International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 7, July 2014)

Prediction of Academic Performance Using Bayesian Estimation and Normal Distribution

M.Rebekah, Assistant Professor
Department of Mathematics
Chilkur Balaji Institute of Technology
Hyderabad, India

Dr A. Ramakrishna Prasad, Professor
Department of Mathematics
Jawaharlal Nehru Technological University
Hyderabad, India

***Abstract*—Predictive analytics is a statistical analysis where we extract information from the data and identify patterns and predict future outcomes with some level of reliability. Data mining and statistical techniques are used in predictive analytics. The main objective of the study is to predict the final result in mathematics of the engineering students studying in the first year by using the internal marks. Bayesian estimation is used as a technique to predict the final result of the students. By normal distribution the students who are below average, average and above average are predicted. This study will help in predicting the academic performance of the student class wise and in turn predicting performance of faculty.**

***Index Terms*—Prediction, Academic Performance, Bayesian Estimation, Normal Distribution, Mathematics, Statistical Analysis, Data mining**

I. INTRODUCTION

The performance of the students in India especially in higher education is a turning point in the academics. The main aim of the higher educational institutes is to offer highest level of quality education to its students. The amount of data stored in educational database is growing rapidly. These data bases include hidden information for development of students' academic performance. Data mining is used to bring out significant information and to build up important relationships among variables stored in large data-set / data-warehouse. A variety of techniques are used in data mining like classification, clustering and rule mining and can be used to bring out unknown hidden information from the educational records.

II. DATA MINING

Data mining techniques and methods can be applied in various fields like cross selling, fraud detection, health care, academics, stock market, customer relationship management, engineering, web mining etc. Educational data mining is the latest upcoming technique of data mining that can be applied on the data related to the field of education. Educational data mining is the method of transforming raw data compiled by education systems in useful information that could be used to take knowledgeable decisions and answer research questions.

III. EDUCATIONAL DATA MINING

Higher education faces a new era as a result of changes in the way people view colleges and universities. Expectations for better performance in terms of teaching and producing competent college graduates are increasing. Educational data mining is used to study the student data available in the university data base and bring out the useful information or knowledge from it. Classification methods like decision trees, rule mining, Bayesian network etc., can be applied on the student data for predicting the students behavior, performance in examination etc. This prediction will help the tutors to identify the weak or about fail students and help them to pass or score better marks.

IV. RELATED WORK

Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students as described by Alaa el-Halees. Mining in educational environment is called Educational Data Mining. Bharadwaj and Pal conducted study on the student performance by selecting 300 students from 5 different degree colleges in India. In their study, it was found that students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, family annual income, and student's family status were highly correlated with the student academic performance. Hijazi and Naqvi conducted a study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance. Bharadwaj and Pal in their another study they used students' previous semester marks, class test grade, seminar performance assignment, performance, general proficiency, attendance in class and lab work to predict students' mark in their end semester. Kovacic used enrollment data to predict successful and unsuccessful student in New Zealand and he found59.4% and 60.5% of classification accuracy while using decision tree algorithms CHAID and CART respectively. Yadav Bhardwaj and Pal conducted study on the student retention based by selecting 398 students from MCA course of VBS Purvanchal University, Juniper, India. By means of classification they show that student's graduation stream and grade in graduation play important role in retention. Galit gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams. Pal conducted study on the student dropout rate by selecting 1650

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 7, July 2014)

students from different branches of engineering college. In their study, it was found that student's dropout rate in engineering exam, high school grade; senior secondary exam grade, family annual income and mother's occupation were highly correlated with the student academic performance. Al-Radaideh, and others applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the Naive Bayer's were used. The outcome of the results indicated that Decision Tree model had better prediction than other models. Yadav and Pal conducted a study using classification tree to predict student academic performance using students' gender, admission type, previous schools marks, medium of teaching, location of living, accommodation type, father's qualification, mother's qualification, father's occupation, mother's occupation, family annual income and so on. In their study, they achieved around 62.22%, 62.22% and 67.77% over all prediction accuracy using ID3, CART and C4.5 decision tree algorithms respectively. In another study Yadav and pal used students' attendance, class test grade, Seminar and assignment marks, lab works to predict students' performance at the end of the semester with the help of three decision tree algorithms ID3, CART and C4.5. In their study they achieved 52.08%, 56.25% and 45.83% classification accuracy respectively. Merceron A and others concluded that association technique requires not only that adequate thresholds be chosen for the two standard parameters of support and confidence, but also that appropriate measures of interestingness be considered to Bray, in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socio-economic conditions.

