

Assessment of Performance of Students using Conditional Statistical Technique



Hema Malini B H, Suma V, Suresh L, Shankar M M

Abstract: Educational data mining (EDM) is gaining importance in every field. Due to the competency in every branch of engineering, the institutions are concentrating mainly on improving the performance of students. Efforts are also put towards knowing the reasons for low performance and identifying the factors affecting the student's performance. Researchers are working on preparing predictive models for improving student performance. The present study is considering the educational data of 1186 students. The data is classified as demographic and study related variables. An effort is made to predict the student performance, using a statistical technique – Chi Square test. The attributes affecting and not affecting the performance of students are assessed. The results are plotted using Pie Chart and histograms. The association between demographic and education variables with semester results is tabulated.

Keywords: Student Performance, Demographic, study related, Educational Data Mining.

I. INTRODUCTION

The number of engineering colleges is shooting up day by day. Parents desire to see their wards as engineers are still in its boom. Though there are opportunities to be the entrepreneurs, many parents still prefer their children to get placed in the campus. Because of the competency in professional colleges, it is the responsibility of every faculty to improve the performance of their undergraduates and bring up the results, thus improving the rate of students eligible for placements. The admission of any college depends on the results and placements the institution provides. To maintain, this chain in balance, the educational data mining has extended its importance in the present day. Lot of researchers are working in the area of Educational Data Mining. There is a tremendous growth in the educational field. Present day students have migrated from traditional text book studies and studying from notes. There is lot of information available on internet. Lot of educationalists record videos and post it in YouTube.

Students find it easier to watch the videos and prepare for their examination rather than sitting with the class notes. There is a drastic change in the educational field. The continuous interaction with the undergraduates shows that most of the students prepare for examination only in the last minutes. Only few students study from the beginning of the academic year. The present study is focused on engineering students in a premiere institution in a southern state of India, Karnataka. The institution is affiliated to a premiere University. The university runs semester schemes. Each academic year has two semesters – Odd and Even. For institutions which are affiliated to this university, it takes the responsibility of deciding the courses to be studied in each semester. The Board of Studies (BoS) members decide the scheme for each semester, the detailed syllabus to be studied in any course.

But, the traditional examination system is the same. Students have to prepare for university examinations which are considered as external examination.

The question paper set by any senior professors in the university would come as university question paper. The undergraduate has to get the minimum qualifying marks to clear the course. In each course, internal assessment marks would be awarded. The students are expected to keep the minimum requirement of marks. To clear a course, a minimum of marks has to be maintained by the student including internal and external marks. Keeping all these factors in mind, the academic performance of a student plays a prominent role in deciding the results of any student.

So, researchers in the field of higher education have concluded EDM as a research area wherein predictive models need to be prepared. The models are used to predict the drop outs from engineering course, predicting the best average and weak performers in any class and give the necessary support for them to boost up the performance, the factors affecting the performance, the demographic, social factors, affecting studies etc. All possible parameters affecting the performance of students are listed by the researchers in their works. In the related literature, the researchers have made attempts to apply machine learning algorithms to obtain the predictive models to study the performance of undergraduates. The present work is an attempt using the statistical method to assess the student performance of students. Since the given variable (x and y) in the dataset is categorical data, Chi Square test is used to find out any association between x and y variables. x is a dependent variable. It could be gender, accommodation, branch, selection, board, puc_marks, annual income etc. y is independent variable.

Revised Manuscript Received on March 30, 2020.

* Correspondence Author

Dr. HemaMalini B. H*, Associate Professor, Department of CSE, BMS Institute of Technology, Bengaluru, India.

Dr. Suma V. Professor, Department of ISE, Dayananda Sagar College of Engineering, Bengaluru, India.

Dr. Suresh L. Principal and Professor, Department of CSE, Cambridge Institute of Technology, Bengaluru, India.

Mr. Shankar M. M. Data Scientist, CARES, Bangalore, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The semester end result is considered as independent variable. The effort is to find out which variable contribute the most for student performance. The results are tabulated.

