



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 10    **Issue:** VI    **Month of publication:** June 2022

**DOI:** <https://doi.org/10.22214/ijraset.2022.44032>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Student Performance Prediction using AI and ML

Dr. Ankita Karale<sup>1</sup>, Ayushi Nalawar<sup>2</sup>, Bhushan Bhujbal<sup>3</sup>, Sakshi Bharit<sup>4</sup>

<sup>1</sup>Professor, Sandip Institute of Technology and Research Centre, Nashik, India

<sup>2, 3, 4</sup>Student, Sandip Institute of Technology and Research Centre, Nashik, India

**Abstract:** As we all know that due to this pandemic situation many problems had been faced in the education stream. As we saw that the most of the students got good marks and on the other hand most students got lesser or average marks. In Critical case the good students got average marks or lesser than their expectations. As a result, all students got the admission but if the person is not that capable but taking admission to that college due to good result. So it seems unfair so to resolve this loophole, we decided to build a software for student performance prediction. It is an important desire in most of the educational departments and institutes to predict the student's performance. Machine Learning algorithm and descriptive datasets include school, Sex, Age, Address, Family size, P status, M education, F education, M job, F job, Reason, Guardians, Travel time, Study time, Failures, School sup, Family sup, Activities, Free time, Go out, Health, attendance. We have been considered that the prediction and classification of student performance respectively using three type of machine learning algorithms such as Random Forest, ANN, XG Boost are implemented to predict the student's academic performance. The prediction based models are created to predict academic outcomes of student performance at the end of the year each. The analysis of demographical attributes release that they are also potential indicators of a student's academic success or failure. The results of these case study gives techniques for accurately predicting student performance, and compare the accuracy with MI algorithms Index Terms- ML and AI, ANN, Random Forest, XG Boost prediction model. Index Terms- ML and AI, ANN, Random Forest, XG Boost prediction model.

**Keywords:** Machine Learning, XG-BOOST, Random Forest Algorithm, Artificial Neural Network, Social Media Post.

## I. INTRODUCTION

An education system is one of the most important parts for the growth of any country. Also education is considered as a very important need for motivating self-assurance as well as providing the things needed to be part in today's World. Predicting student performance is an important task for the students in universities, colleges and schools etc. In present time computers and portable devices are used in every phase of daily life and many materials are available online anytime, anywhere. Technologies like Machine Learning and Artificial Intelligence had a surprising evolution in many fields especially in educational teaching and learning processes. Higher education institution have started to adopt the use of technology into their traditional. In this paper, we used a machine learning algorithm and descriptive datasets include school, Sex, Age, Address, Fam size, P status, M education, F education, M job, F job, Reason, Travel time, Study time, Failures, School sup, Fam sup, Paid, Activities, Higher, Internet, Free time, Go out, Health Absence have been considered for the prediction and classification of student performance respectively using three machine learning algorithms (Random forest, Decision Tree, Linear Regression) are implemented to predict the student's performance. Prediction based models were designed to predict academic outcomes of student performance at the end of the school year for each dataset though the attributes factor school, Sex, Age, Address, Fam size, P status, M education, F education, M job, F job, Reason, Travel time, Study 4time, Failures, School sup, Fam sup, Paid, Activities, Higher, Internet, Free time, Go out, Health Absences. They are also possible indicators of a student's academic success or failure.

The results of these case studies give insight into techniques for accurately predicting student performance, and compare the accuracy of MI algorithms. Prediction of student's performance became an important desire in most educational entities and institutes. This is essential to help the students and assure their retention, providing learning resources and experience, and improving the university's ranking and reputation. However, that might be difficult to achieve for start up to mid-sized universities and colleges, specifically people who are professionals in graduate and postgraduate programs, and feature small scholar records for analyze. Machine Learning is one of the most powerful and emerging technologies, so we are going to use machine learning to improve the accuracy in the model.

The categorization of tasks is used to evaluate students' performance and as there are many approaches that are used for data categorization, the decision tree methods used. Information like factor school, Sex, Age, Address, Fam size, P status, M education, F education, M job, F job, Reason, Travel time, Study time, Failures, School sup, Fam sup, Paid, Activities, Internet, Free time, Go out, Health, Absence marks was collected from the students system, to predict the performance at the end of semester year.

### A. Motivation

Prediction of scholar's overall performance have become an vital choice in maximum educational entities and institutes. This is essential to help the students and assure their retention, providing learning resources and experience, and improving the university's ranking and reputation. However, that might be difficult to achieve for start up to mid-sized universities and colleges, mainly people who are experts in graduate and submit graduate programs, and feature small scholar records for analyze. Machine Learning is one of the most powerful and emerging technologies, so we are going to use machine learning to improve the accuracy in the model. The categorization of tasks is used to evaluate students' performance and as there are many approaches that are used for data categorization, the decision tree methods used. Information like factor school, Sex, Age, Address, Fam size, P status, M education, F education, M job, F job, Reason, Travel time, Study time, Failures ,School sup, Fam sup, Paid, Activities, Internet, Free time, Go out, Health, Absence marks was collected from the students system, to predict the performance at the end of the semester/year.

## II. LITERATURE REVIEW

An computerized assessment machine has been proposed to assess scholar overall performance and to investigate the scholar achievement. Here the writer makes use of tree set of rules for predicting scholar overall performance accurately. In the proposed machine Education Data Mining (EDM) is used for the type. Clustering facts mining method is used for studying the huge set of scholar database. This method will accelerate the looking technique and the additionally yield the type end result extra accurately[1]. M.Ramaswami and R.Bhaskaran have used CHAID prediction version to investigate the interrelation among variables which can be used to are expecting the final results of the overall performance at better secondary faculty schooling. The functions like medium of instruction, marks received in secondary schooling, place of faculty, dwelling region and form of secondary schooling had been the most powerful signs for the scholar overall performance in better secondary schooling. The CHAID prediction version of scholar overall performance turned into built with seven magnificencepredictor. [2]

Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt- Thieme have implemented system gaining knowledge of strategies to enhance the prediction effects of instructional performances for 2 the actual case studies. Three strategies were used to address the magnificence imbalance hassle and they all display high-quality effects. They first re balanced the datasets after which used each cost-insensitive and touchygaining knowledge of with SVM for the small datasets and with Decision Tree for the bigger datasets. The fashions are to beginwith deployed at the neighborhood web. [3]

Arockiam et al. used FP Tree and K-approach clustering method for locating the similarity among city and rural college students programming abilities. FP Tree mining is implemented to sieve the styles from the dataset. K-approach clustering is used to decide the programming abilities of the scholars. The take a look at actually suggests that the agricultural and the city college students range of their programming abilities. The massive proportions of city college students are proper in programming talent in comparison to rural college students. It divulges that academicians provide extra training to town university college students within side the programming subject.[4]

Cortez and Silva tried to are expecting failure withinside the middle classes (Mathematics and Portuguese) of secondary faculty college students from the Alentejo location of Portugal with the aid of using utilizing. Four facts mining algorithms including Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Support Vector Machine (SVM) had been implemented on a facts set of 788 college students, who regarded in 2006 examination. It turned into said that DT and NN algorithms had the predictive accuracy of 93%and 91% for 2-magnificence dataset (pass/fail) respectively. [5]

San Pedro et. al evaluation an internet primarily based totally tutoring machine for arithmetic from 3747 faculty college students and expected whether or not a scholar will (five years later) attend university. Authors found out the scholars are a hit in center faculty arithmetic as measured with the aid of using the schooling machine are much more likely to enrol five years later in university. On the opposite hand the scholars who confirmed confusion, carelessness withinside the machine have decrease opportunity of university enrollment. For the prediction they used logistic regression classifier. [6]

Vihavainen et. al labored with a image facts from Computer Science college students programming path of Helsinki University and attempted to are expecting whether or not a scholar will fail introductory arithmetic path. [7]

Bayer et. al expected whether or not a bachelor scholar will drop-out from university. They labored with the facts of Applied Informatics bachelor college students from Masaryk University and predicted pupil's studies, sports activities with special university college students thru email or discussion. They located college students who talk with college students having proper grades can effectively graduate with a better opportunity than college students with comparable overall performance however now no longer speaking with a hit college students. In this case, J48 choice tree learner, IB1 lazy learner, PART rule learner, SMO guide vector machines were used. [8]

Bhardwaj and Pal predicts college students overall performance and located out dwelling place has excessive have an effect on on college students very last grade. They used Purvanchal University's Department of Computer Applications scholar's facts and used Bayesian Classifier for predicting.[9]

studied the C++ path in Yarmouk University, Jordan. Three exclusive type strategies specifically ID3, C4.5 and the Naive Bayes are used. The effects indicated that Decision Tree version had higher prediction than different fashions. [10]

### III. PROBLEM STATEMENT

To develop an application to provide information Prediction of students. Educational organizations are one of the important parts of our society and playing a vital role for growth and development of any nation. Educational MI is the application of MI. This paper investigates the accuracy of MI techniques for predicting student performance. The faculty cannot find out students abilities and their interest easily so that they can enhance them in it. Thus it is able to have an effect on with poor college results, placement and career of individual. The effect is- it assist us from satisfying mission and vision of the institute. If the mission get a success then it'll be wonderful assist for school to enhance training system.

### IV. OBJECTIVE

- 1) To design a user-friendly web interface/ website on which the system can be implemented.
- 2) To be able to make the performance prediction mechanism more efficient and accurate.
- 3) To be able to predict the student performance using the MI algorithms
- 4) The objective of this project is to use MI algorithm to study students' performance in the year. MI provides many tasks that could be used to analyze the student performance. In this research, the classification algorithm is used to evaluate students' performance and as there are many processes that are used for data classification, the decision tree method is used here.

### V. METHODOLOGY

We are proposing an automated solution for the overall performance evaluation of the college students the usage of system mastering. we used a system mastering set of rules and descriptive datasets attribute /thing Encompass school, Sex, Age, Address, Famsize, Pstatus, Medu, Fedu, Mjob, Fjob, Reason, Gaurdian, Traveltime, Studytime, Failures, Schoolsup, Famsup, Paid, Activities, Higher, Interne, Farmel, Freetime Goout Health, Absence. had been taken into consideration for the prediction and class of pupil overall performance respectively the usage of 3 system mastering algorithms inclusive of Random forest, ANN, XG Boost are applied to are expecting the pupil's educational overall performance. This dataset for the current look at turned into accumulated from the internet site <https://www.kaggle.com>. It consists of many independent Attribute or Factors. The attributes may be divided into 5 classes which can be non-public and lifestyle, reading style, own circle of relatives related, instructional surroundings satisfaction, and pupil's grades. Following suggests the attributes used with a view to assemble the dataset. Each pupil has been intended as Weak or Good primarily based totally on there very last results. The susceptible pupil who has a very last grade much less than sixty out of hundred. On the opposite hand, the Good pupil who has a very last grade identical or more than sixty. Recognizing the susceptible college students is extra critical than Recognizing the coolest college students, consequently the susceptible pupil is taken into consideration a high quality cost of the goal attribute.

### VI. ALGORITHM USED

- 1) *ANN*: Artificial Neural Networks are special type of machine learning algorithms that are designed after the human brain. ANNs are nonlinear statistical models which is use for solution of complex relationship between the inputs and outputs to discover a new pattern. The variety of tasks such as image recognition, speech recognition, machine translation as well as medical diagnosis makes use of these artificial neural networks. They are just like how the neurons in our nervous system are able to learn from the past data, similarly, the ANN is able to learn from the data and gives responses in the form of predictions or classifications. The algorithm goes in that way like the first step is we gives a data input to algorithm after that in next step the data is divided into parts like dataset training and dataset testing process where the dataset is trained and tested for the next process. In the next process we install neural net package after that we follow the process of network training, plotting network, testing network and the output printing after that if error appears we again follow the same sequence from network training to output printing and if error is not appeared and we go for ANN output mapping. Features Of Artificial Network (ANN).

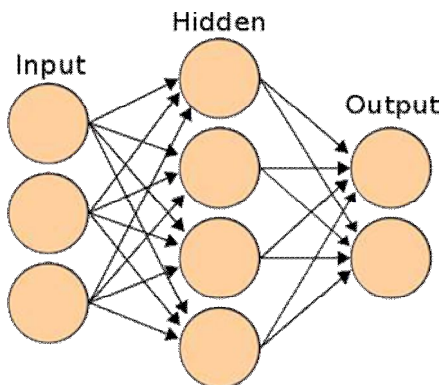


Fig 1 ANN

- 2) *Random Forest*: The random forest is the classification and prediction algorithm consisting of many decisions trees. It uses feature randomness and bagging when building individual tree try to design uncorrelated forest of trees whose prediction is more accurate than any individual tree. We use Random Forest for student performance prediction of result. Random Forest uses Decision trees for accuracy of student performance prediction. The random forest consists of huge number of decision trees. Decision trees are sensitive to the data to avoid this condition we use Bagging. Bagging is a process where we train decision tree for that we take random sample from the dataset.
- 3) *XG Boost*: XG Boost has widely used and really very popular tool among Kaggle competitors and Data Science industry, as it has been used for production on large-scale problems. It is a very flexible and versatile tool that can work through most regression, classification and ranking problems as well as user-built objective functions. As an open-source software, it is easily accessible and it may be used through different platforms and interface. The primary purpose of using XG Boost is due to its execution speed, and it's model performance. XG Boost uses ensemble learning methods it uses a combination of different algorithms and produces output as a single model. XG Boost supports parallel and distributed computing while offering efficient memory usage.



Fig 2 XGBoost

## VII. DATA PRE-PROCESSING AND CLEANING

- 1) *Datasets*: We're taking information from the kaggle.com . In this dataset includes many attributes and elements .The elements are in dataset like School sup, Family sup, Activities, Free time, Go out, Health, attendance Sex, Age, Address, Family size, school, P status, M education, F education, M job, F job, Reason, Guardians, Travel time, Study time, Failures. Where we take the information of greater than four hundred college students in that way. Data is include within side the shape of binary and the numeric type.
- 2) *Data pre-processing*: Data pre-processing is a unique sort of processing in which it's far completed on uncooked information to put together it for some other processing procedure. For information mining procedure it's far used as a initial step. These strategies is used for education gadget getting to know and AI fashions and for going for walks inferences towards them.

3) *Data cleansing*: Data cleansing could be very critical procedure due to the fact the smooth information mitigate information mining and assist in growing a hit strategic choice and predictions. The information cleansing includes smoothing the noisy information and lacking information. The noisy information may be smoothen with the aid of using the use of a few strategies like binning technique, analyzing the outliers information and regressing Here we pre-processed and wiped clean information the use of python libraries like numpy and pandas. In our dataset there are numerous inappropriate information and information is imbalanced which influences our version accuracy, so for greater accuracy we wiped clean our information for cleansing we use tree primarily based totally version and we also are seeking to do away with outliers. After all this procedure we in the end balanced and wiped clean our information that is see in table.

### VIII. SYSTEM ARCHITECTURE

A system architecture is the embodiment of concepts and the distribution of correspondences between the functions of things or information and formal elements. The purpose of system architecture is to define a comprehensive solution primarily based totally on principles, ideas, and residences logically associated with and constant with every other. In it the first part is dataset sort listing, after that we do data pre- processing using python libraries after the pre-processing the data is split into two parts Training data and Testing data. The training data then goes for feature extraction in this data processed in the algorithms like Random forest , ANN , XG Boost and the Testing data goes to the feature extraction part and then it move forward for the feature matching where the data sent to the algorithms for the processing after the compression of Training and Testing data it sent to the prediction model and after that it gives accurate result.

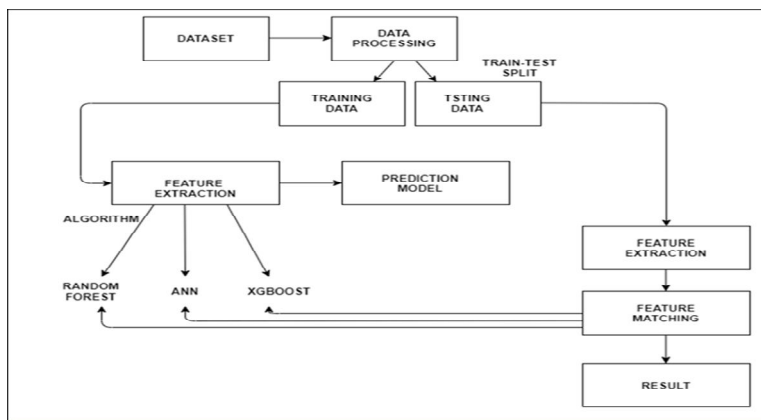


Fig 3 System Achitecture

### IX. EXISTING SYSTEM

The existing system is created using algorithms like Random Forest, Decision Tree, Linear Regression. They use only one algorithm for the result and not compare with all three algorithms result, this result is only based on the single algorithm because of that the existing system gives less accurate result. They only used grade system and their prediction factors are fixed because of this the accuracy of predicted result differs. In the existing the performance is very less.

### X. PROPOSED SYSTEM

The proposed system predicts the result of the students based on their current and previous performance. In this system we are using Random Forest, XG Boost, ANN for more accuracy in result and for the better performance. In our system first we go through the admin login page where it is pass through the authentication process for validation. After that the system split in to two part teacher login and student login process after the success logged in by the teacher or student in the application the data given by the student and the teacher is collected in the data set after that the data goes for the data pre-processing where the data gets cleaned than the data sent to the training and testing process where the data split into two parts data training and data testing where the data trained and tested by the algorithms after this process the data sent to the feature extraction model and after that this data sent to the prediction model and feature extraction after that the data sent to the algorithm for the prediction and compression ( Random Forest, XG Boost, ANN) after this process the predicted and compared shown in the form of result. In our proposed system the accuracy of the predicted result is more than the existing system and it gives better performance.

**XI. RESULT**

**A. Existing System Accuracy Result**

Table 1

Dataset Size	9255
Random Forest Accuracy	70.04
Decision Tree Accuracy	31.52
Linear Regression Accuracy	34.70

In the above existing system they used Linear Regression, Decision Tree & Random Forest. Result shows less accuracy in Linear Regression and Decision Tree as compared to Random Forest. So In our proposed system we selected the Random Forest algorithm to work for increasing accuracy and it gives us 80.29% accuracy. Also we tried two more algorithms ANN & XG-boost gives accuracy 73.11% & 74.38% respectively. For finding accuracy we used the Mean Absolute Error function. So according to our research Random Forest will be more suitable for this project. Also mentioned accuracy in the following table.

**B. Proposed System Accuracy Result**

Table 2

Dataset Size	400
Random Forest Accuracy	80.29
ANN Accuracy	73.11
XG-Boost Accuracy	74.38

**XII.FLOWCHART**

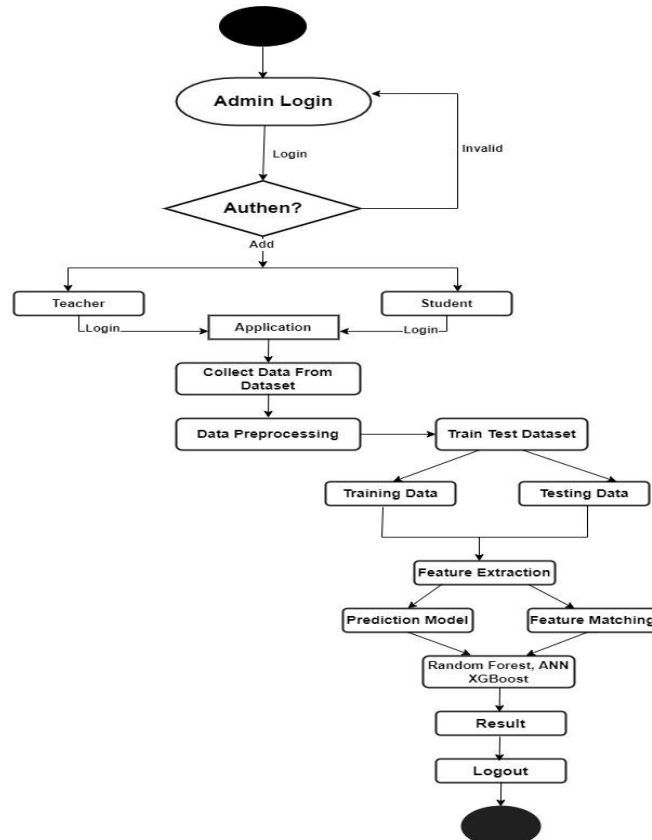


Fig 4 Flowchart

### XIII. CONCLUSION

Predicting a student's educational overall performance is fairly useful to help the lecturers and novices to decorate their learning training approach schematically. This paper analyzed the student's educational ordinary overall performance with numerous device learning algorithms. To treatment the trouble of identifying the students who have a awful educational ordinary overall performance, three kind models have been built to are watching for the general overall performance of the students. Three device learning techniques, Random forest, ANN and XG Boost, have been used. This set of policies gives better ordinary overall performance for predicting the student's educational ordinary overall performance. In conclusion, student's educational dataset assessment on predicting student's educational ordinary overall performance has stimulated us to carry out further research to be achieved in our domain. It will help the educational system to track the student's educational ordinary overall performance in a based way.

### REFERENCES

- [1] Pushpa S.K, Manjunath T.N, "Class end result prediction the use of system learning", International Conference on Smart Technology for Smart Nation, 2017 p1208-1212.
- [2] M.Ramaswami and R.Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, No. 1, January 2010.
- [3] Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt- Thieme, "Improving Academic Performance Prediction with the aid of using Dealing with Class Imbalance", 2009 Ninth International Conference on Intelligent Systems Design and Applications.
- [4] L.Arockiam, S.Charles, I.Carol, P.Bastin Thiyagaraj, S. Yosuva, V. Arulkumar, "Deriving Association among Urban and Rural Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 687-690
- [5] P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12
- [6] M. O. Pedro, R. Baker, A. Bowers, and N. Heffernan, "Predicting university enrollment from scholar interplay with an sensible tutoring gadget in center school," in Educational Data Mining 2013.
- [7] A.Vihavainen, M. Luukkainen, and J. Kurhila, "Using students' programming conduct to are expecting achievement in an introductory arithmetic course," in Educational Data Mining 2013.
- [8] J. Bayer, H. Bydzovsk'a, J. G'eryk, T. Obsivac, and L. Popelinsky, "Predicting drop-out from social behaviour of students." International Educational Data Mining Society, 2012.
- [9] B. K. Baradwaj and S. Pal, "Mining instructional records to investigate students' overall performance arXiv preprint arXiv: 1201 3417, 2012.
- [10] Q. A. Al-Radaideh, E. M. Al-Shawakfa, and M. I. Al- Najjar, "Mining scholar records the use of choice trees," in International Arab Conference on Information Technology (ACIT 2006), Yarmouk University, Jordan, 2006.
- [11] Premalatha, K. Course and program outcomes assessment methods in outcome-based education: A review. J. Educ. 2019, 199
- [12] [12] Rajak, A.; Shrivastava, A.K.; Shrivastava, D.P. Automating outcome based education for the attainment of course and program outcomes. In Proceedings of the 2018 Fifth HCT Information Technology Trends (ITT), Dubai, UAE, 28–29 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 373–376.
- [13] Zohair, L.M.A. Prediction of student's performance by modelling small dataset size. Int. J. Educ. Technol. High. Educ. 2019.
- [14] Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán- Domínguez, A. Analyzing and predicting students' performance by means of machine learning: A review. Appl. Sci. 2020, 10, 1042.
- [15] Yadav, A.; Alexander, V.; Mehta, S. Case-based Instruction in Undergraduate Engineering: Does Student Confidence Predict Learning. Int. J. Eng. Educ. 2019, 35, 25–34.





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)