University Result Prediction Using Machine Learning Algorithms

Er.SATHISH R[1],Dr.UMA V[2],Ms.Ruby Mythili M[3],Ms.Geedha K[4],Ms.Pavithrayani K[5]

[1]Assistant Professor, [2]Professor & Head,[1345]UG Scholor,
[1345]Department of Information Technology,KGiSL Institute of Technology
[2]Department of Information Technology,Hindusthan Institute of Technology

Abstract— In the realm of Open Online Course and open training frameworks, students have adaptability to get the hang of anything effortlessly as the learning content is effectively accessible. Yet, this office can make students careless. Along these lines, it ends up hard to foresee student's execution ahead of time. In this examination an endeavour is made to assist the understudy with knowing his/her execution ahead of time by utilizing univariate direct relapse demonstrates. We gathered the signs of internal test parts of one subject to anticipate the last grade in that subject. The inside imprints are standardized to 100 (rate) to have precise outcomes. The model gives anticipated evaluation of conclusive examination specifically subject. It additionally encourages understudies to know what number of imprints in the inside examination are required to get specific evaluation.

Keywords : Machine Learning, Result Prediction, Internal Marks, University Grades, College Result, Student Performance.

I. INTRODUCTION

In present instructive frameworks, the performance prediction of the students is getting complex each day. Foreseeing performance prediction ahead of time can support the students and their educator to monitor advancement of the student. Numerous establishments have received consistent assessment framework today. Such frameworks are helpful to the students in improving them. The motivation behind ceaseless assessment framework is to support normal students . In persistent assessment framework, unit tests or class tests are led at standard period. To have consistence execution in the last graduate of the subject it is required to show up in all the unit tests or class test. In this paper we have considered the examination example of our Anna University. According to the inner examination example of Anna University, Minimum of three unit trial of any subject ought to be directed. At that point after, one sessional examination is directed. The weightage of interior examination is 20% and outer examination in 80%.

II. RELATED WORKS

Our study employs statistical methods to predict the university result from internal exam analysis and evaluate the accuracy of this detection. Hence, we present our literature review on studies that applied statistical methods.

In the examination work done by G. N. Pandey, SonaliAgarwal and M. D. Tiwari [1] they have connected gathering of classifiers to the dataset taken and among those classifiers they have made endeavors to discover the best classifier. Relative examination is done to dissect the best or ideal arrangement. They have considered an information block. This information solid shape contains name, verbal capacity and scores in maths. The signs of the understudy are put away independently in another cell. Taking those 3 traits and 2000 records of the understudy's characterization examination is performed. So as to perform information pre-handling and order WEKA programming is utilized by them. WEKA programming handles missing qualities all around proficiently. It likewise helps in advantageous treatment of information. Characterization goes under directed realized which implies that class names are available.

In the exploration work done by MounikaGoyal and RajanVokra [2] they have executed different information mining strategies like group investigation, exception examination, order. They have gathered the student's into 4 branches dependent on their outcomes utilizing the strategies like grouping and arrangement. Consequently they had connected grouping strategies in advanced education. They examined how OLAP and OLTP are helpful to the general public and how OLTP encourages the understudies to learn things.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

They likewise contemplated how understudies can utilize biometric framework, coordination of instalment passage, reconciliation of messages, correspondence through email, access through versatile mail and so on the primary advantages they proposed are possession cost is low, security is high and arrangement is adaptable. The fundamental contrast among OLAP and OLTP is OLAP is just for recovery reason, information show is multi-dimensional, star blueprint is pursued, chiefs and officials are the ordinary clients. In the event of OLTP it is utilized to refresh, element relationship information demonstrate is pursued, standardized outline is utilized and the common clients are the staff individuals in an association. An information distribution centerframework streamlines the intricate framework into basic one. Information stockroom development is arranged into four sections to be specific information sources, organizing zone, information stockpiling, conveyance of data. After that different information mining systems are connected.

