iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

Analytics in Higher Education System

Ankita Chopra[1]
[1]*IT Department, Jagannath Institute of Management Studies (Affiliated to IP University)*

Abstract*: As per the national employability report 2016 and looking at the current trend it is very clear that unemployability among young IT graduates is increasing and the major reason can be related to the lack of skill set among IT students required by the companies .To better understand and solve the problem predictive analytics can be used. As huge amounts of data is collected and maintained by the educational institutions applying analytics on this data can derive valuable patterns and results. This paper tries to explain the major factors that affect the employability of an IT student and how analytics can help understand this factor. In this paper section I describes introduction, Section II explains how predictive analytics can help identify the factor that affects employability, Section III highlights results & discussion, Section 5 concludes the paper with references.
Keywords*: Employability, Analytics, Classification Algorithm, Skill Set, Patterns

I. INTRODUCTION

A. Current Scenario

In current scenario students taking up higher education courses have considerably gone down. The loss of interest can be contributed to many factors. In developed countries the percentage of students taking higher education is 28% to 90% whereas in developing country like India the percentage is as low as 13%. Students showing less interest in higher education have a direct effect on global economy of the country as the country does not produce employable graduates.

This scenario categorically speaks about how much we are lagging behind in terms of percentage of higher education institutions, number of students pursuing higher education, number of employable graduates to be produced .As per the figures we claim that in 2020 India will be ranked third in terms of higher education over all the other countries but the current scenario is that only one institution from India being featured into list of relevant institutions across the globe. The situation is alarming. To better understand and analyse the situation we need to first observe the challenges that the higher education system is facing and take relevant steps to improve it.

B. Challenges

These are the underlying challenges faced by higher education system today

The first challenge the system is facing is student who get enrolled think from, the perspective of return for the investment made. As many think of graduates are unemployed they think that the investment made to enrol for higher education system is a loss henceforth, there is a decrease in the number of students getting enrolled for higher education courses. Lack of un employability among graduates can be related to the skill set gap among students.

The second challenge the system is facing is the fact that with the advancement of technology and internet boom more & more students are applying for online courses to stay up to date with the industry standards bypassing the non –academic expenses which the higher education system imparts on students. The third challenge the system is facing is related to the worth of a college degree. As students feel that the salary of a college graduate relative to a diploma holder is not rising as the employers find that the college graduates lack distinctive qualities they want in new employee. The fourth challenge the system is facing is on the teaching/learning model with the information technology sector increasing in leaps and bounds the requirement of the industry are more advanced So, there is a need to improve the teaching and learning methods by incorporating new ideas and techniques and make the higher education system more innovative and flexible to changes. The fifth challenge the system is facing is there is no project based learning that means students need to learn and apply the skill to get a job which is lacking in the system today. Due to this reason the intake of students in higher education courses has reduced considerably as there is no output of the programme.

All the above challenges can be reduced if we start analysing our past student data. Analytics is emerging field in science and technology .In order to overcome the challenges. We need to analyse the rends that led to the dynamic change in higher education sector. In today's time huge amounts of data is generated and maintained by the higher educational institutions. We need to analyse the trend of past 5 years student data to reach a conclusion as to what is leading students to not opt for higher educational courses and how it can be improved.

With help of analytics institute can better identify and address the challenges faced by the institution in terms of candidate retention , higher number of in take for the courses etc.

II. ROLE OF ANALYTICS

A. Data Mining

Data mining is a step in knowledge discovery process. .In today's time collecting data is not a problem but to derive knowledge from data we need to analyze the data. The process of deriving knowledge from data is called as data mining. It is also called as knowledge mining. Data mining can be clearly understood with the help of KDD process.

B. Knowledge discovery process :

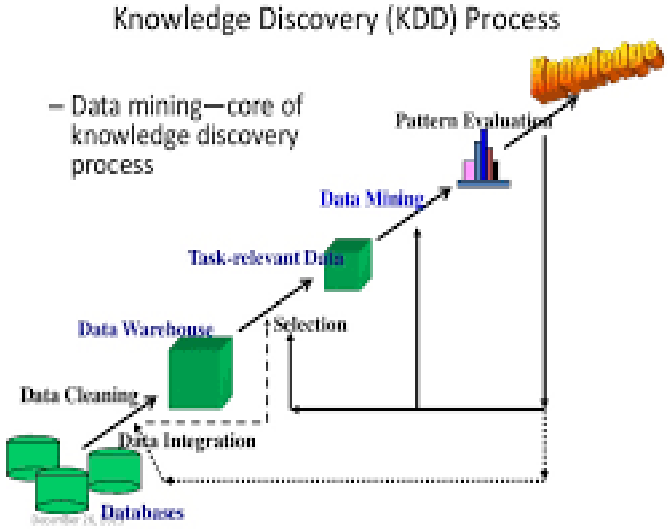


Fig 1:KDD Process

KDD is called as knowledge discovery process. The process is described as follows

Step -1: Data Pre processing: In this step data from heterogeneous sources like databases ,flat files etc is put into a centralized repository which is generally data warehouse (OLAP) sever. The data is cleaned, transformed and integrated into the OLAP server which is termed as preprocessing of data.

