iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52851

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 11 Issue V May 2023- Available at www.ijraset.com

Machine Learning Based Result Analysis Prediction of Student's Academic Performance

Akshay Bodkhe[1], Niranjan Zingade[2], Bhavesh Mundada[3]
[1, 2, 3]*Electronics & Telecommunication Engineering Department, Pune Institute of Computer Technology, Pune*

Abstract: ***Failure at any phase of education happens frequently. The rise in drop - out rates is a result of numerous reasons. Poor grades are one of the biggest causes of school abandonment. This has an influence on performance because so many students find it difficult to adjust to the institution's learning environment once they get there. Other factors include student participation in extracurricular tasks and politics. Learners' performances frequently tend to be unsatisfactory for these different predictable and unpredictable reasons, which have an impact on development. As a result, it's important to examine undergraduate results to identify the real reasons for students' varied level of performance. The primary goal of our research work is to identify the numerous variables that affect achievement at the under-graduation level. Therefore, the main motivation behind this effort is to help students identify the factors that lead to their performance so that they can take action to change their results. The learners, course teachers, and others will have the opportunity to improve the environment once the major elements have been recognized and assessed. This paper highlights the importance of using student data to drive improvement in education planning. It then presents techniques of how to obtain knowledge from databases such as large arrays of student data from academic Institution databases. To early predict the student's academic performance, we have proposed deep learning model of Recurrent Neural (RNN) classifier. This proposed methodology is compared with various traditional machines learning classification models and deep learning classifier.***

Keywords: RNN Recurrent Neural Network, EDM Educational Data Mining, LA Learning Analytics, DT Decision Trees, KNN K-nearest Neighbour, NB Naïve Bayes, SVM Support Vector Machine, RF Random Forests

I. INTRODUCTION

A. *Introduction*

1) *Background:* Student result in educational institutions is an important aspect and parameter to check the performance of students in their academic subjects. Technological development impacts every aspect of human life. Deep learning or Machine learning is the best technological tool for result analysis and prediction. Result analysis gives clear picture of the student's overall academic performance. We can define trends of students in specific subjects by using Result analysis. To early predict the student's academic performance, we have proposed deep learning model of Recurrent Neural (RNN) classifier. This proposed methodology is compared with various traditional machines learning classification models and deep learning classifier.

2) *Relevance:* Computer-supported collaborative learning, computer-adaptive testing, individual learning via educational software, testing in general, and the variables that contribute to student failure or non-retention in courses are just a few of the topics that researchers in educational data mining. Education and associated trans disciplinary sectors are impacted by educational data mining techniques (such as artificial intelligence in education, intelligent tutoring systems, and user modelling). EDM has been used to enhance student models, among other things. Student models deal with details about a student's traits or states, such as their knowledge, metacognitive abilities, and learning mindset. The programmer's power to handle individual variances is made possible by modelling individual student differences in these areas, significantly enhancing student learning. The EDM taxonomy by Baker deals with model-based discovery. Any procedure they can verify, most often via prediction, is used to construct a model of a learning activity. This model is then utilized as one of the modules in another study, such as relationship mining or prediction. When using Discovery with Models, you may analyse which educational resources will work best for specific student subgroups. Student learning behaviour's effects on learning. Data Mining Massive data sets are filtered using data mining to discover links and patterns they may use for data assessment to solve business problems. Businesses may use data mining techniques and technologies to predict future trends and make better-informed management choices. It is a core part of data science and one of the fundamental subfields of data analytics, which uses sophisticated analytics techniques to extract information from data sets. DM is a step in the Knowledge Discovery in Databases (KDD) protocol, a data science

©IJRASET: All Rights are Reserved

method for collecting, processing, and analysing data. DM and KDD are often considered independent concepts, although they are commonly employed together.

B. *Scope and Objectives*

The scope of machine learning-based result analysis for predicting student's academic performance can include the following:

1) *Data Collection and pre-Processing*: Collecting data from various sources such as student records, attendance records, and exam results, and cleaning and pre-processing the data to make it suitable for analysis.
2) *Feature Engineering:* Extracting relevant features from the data, such as student demographics, previous academic performance, and socio-economic factors, to build a predictive model.
3) *Model Selection and Training:* Selecting an appropriate machine learning model, such as regression or decision trees, and training it on the pre-processed data to predict student academic performance.
4) *Evaluation and Interpretation:* Evaluating the performance of the predictive model using various metrics such as accuracy, precision, and recall, and interpreting the results to identify the factors that influence student academic performance.
5) *To Identify at-risk Students:* Predictive models can identify students who are at risk of failing or dropping out of school. This information can help teachers and administrators intervene early to support struggling students.
6) *To Improve Teaching Methods:* Analysis of the factors that influence academic performance can help teachers and administrators identify areas where teaching methods can be improved to better support students.

II. DESIGN FLOW/PROCESS

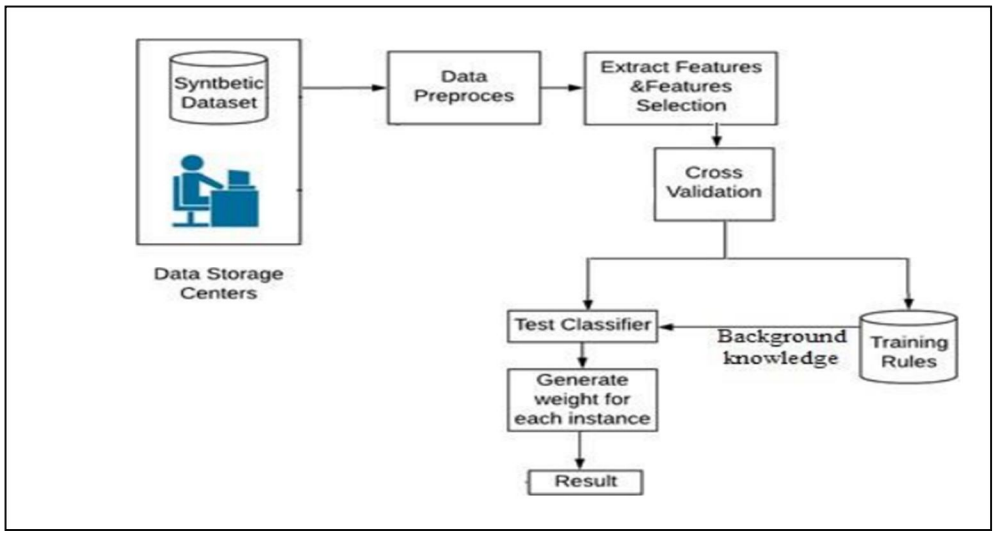


Fig 1: Flow of model Design

1) Data Collection: We gather data from a variety of sources, including UCI ML repository, Kaggle and several real-time data sources. Prior to executing the classification activity, the data must be pre-processed to enhance the outcome by addressing the missing values and removing the redundant features contained in the chosen dataset. To get the greatest results during the DM process, the dataset must be treated swiftly.

a) *Data Cleaning:* It addresses noisy data, missing information, etc. Different strategies have been adopted when some data in the information is incomplete, such as filling in the gaps or disregarding the tuples. Data may contain null values that are incomprehensible to machines. This noisy data may result from poor data collecting, incorrect data input, etc. Regression, clustering, and the binning approach are used to address it.

b) *Data Transformation:* This technique is used to change the data into the form that is suited for the mining procedure. Normalization, attribute choice, and discretization are involved in this technique.

c) *Stop word Removal and Stemming (Porter's algorithm):* Then we will apply various pre-processing steps such as lexical analysis, stop word removal, stemming (Porter's algorithm), index term selection and data cleaning in order to make our dataset proper.

2) *Feature Extraction and Selection:* From the data input, this procedure retrieves a variety of features. The extracted features are then standardized using a feature selection threshold, which eliminates redundant and unnecessary features for training. The normalized data with relational characteristics is used to extract a variety of hybrid attributes, and training is carried out by selecting an optimization strategy. The hybrid method has been used for feature selection from fully extracted features—selecting the best quality increases classification accuracy. Many irrelevant features appear during the feature extraction, which need to eliminate when we choose the parts. The benefit of this method is that it provides a respective feature selection for the individual feature set.
3) *Classification:* After the module has been successfully executed, the selected features are given as input to the training module, which produces comprehensive Background Knowledge for the overall system. After we get the training model, we can feed the testing data into it and get the prediction of classification. The testing stage includes pre-processing of testing text, vectorization, and classification of the testing text. The module testing evaluates the system's predictive performance using deep learning (RNN) methods. This step assessed the system's performance using different datasets.

III. LITERATURE SURVEY

1) According to Chitra Jalota et al. [2] in 2019, higher education universities frequently have a strong interest in learning about the overall success rate of the graduates. For the purpose of predicting students' achievement, they must thus employ a variety of methodologies, including physical exam, statistical tools, and the most widely used Data Mining (DM) approaches. EDM is a new field of study that makes use of DM methods. It makes use of statistical methods and Machine Learning (ML) methods to assist the user in deciphering a student's study habits, educational success, and future developments. This study discusses numerous DM approaches that can be used to forecast students' quality standards. Weka was performed to examine the DM strategies using the dataset from Kalboard 360.
2) According to Dr. R. Raju et al. [3] in 2020, the world is producing a lot of information these days, and the educational system has none of that. The collection of information needs to be examined because the pursuit of education is on the rise. DM combined with 11 machine learning algorithms are increasingly being employed in studies on how to analyse student details. Analysis of the data can be used to extract pertinent data from educational data and to identify significant relationships between various factors. In order to forecast a student's conduct or exam results, educational analysis of the data is utilized to examine student data. The evaluation of educational data in this research is determined by a variety of ML methods. The major goal is to examine how different ML techniques are applied to educational data that is utilized for forecasting.
3) According to Nabila Khodeir et al. [4] in 2019, EDM uses ML, DM, and statistical methods to analyse various forms of educational data. An EDM application used to suggest or modify the educational system resources is learner simulation. This paper identifies the behavior of several educational systems. Affective state modelling, academic achievement forecasting of student, and modelling of the pupil's style of learning are also covered. Additionally, student profiling, categorization, and collaborative analysis are also addressed.
4) According to Balqis Al Breiki et al. [5] in 2019, from data sets that reveal details about academia, EDM involves the extraction of ideas and related meaningful information. With the help of a toolbox, strategies, and methods for designing research, EDM may immediately detect patterns and associations in sizable data sets collected from educational settings. Given the abundance of information in huge educational datasets, forecasting student achievement has become a serious difficulty. Furthermore, Learning Outcome Assessments (LOA) are essential for determining and implementing improvements in the quality of instruction as well as for directing the development of specific pupils. To provide academic institutions with data essential to improve learning outcomes as soon as feasible, this research strives to make student achievement projections more effective and precise. To create learning methods that can provide precise forecasts of student GPA, this research uses regression and a number of ML approaches. Furthermore, a variety of attribute evaluator approaches were used to identify the components that have a substantial impact on a student's overall performance.
5) According to Muhammad Sammy Ahmad et al. [6] in 2021, over the past 20 years, DM and ML approaches have made astounding development. These advancements should 12 help academia learn more about how individuals learn in various educational environments. In order to predict student achievement focused on the demographic and evaluation data, the research compares ANN and RF machine learning algorithms. The two approaches were assessed following the application of feature-engineering approaches and analysis of the Open University Learning Analytics Dataset (OULAD). The outcomes demonstrated that the efficiency of the ANN technique surpassed that of the RF model by 91.08 percent to 81.35 percent. When used to predict academic achievement and in early warning method, ANN operates well on data sets.

6) According to Usman Ali et al. [7] in 2019, huge numbers of records in educational data sets can make it difficult to produce data of high quality. Many scholars in the field of education are now analysing data utilizing the DM technique. However, instead doing feature extraction on data, several research papers concentrate on choosing the best learning method. As a result, completing categorization on the dataset demands a significant amount of processing time due to its high computational effort. This document provides an overview of the feature selection techniques used for the evaluation of data characteristics. To improve the quality of the students' data collection, the suggested hybrid method integrates feature extraction with wrapper-based methods.

7) According to Muhib Al-kmali et al. [8] in 2020, making the right choices in academic procedures at a university has a significant impact on raising the standard of instruction and may be very advantageous for students, staff members, and the whole academic world. To help decision-makers increase the performance of learning systems, a decision support technology is provided in this study that offers accurate assessment, good decision assistance, and monitoring and scheduling capabilities. A collection of ML techniques is utilized to accomplish this. In a real case, the findings demonstrate that it can forecast student achievement to aid in decision-making. Additionally, utilizing precision, recall, accuracy, and F-measure, four ML techniques—SVM, NB, DT, and RF are compared.

8) According to Gabriella Casalino et al. [9] in 2020, Virtual Learning Environments (VLEs) are Web-based systems that offer instructional materials as well as study aids. Daily logs of student engagement with VLEs are gathered, hence computerized 13 solutions are expected to handle and evaluate such enormous amounts of data. The insights derived from educational data can be utilized by pupils, instructors, administrators, and generally all stakeholders participating in the VLEs' learning activities. Valuable information can be found utilizing ML approach. Historically, static ML techniques have been used to analyse educational data sets. Nevertheless, because they are by their very nature non-stationary, educational dataset are best handled as data streams. The findings of a classification research in which the RF method was employed to create a model for forecasting the passing or failing of students' tests were presented in this publication. The strongest discriminating qualities for the prediction task are also found via a feature importance assessment. On the Open University Learning Analytics Database (OULAD), tests were conducted to demonstrate the accuracy of adaptive RF in producing classification techniques from changing data sets.

9) In 2020, according to Latifa Estrelita Indi Pramesti Aji et al. [10] forecasting a student's success is more challenging due to the large amount of data in educational databases. As a result, it is advised to do a thorough literature analysis to predict student achievement using DM approaches, particularly using the C4.5 method. The major goal of this study is to describe the data mining methods utilized to forecast student performance. This research article also considers how the classification technique might be applied to categorize the key elements of a student's data. In short, using educational DM techniques, especially the C4.5 classifier, may help students perform better and advance more successfully. The efficiency of the C4.5 method was 71.9 percent according to the dataset's findings. Students, professors, and academic institutions may benefit from this and gain influence.

IV. SYSTEM REQUIREMENTS

A. *Functional Requirements*

1) *Dataset:* Training Deep learning models requires extensive data to achieve high accuracy, low loss of features and increasing efficiency.
2) *Data Pre-processing:* Training Deep Learning models on large datasets requires effective data pre-processing under consideration for effective feature extraction. We have use various data Pre-processing techniques for same.
3) *Model Interpretability:* for the effective recognition, effective models like RNN are used to explain the working of Models used.

B. *Non-Functional Requirements*

1) *Testability:* for effective working of model, modular architecture is used where each part is divided into multiple modules for correct working
2) *Reliability:* Feature tests are performed to ensure reliability of dataset and quality of dataset, So the accuracy and performance of system are satisfactory.

C. Software Requirements

1) Front end

a) Operating System: -Windows XP/7/8
b) Programming Language: JAVA/J2EE/
c) Tools: Eclipse or Higher, Heidi SQL, JDK 1.7 or Higher
d) Database: MySQL 5.1

2) Backend

a) MySQL 5.1

D. Hardware Requirements

1) Processor: - Intel Pentium 4 or above
2) Memory: - 2 GB or above
3) Other peripheral: - Printer
4) Hard Disk: - 500gb

V. CONCLUSIONS

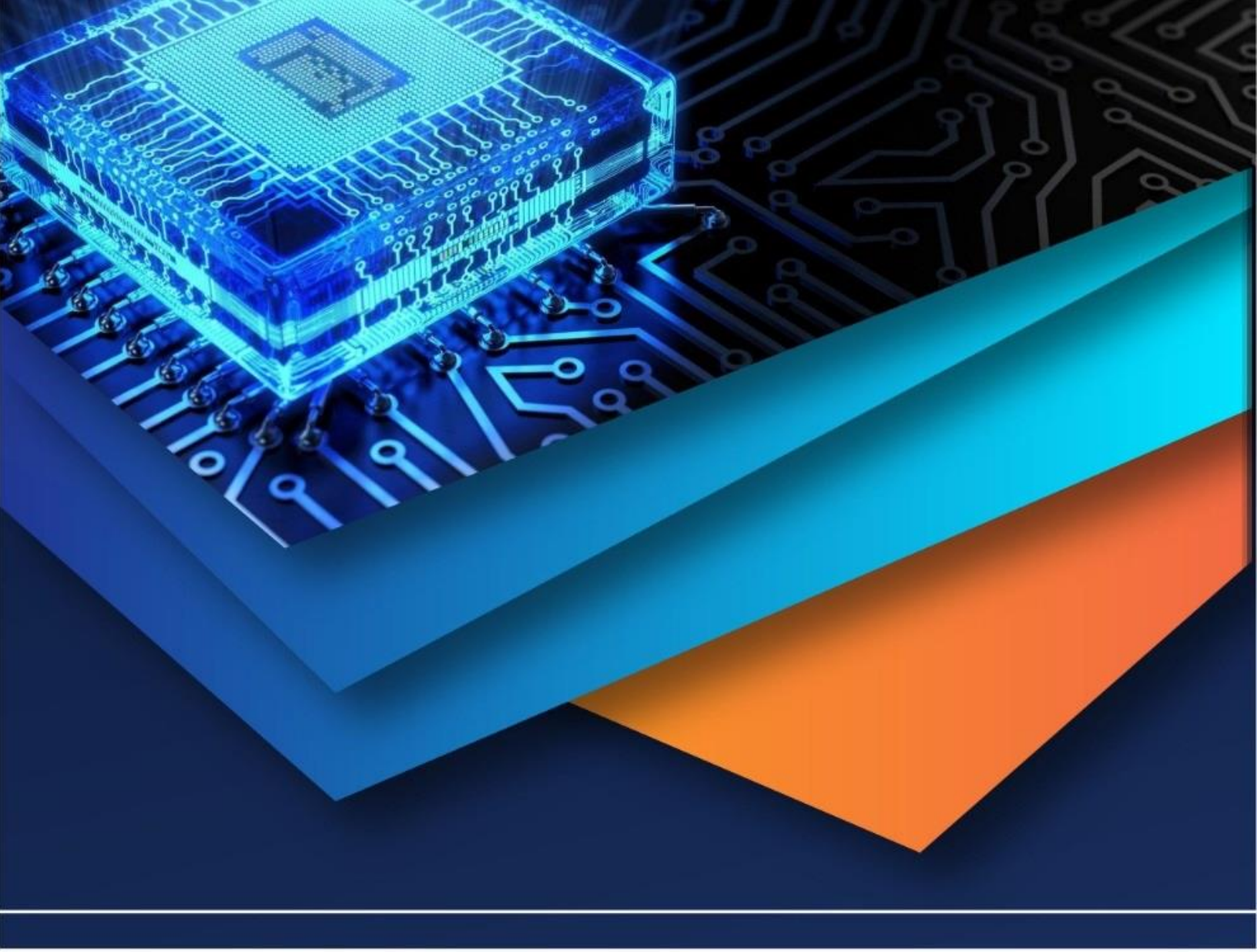
In conclusion, machine learning has proven to be a powerful tool in analysing and predicting student performance. By utilizing various data sources and algorithms, machine learning models can identify patterns and relationships between factors such as attendance, study habits, and assessment scores to make accurate predictions about a student's academic performance. This information can be used by educators and institutions to develop targeted interventions and support systems to help students succeed. However, it is important to note that machine learning is not a silver bullet and should be used in conjunction with other forms of data analysis and human expertise to ensure the best outcomes for students. Overall, machine learning has the potential to revolutionize the way we approach education and student success.

VI. ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion and it asks would be incomplete without the mention of the people who made it possible and whose constant encouragement and guidance have been a source of inspiration throughout the course of this project. We take this opportunity to express my sincere thanks to my guide respected Dr. M.V.Munot ma'am for their support and encouragement throughout the completion of this project. Finally, I would like to thank all the teaching and non-teaching faculty members and lab staff of the department of electronics and telecommunications engineering for their encouragement. I also extend our thanks to all those who helped us directly or indirectly in the completion of this project.

REFRENCES

[1] Kelly J. de O. Santos, Angelo G. Menezes, Andre B. de Carvalho and Carlos A. E. Montesco. "Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout", 2019, 19th International Conference on Advanced Learning Technologies (ICALT), IEEE.
[2] Chitra Jalota and Rashmi Agrawal. "Analysis of Educational Data Mining using Classification", 2019, International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), IEEE.
[3] Dr. R. Raju, Mrs. N. Kalaiselvi, Aathiqa Sulthana M, Divya I and Selvarani A. "Educational Data Mining: A Comprehensive Study", 2020, IEEE.
[4] Nabila Khodeir. "Student Modeling Using Educational Data Mining Techniques", 2019, IEEE.
[5] Balqis Al Breiki, Nazar Zaki and Elfadil A. Mohamed. "Using Educational Data Mining Techniques to Predict Student Performance", 2019, International Conference on Electrical and Computing Technologies and Applications (ICECTA), IEEE.
[6] Muhammad Sammy Ahmad, Ahmed H. Asad and Ammar Mohammed. "A Machine Learning Based Approach for Student Performance Evaluation in Educational Data Mining", 2021, IEEE.
[7] Usman Ali, Khawaja Sarmad Arif and Dr. Usman Qamar. "A Hybrid Scheme for Feature Selection of High Dimensional Educational Data", 2019, International Conference on Communication Technologies (ComTech 2019), IEEE
[8] Muhib Al-kmali, Hamzah Mugahed, Wadii Boulila, Mohammed Al-Sarem and Anmar Abuhamdah. "A Machine-Learning based Approach to Support Academic Decision-Making at Higher Educational Institutions", 2020, IEEE
[9] Gabriella Casalino, Giovanna Castellano, Andrea Mannavola and Gennaro Vessio. "Educational Stream Data Analysis: A Case Study", 2020, IEEE. 29
[10] Latifaestrelita Indi Pramesti Aji and Andi Sunyoto. "An Implementation of C4.5 Classification Algorithm to Analyze Student's Performance", 2020, 3rd International Conference onInformation and Communication Technology (ICOIACT), IEEE



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)