In the examination paper by J. K. JyothiKalpanaand K. VenkataLakkshmi [3] researched the improved scholarly area of information mining study the alumni understudies information gathered from the Engineering school and Technology,Villupuram. The information contains five years time frame [2008-2013]. The strategies they have utilized are centroid based, thickness based and appropriation based grouping. The product utilized is matlab to apply the methods on the Engineering understudy's informational index. They have learned about the centroid based, thickness based and circulation based grouping calculations. They connected these calculations to group understudies and made an endeavor to raise their execution. They have utilized basic k-implies calculation to dissect the execution of understudies.

III. METHODOLOGY

The purpose of this work is to predict students' mark based on KGiSL College internal marks datasets. In this study we used RAPIDMINER for data analysis. The RapidMiner is an easy to use graphical user interface that gives the power of the RAPIDMINER software. It is a freemium software that implements a large collection of machine learning algorithm and is widely used in data mining application. In this experimentation we used Machine Learning and Deep Learning build the models.

TABLE I

S.NO	List of all variables with their description and sources.		
	Variables name	**Description and possible values**	**Source**
1	IM1	Internal Mark 1 (0-100)	KGiSL Institute of Technology
2	IM2	Internal Mark 2 (0-100)	KGiSL Institute of Technology
3	IM3	Internal Mark 3 (0-100)	KGiSL Institute of Technology
4	Internal	Internal overall mark(PASS,FAIL)	KGiSL Institute of Technology
5	External	External mark (PASS,FAIL)	KGiSL Institute of Technology

IV. RESULTS & DISCUSSIONS

Naive Bayes

The Naive Bayesian classifier depends on Bayes' hypothesis with the freedom presumptions between indicators. A Naive Bayesian model is anything but difficult to work, with no convoluted iterative parameter estimation which makes it especially valuable for huge datasets. Regardless of its straightforwardness, the Naive Bayesian classifier regularly does shockingly well and is broadly utilized on the grounds that it frequently outflanks increasingly complex arrangement techniques. In result prediction Naive Bayes produces 95% accuracy. The confusion matrix for the predictor is given below:

	true FAIL	true PASS	class precision
pred. FAIL	4	1	80.00%
pred. PASS	1	30	96.77%
class recall	80.00%	96.77%	

Generalized Linear Model

In its most straightforward structure, a direct model determines the (straight) connection between a ward (or reaction) variable Y, and a lot of indicator factors. The accuracy of the model is calculated as 89%. The confusion matrix for the predictor is given below:

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

	true FAIL	true PASS	class precision
pred. FAIL	1	1	50.00%
pred. PASS	3	30	90.91%
class recall	25.00%	96.77%	

Logistic Regression

Logistic regression is the fitting relapse investigation to lead when the reliant variable is dichotomous (paired). Like all relapse investigations, the strategic relapse is a prescient examination. Logistic Regression is utilized to portray information and to clarify the connection between one ward double factor and at least one ostensible, ordinal, interim or proportion level autonomous factors. The accuracy of the model is 91%. The confusion matrix for the predictor is given below:

	true FAIL	true PASS	class precision
pred. FAIL	2	1	66.67%
pred. PASS	2	30	93.75%
class recall	50.00%	96.77%	

Fast Large Margin

The Fast Large Margin administrator applies a quick edge student dependent on the direct help vector learning plan. The accuracy of the model is 55%. The confusion matrix is given below:

	true FAIL	true PASS	class precision
pred. FAIL	3	16	15.79%
pred. PASS	0	17	100.00%
class recall	100.00%	51.52%	

Deep Learning

Deep learning is a class of AI calculations that, utilize a course of numerous layers of nonlinear handling units for highlight extraction and change. Each progressive layer utilizes the yield from the past layer as info, learn in administered (e.g., order) and additionally unsupervised (e.g., design examination) habits, get familiar with various dimensions of portrayals that relate to various dimensions of deliberation; the dimensions structure a chain of importance of ideas. The accuracy of deep learning model is 89%. The confusion matrix for the same is given below:

	true FAIL	true PASS	class precision
pred. FAIL	1	1	50.00%
pred. PASS	3	30	90.91%
class recall	25.00%	96.77%	

Decision Tree

Decision trees are used for classification and regression problems, this story we talk about classification. The accuracy of decision tree is 92%. The confusion matrix for the same is given below.

