

Data mining to reduce Drop Out rate of student: Use of Classification and Prediction method

Shubhangi Bhatambrekar, Savita Mohurle

Department of Computer science, Modern College, Ganeshkhind, Pune 411016,
University of Pune

Abstract: Dropout rate of student is a big risk for both the educational institution as well as for the upcoming career of student. It is the need of today's era to find out reason behind increasing drop out rate of students from the courses. This paper tries to identify the reasons behind increasing dropout rate of student. Taking into consideration the three reasons viz. average learners, family background and interest in subject, the dropout rate of student can be predicted and analyzed to retain students in the institution. Classification and prediction method of data mining is used to find the exact reason. Through mining available data of student, the factors that influence the academic development of student can be recognized.

Keywords: dropout, classification, prediction, influence, student.

1. Introduction:

Data mining (sometimes called data or knowledge discovery) has become the area of growing significance because it helps in analyzing data from different perspectives and summarizing it into useful information [1]. Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well-mannered [2]. There are various reasons why student want to abandon the course in between, due to which not only the career and future of student affect as well as it affect the development of Nation. More the educated people more is the nation developed. The following three reasons are equally responsible for the dropout of students from the course:

- Average learners

- Family background
- Interest in subject

Educational data or data available with the institution can discover the unidentified and unpredicted fact why the student dropout rate has increased. Advance study says that it is desirable to identify the students who tend to dropout if the performance of student degrades. It has become necessary to analyze the results and performance of each and every student and prevent them from abandoning education. Mining educational data available with the institution can trace out the student who can leave the institution by any of the above three reasons. Fig 1.1 shows Requirement and Result that gives reasons behind students drop out.