Step -2: Data Mining Engine : There are various data mining functionalities available to extract knowledge from the data put into the repository. The functionalities are generalization, comparison, association, classification, prediction , clustering & outlier analysis. The result of these algorithms is patterns which are further interpreted.

Step -3: Pattern Evaluation: Not all patterns generated by data mining algorithms are interesting. To understand interestingness of pattern . We need to measure its support and confidence based on these parameters patterns is evaluated.

Step -4: Knowledge Representation: After evaluating the interesting patterns knowledge is derived and represented using bar graphs, pie- charts, tables etc.

The above steps are also termed as primitive tasks of task relevant data.

C. MINING FUNCTIONALITIES (CLASSIFICATION & PREDICTION)

Classification and prediction are the most important data mining functionalities of all . They are basically used in all real world applications.

Classification is a two step process in which data is classified as per class labelled attribute this data set is called training data set. A test data set is taken as ample from training data set to check if the classification is done correctly or not . Classification rules are written and evaluated on the data set for accuracy. Classification is also called as supervised learning.

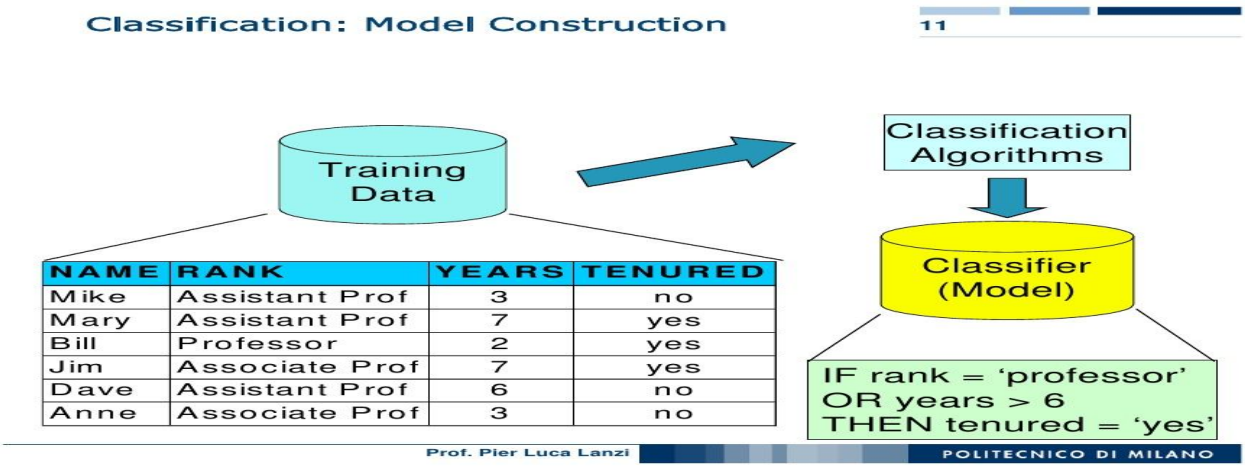


Fig 2 Classification model

Prediction is a data mining task which is used to classify future or unknown labels. Prediction is generally used when we have a training data set and we need to test the accuracy of this data set to predict the outcomes. There are various algorithms for prediction they are

1) Bayes theore
2) Back Propagation

Machine Learning & neural network approaches are also used for prediction.

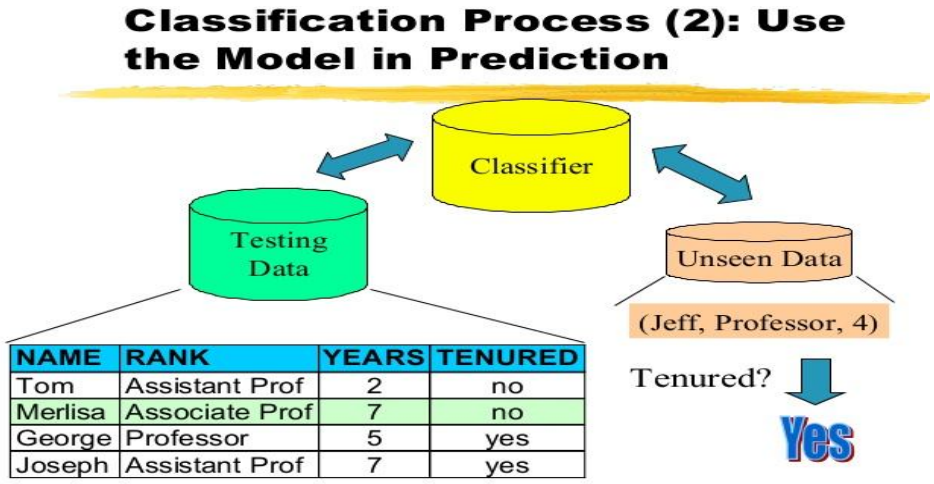


Fig 3:Prediction model

III.DATA MINING APPLICATIONS IN HIGHER EDUCATION

The major challenge faced by the higher education system as discussed above is decrease in number of students for higher educational courses. This can be further divided into sub challenges like