	true FAIL	true PASS	class precision
pred. FAIL	1	0	100.00%
pred. PASS	3	32	91.43%
class recall	25.00%	100.00%	

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
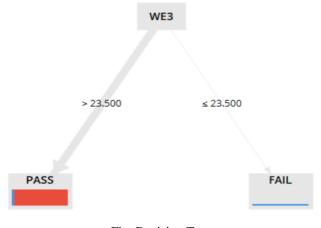*Volume- 5, Special Issue- March, 2019*

Fig: Decision Tree

Support Vector Machine

A Support Vector Machine (SVM) is a discriminative classifier formally characterized by an isolating hyperplane. At the end of the day, given marked preparing information (administered learning), the calculation yields an ideal hyperplane which arranges new precedents. In two dimensional space, this hyperplane is a line isolating a plane in two sections where in each class lay in either side. The accuracy of SVM is 89%. The confusion matrix can be given as:

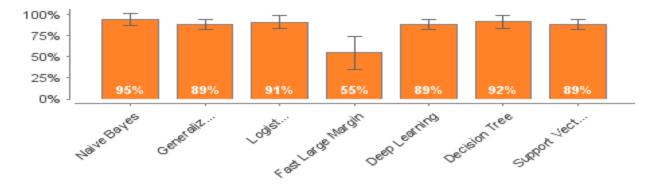|	true FAIL	true PASS	class precision
pred. FAIL	1	1	50.00%
pred. PASS	3	30	90.91%
class recall	25.00%	96.77%	

Fig: Accuracy of Models

V. CONCLUSIONS

This investigation will help to the understudies and the instructors to improve the execution of the understudies. This examination presents understudies mark forecast show advancement approaches dependent on institutional inside and outside open information sources that can be utilized in functional settings to foresee understudies' scholastic execution. The consequence of this investigation demonstrates that display dependent on institutional inward databases and outer open information sources performs superior to anything the model dependent on just institutional inner databases. Besides, the outcome firmly underpins that understudies' present scholastic execution is the best indicator in anticipating understudies' imprint. Among different indicators A dimension point, NSS results are additionally exceedingly suggested. This investigation underlines the significance of connected open information sources in creating prescient models. Along these lines, this investigation proposes more research contemplate utilizing outside information sources here.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

REFERENCES

[1] SonaliAgarwal, G. N. Pandey and M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e Education, e-Business, e-Management and e-learning , Vol. 2, No. 2, April 2012.

[2] MounikaGoyal and RajanVokra, "Applications of Data Mining in Higher Education", International Journal on Computer Science Issues (IJCSI), ISSN (Online): 1694-0814 ,Vol 9, Issue 2, No 1, March 20

[3] M.R. Thanasekhar and N. Balaji, "Performance Analysis of Students using Data Mining techniques", International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), ISSN (Online) : 2319 – 8753, Volume 3, Special Issue 3, March 2014.

[4] L. Benko, J. Reichel, and M. Munk.Analysis of student behavior in virtual learning environment depending on student assessments. In 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA), page 16, Nov 2015.

[5] Y. Y. Chen, S. MohdTaib, and C. S. CheNordin. Determinants of student performance in advanced programming course. In 2012 International Conference for Internet Technology and Secured Transactions,page 304307, Dec 2012.

[6] T. Devasia, Vinushree T P, and V. Hegde. Prediction of students performance using educational data mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), page 9195, March 2016.

[7] H. Goker and H. I. Bulbul.Improving an early warning system to prediction of student examination achievement. In 2014 13th International Conference on Machine Learning and Applications, pages 568–573, Dec 2014.

[8] S. Anupama Kumar and M. N. Vijayalakshmi. Mining of student academic evaluation records in higher education. In 2012 International Conference on Recent Advances in Computing and Software Systems, page 6770, April 2012.

[9] A. Zafra, C. Romero, and S. Ventura.Predicting academic achievement using multiple instance genetic programming. In 2009 Ninth International Conference on Intelligent Systems Design and Applications, page 11201125, Nov 2009.