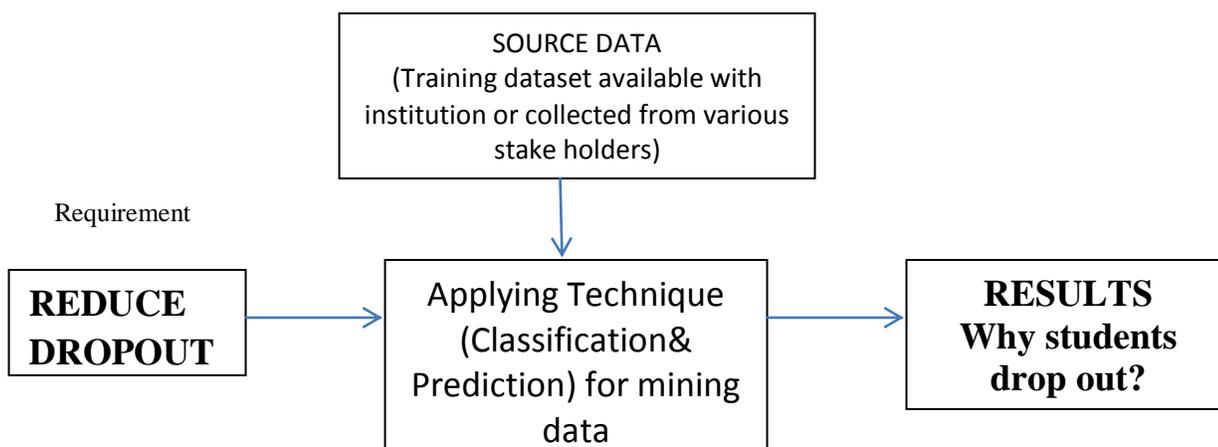


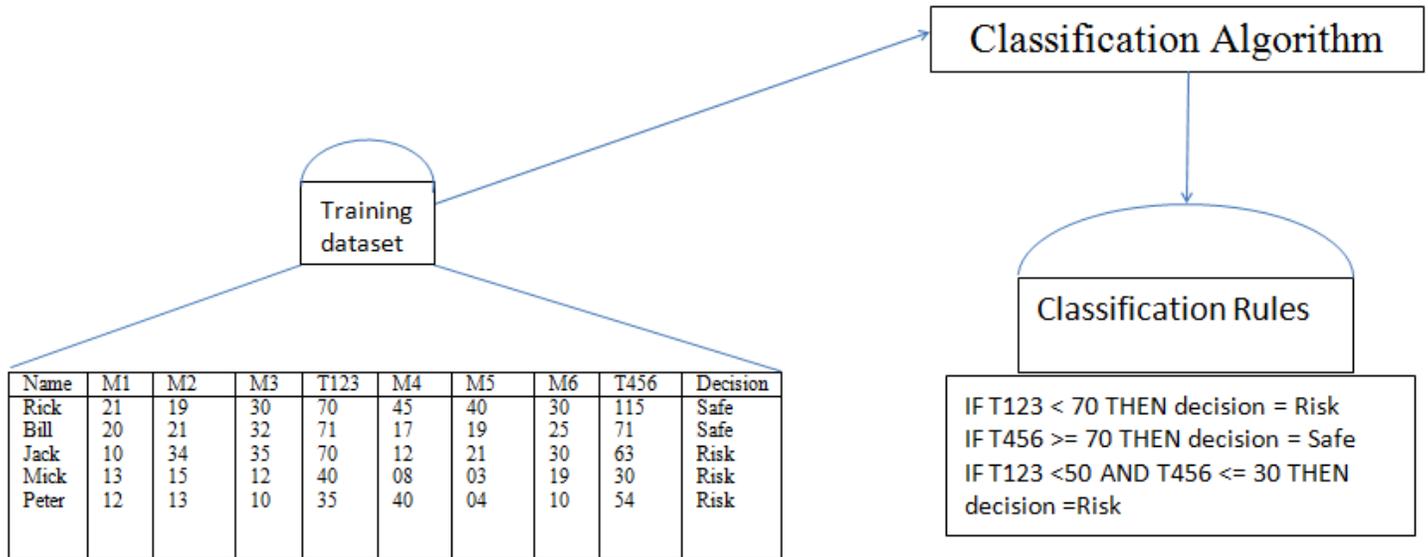
Fig. 1.1 Requirement and Result Model

2. Reasons

2.1 Applying Classification method for average learners:

Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example predicts with 85 percent accuracy which student will or will not

graduate. The university could use this information to concentrate academic assistance on those students most at risk[3].The understanding ability of average learners can be less. The learners may not imbibe things faster. By using classifier model of classification technique of data mining, we can decide which student is at risk and which student is safe from drop out. Fig. 2.1 shows Classifier Model.



subject is out of 50. M1, M2, M3, M4, M5, M6 are marks.T123 and T456 is total of M1, M2,M3 and M4,M5,M6 respectively.

Fig. 2.1 Classifier Model

From the above Classifier Model shows the marks obtain by student in six subjects in first year of graduation. We can predict that if the students who are at risk could not do well in upcoming exam or assignment,they are likely to drop out due to weak performance. For those students, teachers, family and society shouldtake proper steps and give proper guidance.

2.2 Family background

India is a diverse country having families with various cultural backgrounds in areas like family size, occupation and in extreme case religion. These factors pose their own restrictions on educational upbringing of children .When the scores of students are traced it is more often discovered that family background is major contributory factor [4]. Orhunger (1990) asserts that the cultural background of the family is determined to a larger extent by its socio-economic background.

Following are the rule that can be applied to predict the academic performance of the child.

Rule 1:

Parental occupation: Father – employed/Mother – employed
 Marital status of parents: divorced
 Annual income: above 10 Lakh
 Homely Environment: Unhappy

Rule 2:

Parental occupation: Father – employed/Mother – Housewife (unemployed)
 Marital status of parents: Stay together
 Annual income: 5 Lakh
 Homely Environment: Happy

Rule3:

Parental occupation: Father – Unemployed/Mother – Employed
 Marital status of parents: Stay together
 Annual income: 3 Lakh
 Homely Environment: Frequent quarrel

Rule 4:

Parental occupation: Father – employed/Mother – Housewife (unemployed)
 Marital status of parents: Stay together.
 Annual income: 10 Lakh.

Homely Environment: Unhappy.

Rule 5:

Parental occupation: Father – Employed/Mother – Employed
 Marital status of parents: Stay together
 Annual income: 4 Lakh
 Homely Environment: Happy but financial problems.

Many more rules can be applied to trace the family status of student. From **Rules 1 to 5**, we can predict that all the rules can directly or indirectly influence the interest of student thus degrading their academic performance. Above five rules states that due to parental occupation, marital status of parents or annual income, the homely environment of the family get disturb. Even though students suffer from family background problems, organization or teacher can improve the performance of student by taking him/her into confidence and proper guidance.

