iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43073

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: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Students Performance Analysis System Using Cumulative Predictor Algorithm

Mr. K. Praveen Kumar[1], K. Sai Pranav[2], D. Gowtham[3], S. Abhishek[4]
[1]*Assistant Professor, Dept of IT, Anurag Group of Institutions, Telangana*
[1, 2, 3, 4]*Bachelor of Technology, Information Technology, Anurag Group of Institutions, Telangana, India*

Abstract: ***Towards automation to do mundane tasks and the expectations for students already equipped with good programming skills is on the rise. In parallel, there has been a rising number of students who find it difficult to attain the skills necessary in order to get the dream IT job they desire. The aim of this project is to bridge the gap between the employer and the future employee of the company by the use of SPAS at college level. Student performance analysis system (SPAS) is an online web application system which enables students to know prior hand if their level of skills for the placement is enough to get placed or not, given the necessary inputs. SPAS has an intelligent learning algorithm which utilises a rich database, analyses the records of previous students' traits and develops a model for further prediction. The performance evaluation of students by SPAS is by the cumulative predictor algorithm involving generation of several random forest trees on the available data. SPAS learns and creates its model reaching higher accuracy with increasing data availability.***

I. INTRODUCTION

Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. There are several data regarding the students which stay unused with untapped potential of data mining which could revolution is the field of education. Since the ultimate aim of an educational institution is to create a pool of skilled professionals to take on the society to a next upgraded level, they need to create an environment for their students to grow in every vertical by giving them right exposure and training. Most of the educational institutions, maintain huge databases of students and the information keeps on increasing with time, but there is no action taken to gain knowledge from it. DM has the suitable techniques in mining the data to discover new information and knowledge about students. DM provides various methods for analysis which include classification, clustering, and association rules. Classification, one of the prediction algorithms, classifies the data (constructs a pattern) based on the training set and uses the pattern to classify a new data (testing set). In this paper, we consider the students' academic performance (SAP) system in University Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Malaysia as our existing system. IHL faces a major challenge in order to improve and manage the organization to be more efficient in managing students' activities. To achieve this target, DM is considered as the one of most suitable technique in giving additional insights to the IHL community to help them make better decisions in educational activities. The IHL make use of WEKA tool in order to build a model and predict the SAP in order for the professors to provide the students with individual attention. In SAP system, the classification method is selected to be applied on the students' data.

II. LITERATURE SURVEY

A. Model prediction of academic performance for first year students

Authors: García, E.P.I. and Mora, P.M.

The aim of this paper was to obtain a model to predict new students' academic performance taking into account socio-demographic and academic variables. The sample contained records of first semester students at a School of Engineering from a range of students' generations. The data was divided into three groups: students who passed none or up to two courses (low), students who passed three or four courses (middle), and students who passed all five courses (high). By using data mining techniques, the Naïve Bayes classifier and the Rapid miner software, we obtained a model of almost 60% accuracy. This model was applied to predict the academic performance of the following generation. After checking the results of the predictions, 50% were classified as correct. However, we observed that, for students of certain engineering majors of high and low groups, the model's accuracy was higher than 70%.

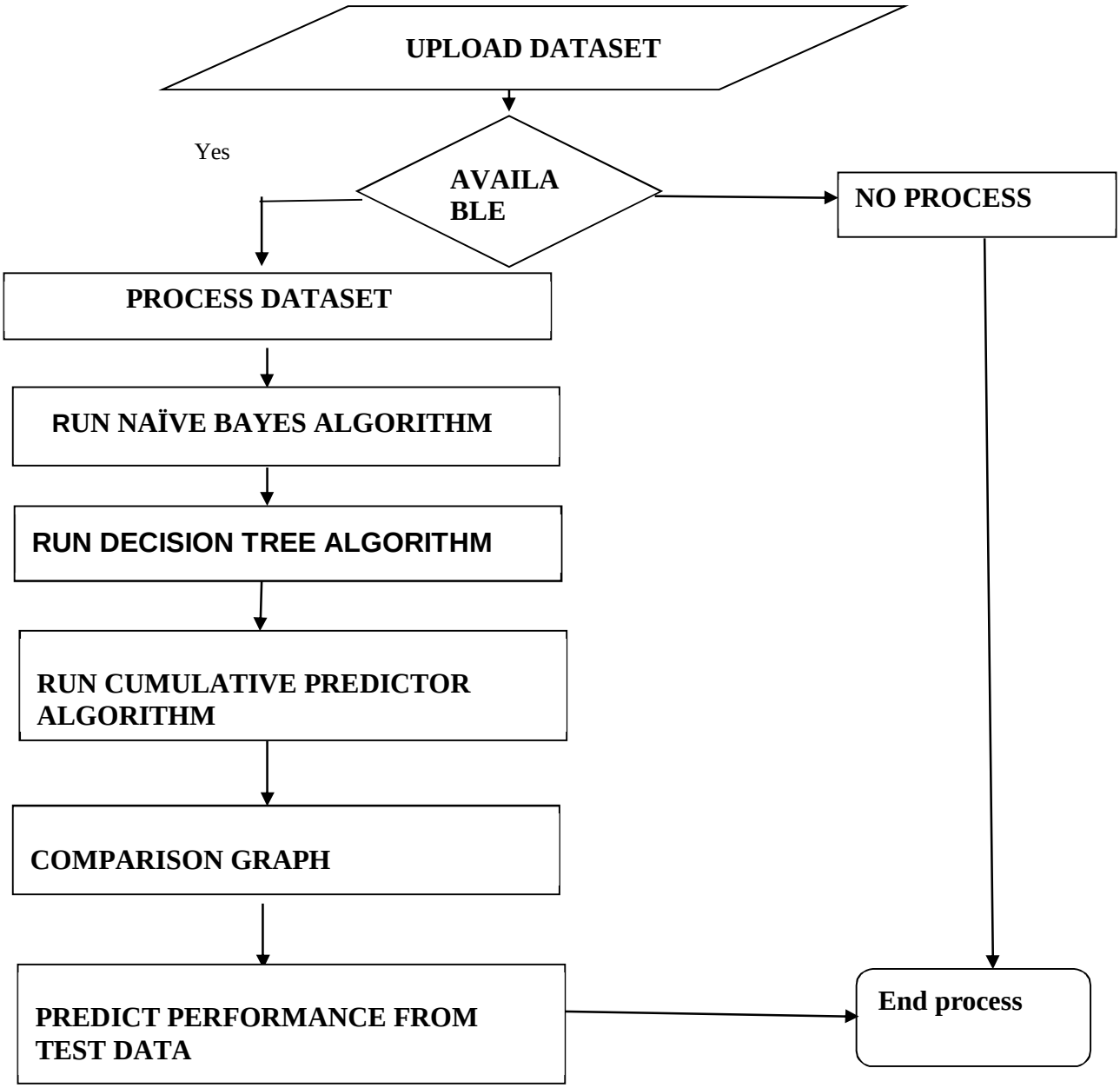
©IJRASET: All Rights are Reserved

B. Predicting Student Performance by Using Data Mining Methods for Classification

Authors: Kabakchieva, D.

Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making. This paper presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management. Scope- based This system can be very easily implemented and utilized by any educational institution. It can be used by faculties and students who do not have any knowledge on data mining techniques. Although there are so many benchmarks comparing the performance and accuracy of different classification algorithms, there are still very few experiments carried out on educational datasets. In this work, we compare the performance and the interpretation levels of the output of different classification techniques applied on educational datasets and finally develop a much more efficient algorithm called the cumulative predictor algorithm..

Fig 2.1: Flow diagram of student's performance analysis system using cumulative predictor algorithm



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

III. MOTIVATION

A country's growth is strongly measured by quality of its education system. Education sector, across the globe has witnessed sea change in its functioning. Today it is recognized as an industry and like any other industry it is facing challenges, the major challenges of higher education being decrease in students' success rate and their leaving a course without completion. An early prediction of students' failure may help the management provide timely counseling as well coaching to increase success rate and student retention..

IV. EXISTING SYSTEM

The data were collected from eight year period intakes from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records, and family background information. DT, NB, and RB classification techniques are applied to the students' data in order to produce the best SAP prediction model. The experiment result shows the RB is a best model among the other techniques by receiving the highest accuracy value of 71.3%. The extracted knowledge from prediction model will be used to identify and profile the student to determine the students' level of success in the first semester. This project acts as the basis of SPAS and gives a clear idea regarding the parameters involved in predicting students' performance.

V. PROPOSED SYSTEM

The aim of this project is to bridge the gap between the employer and the future employee of the company by the use of SPAS at college level. Student performance analysis system (SPAS) is an online web application system which enables students to know prior hand if their level of skills for the placement is enough to get placed or not, given the necessary inputs. SPAS has an intelligent learning algorithm which utilises a rich database, analyses the records of previous students' traits and develops a model for further prediction.

VI. ARCHITECTURE

The architecture provides the entire process flow of the system.

