INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 7, Issue 3 - V7I3-2125)
Available online at: https://www.ijariit.com

Student advisory system

Sinchanaraj H. N.
sinchana.hn2103@gmail.com
JSS Science and Technology, Mysore, Karnataka

Disha O. V.
dishavijay2699@gmail.com
JSS Science and Technology, Mysore, Karnataka

Nihaal N.
nihaalniranjana@gmail.com
JSS Science and Technology, Mysore, Karnataka

Keerthi D. C.
devarakondikeerthi@gmail.com
JSS Science and Technology, Mysore, Karnataka

Dr. G. Madhusudhan
madhusudan@gmail.com
JSS Science and Technology, Mysore, Karnataka

ABSTRACT

The education crisis is now widely spread in global in term of decreasing number of student and decreasing degree requirements for some jobs. Educational institutes look for more efficient technology that assist better management and support decision making procedures or assist them to set new strategies. One of the effective ways to address the challenges for improving the quality is to provide new knowledge related to the educational processes and entities to the managerial system. With the data science techniques the knowledge can be extracted from operational and historical data that resides within the educational organization's databases using. The datasets for system implementation contains information about past data of students. Educational data mining (EDM) is recently interested in data mining area to discover useful knowledge in educational data to help educators improve their administration planning and student services. This system proposes applying of data science techniques in educational data. Association rule applied in students data to find some knowledge for supporting management planning. Data Science algorithms applied in course grades and job data of graduated student to predict job after graduated. The results of these studies give good knowledge for student management planning and job prediction. The main objective of the proposed system is to find the correlations between the student educational parameters with the types of the job.

Keywords: *Eclat Algorithm, Support Count, Confidence, Association Rules*

1. INTRODUCTION

The primary aim of students who join professional courses in higher learning institutions is to secure a well-paid job in a reputed organization. Professional education can be either completely technical or it can be managerial as well. Technology courses provides technical education to students in various fields such as Computer Science and Engineering, Electronics and Communication Engineering, Civil Engineering Mechanical Engineering, etc. This degree/ post graduate is aimed at making students experts in state of the art conjectural as well as practical knowledge in various engineering branches. The prediction of job status that students are most likely to achieve will help students to put in more hard work to make appropriate progress in stepping into a career in various technical fields. It will also help the teachers as well as others in an institution to provide proper care towards the improvement of students in the duration of course. A high placement rate is a key entity in building the reputation of an educational institution. Hence such a system has a significant place in the educational system of any higher learning institution.

The occurrence of disrupt technology in many areas has conducted the large volume of data in various format and unprecedented speed of generated data. These data phenomenon known in term of big data [1] (volume, various, velocity). The big data required to analyses with appropriate approach for extracting useful knowledge. The data mining aims to discover pattern or useful knowledge from large collection of data [2]. The major functions of data mining are applying various algorithms to discover useful pattern and catch patterns which hidden in data. Educational Data Mining (EDM) is a new research direction in data mining. The focuses of EDM are focus on discovering useful knowledge and mining the helpful patterns from educational data, such as student profile data, student registration data, student job data and other occurrence of student data which can collect during college [3].

There are a lot of objective of EDM depend on data source and educational problem. In 2011, B.K. Baradwaj and S. Pal [4] studied on classification task of student database to predict the student's performance in end semester examination. The student' performance was divided into four stages (first, second, third and fail). It helps earlier in identifying the dropouts and student who need special attention from teacher. In 2015, N. Buniyamin et al. [5] presented the use of Neuro-Fuzzy classification in a student's academic data in an electrical engineering faculty of Malaysian public university. The study showed that the output of system can determine probability of student to achieve excellent grade even if the student achieved weak in certain course or subject. In 2016, A. A. Saa [6] used multiple data mining tasks to create qualitative predictive models to predict the students' grades from educational dataset. Four decision tree algorithms have been implemented and Naïve Bayes algorithm. The results can motivate the university to perform data mining task on their student data, as well as student to improve their performances. In 2017, F. Matsebula and E. Mnkandla [7] proposed an architecture for big data analytics in higher education. The architecture composed of five parts; data gathering device which collected student data from various data source (student's card, social networking and student information system), data storage and management system which consists of bigdata management, data analytics system which process algorithms from data, data visualization which help in decision making process, and action system for providing alerts, warning or guiding to student or administrators.