V. METHODOLOGY

Data is collected from 300 students studying in the first year of the engineering college in Andhra Pradesh, India. These students belong to five different branches namely ECE, CSE, MECH, CIVIL and EEE. There are 60 students in each branch. The analysis is done to predict the final result of the students in the paper called Mathematics–1 by using the internal marks. The university conducts three internal exams; each for 25 marks and the average of these three internal marks is taken and added to the final result. The marks are given as shown below:

	Descriptive (10)	Objective (10)	Assignment (5)	Total (25)
MID - 1	06	04	05	15
MID -2	07	05	05	17
MID -3	06	05	05	16
				48/3=16

The mean and standard deviation is calculated for each branch separately and tabulated as shown below:

Branch	*X*	*σ*
ECE	*15.4*	*4.1*
CSE	*13.8*	*3.7*
MECH	*16.06*	*5.3*
CIVIL	*10.8*	*5.5*
EEE	*16.8*	*4.4*

X=Mean mark obtained for each branch separately
σ= Standard deviation
The average of the population that is N=300 is calculated as $\mu = 14.5$
The sample size n=60

VI. BAYESIAN ESTIMATION:

In Bayesian Estimation prior feelings about the possible values of μ combined with the direct sample evidences. Suppose the prior distribution has mean μ0 and standard deviation σ. The posterior distribution can be approximated by normal distribution with

Mean of the posterior distribution, $\mu_1 = \frac{nX\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$

n = sample size
X =sample mean
σ^2 *=variance of the population*

Standard deviation of the posterior distribution $\sigma_1 = \sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}}$

THE POSTERIOR DISTRIBUTION IS SHOWN BELOW:

Branch	*X*	*σ*
ECE	*15.38*	*0.52*
CSE	*13.81*	*0.47*
MECH	*16.03*	*0.68*
CIVIL	*10.86*	*0.704*
EEE	*9.98*	*0.56*

From the above table population mean = 13.2
Population S.D. = 6.5

The No of students who are expected to get below 40%, 40% to 60% and above 60 % is calculated from the normal distribution table as shown below
The final exam marks are for 75 but here on the basis of the internal marks we are calculating for 25 marks and the % is taken
(1) The number of students who got less than 10 marks that is below 40% is

P(X<10)

$Z = \frac{X - \mu}{\sigma}$

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 7, July 2014)

$= \frac{10-13.2}{6.5}$

= -0.492

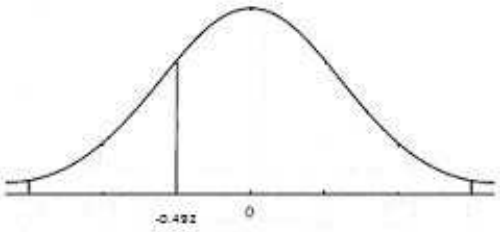
P (Z<-0.492) = 0.5-A(0.492)

= 0.5-0.1879

= 0.3121

Total No of students who are expecting less than 40% are 0.3121*300 = 93.63

94 students got below 40% marks



(2) The number of students who got marks between 10 to 15 that is between 40 to 60% is

P (10<X<15)

When X = 10, Z = 0.492

When X = 15, $Z = \frac{15-13.2}{6.5}$

= 0.2769

P (-0.492<X<0.276) = A (0.49) + A(0.27)

