful to the Institutes & Organizations. As such, research results provide useful insight into both micro and macro level aspects of educational process, which can benefit both students and academic Institutions. Data mining has shown promising results in educational domain and a substantial potential to serve as a tool for improvement of quality in education.

Keywords - educational data mining, academic performance, classification analysis , KNN algorithm, educational strategy.

ARTICLE INFO

Article History

Received: 9th May 2018

Received in revised form :
9th May 2018

Accepted: 13th May 2018

Published online :

15th May 2018

I. INTRODUCTION

Students profile analysis system has become important factors in modern education field. The goal of higher education institutions is to provide a quality education process to all enrolled students. One way to achieve this goal is by discovering knowledge about student's performance and student's characteristics. For a common user, it is extremely difficult task to analyse each and every student that he/she is applicable for some event. This system should help the institutions, organizations and universities administrative task and provide real- time analysis of the data. This knowledge can in many cases be extracted from different data sets by using various data mining techniques. In this paper a research that includes evaluation of classification analysis using KNN algorithm, a data mining approach, for the purpose of IT & CS student's performance

description is presented and elaborated. Building this individual and group wise analysis system by data mining approach will further help the ease of categorization of any student's quality and performance. This chapter will provide a brief understanding about the student's profile, where the project has been executed and the current problem that they face in managing student's information. The study analysis enables the definition of the project problem statement, its objectives, scopes and advantages of the estimating student profile system.

II. MOTIVATION

Our concept is to categorize students by their performances and qualities to reduce common user manuals work using analysis. It makes perfect sense

to isolate the students individually and by grouping. This system might enhance student performance and faculty can allocate students for particular events.

III. LITERATURE SURVEY

Educational data mining is an newly developed research technique that mainly deals with the development of new techniques regarding data originating and exploring to the educational context.

It is an computerized approach to analyse educational data to study educational question. This data ming technique developed through EDM are most promising future line of research.

Through the research done by Pilot Mudhovozi, in university of venda, Thohoyandou, South Africa,2012. This paper was mainly done on university students of different years and the survey conclude that the first year student had experienced academic and social adjustment problem, they are over relied to social network. The student will be exposed to various coping resources to enable them quickly and smoothly to adjust their university life.

Through this research paper technique we developed our project which is analyze student data regarding first, second,third and final year performance and then show the student performance individually and group wise.

IV. METHODOLOGY

KNN - Algorithm

1. Determine parameter k= number of nearest neighbour
2. Calculate the distance between the input parameter set and all training samples.
3. Sort the distance & determine nearest neighbours on the Kth minimum distance.
4. Gather the category of nearest neighbours.
5. Use simple majority of category of nearest neighbours as the prediction value of query instance.

Euclidean

$$\text{Distance} = \sqrt{(-1)^2 + (-1)^2 + \dots + (-1)^2}$$

Where,

x1 y1= input parameters.

x1,y1,n1= Respective training input parameters. e.g.

- 1) K=2 ->Determine K parameter
- 2) Calculate distance domain -> Networking
Input value = 55
Sport rating = 4
Certificate rating = 4
Soft skill rating = 5

Input Query Value = 55,4,4,5,2

- 3) Suppose input history set :

Student Name	Year	Domain (Networking)	Rating 1	Rating 2	Rating 3
a	FE	55	4	4	5
b	SE	70	3	4	2
c	TE	60	4	3	5
d	BE	85	4	2	5

initial input Query set = (55,4,4,5) K = 2

- 4) Calculate Distance:

a. $(55-55)^2 + (4-4)^2 + (4-4)^2 + (5-5)^2 = 0$

b. $(70-55)^2 + (3-4)^2 + (4-4)^2 + (2-5)^2 = \sqrt{235} = 15.32$

c. $(60-55)^2 + (4-4)^2 + (3-4)^2 + (5-5)^2 = \sqrt{26} = 5.09$

d. $(85-55)^2 + (4-4)^2 + (2-4)^2 + (5-5)^2 = \sqrt{904} = 30.06$

- e. Sort the distance if it is included in KNN:

1. 0 → Yes → Record No. 1
2. 5.09 → Yes → Record No. 3
3. 15.32 → No → Record No. 1
4. 30.06 → No → Record No. 4

- f. Display the result of suggested student list:

Name of student:

1)	a
2)	c

- g. View information of recommended student and Exit.

V. PROPOSED SYSTEM

In this proposed work we are using K-nearest neighbor algorithm. Through this algorithm frequent information is generated which would be used for analysis of data. Collected data is being analysed in which rules generated through algorithm are applied.

Analysed profile further classified. The purpose of this project is to reduce the complexity of administrator to isolate students for particular technical and nontechnical event.

