

# Educational Data Analytics using Association Rule Mining for Student Job Prediction

K. J Ashika<sup>1</sup>, Jagruthi M. N<sup>2</sup>, Nikhil N<sup>3</sup>, Pooja S<sup>4</sup>

<sup>1,2,3,4</sup> U.G Student, Dept. of Computer Science and Engineering, The National Institute of Engineering, Mysuru, Karnataka, India.

**Abstract:** The competition for employment is increasing every day and placement has become a challenging task. Students in the third/final year of Engineering college start feeling the pressure of placement season happening around them and they feel the need to know where they stand and how they can improve their chances of getting a job. To resolve such issues, we present the design and development of a browser-based application for student job prediction using Data mining. This paper proposes applying the ECLAT algorithm for student job prediction, as a way to determine the characteristics of the student with different job types. Result of these studies will be displayed to students to assess their capabilities and identify their interests and which career area they are going to put them in.

**Keywords:** Educational Data Mining, Classification, Association Rule Mining, Éclat Algorithm

## 1. INTRODUCTION

Educational Data Mining (EDM) uses Data Mining Algorithms to discover useful knowledge from educational data sets. This result helps to improve their administration planning and student performance.

There are many objectives in EDM that depends on student data. There are different classifications in the student database to predict the job type based on the student's skills and their performance in academic subjects. Job Prediction is divided into IT, NON-IT sectors like Bank Sector, Teacher and Govt, etc. It helps students and administrators to identify their

weak points to give special attention. The Study of the Association rule shows that the result of the Educational System can determine the probability of students achieving good grades even if the student achieved weak in some courses or subjects.

Big data analytics architecture is composed of five parts: data gathering devices collects the student data from different source (student's grade card, placement office, student information system, and social networking), data storage and management system, data analytics system it helps in process algorithms from the data, data visualization.

The main objective is to provide the answer to two main questions using data mining algorithms. First, how Data Mining takes specified attributes from the data set. Second, how Data Mining can predict a student's job. First, the association rule mining is used to discover the specified or interesting patterns from the Education data set such as USN, Aptitude grades, Subject's grades, Communication Skills, etc. Second, by using the Eclat algorithm and mapping the student's features with jobs to find a strong association set from the data set. The result from Data Mining shows the accuracy and precision for student's job predictions.

## 2. EXISTING

In the presently existing system, Student details such as marks, field of job each student has settled in, and other academic information are maintained by Institutions. The maintenance of this information in the Database is made easy by various tools available

in the market. However, these tools are expensive and also, they lack customized services. Analyzing this stored information is very tedious and consumes a large amount of time, making it complex for Educational Institutions to take efficient measures to improve the skills of each individual student to match the desired job field.

### 3. PROPOSED

The Proposed system is an Educational system which analyses the Educational data. The Proposed system makes use of the Data mining techniques for student job prediction by processing previous student's data. The research is conducted taking

into considerations the various parameters such as results of SSLC, PUC, and grades in CPP, Java, Communication skill, aptitude. The system uses the Data science approach ECLAT [Equivalence Class Clustering and bottom-up Lattice Traversal] for student job prediction.

The main objective is to predict the correlation between student parameters with job types. In the proposed system we have considered the job types such as "IT", "BANK SECTOR", "BUSINESS", "TEACHING". The system provides valid information from existing students to manage relationship with upcoming students to improve the performance. The main objective of the research is to answer two main questions by using the Data mining algorithm. First, how data mining can help Education institutions to increase their reputations. Second, how Data mining can predict students' job.

Proposed system consists of 2 End-users:

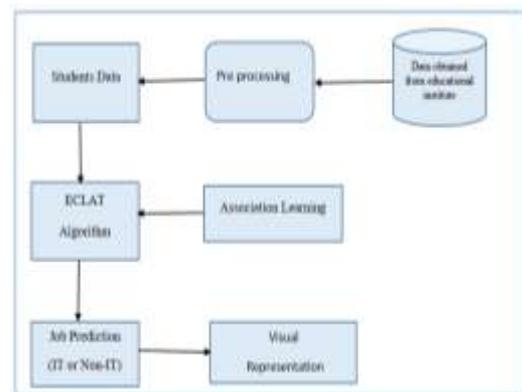
#### 3.1. Admin

- management of Student data.
- management of student Account.

- resolving the issues.
- Run the prediction module.

#### 3.2. Student

- Post queries.
- Managing Account.
- View prediction.