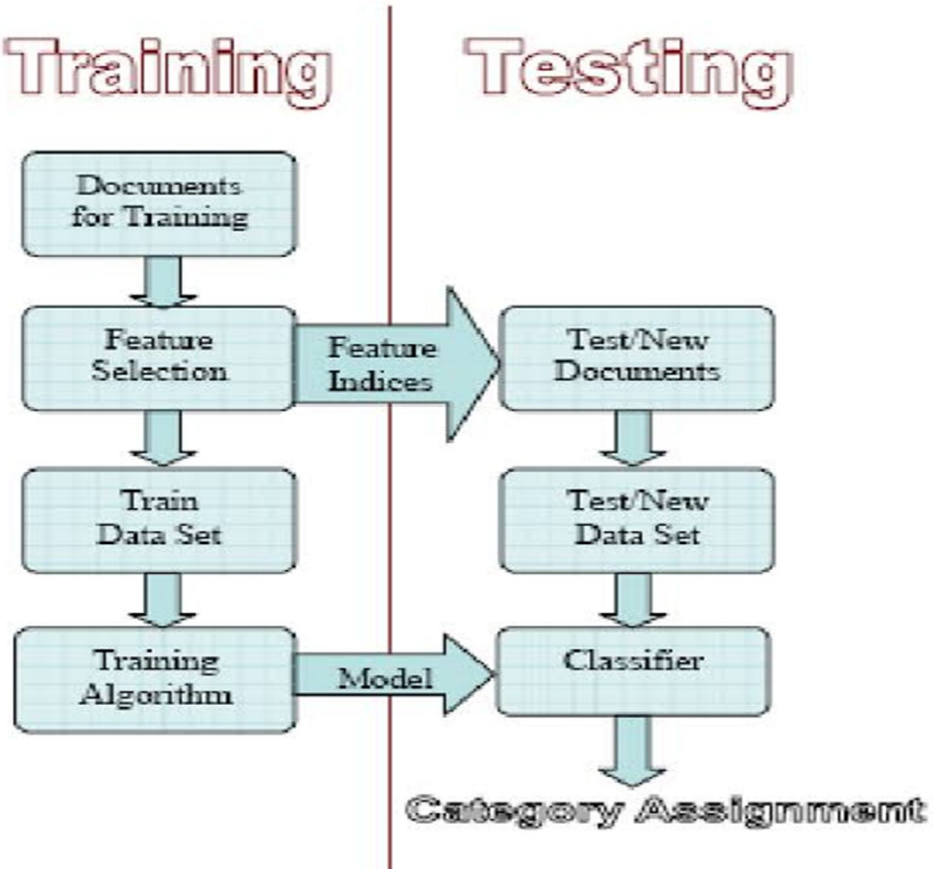


Fig-5 Architecture diagram

©IJRASET: All Rights are Reserved

VII. IMPLEMENTATION

A. *Algorithms Used*

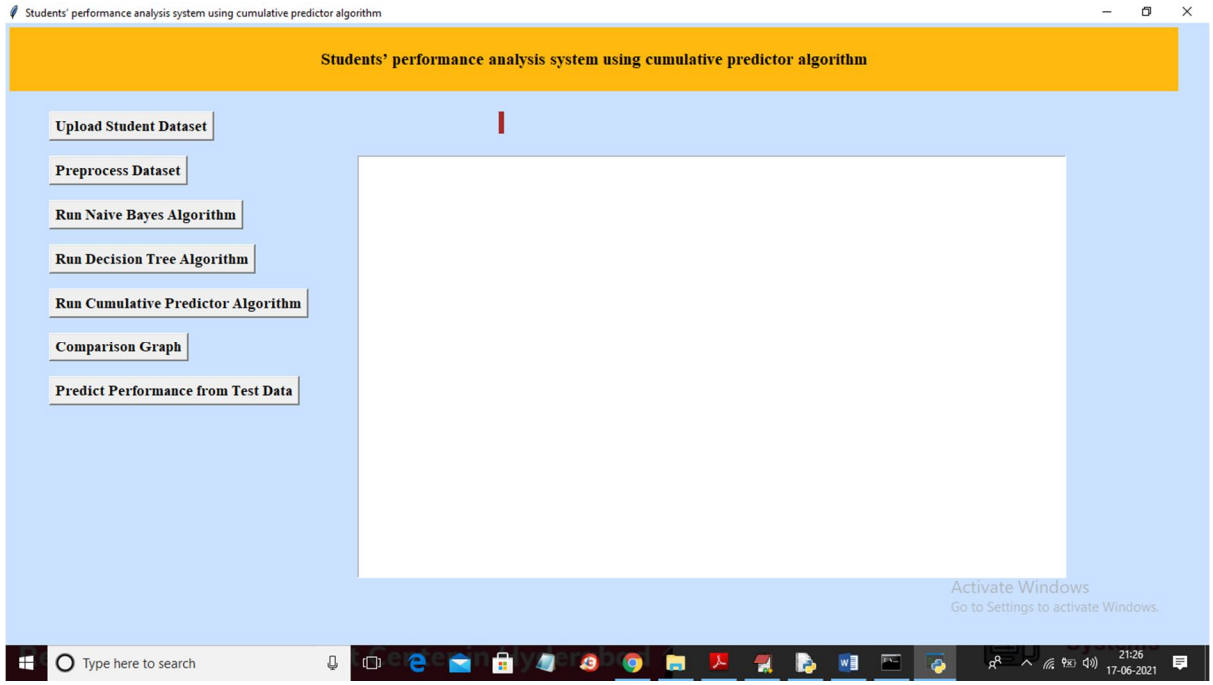
1) *Decision Tree:* Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. A decision tree is a very specific type of probability tree that enables you to make a decision about some kind of process. For example, you might want to choose between manufacturing item A or item B, or investing in choice 1, choice 2, or choice 3.
2) *Naive Bayes:* The Naive Bayes classification algorithm is a probabilistic classifier. It is based on probability models that incorporate strong independence assumptions. The independence assumptions often do not have an impact on reality. Therefore they are considered as naive. Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

B. *Methodology*

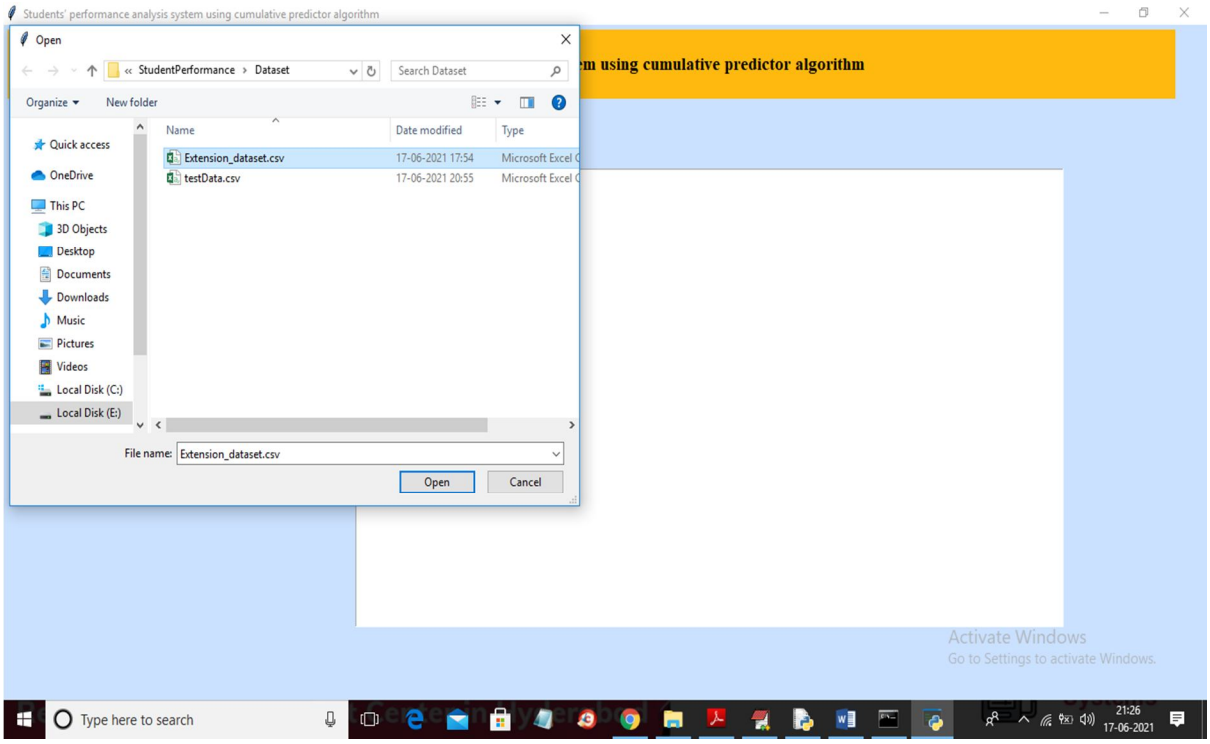
Testing is a process of executing a program with the aim off inding error. To make oursoftware perform well it should be error free. If testing is done successfully it will remove all the errors from the software.

1) *Unit Testing:* Software verification and validation method in which a programmer tests if individualunitsof source code are fit for use. It is usually conducted by the development team.
2) *Integration Testing:* The phase in software testing in which individual software modules are combined andtested as a group. It is usually conducted by testing teams.
3) *Alpha Testing:* Type of testing a software product or system conducted at the developer's site. Usuallyit is performed by the end users.
4) *Beta Testing:* Final testing before releasing application for commercial purpose. It is typically doneby end- users or others.
5) *Performance Testing:* Functional testing conducted to evaluate the compliance of a system or component with specified performance requirements. It is usually conducted by the performance engineer.