The main objective of this research is to answer two main questions using data science algorithms. First, how data mining can help student management working process. Second, how data mining can predict the student's jobs. To answer these questions, we provide two task of data mining with two difference sources of data. First, the association rule mining [8] is used to discover interesting relation between feature in student data and job types. The result of association rule mining could be used to help for student management planning.

2. LITERATURE SURVEY

[1] An Intelligent Student Advising System Using Collaborative Filtering

We propose a web based intelligent student advising system using collaborative filtering, a technique commonly used in recommendation systems assuming that users with similar characteristics and behaviors will have similar preferences. With our advising system, students are sorted into groups and given advice based on their similarities to the groups. If a student is determined to be similar to a group students, a course preferred by that group might be recommended to the student.

Inference: We got the idea that grouping the students and predicting the suitable course for the student lacks over data for prediction and not all student behaviours are connected to course advising.

[2]Mining Students' Data for Performance Prediction

Students' academic performance is based upon diverse factors. A very promising tool to attain this objective is the use of Data Mining. Data mining techniques are used to discover hidden information patterns and relationships of large amount of data, which is very much helpful in decision making.The type of information is produced by the data and it decides the processing method of data. A lot of data that can produce valuable information, in education sector contains this valuable information. Which helps the education sector to capture and compile low cost information for this information and communication technology is used. Now-a-days educational database is increased rapidly because of the large amount of data stored in it. The loyal students motivate the higher education systems, to know them well; the best way is by large valid management and processing of the students' database. Data mining approach provised valid information from existing student to manage relationships with upcoming students.

Inference: We got the idea that mining student's data provided valid information from existing students to manage relationships with upcoming students.

[3]Mining Frequent Patterns in Data Using Apriori and Eclat

In a world where the data is continuously generated by a multitude of sources and in extremely large quantities, it becomes crucial to transform it into valuable and actionable knowledge. The process of transforming raw data into knowledge becomes especially important for large companies that strive to understand as much as possible the behavior of their customers and build a strong competitive advantage by offering a better targeted and a more personalized experience.This is often transformed into a problem of finding the most frequent patterns in large datasets and create a generalizable and interpretable picture of reality.

Association rules mining is commonly used for market basket analysis problems with the goal to identify combination of items more likely to appear together in the transaction of a dataset. It helps to identify interesting rules, frequent patterns, correlations or just casual data structures in transactional datasets The Apriori algorithm employs level-wise search for frequent item-sets, using bottom-up research and moving upward levelwise in the lattice. In order to identify the association rules, the algorithm runs through two phases: Candidate Generation and Pruning. ECLAT or Equivalence Class Transformation Algorithm, is a frequent pattern mining algorithm that mines efficiently frequent patterns by performing a bottom like depth first search or in other words, a bottom up Lattice traversal. In order to mine frequent patterns, it requires to be applied on a vertical database.In order to mine the most frequent and potentially interesting patterns, association rule mining algorithms come to be perfectly suitable for the task. By comparing Apriori and Eclat in this paper, it was possible to highlight how it becomes more computationally expensive for Apriori compared to Eclat to mine frequent patterns at very low support threshold levels.

Inference: We got the idea that the performance of Apriori algorithm is slower and requires more computational power compared to the performance of Eclat algorithm especially when lowering the minimum support threshold at which the algorithms are applied.

3.PROPOSED SYSTEM AND ARCHITECTURE

Student Performance improvement with job oriented in education sector is helpful for education institutes to improvise their reputations. Identification of different factors which affects a student's learning behavior and performance during academic career. Construction of a prediction model using classification data science techniques on the bases of identified predictive variables. Proposed system is an educational system which uses data science to process educational data. Here system process previous students data such as grades on top 3 subjects, area of interest and the aggregate using data mining technique. The main objective is to predict the correlation between student parameters with the job types. In the Proposed system we can consider the job types such as "IT", "NON IT", "BANK", "EDUCATION","GOVT" etc. System provides the useful information which helps education sector to identify the factors related to the student jobs. The system uses Data Science approach which provides valid information from existing students to manage relationships with upcoming students and to identify the most effective factor to determine a student's job and then adjusting these factors to improve the students performance. Proposed system is a web based application which makes use of data mining technique for the extraction of useful information.