**Figure 1:** Architecture Diagram

### 4. METHODOLOGY

#### Association Rule Mining

##### 4.1 Meaning ECLAT Algorithm

ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal) is one of the well-known methods of Association Rule Mining. Éclat is also considered a scalable version of the Apriori Algorithm. Apriori Algorithm proceeds in a horizontal way performing Breadth-First Search (BFS) of a formed graph. On the other hand, ÉCLAT proceeds in a vertical manner imitating Depth-First Search (DFS) of a graph. The vertical approaching nature of ECLAT makes it faster than an Apriori Algorithm.

##### 4.2 ECLAT Algorithm

STEP 1: Get transactional ID List (Tidlist) for each item (DB scan).

STEP 2: Tidlist of {a} is exactly the list of transactions containing {a}.

STEP 3: Intersect Tidlist of {a} with the Tidlists of all other items, resulting in Tidlists of {a, b}, {a, c}, {a, d}, ...={a}-conditional database (if {a} removed).

STEP 4: Repeat from STEP1 on {a}-conditional Database.

STEP 5: Repeat for all other items.

### 4.3 Sample Example

**Table 1: Dataset:**

TID	Itemset
T1	A, B, E
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C,
T7	A, C
T8	A, B, C, E
T9	A, B, C

**Minimum Support count: 2**

**Minimum Confidence = 80%**

Itemset A, B, C, D, and E.

**Table 2: Transforming Dataset into Vertical format:**

Itemset	TID
A	T1, T4, T5, T7, T8, T9
B	T1, T2, T3, T4, T6, T8, T9
C	T3, T5, T6, T7, T8, T9

D	T2, T4
E	T1, T8

**Table 3: Frequent-1 Itemset:**

Itemset	TID
A	T1, T4, T5, T7, T8, T9
B	T1, T2, T3, T4, T6, T8, T9
C	T3, T5, T6, T7, T8, T9
D	T2, T4
E	T1, T8

**Table 4: Frequent-2 Itemset:**

Itemset	TID
A, B	T1, T4, T8, T9
A, C	T5, T7, T8, T9
A, D	T4
A, E	T1, T8
B, C	T3, T6, T8, T9
B, D	T2, T4
B, E	T1, T8
C, E	T8

**Table 5: Frequent-3 Itemset:**

Itemset	TID
A, B, C	T8, T9
A, B, E	T1, T8

A, B, D	T4
A, C, E	T8
B, C, E	T8

Table 6: Frequent-4 Itemset:

Itemset	TID
A, B, C, E	T8

### 5. DATA UNDERSTANDING AND PROCESSING

In this, Educational Data Mining (EDM) we focus on discovering useful Knowledge and mining, the helpful pattern.

Patterns from Educational data, such as Student Profile data, Student registration data, Student job data and other occurrences of Student data, which is collected from college.

Specifically, the sample datasets collect during college/specific branch containing datasets (shown in figure 1) exactly 647 records described by 8 attributes (figure 2), the volume of collected data will increase in the future years.

Figure 2 shows the attributes of student’s data set, which we include in our project, it displays the different parameters of the student like roll no, Before applying Eclat Algorithm to our student data set, we use five list boxes to store the data, each list boxes contains different data. The First list box contains the Transaction id (T-id). The second list box contains the dataset.

The third list box contains a distinct item. And the fourth list box contains the frequent itemset (L). The fifth list box contain the final output

subjects’ grade and result, where result consist the job types.

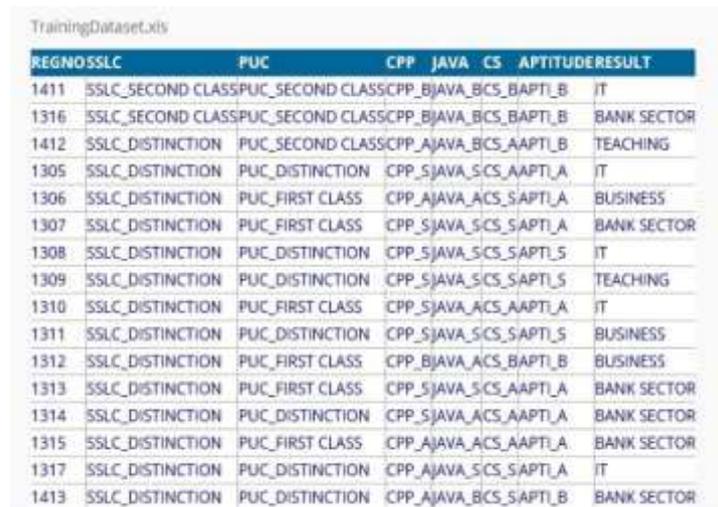


Figure 2: Student Dataset.