II. BACKGROUND

Akansha Mishra et. al [1], have used clustering method to categorize 84 students from North Spain studying under graduation in Psychology.

The students are divided into two groups: passed and failed students. Three clusters are created for grouping all the students. All clusters contain passed and failed students. Three models were obtained for studies. The models were useful in analyzing the characters of students and planning classes, providing necessary support to weak students.

Amirah Mohamed Shahiria et. al [2], have taken up the survey of forecasting academic performance of undergraduates in Malaysia. The extensive survey covers the works involving use of machine learning methods to analyze the performance of students in Malaysia. An extensive survey is conducted considering almost all the attributes that affect the performance of any student. The attributes considered are student demographic information, high school background, internal assessment marks, CGPA (Cumulative Grade Point Average), social network, scholarship information-if any, extra-curricular activities-if any, psychometric factors and external assessments. The research work under discussion involved techniques applied for assessment: Neural Networks, Decision Tree, K-Nearest Neighbour and Support Vector Machines.

Ankita Katare, et al [3] discuss about EDM. The authors revisit the various Knowledge Data Discovery processes, the various questions that can be posed in studying the performance of undergraduates. The researchers also discuss about the various data mining techniques available. Decision Tree Classifier, Bayesian Classifiers, K-Nearest Neighbor Classifier (K-NN), Neural Network and Support Vector Machine are discussed in detail. Their research discuss about related work on the performance of students using different data mining techniques and analytical techniques.

Anoushka Jain et. al [4] has made an attempt to get a predictive model using data mining methods to measure the intellectual performance of undergraduates. The researchers have considered the social factors to correlate the academic performance and an attempt is made to get the best predictive model for the available dataset. The researchers have considered 212 records. The data collection was done for the following attributes: Gender, Percentage, Class studying, living location, mode of transport-public or walking, size of the family, income of the family, marital status of parents (married or widowed), qualification of father, qualification of mother, occupation of father, occupation of mother, no friends, weekly study hours and percentage scored are taken as input. The data mining techniques used were Decision Tree, ID3, C4.5 Decision Tree, Random Forest, Random Tree and REPTree. A comparison analysis is made. Naïve Bayes Classification is also applied. Anwar Ali Yahya et. al [6], have carried out their research on the kind of questions that a faculty rises in the classroom to assess the performance of students. The weightage is given to the highly famous Blooms Taxonomy cognitive levels, which is very helpful for NBA (National Board of Accreditation) and almost all

colleges follow the idea these days. The six levels in Blooms Taxonomy namely Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation is explained in detail.

Brijesh Kumar Bhardwaj et. al [7], have collected data from five degree colleges in Uttar Pradesh. The details of students pursuing BCA were collected for research. MATLAB is used for implementation purpose. The attributes used for research were grade of the student in senior secondary education, locality of living, medium of teaching, qualification of mother, other habits of the student, income of the family and family status of the student. Bayesian classification method was applied on the available dataset. The authors have concluded that in addition to the efforts that they put up in studies, there were other parameters in personal life affecting the performance. If the failure rate could be predicted at an earlier stage, the proper attention could be given and results could be improved.

Fatma Chiheb et. al [8], have considered the case of Algerian University, to forecast the performance of undergraduates. The technique used is decision trees. The CRISP-DM model is followed. The model involved Business understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. The data was collected from the 2009/2010 to 2014/2015 batches. Decision Tree and J48 algorithms were applied for the study. Established on the current results of Under Graduate and Post Graduate students in the department, an attempt was made to forecast the performance of undergraduates in the next semester. Their future aim was to increase the size of data set and reduce the error rate by improving the accuracy of the models.