3.1 Objectives of the Proposed System

- Proposed project is a student job prediction and management system which is meant for educational institute.
- Proposed project makes use of Data Science technique for the job prediction of future.
- To improve the performance of the student.
- The main objective is to predict the correlation between student parameters with the job types.
- To provide valid information from existing students to manage relationships with upcoming students.
- Identification of different factors which affects a student's learning behavior and performance during academic career.
- Construction of a prediction model using Data Science technique on the bases of identified predictive variables.
- Visual Studio and SQL Server used to develop this project

3.2 Architecture

Three tier Architecture:

Three tier architecture consists of three layers. They are:

a) The Data Layer:

The key component to most applications is the data. The data has to be served to the presentation layer somehow. The data layer is a separate component (often setup as a separate single or group of projects in a .NET solution), whose sole purpose is to serve up the data from the database and return it to the caller. Through this approach, data can be logically reused, meaning that a portion of an application reusing the same query can make a call to one data layer method, instead of embedding the query multiple times. This is generally more maintainable.

b) Business Layer:

Though a web site could talk to the data access layer directly, it usually goes through another layer called the business layer. The business layer is vital in that it validates the input conditions before calling a method from the data layer. This ensures the data input is correct before proceeding, and can often ensure that the outputs are correct as well. This validation of input is called business rules, meaning the rules that the business layer uses to make "judgments" about the data.

One of the best reasons for reusing logic is that applications that start off small usually grow in functionality. The business layer helps move logic to a central layer for "maximum reusability."

c) Presentation Layer:

The ASP.NET web site or windows forms application (the UI for the project) is called the presentation layer. The presentation layer is the most important layer simply because it's the one that everyone sees and uses. Even with a well structured business and data layer, if the presentation layer is designed poorly, this gives the users a poor view of the system.

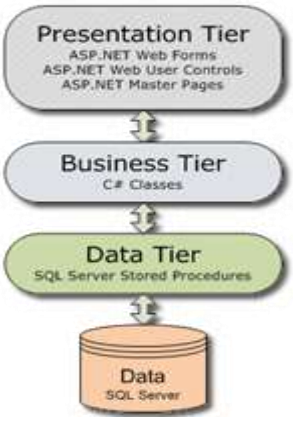


Fig. Three tier Architecture

The **presentation tier** contains the UI (User Interface) elements of the site, and includes all the logic that managers the interaction between the visitor and the client's business. (ASP.NET Web Forms, Web User Controls, ASP.NET Master Pages)

The **business tier** receives requests from the presentation tier and returns a result to the presentation tier depending on the business logic it contains. (C# Classes)

The **data tier** is responsible for storing the application's data and sending it to the business tier when requested. (SQL Server Stored Procedures or Queries).