A. lack of un employability
B. ack of project based learning method sutdated teaching/learning methods

To identify the factors and to overcome the challenge decision tree algorithm is used

Decision Tree Algorithm: In this algorithm the database with its attributes and values is taken for analysis. Among all the attribute's present attribute selection criteria is applied to identify the splitting attribute for the tree. Measures like information gain, gain ration & gini index are calculated. Depending on the criteria that information gain > attribute value .If this criteria is me then that attribute is taken as splitting attribute and decision tree is constructed . On traversing the tree it is very clear that if a student has done his

secondary education from science background then that student is sure to get enrolled for higher education courses. As the outcome of the algorithm is enrolled or not enrolled. Decision rules are formed which are then evaluated to derive patterns. The algorithm can be illustrated as follow

```
INPUT: S, where S = set of classified instances
OUTPUT: Decision Tree
Require: S ≠ ∅, num_attributes > 0
 1: procedure BUILDTREE
 2:     repeat
 3:         maxGain ← 0
 4:         splitA ← null
 5:         e ← Entropy(Attributes)
 6:         for all Attributes a in S do
 7:             gain ← InformationGain(a, e)
 8:             if gain > maxGain then
 9:                 maxGain ← gain
10:                 splitA ← a
11:             end if
12:         end for
13:         Partition(S, splitA)
14:     until all partitions processed
15: end procedure
```

Fig 3: Decision tree algorithm

This algorithm is applied on the student database with the following features

TABLE I
ATTRIBUTES OF STUDENT DATABASE

Sno	Name of feature	Values
1	Student Name	{a,b,c,d,e,f}
2	Sex	{Male,Female}
3	Enrollment No	{1,2,3,4,5}
4	Enrollment year	{2000,2001,2002,2003,2004}
5	Date of Birth	{x,y,z,g,h}
6	Primary Education	{60%,40%,55%,23%,72%}
7	Secondary education	{70%,60%,50%,78%,80%}
8	Father's Name	{ab,cd,ef,gh,ij}
9	Occupation	{sal,buss,sal,sal,sal}

From the above factors it is identified that secondary education has highest information gain and it is taken as the splitting attribute.