HemaMalini B. H et. al [9], have started their research by considering the feedback students on a particular course, used various machine learning algorithms and analysed the performance. The authors have framed a questionnaire pertaining to the course taught, circulated through social media, obtained responses and have carried out the data analysis using WEKA tool. It was concluded that Naïve Bayes algorithm gave better performance. It gave 100% accuracy for the given dataset. HemaMalini B. H et. al [10], conducted a survey on 255 students. A questionnaire was circulated and responses were collected. Various machine learning techniques were applied. The implementation was done using Python. The algorithms applied were Random Forest, SVM and Logistic Regression. The Random Forest classifier showed 88% accuracy for the given dataset. It was concluded to enhance the size of dataset and apply other techniques to obtain the predictive models. Iti Burman et. al [11] have made an attempt to apprehend an individuals' learning behaviour. A survey was conducted to get the inputs from the youngsters. The data mining algorithms are used to obtain predictive models. The research has proved the importance of psychological factors in assessing the performance of students. M. Amala Jayanti et. al [13], bring about the various stakeholders in Educational Data Mining. The authors discuss about the various stakeholders like students, teachers, parents, Institution and government and also the corporate business community which provide placement to the students.

Pratiyush Guleria et. al [14], have collected data and performed analysis for two educational institutions. The parameters considered for assessment are teaching abilities, course contents delivery and infrastructure. The authors concluded that in performance of faculty and the infrastructure provided, there is difference between the institutions. Infrastructure is better in a particular institution whereas faculty performance is better in the other.

III. RESEARCH METHODOLOGY

In the related literature, the researchers have made attempts to apply machine learning algorithms to obtain the predictive models to study the performance of students. The present work is an attempt using the statistical method to assess the student performance of students. Since the given variable (x and y) in the dataset is categorical data, Chi Square test is applied to discover out any association between x and y variables. x is a dependent variable. It could be gender, accommodation, branch, selection, board, puc_marks, annual income etc. y is independent variable. The semester end result is considered as independent variable. The effort is to find out which variable contribute the most for student performance. The results are tabulated.

The data is collected from a premiere engineering college in Karnataka, in India. The institution has a data repository which stores the information of all students in the institution. From this repository, the data was retrieved and filtered manually. The process took abundant time. The data collection process focused on admission related data for the students admitted during the two academic years 2017-18 and 2018-19. The students are from the core branches of engineering like CSE, ISE, EEE, ECE, TCE, Mechanical Engineering and Civil Engineering. The following details of all students were collected: Gender, Date of birth, admission category CET/ COMED-K /Management), reservation-if any, pre-university board (State/CBSE (Central Board of Secondary Education)), total marks scored in pre-university examinations, percentage, accommodation (hostel or day scholar), occupation of both parents and annual income.

After collecting data, the data had to be pre-processed. The repository maintained all semester student's data. It included passed out data of students also. So, precaution was taken to filter the data and collect only student information of two recent batches. In pre-processing phase, the higher semester data of students was filtered off. The students who have opted change of college was also dropped out. The students who have not appeared for examinations were also filtered off. In the initial phase, data collection was done from around 3000 students. After pre-processing, the count came down to 1186.

IV. CASE STUDY

The institution chosen for research is a premiere institute in Karnataka in India affiliated to a premiere university. The admission of students happens through mainly CET or COMED-K quota. There are other quotas like Management, PIO (People of Indian Origin), NRI (Non-Resident of India), FN (Foreign Nationals), SNQ (Super Numerical Quota), SC/ST (Scheduled Caste/Scheduled Tribe) based on the percentage scored in 12th Grade and CET score. It is

accredited by NBA and NAAC. The NIRF (National Institutional Ranking Framework) has ranked it below 100 indicating that it is a top ranked institution. The NIRF is framed by MHRD, Government of India, to assess the quality of engineering colleges across the country under a common framework. The admission is sought by students with good ranking. The engineering course is planned by the university for four academic years (eight semesters). All the colleges under VTU run semester schemes. The colleges will have two semesters – Odd and Even in an academic year. The first year students will have two cycles: Physics Cycle and Chemistry Cycle. Every student has to take up both cycles on a rotational basis. It means, if a student has studied in Physics cycle in first semester, second semester would be Chemistry cycle for him/her and vice-versa. The student enters the core branch in third semester.