4. IMPLEMENTATION

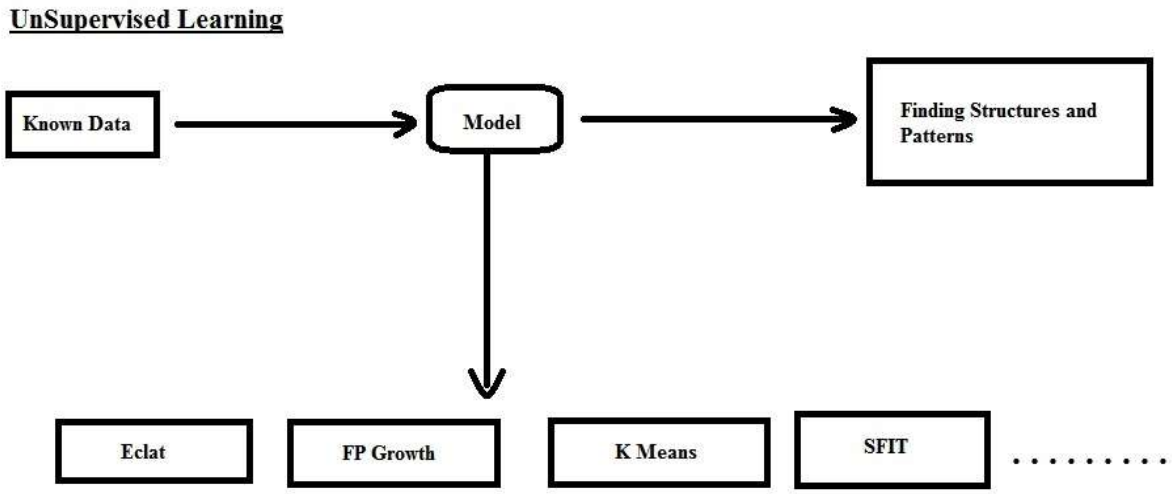
4.1 Machine Learning

Machine learning is a process of studying a system based on data. Machine learning is a part of data science where we use machine learning algorithms to process data.

4.2 Unsupervised Learning

A Descriptive model is used for tasks that would benefit from the insight gained from summarizing data in new and interesting ways. There are no predefined labels in unsupervised learning technique.The goal is to explore the data and find some structure with in. Unsupervised learning works well on transactional data.

Descriptive model developed using clustering techniques and association learning techniques. We have many efficient algorithms such as "eclat algorithm", "AIT algorithm", "SFIT algorithm", "STEM Algorithm", "FP Growth algorithm", "K Means algorithm", "Fuzzy C Means algorithm" etc…



Descriptive Model

4.3 Job Prediction Process

Step 1: Data Collection

We are working on real time application, we build a new application which contains data servers (used to store data). Data collection means collecting data from different sources. Data includes student parameters such as communication skills, aptitude results, 10^{th} results PUC results etc.. and job types.

Step 2: Data Preparation

Here data from servers extracted and analyzed. Complete data extracted and analyzed where we remove irrelevant data and retain data required for processing. According to the project only student characteristics and job types are required to generate outputs.

Step 3: Specify Constraints

SUPPORT COUNT

The relationship between the total number of transaction containing that item (A) with the total number of transaction in data set.

CONFIDENCE

Confidence of item set defined as total number of transaction containing the item set to the total number of transaction containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm)

Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns.

For example, Market-basket analysis, where we track people's buying habits, we might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream. We use eclat algorithm to process educational data and to find the patterns. Here we generate patterns related to student characteristics and job types.

Eclat algorithm is selected because of the following reasons.

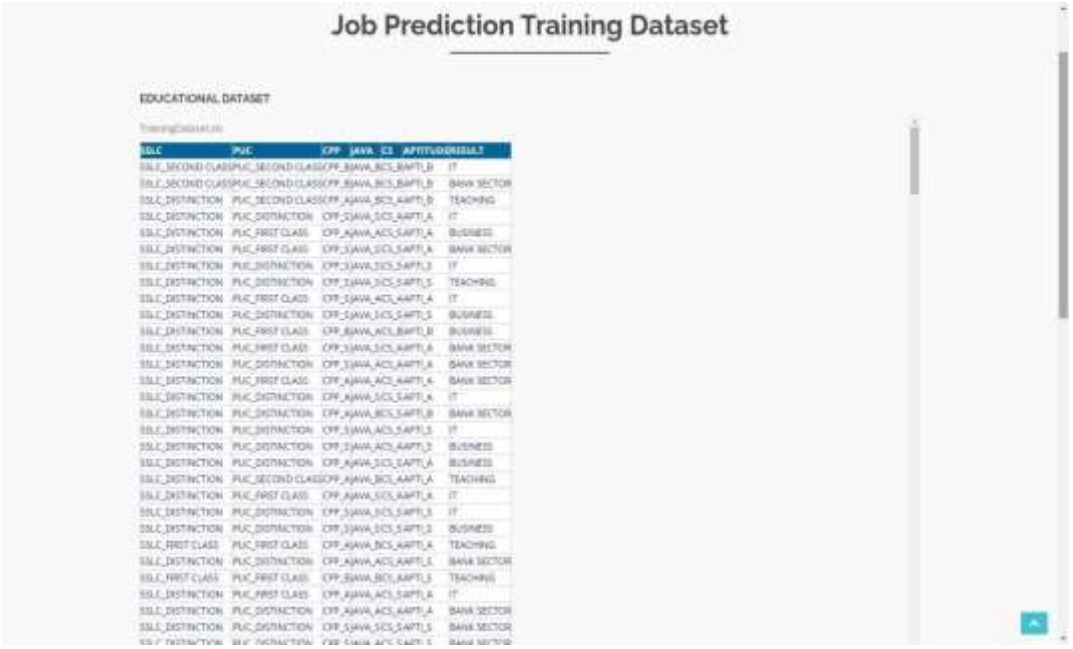
1. Quicker Results (takes less time for Prediction)
2. Works fine for small data set as well as Huge data set.
3. One scan of Database is Enough.
4.Works fine for multiple constraints.