VI. TECHNOLOGIES USED

About JAVA:

Java has been tested, refined, extended, and proven by a dedicated community of Java developers, architects and enthusiasts. Java is designed to enable development of portable, high-performance applications for the widest range of computing platforms possible. By making applications available across heterogeneous environments, businesses can provide more services and boost end-user productivity, communication, and collaboration—and dramatically reduce the cost of ownership of both enterprise and consumer applications.

The original and reference implementation Java compiles, virtual machines, and class libraries were originally released by Sun under proprietary licenses. As of May 2007, in compliance with the specifications of the Java Community Process, Sun relicensed most of its Java technologies under the GNU General Public License. Others have also developed alternative implementations of these Sun technologies, such as the GNU Compiler for Java (byte code compiler), GNU Class path (standard libraries), and Iced Tea-Web (browser plugin for applets).

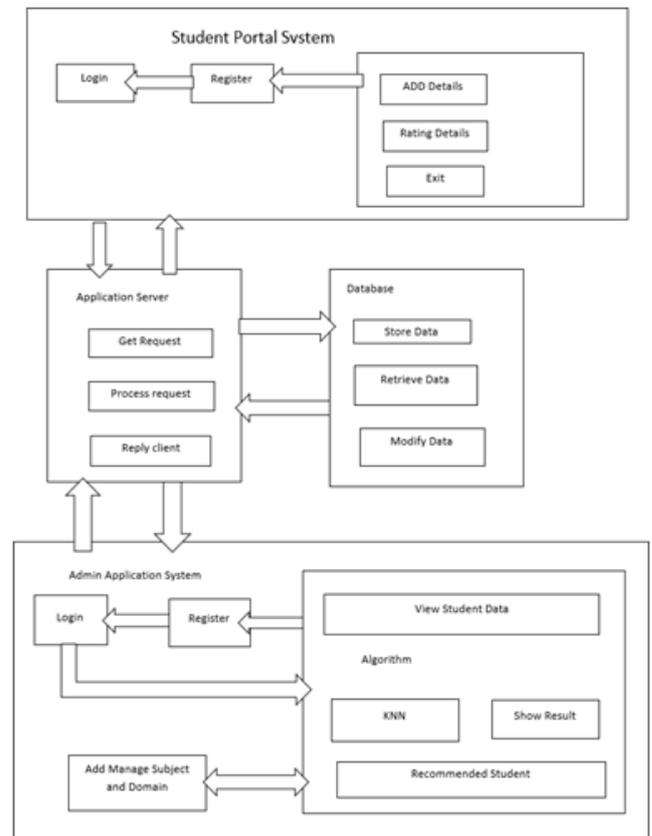
Net beans :

NetBeans is an open-source integrated development environment (IDE) for developing with Java, PHP, C++, and other programming languages. NetBeans is also referred to as a platform of modular components used for developing Java desktop applications.

MySQL :

MySQL is an open source RDBMS that relies on SQL for processing the data in the database. MySQL provides APIs for the languages C, C++, Eiffel, Java, Perl, PHP and Python. In addition, OLE DB and ODBC providers exist for MySQL data connection in the Microsoft environment. A MySQL .NET Native Provider is also available, which allows native MySQL to .NETaccess without the need for OLE DB.

VII. PROPOSED SYSTEM



VIII. CONCLUSION

The research presented in this system deals with developed student clusters, based on student's previous education, which provide a basis for IT and computer science student's success prediction. Interpretation and employment of the obtained results can be very useful for administrator to predict the performance of IT and Computer science students With respect to multiple domain Like Networking, Testing, Development etc. It makes precise decisions that which student is pioneered in particular domain. Because of this System, Students and Admin Knows that in which field student has to improve their performance and also know that on which field he/she has to find job. This system gives decisions concisely.

IX. FUTURE SCOPE

The product scope of designed project is efficient categorization and isolation of students by analysing their profile.

To enhance the quality of student individually. Our main technology is Data mining which is responsible for database addition and holds great potential to improve categorization of database system. This application can be used by any educational organizations, institutes and universities for some events and many IT companies to classify the employers in their particular job profile domain.

REFERENCES

- 1) W. Camara, and M. Michaelides. "AP use in admissions: A response to Geiser and Santelices". College Board Research Note, 2005.
- 2) K. McKenzie, and R. Schweitzer. Who succeeds at university? Factors predicting academic performance in first year Australian university students. Higher education research & development, 20(1), pp. 21-33, 2001.
- 3) I.H., etersen, J. Louw, and K. Dumont. Adjustment to university and academic performance among disadvantaged students in South Africa. Educational Psychology, 29(1), pp. 99-115, 2009.
- 4) K. Hillman. "The first year experience: the transition from secondary school to university and TAFE in Australia". LSAY Research Reports. Longitudinal surveys of Australian youth research report; http://research.acer.edu.au/lsay_research/44,2005.
- 5) C. Romero, and S. Ventura. "Educational data mining: a review of the state of the art". IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 40(6), pp. 601- 618, 2010.