The university follows a carryover scheme. To enter into higher academic year, the student can carry a maximum of four subjects as backlogs. Detention in more than four subjects leads to year back. The student gets detained and has to clear the backlogs and can maintain a maximum of four subjects to enter into next academic year. Before entering third semester, the student can carry a maximum of four subjects as back logs from first and second semesters. To enter into third year (fifth semester), student has to clear all subjects from first year (first and second semesters) and can maintain a maximum of four subjects as back log from second year (third and fourth semesters). Similarly, to enter into final year (seventh semester), the student should have cleared second year (third and fourth semesters) and can carry a maximum of four subjects from fifth and sixth semesters.

Every student gets a unique admission number at the time of admission. The university allots a unique USN (University Seat Number) for each student later which is used till graduation of the student. Till the student gets the USN, the student is identified by the admission number. The admission number is maintained as primary key.

At a later stage, before the first semester examinations, once all the student admissions are completed, the university releases the USN of students. The office releases the list of students along with USN. The pre-processed data file was prepared before the USN generation. So, now keeping the admission number as primary key, a comparison is made. The students having the same admission number entry is checked with the USN. A new USN column gets added to the data file. The two files were merged. The resultant file had both admission number and USN. After the USN is generated, the admission number is no longer useful, since the admission number is known only within the institution. Once the students get the USN, it would be used for student identification till the award of degree from the university.

After pre-processing of the data, the researchers had to wait for the announcements of the results on the university portal. The university releases the student results on the website. Once the results were announced, the results were downloaded from the website.

Assessment of Performance of Students using Conditional Statistical Technique

The marks of all courses were added and gross mark was calculated. The gross marks were used as outcome variable for the data analysis.

Few of the students who had admission never got a USN. So, it was assumed that the student has either left the college or might have taken change of college. Such student details also were filtered. In few cases, the gross marks scored in university examination were either zero or single digit mark.

Assuming that it is not a proper data, it was filtered off. After all the pre-processing process was completed, the cleaned data file was obtained. The data file was saved as .CSV file (comma separated file).

Using the USN, the results are fetched from the university website. This file was used as input file for the statistical analysis. The gross mark is used as the outcome variable. For implementation, the programming language used is R.

The sample data file is shown in figure 1. It consists of the following attributes: Branch, Selection, Gender, Blood Group, Reservation, PUC_Board, PUC_Marks, Accommodation, occupation of Father, Occupation of Mother, Annual Income, PUC_Percentage and PUC Results.

V. ALGORITHM

Step 1: Data collection from institutional repository.

Step 2: Data pre-processing by deleting the students from higher semesters and collecting students' data of 2017-18 and 2018-19 batch.

Step 3: Fetching the admission number of students of these batches.

Step 4: Mapping the admission number to USN

Step 5: Fetching the university results using USN from university website.

Step 6: Separating the variables as Study related variables and Demographic variables.

Step 7: Preparing the predictive models using statistical method and by applying Machine Learning Techniques.

Step 8: Generating the results, to study the impact of various parameters on the performance of students.

VI. IMPLEMENTATION

The input variables are classified as:

1. Study related variables
 - a. Branch
 - b. Selection
 - c. Reservation
 - d. Board
 - e. Accommodation
2. Demographic variables
 - a. Gender

- b. Blood Group
- c. Occupation of Father
- d. Occupation of Mother

VII. PROCEDURE TO CONDUCT CHI SQUARE TEST

Step 1: Examine the collected data on the outlier; eliminate extreme cases from the analysis

Step 2: Exploratory data investigation is attempted to understand the patterns in the data, charts like pie and bar chart is used

Step 3: Selected independent variables such as demographics and study related attributes. Semester result is considered as dependent variable or outcome variable

Step 4: Test of independency - Chi square test is applied, which comprehend whether between two categorical variables, there is dependency or independency

Step 5: Chi square is applied on observed frequency

Step 6: Computed expected frequency based on row and column total by grand total or total observation

Step 7: Chi square value is computed by difference of square of observed frequency and expected frequency, divided by expected frequency. The summation will give chi square value

Step 8: Degrees of freedom is computed based on total row minus 1 and multiplied by total column minus 1