*A mother does homework with her children.
 It is generally acknowledged that family environment is the most powerful influence in determining a child's academic motivation and achievement.* HUREWITZ CREATIVE/CORBIS[5].

2.3 Interest in subject

There are many factors that affect the interest of student in subject.

- i. Lack of interest created by classmates.
- ii. Teachers demotivation.
- iii. Method of teaching.
- iv. Resources available for learning process.
- v. Preference given by parents to the subject (i.e. Subject is chosen by parents).
- vi. Level of organization where student is learning.

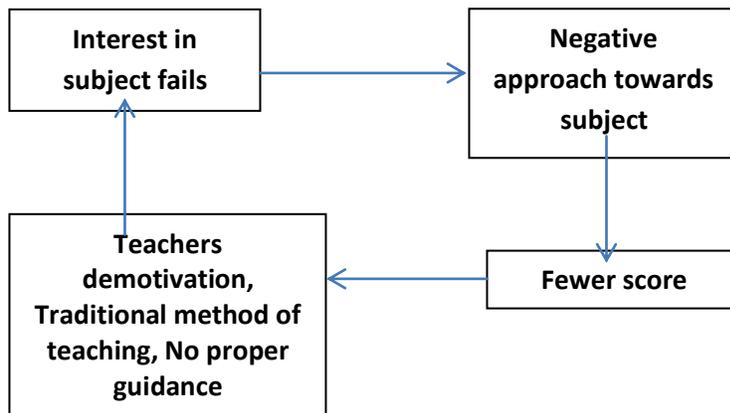


Fig. 2.3 Interest Cycles

The outcome of the above Fig.2.3.is that the negative approach is passed down from one generation of student to another and interest cycle keeps on repeating. Form the fig. above it is predicted that, the above factors can be overcome by providing following excerpts[6]:

- **eMpowerment:** Students feel empowered when they feel that they have some control over some aspects of their leaning. By giving students certain choices and taking inputs from them on some topics related to subject, will feel students dignified.
- **Usefulness:** Student should be explained the usefulness of course. It should be make known that the skills that students acquire from the subject will directly contribute to their success.
- **Success:** Teachers or parents should make student feel that they can succeed in the course if they make reasonable effort. Teachers should help student by

providing additional resources if needed, setting expectations, giving feedbacks, etc.

- **Interest:** Teachers should trigger their interest so that they pay attention and are interested enough while they are engage in teaching learning process.
- **Caring:** Caring is a big motivation for student. By understanding the grasping ability of students, teachers should teach subject accordingly.

Data mining combined with student demographics and other information, enables the academic institution to improve its understanding of its student types [3].

3. Conclusion:

In this paper we had tried to present problem of how factors like average learners, family background, no interest in subject

is responsible for academic development of student. With the help of classifier model and predictions rules we have tried to analyze the available data. This paper may help parents, society as well as the institution to be aware of the effect the surrounding and family has on children and how much it affect the child positively and negatively on academic performance. From the above classification and predictions, measures can be adopted for bringing up children in such a way that they will perform well academically.

4. Future enhancement.

This study can be replicated using different population to determine whether or not family structure and functioning plays a basic role in determining the wellbeing of a student. More study about the other factors like Socio economic, friend circle and some more reasons, etc. for why student drop out can be identified.

5. References:

- [1] Jing Luan. Data Mining Applications In Higher Education. 2004 Spss Inc. Dmhewp-1004.
- [2] Dr. Varunkumar, Anupamachadha. An Empirical Study Of The Application Of Data Mining Techniques In Higher Education. (*Ijacs*) *International Journal Of Advance Computer Science And Applications*. Vol.2, No.3, March 2011.
- [3] Shaeela Ayesha, Tasleem Mustafa, Ahsanrazasattar, M. Inayat Khan. Data Mining Model For Higher Education System. *European Journal Of Scientific Research*, Issn 1450-216x Vol.43 No.1(2010), Pp-24-29.
- [4] Simononowaowizy. The Influence of Family Background On Students' Academic Performance In Physics In Boko Local Government Area Of Benue State.
- [5] www.family.jrank.org, www.facultyfocus.com, www.tutorialspoint.com