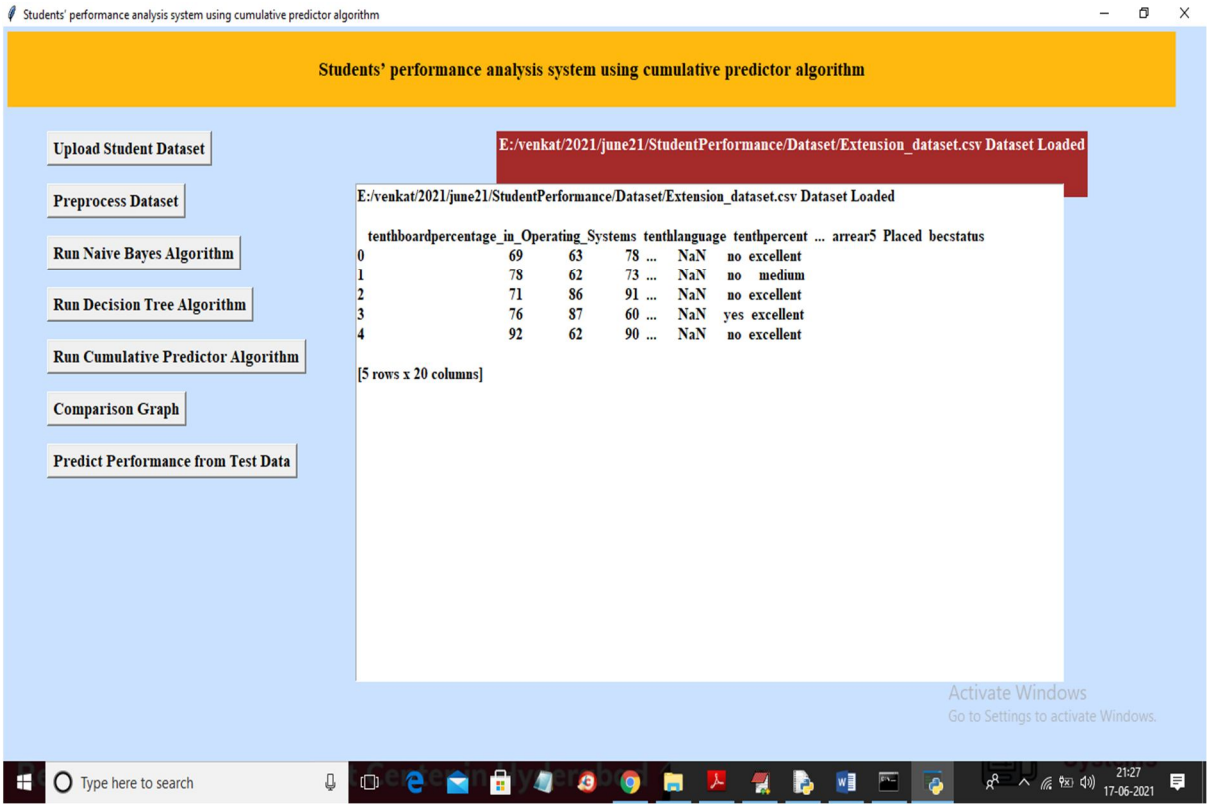
VIII. RESULTS



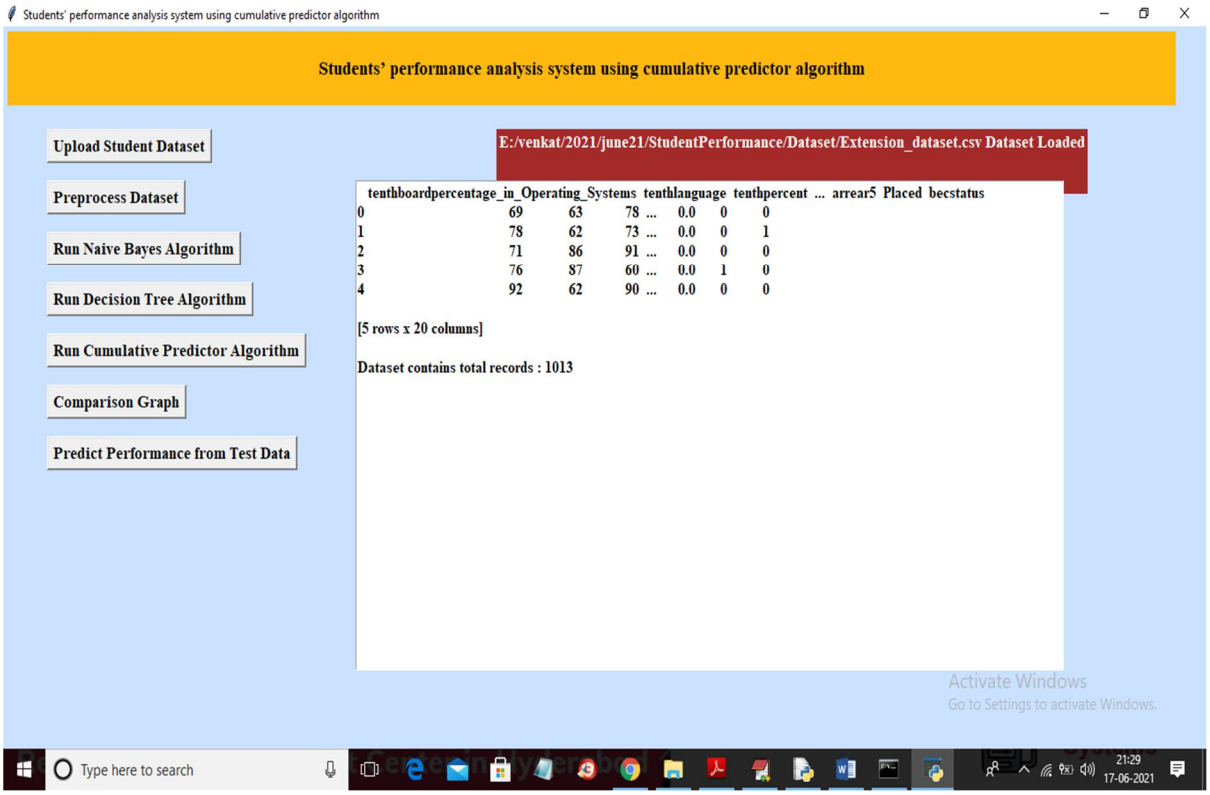
In above screen click on 'Upload Student Dataset' button to upload dataset