Step 9: When P value is < 0.05 , reject null hypothesis and accept alternative hypothesis

Step 10: Higher the chi square value, it indicates lower the p value

Table 1: Association between Demographic and education variables with Semester Results

	Chi square	P value	Remarks
Selection	205.73	0.000	Sig at 1%
PUC	187.38	0.000	Sig at 1%
Branch	81.95	0.000	Sig at 1%
Gender	35.8	0.000	Sig at 1%
Reservation	21.2	0.000	Sig at 1%
Father Occupation	17.02	0.017	Sig at 5%
Accommodation	7.15	0.028	Sig at 5%
Board_PUC	6.74	0.081	Sig at 10%
Blood Group	6.36	0.384	Non sign
Mother Occupation	2.73	0.842	Non Sign

P value – Probability value

	A	D	F	G	H	I	J	K	L	M	N	O	P
1	Branch	Selection	Gender	BloodGroup	Reservation	Board_PUC	Marks_PUC	Accommodation	Fr.Occupation	MR.Occupation	AnnualIncome	Percent_PUC	Results
2	CSE	cet	M	O+	GM	State	298	Day Scholar	PRIVATE	Professionals	6000000	90.03	550
3	CSE	cet	M	O+	GM	State	298	Day Scholar	PRIVATE	Professionals	6000000	90.03	550
4	CSE	comedk	M	O+	GM	State	270	Day Scholar	PRIVATE	Professionals	1500000	88	556
5	CSE	comedk	M	O+	GM	CBSE	280	Hostel	BUSINESS	HomeMaker	400000	86.6	527
6	CSE	cet	M	AB+	SC/ST	State	212	Day Scholar	BUSINESS	PRIVATE	80000	75.8	444
7	MECH	management	M	O+	OBC	State	193	Day Scholar	BUSINESS	HomeMaker	600000	76.5	526
8	CSE	cet	M	O+	GM	State	280	Day Scholar	PRIVATE	HomeMaker	800000	90	526
9	CSE	cet	F	O+	GM	State	288	Hostel	PRIVATE	HomeMaker	800000	95	567
10	CSE	cet	M	O+	GM	State	274	Day Scholar	PRIVATE	GOVT SERVICE	3500000	83.33	571
11	CSE	management	M	O+	GM	CBSE	214	Hostel	GOVT SERVICE	BUSINESS	1000000	76.2	416
12	CSE	cet	M	B+	GM	CBSE	281	Day Scholar	PRIVATE	HomeMaker	700000	93	545
13	CSE	cet	F	B+	OBC	State	290	Day Scholar	BUSINESS	HomeMaker	480000	93.5	593
14	CSE	comedk	M	A+	OBC	State	259	Hostel	BUSINESS	HomeMaker	200000	83.6	423
15	CSE	management	F	A+	J&K	State	255	Day Scholar	BUSINESS	HomeMaker	200000	81.2	544
16	CSE	comedk	F	O+	GM	State	298	Hostel	PRIVATE	HomeMaker	600000	97.7	558
17	CSE	management	F	O+	GM	State	266	Day Scholar	BUSINESS	PRIVATE	600000	88.66	442
18	CSE	comedk	M	AB+	GM	CBSE	277	Day Scholar	PRIVATE	HomeMaker	1000000	92.33	560
19	CSE	comedk	M	A+	GM	CBSE	215	Day Scholar	Professionals	Professionals	1000000	79.8	560
20	CSE	comedk	M	O+	OBC	State	183	Day Scholar	PRIVATE	HomeMaker	518228	61.33	374
21	CSE	comedk	M	O+	GM	CBSE	259	Day Scholar	PRIVATE	GOVT SERVICE	5000000	89.2	463
22	CSE	comedk	M	O-	OBC	State	266	Hostel	PRIVATE	HomeMaker	5000000	88.67	508
23	CSE	management	F	A+	GM	State	249	Day Scholar	BUSINESS	HomeMaker	600000	86.16	550

Figure 1: Sample Data File

Table 2: Role of Education and demographic variables on Semester Results of the students

Variable	Information value
Selection	79.50%
PUC	75.76%
Branch	30.91%
Gender	13.13%
Reservation	7.77%
Father Occupation	6.06%
Accommodation	2.53%
Board_PUC	2.39%
Blood Group	2.30%
Mother Occupation	0.61%



Fig 6 Father Occupation

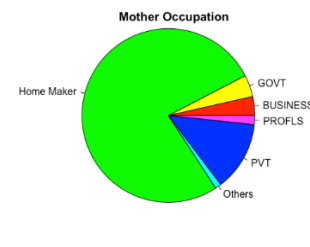


Fig 7 Mother Occupation



Fig 8 Selection

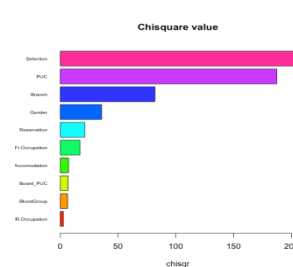


Fig 9 Chi Square Value

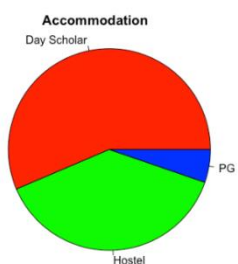


Fig 2 Accommodation

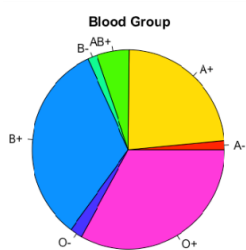


Fig 3 Blood Group

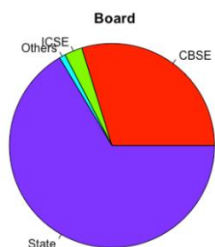


Fig 4 Board

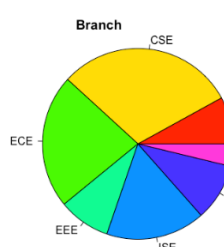


Fig 5 Branch

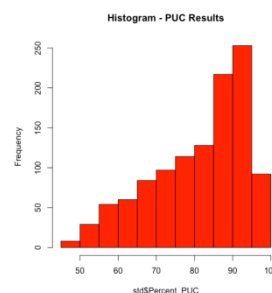


Fig 10 PUC results

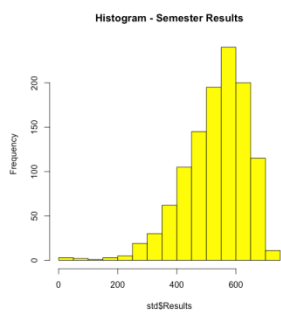


Fig 11 Semester Results

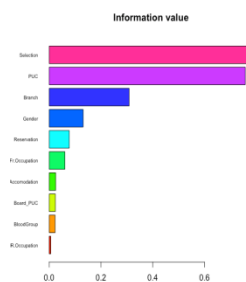


Fig 12 Information Value

The set of graphs shown in figure 2 to figure 9, show the pie charts drawn for the distribution of various attributes considered for research. The data set consists of 1186 entries.

Figure 2 shows the accommodation of students. Out of 1186 students, 665 students are day scholars, 461 students stay in college hostel and 60 stay in paying guest accommodation.

Figure 3 shows blood group of students. 386 students have O+, 65 students have AB+, 383 students have B+, 288 students have A+, 26 students have O-, 17 students have A-, 18 students have B- and 3 students have not mentioned blood groups.

Figure 4 shows Board in which the student completed the twelfth grade. The boards considered were state board, central board (CBSE) and international board (ICSE). 788 students are from state boards of various states, 343 students are from the central board, 33 students are from ICSE, 13 students are from Nepal Board and 4 are from other boards.

Figure 5 shows the Branch of student. 365 students are from Computer Science and Engineering, 205 students are from Information Science and Engineering, 111 students are from Mechanical Engineering, 91 students are from Civil Engineering, 108 students are from Electrical and Electronics Engineering, 262 students are from Electronics and Communication Engineering and 44 students are from Telecommunication Engineering.