Step 5: Patterns Prediction

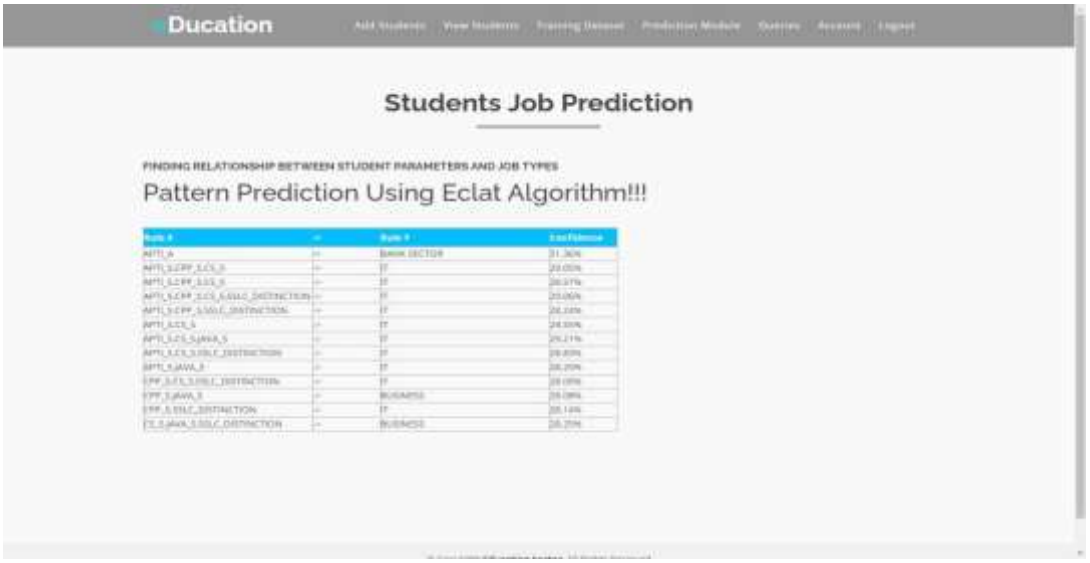
Here system predicts the relationship between student characteristic with job types.

4.4 Results

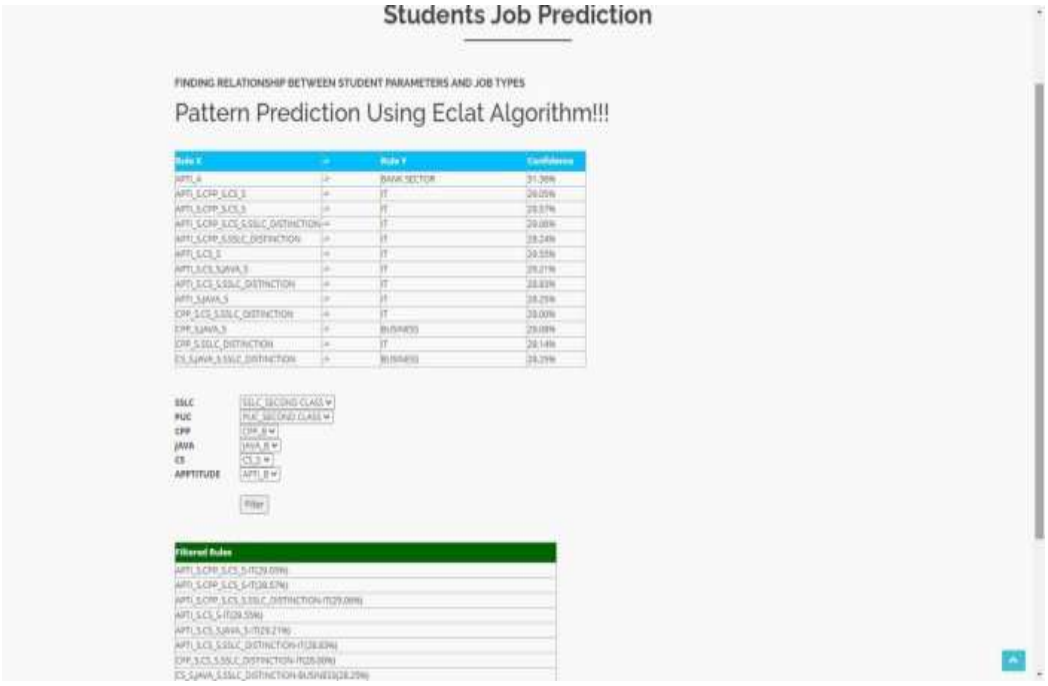
4.4.1 Training Datasets



4.4.2 Job Prediction



4.4.3 Job Prediction with Filtration



5. CONCLUSION

In this work, we applied data science algorithm to discover useful knowledge from educational datasets. First, the association rule mining is used on student datasets to answer the question "how data science can help student management working process". The result shown the significant relationship between student parameters and job types. This result might be help educators who response for management working process to plan their admission promotion. Then algorithm is applied to student's course grade with job datasets to answer the question "how data science can predict the student's jobs". The result rule show that the significant subject which student should be important for future career.

6. FUTURE ENHANCEMENTS

Further we can add more training datasets to get better results. We can also use more algorithms for pattern prediction and can compare and can find the better algorithm.

7. REFERENCES

[1] N. Elgendy and A. Elragal. Big Data Analytics: A Literature Review Paper, Industrial Conference on Data Mining (ICDM), 2014, pp214-227.
[2] Heikki, Mannila, Data mining: machine learning statistics, and database, IEEE, 1996.
[3] Jiawei Han and Micheline Kamber (2011) Data Mining: Concepts and Techniques. 3 editions. Morgan Kaufmann.
[4] N. Elgendy and A. Elragal. Big Data Analytics: A Literature Review Paper, Industrial Conference on Data Mining (ICDM), 2014, pp214-227.
[5] Heikki, Mannila. Data mining: machine learning statistics, and database, IEEE, 1996.
[6] AlejandroPeña-Ayala. Educational data mining: A survey and a data mining-based analysis of recent works, Expert Systems with Applications, March, 2014, Vol. 41(4).
[7] B.K. Baradwaj, S.Pal. Mining Educatinal Data to analyze Students' Performance, International Journal of Advanced Computer Science and Applications (IJACSA), 2011, Vol. 2(6).
[8] N. Buniyamin, U.B. Mat, P.M. Arshad. Educational data mining for prediction and classification of engineering students achievement, The 7 th International Conference on Engineering Education (ICEED), 2015, Kanazawa, Japan, pp. 49-53.
[9] A. A. Saa., Educational Data Mining & Student's Performance Prediction., International Journaal of Advanced Computer Science and Application (IJACSA), 2016, Vol.7(5).
[10] F. Matsebula, E. Mnkandla., A big data architecture for learning analytics in higher education, IEEE Africon, 2017, pp.951-956.
[11] J. Han, M. Kamber, J. Pei,. Data Mining: Concepts and Techniques, Morgan Kaufmann , 2011.
[12] Quinlan, J. R. Induction of Decision Trees. Mach. Learn. 1, 1 ,Mar. 1986, pp. 81–106.
[13] S. Rao, R. Gupta,. Implementing Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm., International Journal of Computer Science And Technology, Mar. 2012, pp. 489-493.