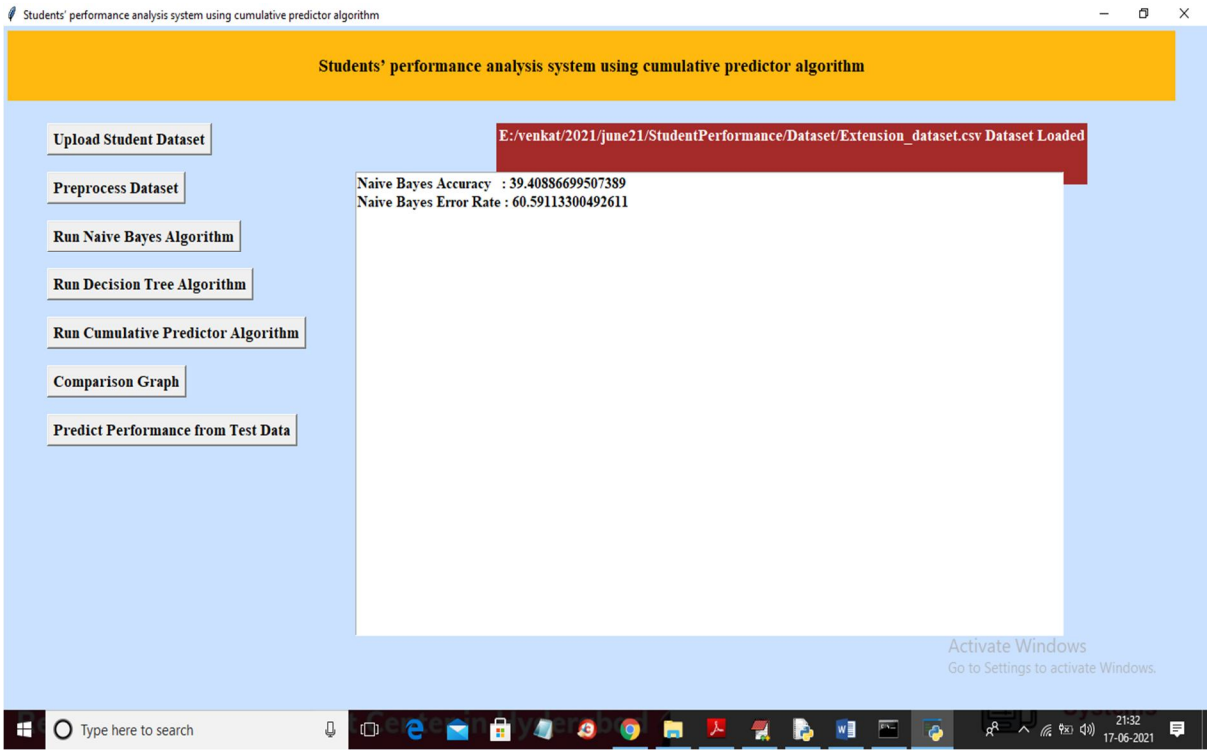
In above screen selecting and uploading 'Extension_dataset.csv' file and then click on 'Open' button to load dataset and to get below screen



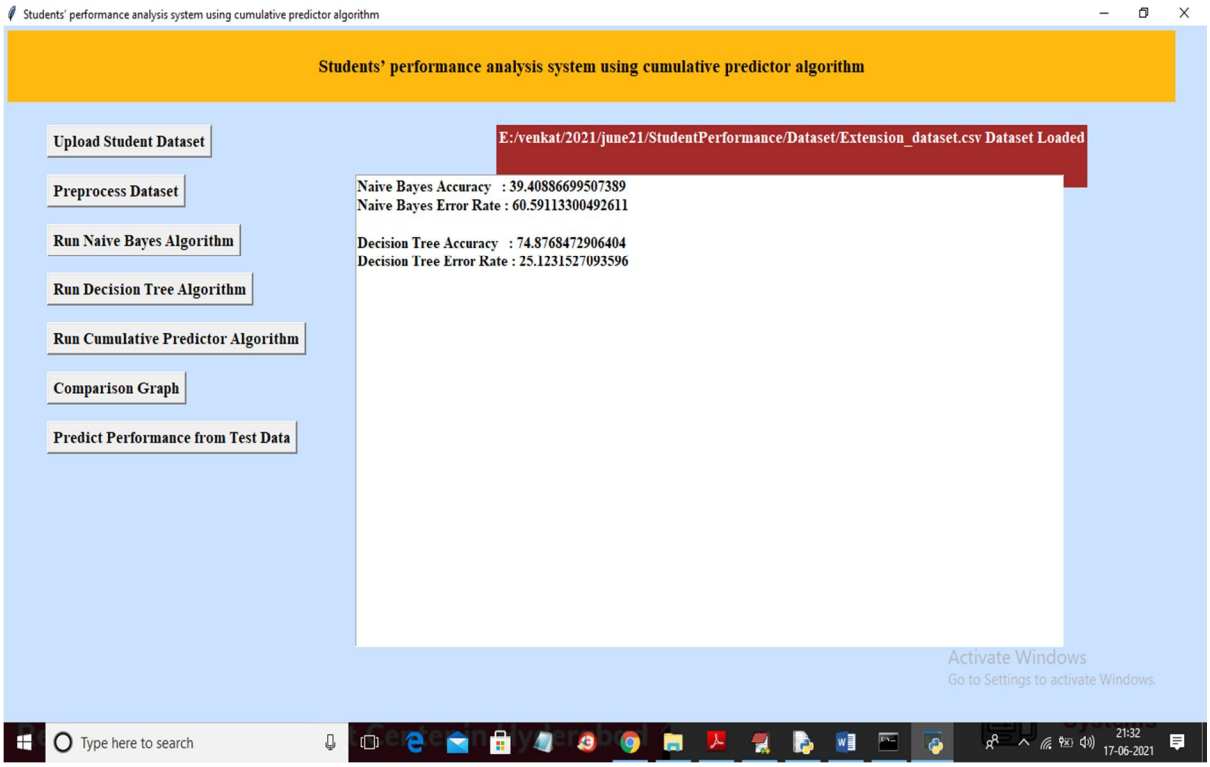
In above screen dataset loaded and we can see dataset contains some missing values 'NAN' and non-numeric values and machine learning algorithms won't accept non-numeric values so we need to preprocess dataset by clicking on 'Preprocess Dataset' button



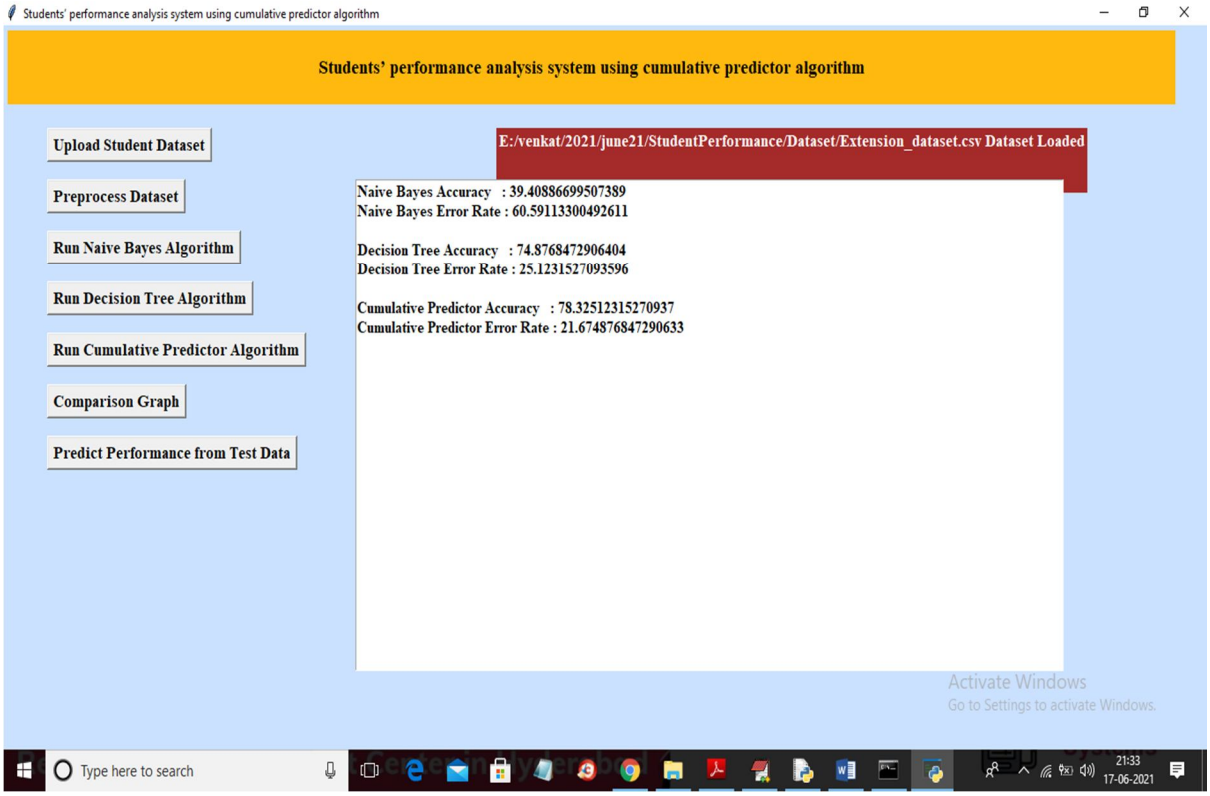
In above screen we can see dataset converted to numeric format and we can see dataset contains total 1013 records and now dataset is ready and now click on 'Run Naïve Bayes Algorithm' button to train Naïve Bayes with above dataset



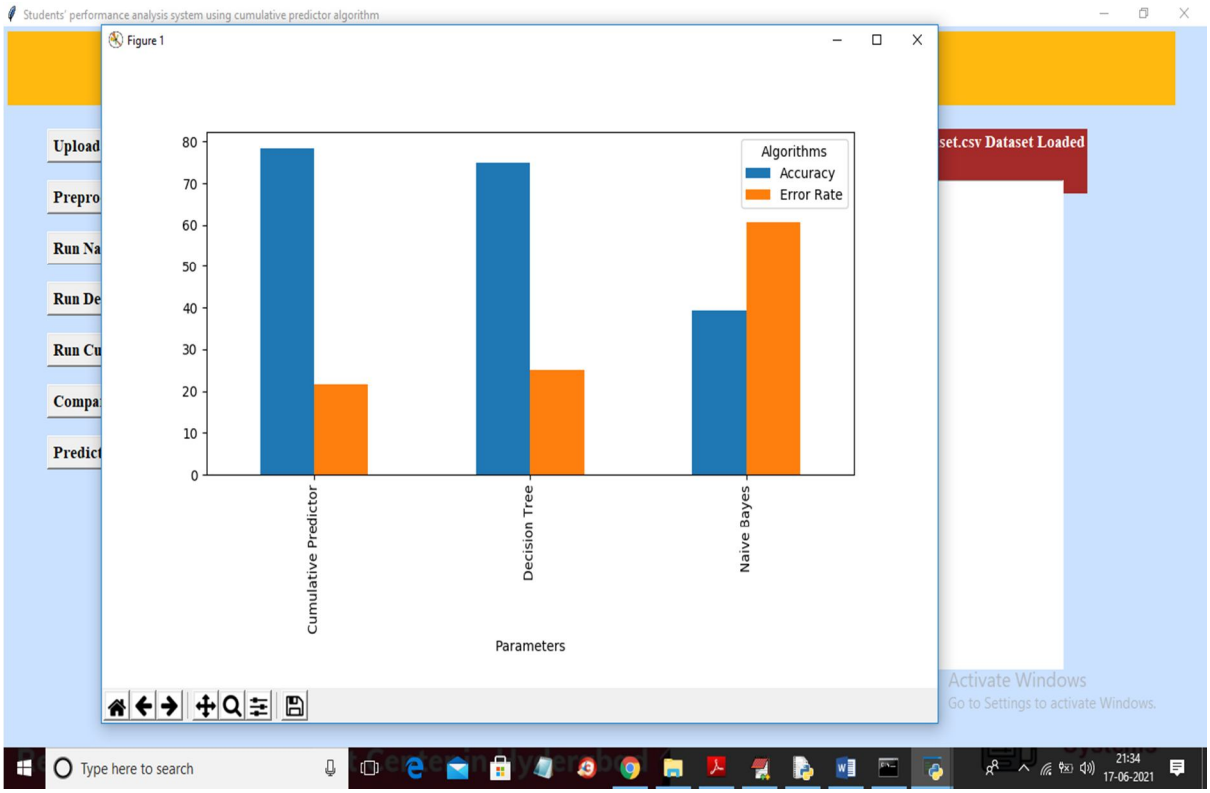
In above screen with Naïve Bayes we got 39% accuracy and now click on 'Run Decision Tree Algorithm' button to train above dataset with decision tree



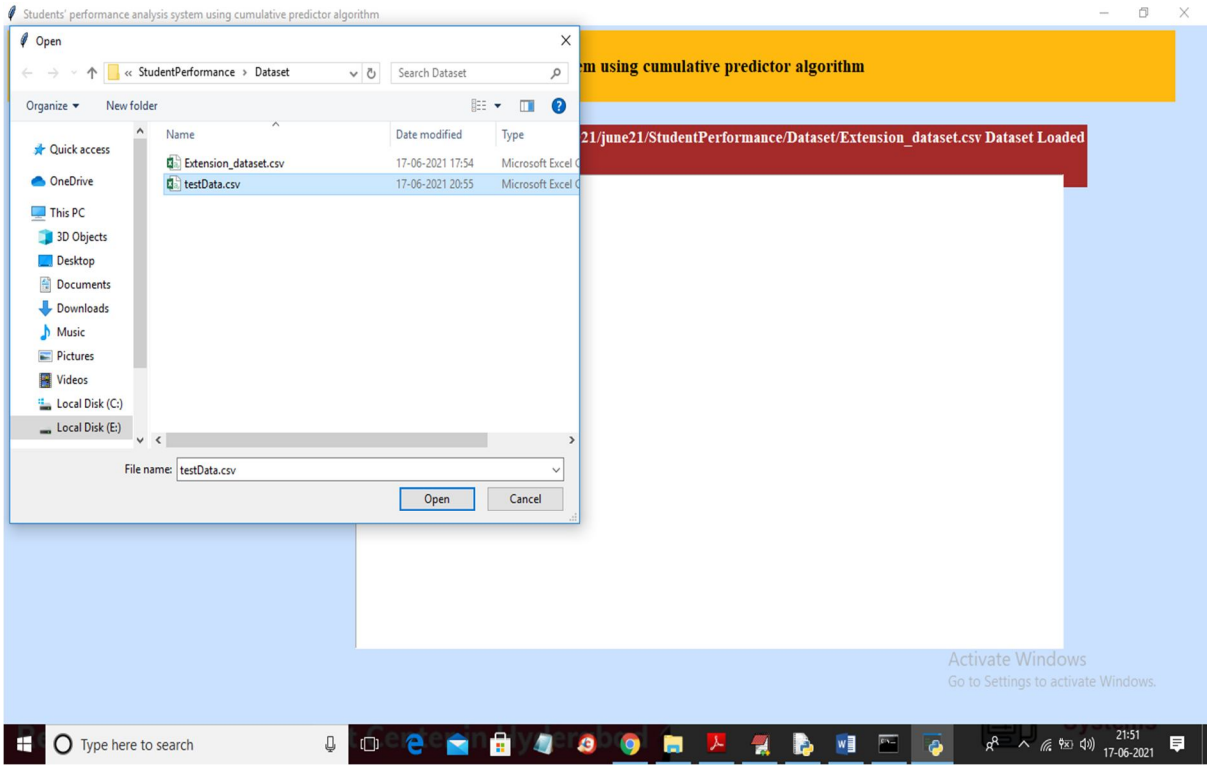
In above screen with decision tree we got 74% accuracy and now click on 'Run Cumulative Predictor Algorithm' button to train CP with above dataset



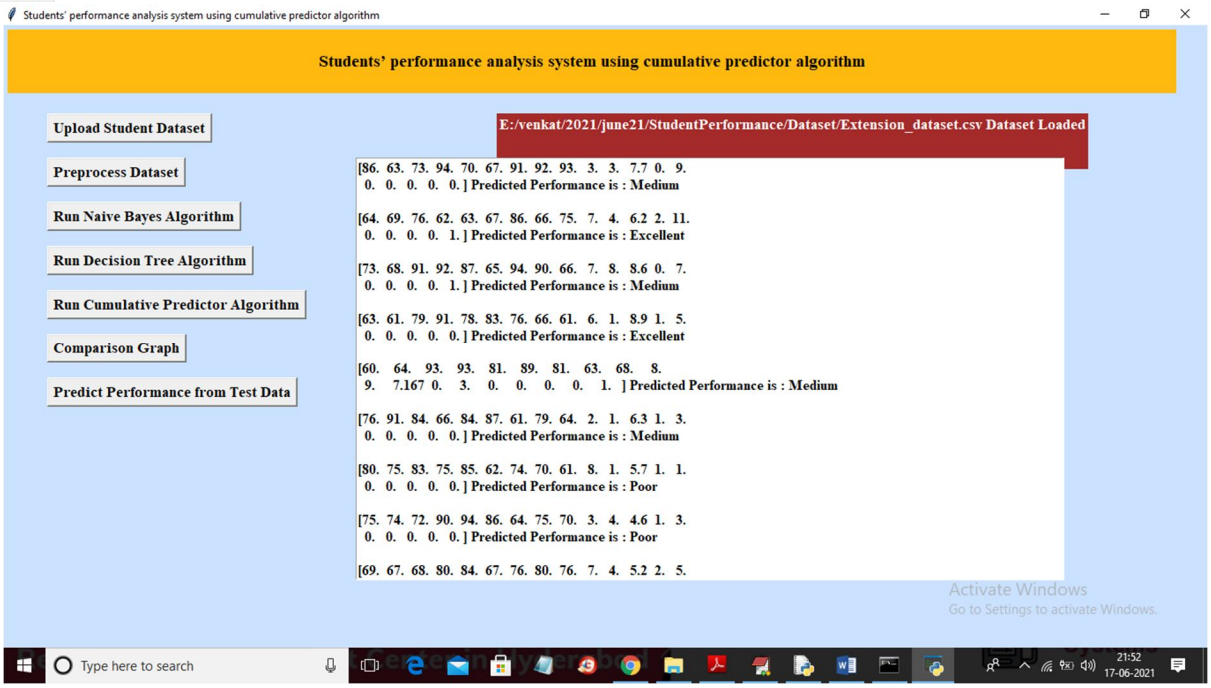
In above with CP we got 78% accuracy and now click on 'Comparison Graph' button to get below graph



In above graph x-axis represents algorithm name and y-axis represents accuracy and error rate and in above algorithm we can see CP got high accuracy with less error rate. Now click on 'Predict Performance from Test Data' button to upload test dataset and to get prediction result



In above screen selecting and uploading 'testData.csv' file and then click on 'Open' button to get below prediction result



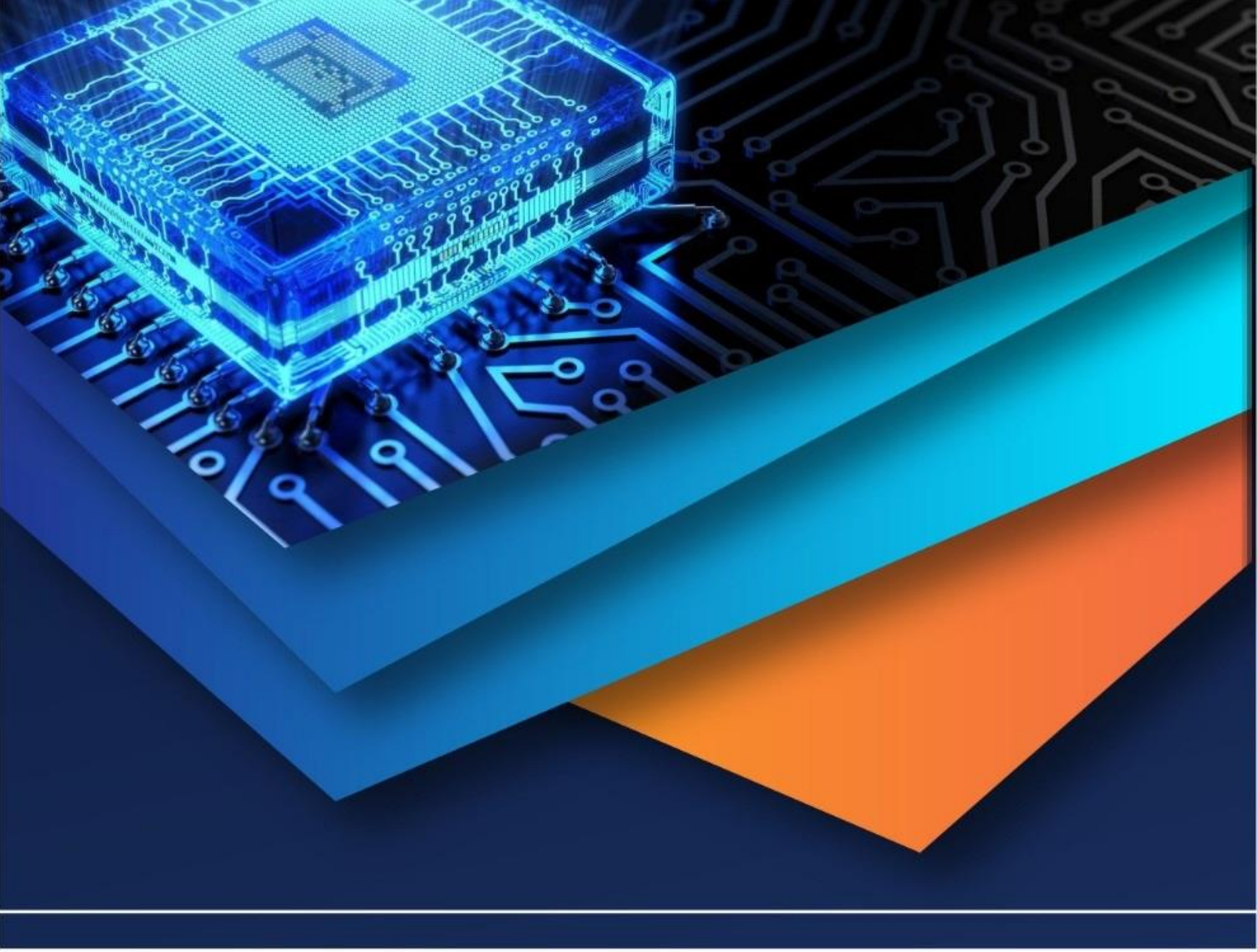
In above screen inside square bracket we can see student performance data and after square bracket we can see the prediction result from CP algorithm.

IX. ACKNOWLEDGMENT

We express our sincere gratitude to our guide, Assistant Professor **Mr. K. Praveen Kumar** for suggestion and support during every stage of this work. We also convey our deep sense of gratitude to Professor **Dr. K. S. Reddy**, Head of Information Technology department.

REFERENCES

[1] Ahmad, F., Ismail, N.H. and Aziz, A.A. (2015)'The prediction of students' academic performance using classification data mining techniques', Applied Mathematical Sciences, Vol. 9, No. 129, pp.6415–6426, HIKARI Ltd.
[2] García, E.P.I. and Mora, P.M. (2011) 'Model prediction of academic performance for first year students', Proceedings of 10th Mexican International Conference on Artificial Intelligence, pp.169–174.
[3] Jain, R. and Minz, S. (2008) 'Drawing conclusion from forest cover type datathe hybridized rough set model', Journal of the Indian Society of Agricultural Statistics, Vol. 62, No. 1, pp.75–84.
[4] Kabakchieva, D. (2013) 'Predicting student performance by using data mining methods for
[5] classification', Cybernetics and Information Technologies, Vol. 13, No. 1, pp.61–72 [online]
[6] http://dx.doi.org/10.2478/cait-2013- 0006 (accessed 16 December 2016).
[7] Mishra, T., Kumar, D. and Gupta, S. (2014) 'Mining students' data for performance prediction', Proceedings of Fourth International Conference on Advanced Computing and Communication Technologies, pp.255– 262.
[8] Oladokun, V.O., Adebanjo, A.T. and Charles-Owaba, O.E. (2008) 'Predicting students' academic performance using artificial neural network: A case study of an engineering course', The Pacific Journal of Science and Technology, Vol.
[9] No. 1, pp.72–79. 9. Patel, P.S. (2015) 'Various data mining techniques used to study student's academic performance', International Journal of Computer Science and Mobile Applications, June, Vol. 3, No. 6, pp.55– 58.
[10] Patidar, P., Dangra, J. and Rawar, M.K. (2015) 'Decision tree C4.5 algorithm and its enhanced approach for educational data mining', Engineering Universe for Scientific Research and Management



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)