Figure 6 shows the Father's Occupation. 428 fathers work in Private organizations, 389 fathers do various businesses, 88 fathers were in Government Services, 77 were Professionals, 117 of them were Farmers, 40 of them served in Defence and 16 were in other professions.

Figure 7 shows the Mother's Occupation. 899 mothers were home makers. 157 of them worked for private organizations, 46 were in different businesses, 46 of them were in Government Services, 20 of them were Professionals and 15 of them performed other chores.

Figure 8 shows the selection of seats under which the student has taken admission. 471 students have taken admission from CET quota, 235 students were from COMED-K, 381 students were from management quota, 74 students were from PIO quota, 20 students were from CET-SNQ quota and 3 students were from CET-J&K quota.

Figure 9 shows the Chi Square Value. The values for the histogram are shown in Table 1.

Figure 10 shows the PUC results. The x axis shows the percentage secured by students in pre-university examinations and y axis shows the frequency distribution of the students.

Figure 11 shows the Semester Results. The x axis shows the percentage secured by students in semester end university examinations and y axis shows the frequency distribution of the students. The values for the histogram are shown in Table 2.

Figure 12 shows the Information Value. This histogram depicts the distribution of each of the attribute over y axis versus the information value on x axis.

Chi square test shows that the information value analysis has tried to establish explanatory power of each predictors or independent variable with outcome variable as semester result of the students. Selection criteria of the student to the course secured highest score, which is, 79.5% and followed by percentage of PUC marks secured by the students, which is, 75%, third place is occupied by the branch the student belongs to, which is, 30% and followed by Gender which is, 13.13%, rest of the variables are less than 10% of explanatory power with outcome variable.

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

O is Observed frequency

E is Expected frequency

E is computed by (Row total * Column total) / Grand total

Higher the value of chi square test, higher the association between two categorical attributes, lower the value of chi square test, lower the association between two categorical attributes. Relation between chi square value and P (Probability) value is indirect. Chi square test is conducted between predictor which included both education related and demographic attributes of the students with semester results as outcome variables. Higher the Chi squares the association between predictor and outcome is high. Variables such as Selection through Reservation, the association is significant at 1%, Father Occupation and Accommodation is Significant at 5% level. Board of PUC is significant at 10% level, last two attributes, Blood group and Mother Occupation does not have any significant association with Outcome variables..

IX. CONCLUSION

Lot of work is happening in educational data mining. The researchers have made an attempt to apply various machine learning techniques and data mining algorithms to prepare the predictive models on the available data sets. The main intention of this paper is to assess the performance of undergraduates at an earlier stage and try to improve the academic performance and bring better results to the institution.

For the purpose of investigation, a case study comprising of undergraduate students database of one of the premiere universities in Karnataka is considered.

In this paper, Chi square test which is one of the statistical methods is applied. The variables are classified into study related variables and demographic variables and the results are thus obtained. From the work, it is apparent that the selection and PUC marks have higher impact over the other attributes considered for study. The limitations of this paper are the given data set has only demographic and study related variables. Other attributes like social life, educational background of parents and family details are not considered. The forth coming work is to provide a comparative analysis among various statistical techniques and to assess the performance of students.