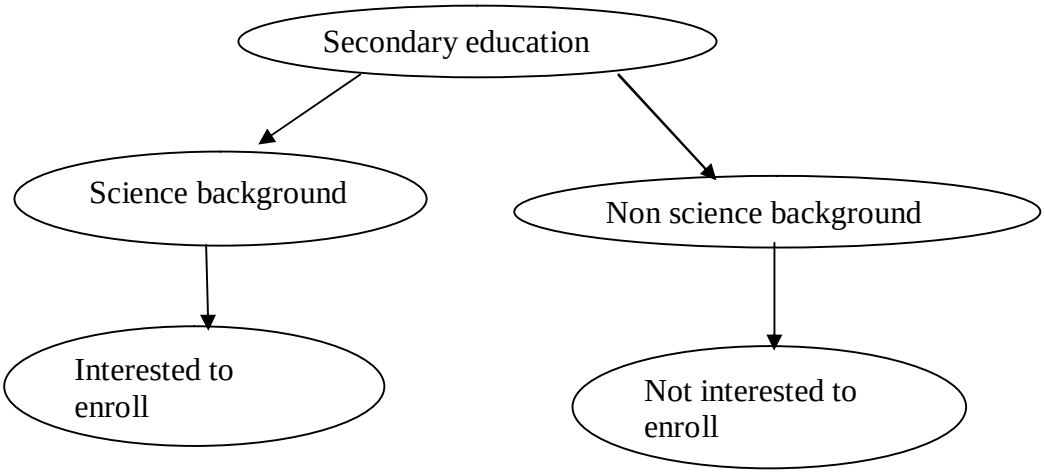


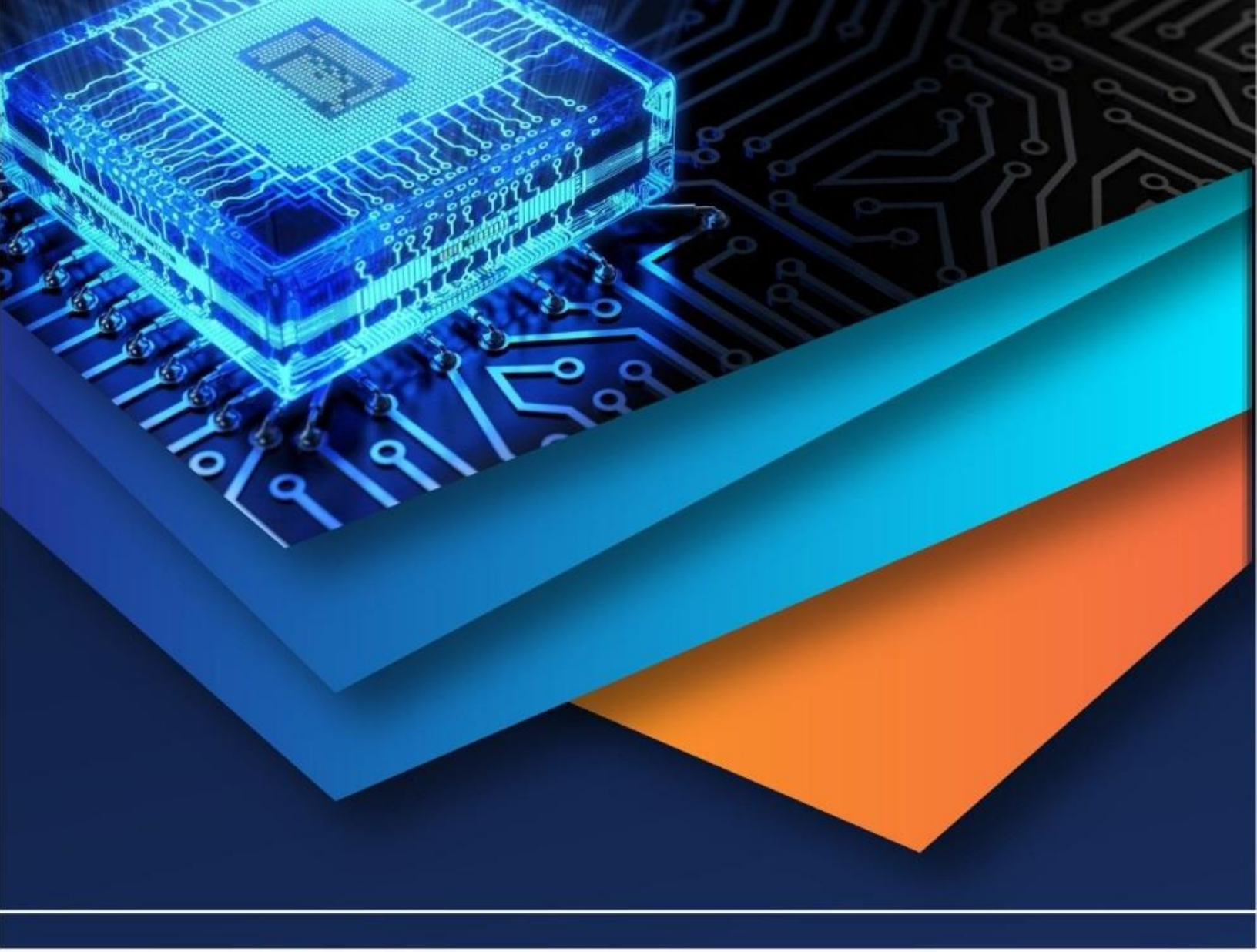
Fig 4: Decision Tree

IV. CONCLUSION

There are many challenges in higher education system and if we want to overcome them we need to start analyzing the student database and make predictions for future. In this paper we tried to analyze student's interest in higher education courses and the factors effecting it. From, the results we can conclude that the student with average marks in secondary education and science background is more likely to get enrolled for higher education course. Future scope would be to understand the factor effecting employability of a graduate student.

REFERENCES

[1] Shaeela Ayesha,(2010),data mining model for higher education system, J. of scientific research, vol-43, pp24-29
[2] Sunita B. Aher,(2011), data mining in education system using WEKA, International J. of Computer Applications, pp20-25.
[3] Alaa el- Halees,(2009), Mining Students Data to Analyze e-learning behavior. A case study.
[4] Connolly T., C. Begg et al,(1999) Database System: A practical approach to design, Implementation and management(3 rd edition), Harlow; Addison-Wesley,687
[5] Erdogan and Timor (2005) A data mining application in a student database. Journal of Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)
[6] Han,J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor
[7] ZhaoHui. Maclennan.J, (2005). Data Mining with SQL Server 2005 Wihely Publishing, Inc
[8] Bhise R.B, Throat S.S, Supekar A.K, "Importance of data mining in higher education system ", IOSR Journal, ISSN:2279-0837, vol 6, Issue 6(jan-feb 2013), pp 18-21.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY