International Journal Of Engineering Research And Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 13, Issue 11 (November 2017), PP.60-65

A Framework For Analytical Services Using Data Mining Technuiques to Predict the Enrollment of Student At A University – A Case Study

S.Radhika*, B.Aravind Reddy
[1]*Asst.professor,Dept of CSE Nalla Narsimha Reddy educational society's Group of institutions,Chowdariguda, KorrrumlaX,Medchal,Hyderabad*
[2]*Senior quality Analyst,IBM,Arena towers,Ramanthapur,Hyderabad*
[1]*radhikareddysb@gmail.com*
*Corresponding Author: S.Radhika**

ABSTRACT: Major problem related to admissions in universities and affiliated colleges is that most of the universities not able to predict the likelihood of an applicant that whether the candidate is willing to enroll in an academic program or not. Universities usually spent more expenditure in promoting their programs. Identifying candidates of higher chances of enrollment into the academic program will help the university to reduce promotional expenditure. A candidate in general applies to more than one university to extend their chances of getting enrolled within that academic year. If Universities can give their decision quickly to the candidate then there will be higher possibilities of getting acceptance from registered candidates. Most of the Universities collect the data from applicants as part of their admissions process. In this paper we used the data available, which is collected from 'X'
University to predict whether an candidate is willing to join an academic course or not. And also focused on analyzing the crucial characteristics that influence the admission process. A model is built to solve a current problem facing in the admissions department at almost in all academic institutions. We used a classification technique which can solve the current problem and derives the various steps that are performed by Analytic Services

Keywords: OLAM, Naive Bayes, Accuracy, classification.

Date of Submission: 05 -09-2017 Date of acceptance: 23-11-2017

I. INTRODUCTION

Data Mining is the process of discovering hidden patterns in data by applying different techniques like classification, prediction, and association's techniques among the data objects based on class and concept and presenting mining results. For example if customer is purchasing a product, then customer 'X' represent class and the concept, of customer represent big spender or less spender. Data Mining integrates with Online Analytical Processing (OLAP) and provides extremely flexible and extensible for the users. On-line analytical mining(OLAM) extensively increase the power of analysis by providing various available services, applying on different subsets of data at different levels of abstraction in combination with various analytic services of OLAP like drill down, drill up, filtering ,pivot, , slicing and dicing.

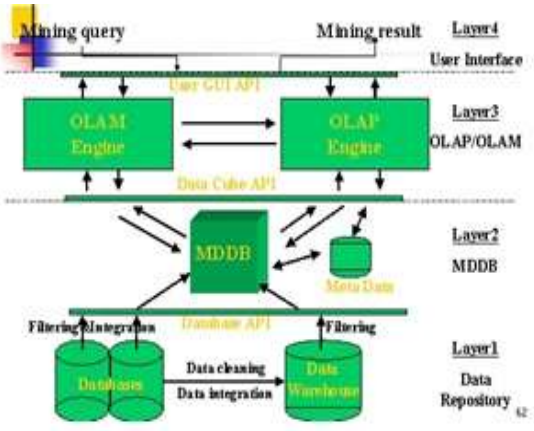


Fig1: OLAM Architecture

II. PROBLEM STATEMENT

Identifying candidate of higher possibility of getting enrollment into academic program will be the major problem. A candidate in general applies to more than one university to extend their chances of getting enrolled within that academic year. If Universities can give their decision quickly to the candidate then there will be higher possibilities of getting acceptance from registered candidates. Most of the Universities collect the data from applicants as part of their admissions process. Analyzing the complex characteristics influencing the enrollment process can help in adjusting the admissions policies at the university and it ensures efficient and effective cost management in admission process at universities.

III. LITERATURE SURVEY

In paper [9], author used an approach based on extraction of features from web based educational system to classify students to achieve the result of their grades. The author used various classification techniques and compared their performance on particular dataset. Further, proceeded with various combinations of classifiers to improve the accuracy of classifier.

In paper [10], the author used association rules and analyzed about its uses in Educational institutions.

In paper [11], the author focused on how classification techniques can be useful in improving the performance of students in educational institutions. And author applied various mining algorithms on dataset based on the subject considered Moodle e-learning of student.

In paper [12], the author considered the relationship between results of the students appeared in the University entrance examination & their success rate, using clustering techniques students at university are grouped according to their characteristic, forming clusters and calculated mean distance between clusters using K-means algorithm.

IV. PREPARING DATASET

As part of the administration the admin department collects information from candidates. Historical data is also made available to the candidate indicating the status of applicants who are enrolled in last academic year .The following dataset contains 33 attributes for each applicant and there are nearly about 11355 tuples available.

S.NO	Crucial attributes	Description	Type of attribute	Data type
KEY ATTRIBUTE				
1	ID	Identifier for the record	Numerical	number
Identity Key				
2	CITY	Address details	Categorical	String
3	STATE	Address details	Categorical	String
4	ZIP1	Location identity	Categorical	number
5	ZIP2	Location identity	Categorical	number
6	DOB	Applicant Date of birth	Numerical	number
7	GENDER	Applicant male/female	Categorical	String
8	ETHNICITY	Nationality of applicant	Categorical	String
Score secured				
9	SSC	Secondary school marks/percentage	Numerical	number
10	Entrance-1	Verbal exam %	Numerical	number
11	Entrance-2	Non verbal exam %	Numerical	number
12	Inter score	Intermediate score/percentage	Numerical	number
13	Test score	Total marks secured in entrance exam	Numerical	number
14	Total score percentage	Total % of score secured	Numerical	number
Application specific data				
15	Primary source	How did the applicant know about the college	Categorical	String
16	Source date	When did applicant got the news about college	Numerical	number
17	Application date	Date when applicant taken application	Numerical	number
18	legacy	Any Past associations with applicant	Categorical	String
19	First date visited campus	Date when first visited the college camplus	Numerical	number
20	First load date	Date when applicant data is entered in to system	Numerical	number
21	Multiple contacts	Multiple contacts of applicant	Categorical	String
22	Application type	Type of application	Categorical	String
23	Application status	Status of application	Categorical	String
About college				
24	College type	Type of college	Categorical	String
25	Major area	specialization offered by college	Categorical	String
26	Financial Aid interest	Financial required for applicant	Categorical	String
27	Financial aid Received	Financial aid max amount to be given for applicant	Categorical	String
28	Total budget required	Total amount required	Numerical	number
29	Parent contribution	Contributions in fee by the parent	Numerical	number

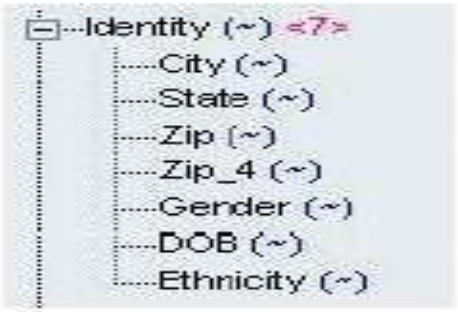
Table 1: Available data

Data is brought in to the cube environment once the preprocessing is done and then be accessed through analytical Framework for predictive analysis. This Framework uses Multi Dimensional expressions (MDX) to identify the blocks within the cube to get input data. This analytical framework will take only ordinary dimensions as mining attributes. The data required for predictive analysis should be measured within a cube and modeled with standard dimensions.

- **Data Pre-Processing**

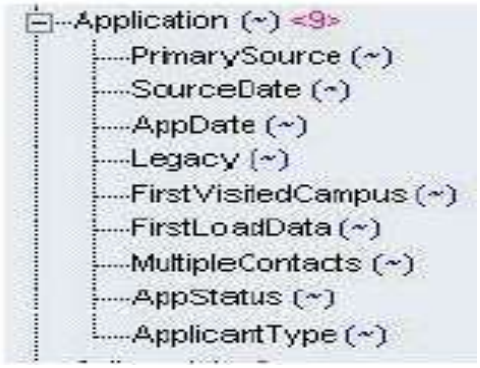
Mostly the input data will be available only two type's i.e...either in number or string. But, in Analytical Services it is necessary to store the data in a numerical format, hence all string type input data has to be encoded into a 'number' format. Example, if the gender information is given as a string with values Male or Female , it needs to be convert into a numerical values like '0' or '1', before storing the data in the Analytic Services. Mining attributes can be of two types 'categorical' or 'numerical'. Attributes that describes discrete information like about gender , zip code , customer category ('Gold', 'Silver', 'Blue'), status information ('Applied', 'Approved', 'Declined', 'On Hold'), etc. can be termed as 'categorical' attributes. Attributes that describe continuous information like about sales, revenue, income, etc. are termed as 'numerical' attributes. The Analytic Services has the capability of handling both categorical and numerical attribute types. The effectiveness of algorithm will be more dependent on the quality and completeness of the source data. The current problem at X University, the data is available in both 'string' and 'number' formats.

- Gender, City, State, and Ethnicity are related to object Identity –



Will be transformed from string to number format

• Application Status, Primary Source of contact, Applicant Type, etc. related to the application process



• Application Date, Source Date, etc. are available as strings, in two different formats – 'yymmdd' and 'mm/dd/yy', hence these values should be encoded into a numeric format.

All string type of attributes will be encoded into corresponding numeric values while loading data into analytical services

State id	State name
1	TG
2	AP
3	AR
4	AS
5	BR
6	AB
7	CR

Applied Status ID	Application Status
1	Applied
2	Offered admission
3	Fee paid
4	Course enrolled

Table 2: string values encoded to numeric

V. IMPLEMENTATION

Selecting an appropriate algorithm major step in mining process. There various algorithms available for Mining process like Naive Bayes, Regression, Decision Tree, Association Rules, Neural Network, and Clustering. Selecting an suitable algorithm for a specific problem needs basic knowledge of the problem and mathematical techniques to solve problems efficiently and effectively in any domain. The problem that is being discussed in this paper is to classify each applicant into a discrete set of classes on basis of certain numerical and categorical data collected from candidate. Class referred to the status of the applicants applications like whether he 'will enroll' or 'will not enroll' the academic course.

- **Building A Model:**

Fig 2: selecting a database to be mined

Fig 3: Selecting a new task

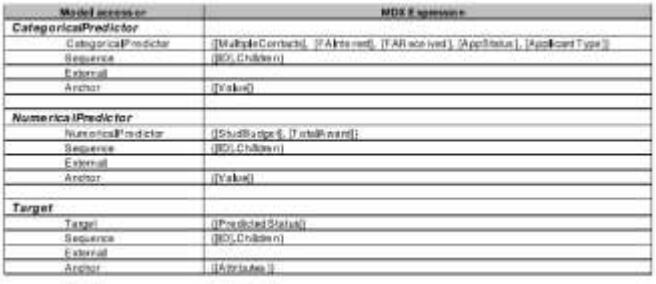


Fig 4: Setting accessors using naïve bayes algorithm

➢ **Testing**

Fig 5: Model based on naïve bayes algorithm

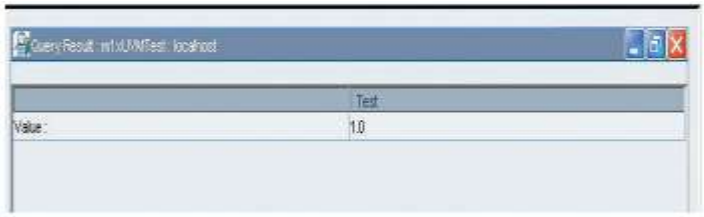


Fig 6: Test results

VI. INTERPRETING THE RESULTS

The model was verified against the set of available dataset (over 11355 instances). And the possible outcomes will be either the candidate will enroll or will not enroll the course.

➢ **The Confusion Matrix**

The model predicted that 1307 (1256+ 51) students will enroll. From that , only 1256 actually enrolled and 51 has not enrolled. This implies that there were 51 false positives. Similarly, 10048(9578 + 470) students will not enroll. Out of that, only 9578 did not enroll, and 470 actually enrolled. This implies that there were 470 false negatives.

N=11355 Actual/Predicted	**Will Enroll**	**Will Nor Enroll**	
Will Enroll	1256	470-->Fn	1726
Will Not Enroll	51->Fp	9578	9629
	1307	10048	**11355**

FP-False positive
FN-False Negative

Table 3: Confusion Matrix

➢ **Anaysing The Result**

Accuracy of classifier

Here
P=TP+FP=1256+51=1307
N=TN+FN=9578+470=10048
TP=1256
TN=9578
Hence
Accuracy=TP+TN/P+N
Accuracy=1256+9578/1307+10048 =10834/11355=95.4%
Accuracy=95.4% correclty classified

4.58% is classified incorrectly

Incorrect predictions	no of cases	% of cases
False positives	51	4.49%
False Negatives	470	4.13%
TOTAL	521	4.58%

Table 4: Accuracy of classifier

VII. CONCLUSION

Data mining framework is one of the comprehensive enterprise class analytical functions offered with Analytic Services. In this paper 'Naive Bayes' algorithm is used to solve a real world problem, we observed there a was 95.4% success rate using Analytic Services Data Mining Framework. In future different classification algorithms can be applied to achieve more accuracy and further comparison can be done with different classification techniques.

REFERENCES

[1]. M. Al-Razgan, A. S. Al-Khalifa, and H. S. Al-Khalifa, Educational data mining: A systematic review of the published literature 2006-2013, in Proc. the 1st International Conference on Advanced Data and Information Engineering, 2013, pp. 711-719.

[2]. F. Siraj and M. A. Abdoulha, Mining enrolment data using predictive and descriptive approaches, Knowledge-Oriented Applications in Data Mining, pp. 53-72, 2007.

[3]. Q. A. Al-Radaideh, A. A. Ananbeh, and E. M. Al-Shawakfa, A classification model for predicting the suitable study track for school students. International Journal of Research and Reviews in Applied Sciences, vol. 8, 2001.

[4]. B. K. Baradwaj and S. Pal, Mining educational data to analyze students' performance .International Journal of Advanced Computer Science and Applications, vol. 2, 2011.

[5]. Nandeshwar and S. Chaudhari. (2009). Enrollment Prediction Models Using Data Mining.[Online].

[6]. D. Kabakchieva, Predicting student performance by using data mining methods for classification, Cybernetics and Information Technologies, vol. 13, 2013.

[7]. D. Garc´ıa-Saiz and M. Zorrilla, Comparing classification methods for predicting distance student's performance, The Journal of Machine Learning Research, 2011.

[8]. C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, Data mining algorithms to classify students, presented at 1st International Conference on Educational Data Mining, 2008.

[9]. Z. J. Kovačić, Early prediction of student success: Mining students enrolment data, present ed at the Informing Science & IT Education Conference, 2010.

[10]. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The weka data mining software: an update, presented at the Special Interest Group on Knowledge Discovery and Data Mining, SIGKDD, 2009.

[11]. E. Boretz, Grade inflation and the myth of student consumerism, College Teaching, vol. 52, 2004.

[12]. W. H¨am¨al¨ainen and M. Vinni, Classifiers for educational data mining," Handbook of Educational Data Mining, 2010.

[13]. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

[14]. Han, J. and Kamber, M., 2006. Data mining, Amsterdam: Elsevier.

[15].

S.Radhika*. " A Framework For Analytical Services Using Data Mining Technuiques to Predict the Enrollment of Student At A University – A Case Study." International Journal Of Engineering Research And Development , vol. 13, no. 11, 2017, pp. 60–65.