REFERENCES

1. Akansha Mishra, Rashi Bansal, Shailendra Narayan. (2017). Educational Data Mining and Learning Analysis. IEEE. 491-494.
2. Amirah Mohamed Shahiria, Wahidah Husaina. Nur'aini Abdul Rashida. (2015). A Review on Predicting Student's Performance using Data Mining Techniques. Elsevier. 414-422.
3. Ankita Katare, Shubha Dubey. (2017) A Study of various Techniques for predicting student performance under Educational Data Mining. International Journal of Electrical, Electronics and Computer Engineering. 6(1): 24-28.
4. Anoushka Jain, Tanupriya Choudhury, Parveen Mor, A. Sai Sabitha. (2017). Intellectual Performance Analysis of Students by comparing various Data Mining Techniques. IEEE. 57-63.
5. Anu Marwaha, Sachin Ahuja. (2017). A Review on Identifying influencing factors and data mining techniques best suited for analysing students' performance. IEEE. 373-378.
6. Anwar Ali Yahya, Ahmed Abdu Alattab. (2013). Educational Data Mining: A Case Study of Teacher's Classroom Questions. IEEE. 92-97.
7. Brijesh Kumar Bhardwaj, Saurabh Pal. (2011). Data Mining: A prediction for performance improvement using classification. (IJCSIS) International Journal of Computer Science and Information Security. 9(4). 2011.
8. Fatma Chiheb, Fatima Boumahdi, Hafida Bouarfa. (2017). Predicting Students Performance Using Decision Trees: Case of an Algerian University. International Conference on Mathematics and information Technology, Adrar, Algeria – December 4 - 5, 2017
9. HemaMalini B. H., L. Suresh. (2018). Assessment of Performance of Engineering Students using Educational Data Mining. Journal of Emerging Technologies and Innovative Research. 5(6). 16-20.
10. HemaMalini B. H., L. Suresh, Mayank Kushal. (2018). Comprehensive Analysis of Students' Performance by applying Machine Learning Techniques. Paper presented at 3rd International Conference on Smart Computing & Informatics (SCI), Springer Conference held at KIIT, Bhubaneswar, India.
11. Iti Burman, Subhranil Som, Mayank Sharma. (2017). Enhancing Student Learning Behaviour using EDM and Psychometric Analysis. IEEE. 360-363.
12. John Jacob, Kavya Jha, Paarth Kotak, Shubha Puthran. (2015). Educational Data Mining Techniques and their Applications. IEEE. 1344-1348.
13. M Amala Jayanthi, R Lakshmana Kumar, Abhijith Surendran, K Prathap. (2016). Research Contemplate on Educational Data Mining. International Conference on Advances in Computer Applications (ICACA). IEEE. 110-114.
14. Pratiyush Guleria, Dr.Manish Arora, Dr.Manu Sood. (2013). Increasing Quality of Education Using Educational Data Mining. IEEE. 118-122.

AUTHORS PROFILE



Dr. HemaMalini. B. H., M.Tech, Ph. D, has more than 23 years of teaching experience, teaches Automata Theory, Compiler Design. Her research interest is Educational Data Mining. She has publications in the area of Cloud Computing, Big Data and Machine Learning.



Dr. Suma V, M.S., Ph. D, Nodal Officer between IPR Cell, DSCE and Karnataka State Council for Science and technology. Head of Advanced Software and Engineering and Cloud Computing Research Group, RIIC, DSI. Chair, Affiliate Institute Programme, RIIC, DSI. Total of 18 years of teaching experience as Professor in Department of Information Science & Engineering Dayananda Sagar College of Engineering, Bangalore.



Dr. Suresh L, M. E., Ph. D, Principal, with more than 30 years of teaching and research experience is a nominee of VTU, Belagavi to the "Academic Council" of Dr. Ambedkar Institute of Technology, Bengaluru (Autonomous). Academic Senate member of Dr. TIT, autonomous institute, period. Nagarjuna Institute of Technology, Autonomous institute, period. Ambedkar Institute of Technology, Autonomous institute. VTU Nominee, KLE University, Hubli. VTU Nominee, VTU Nominee for recruitment.. Member secretary: Cambridge Institute of Technology, Bangalore. Member: Cambridge Institute of Technology, North Campus, Bangalore.



Mr. Shankar M. M., MBA, has background of MBA in Marketing from University of Madras, has 2 decades work experience which included in the field of Academics, Industry, Data analysis Consultancy and Research and Statistical software Training, Undergone various FDP and SFDP programme at IIMA and IIMC, handled more than 500 training programme pan India in both Academics and Industry in last decade. Handled more than 100 clients in the level of Team and Individual on statistical modeling and data analysis using various software such as R, SPSS, AMOS, SmartPLS and Jamovi. His Area of Interest included Marketing Research and Academic Research.