No	ATTRIBUTES
[1]	ROLL NO
[2]	SSLC
[3]	PUC
[4]	CPP
[5]	JAVA
[6]	CS
[7]	APTITUDE
[8]	RESULT
...	.....

Figure 3: List of attributes contains in the student data set

(displaying patterns). The fourth and Fifth list box contains the result of the Eclat algorithm as shown in Figure 3.

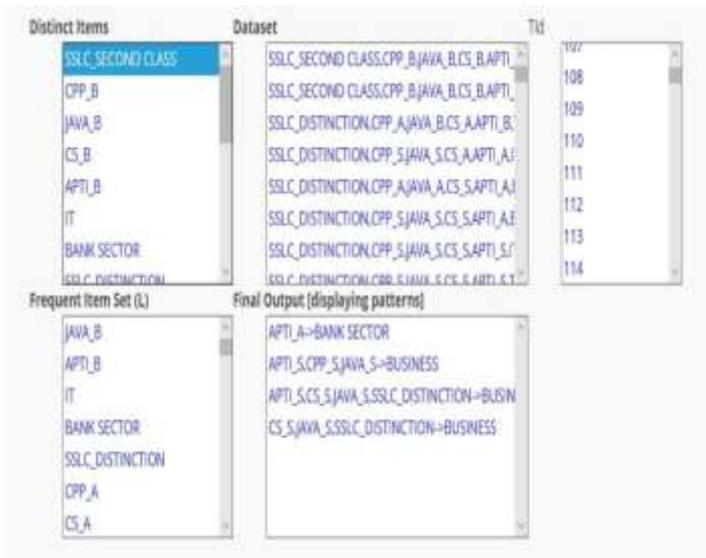


Figure 4: Pattern Details

To the Eclat algorithm, we input the T-id and distinct data, Support count and confidence value, based on the support count and confident value, we find the frequent items using Association rule techniques.

As shown in the figure 3 algorithm provides the frequent itemset (L), and we apply a strong association rule on the frequent itemset (L). The output provided by the algorithm is not in the order of the Student’s parameter relates to job type, so we need to program the project to display the items in the order, then the final output is stored in the list box for the further process.

**Experimental Results**

As we explained in the Methodology and Data Processing and Understanding section in our paper, figure 4 shows the result of the Job prediction module. This result will be visible for both Admin as well as Students. In the Admin portal, the result needs to be displayed in pattern as shown in figure 3 with final prediction result as shown in figure 4, as Admin needs to analyse pattern details to improve their administrator planning.

In student portal, we just displayed the prediction module result.



Figure 5: Student Prediction module result.

**6. CONCLUSION**

In this project, we are using the Eclat algorithm in Data mining techniques to discover the useful pattern or data from the Student dataset. By using association rule, we provide the answer for the question “How Data mining is going to aid the process of student management”. The outcome of this algorithm shows the considerable relationship between student parameters and job types. These impacts help educators who respond to student management working process to plan their Admission promotion. The algorithm is then applied to student’s skills and course grades with job datasets to provide the answer to the question “How Data mining algorithm predicts the job for students”. The result of the Association rule shows that the significant course or subject in which student should be prepared for future the career.

**Future Enhancements**

In the proposed System, Admin assigns Id and Password for Students, which is done manually. But sometimes students are unable to get the updated information. By implementing an SMS/Email module, we can solve these problems, where students get information through an Email or SMS.

## ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Guide **Mr. G. Ramesh**, Assistant Professor, Department of CS&E, NIE for their valuable suggestions and guidance during the entire process of the Paper and for having permitted us to pursue work on the subject.

It also gives us immense pleasure to thank our Co-Guide **Mr. M. J Yogesh**, Assistant Professor, Department of CS&E, NIE for their constant motivation and support during the entire process.

## REFERENCES

[1] 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.

[2] N. Elgendy and A. Elragal. Big Data Analytics: A Literature Review Paper, Industrial Conference on Data Mining (ICDM), 2014, pp214-227.

[3] N. Buniyamin, U.B. Mat, P.M. Arshad. Educational data mining for prediction and classification of engineering student's achievement, The 7th International Conference on Engineering Education (ICEED), 2015, Kanazawa, Japan, pp. 49-53.

[4] Saa., Educational Data Mining & Student's Performance Prediction., International Journal of Advanced Computer Science and Application (IJACSA), 2016, Vol.7(5).

[5] F. Matsebula, E. Mnkandla., A big data architecture for learning analytics in higher education, IEEE African, 2017, pp.951-956.