= 0.18+0.106

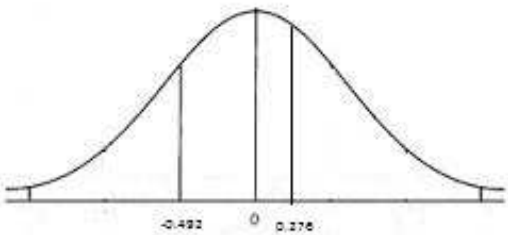
= 0.286

Total number of students = 0.2868*300 = 85.8 = 86

(3) The number of students who are expected to get greater than 15 marks that is

Above 60% is P(X > 15)

= P (Z > 0.27)



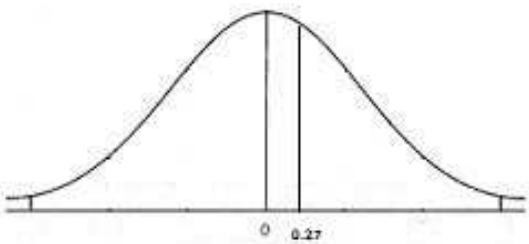
= 0.5 – A (0.27)

= 0.5 – 0.106

= 0.394

Therefore total No of students = 0.394*300 = 118.8

= 118



VII. CONCLUSION:

The first year final result of the students is predicted by using Bayesian estimation. It is predicted that mechanical branch students are having highest mean mark compared to the other branches. By using normal distribution it is predicted that

94 students are below 40%

86 students are between 40% and 60%.

118 students are above 60%.

This prediction will help in identify the number of students who are below average, average and above average. The performance of the faculty can be evaluated.

The same procedure can be used to predict the No of students getting marks in various ranges for different subjects and the faculty performance can be evaluated Necessary changes can be made for further improvement of the faculty performance thereby increasing student's success.

REFERENCES

[1] International Journal of Engineering Education, Vol. 27, No. 6, 2011, TEMPUS Publications, 1343–1351,

[2] B. F. French, J. C. Immekus and W. C. Oakes, An examination of indicators of engineering students' success and persistence, Journal of Engineering Education, 94(4), 2005, pp. 419–425.

[3] S-M. R. Ting, Predicting Asian Americans' academic performance in the first year of college: An approach combining SAT scores and non cognitive variables, Journal of College Student Development, 41(4), 2000, pp. 442–449.

[4] R. L. Rex and D. D. Hein, An educational index to predict academic performance, Southern Journal of Educational Research, 6(4), 1972, 209–218.

[5] Huang, S., & Fang, N. (2010). Prediction of Student Academic Performance in an Engineering Dynamics Course: Development and Validation of Multivariate Regression Models. International Journal of Engineering Education, 26(4), 1-10.

[6] Huang, S., & Fang, N. (2010). Regression models for predicting student academic performance in an engineering dynamics course. Proceedings of the American Society for Engineering Education Annual Conference & Exposition, Louisville, KY.

[7] Huang, S., & Fang, N. (2011). WIP: Prediction of students' academic performance in an introductory engineering course. Proceedings of the 2011 Frontiers in Education Conference.

[8] P. J. Shull and M. Weiner, M., Thinking inside the box: Self efficacy of women in engineering, International Journal of Engineering Education, 18(4), 2002, pp. 438–446.

[9] R. Lynch, N. Seers and S. Gordon, Design of a novel diagnostic tool for student performance in engineering degree courses, International Journal of Engineering Education, 24(6), 2008, pp. 1103–1110.

[10] J. Allen, S. B. Robbins, A. Casella's and I.-S. Oh, Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness, Research in Higher Education, 49(7), 2008, pp. 647–664.

[11] L. Duckworth and M. E. P. Seligman, Self-discipline outdoes IQ in predicting academic performance of adolescents, Psychological Science, 16(12), 2005, pp. 939–944.

[12] International Journal of Electronics Communication and Computer Engineering Volume4, Issue5, ISSN (Online):2249- 071X, ISSN (Print): 2278- 4209

[13] Analysis and Mining of Educational Data for Predicting the Performance of Students