iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.46456

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 VIII Aug 2022- Available at www.ijraset.com

Fake Product Review Monitoring System

Deekshitha K UG[1], Deepa R[2], Ms. P. Padma[3]
[1]*Student, Information Technology, Sri Sairam Engineering CollegeAutonomous institutions Chennai*
[2]*UG Student,* [3]*Guide, Information Technology, Sri Sairam Engineering CollegeAutonomous institutions Chennai*

Abstract: ***In the current scenario, the data on the web is growing to a larger extent. Social Media is generating a large amount of data such as reviews, comments and customer's opinions on a daily basis. This huge amount of user generated data is worthless unless some miningt e c h n i q u e s are applied to it. Nowadays, there are several people using social media reviews to order anything through online. Online spam detection is oneof the herculean problems since there are many faux or fake reviews that are created by organizations or by the people themselvesfor various purposes. Such organizations tend to write fake reviews to mislead readersor automated detection systems by promoting or demoting the targeted productservices. Fake reviews detection has recentlybecome a limelight that's capturing attention. Fake reviews are generated intentionally to mislead readers to believe false data that makes it tough and non-trivialto discover supported content. Hence, it is highly necessary to create a monitoring system which thoroughly checks for fake reviews among various product websites andremoves them promptly.***
Keywords: Fake reviews, spam detection, opinion mining.

I. INTRODUCTION

In current trends, e-commerce has been one of the very happening fields. In General, it provides facility for customers to write reviews concerned with its service. The existence of these reviews can be used as a source of information. Before purchasing anything, it is a normal human tendency to surf reviews on that product. Based on reviews, customers can compare different brands and can finalize a product of their interest. These online reviews can change the opinion of a customer about the product. If these reviews are true, then this can help the users to select proper product that satisfies their requirements. On the other hand, if the reviews are manipulated or unreal, then there are chances that it can mislead users. This resulted in the development of a system which detects fake reviews for a product by using the text and rating property from a review. The honesty value and measure of a fake review is often measured by utilizing the data mining techniques. An algorithm could very well be used for tracking customer reviews. Fake reviews include dishonest or inaccurate information. They are used to misinform consumers, so they make wrong purchase decisions, thus affecting the revenues for products. Spam product reviews are three types: Deceitful reviews, Reviews of a specific brand and non-reviews. 1) Deceitful (fake) reviews of products that are written to mislead customers. They include undeserving positive reviews to promote the online trade of specific products and negative reviews to defame worthy products. This type of spam product review is called hyperactive spam products reviews. 2) Reviews of a brand only: these opinions target the manufacturer brands instead of the product itself. 3) Non- reviews, which have two sub-kinds: (a) announcements and (b) unrelated reviews that contain no opinions, such as interrogations, responses or undefined text

II. LITERATURE SURVEY

Review spam is strenuous to detect unless read manually. Here are some of the worksproposed and implemented. Paper [1] proposes behavioral approach to detect review spammers who manipulate the ratings on some target products wherein an aggregated behavior scoring methods for rank reviewers is derived. Paper [2] proposes that spotting theindividual fake reviews was quite grueling unlike spotting the groups which was comparatively easier. One frequent item setmining (FIM) method is used to analyze thedataset. In paper [3], the approach was to detect the fake review by identifying the IPaddress of the user ID that is recorded multiple times. Paper [4] used linguisticfeatures like unigram presence, unigram frequency, bigram presence, bigramfrequency and review length to build a model and find fake reviews. Although, the main problem is data scarcity, and it requires both linguistic features and behavioral features. Paper proposes new features like review density, semantic, and emotion and givesthe model and algorithm to construct each of these features. Although, it is not a good metric, and the reduction is not substantial. In paper [6], scraping processing is used tobuild the data set from yelp and then Fake Feature Framework for organizing theextraction and characterization of featuresin fake detection. Their framework is composed of two main types of features: review centric and user centric. Review centric features are only related to the text of the review and User centric features show how the user behaves within the site.

©IJRASET: All Rights are Reserved

III. PROPOSED SYSTEM

The system proposed will include methodslike collection of datasets from Kaggle and preprocessing them.

A. *Pre-Processing*

The term Pre-processing the data is definedas the process of converting a data into an understandable format by cleaning it and preparing the text for classification. Texts from online contain usually lots of noise and uninformative parts such as scripts and advertisements. Processing includes certain steps such as online text cleaning, white space removal, expanding abbreviation, Stemming, stop words removal and feature selection. These might reduce the noise in the text which helps to speed up the performance of the classifier. Before carrying out the transformation and vectorization of the sentences of the reviews, pre-processing steps were used to clean the data and remove noise. The goal of text pre- processing is to convert the textsof the reviews to a form that deep learning algorithms can understand and analyze. The pre-processing steps are as follows: a) Removing punctuation: deleting punctuation marks from the reviews. b) Removing stop words: This process cleans articles from the text; for example, 'the', 'a', '' words are removed from text. c)Stripping useless words and characters from the dataset. d) Word stemming: converting each word of a sentence into itsroot; for instance, 'undesired' becomes 'desire' e) Tokenizing: splitting whole sentences in the text into separate words, keywords, phrases, and pieces of tokens. f)Padding sequences: using deep learning neural networks to ensure that the inputdata have equal sequence length. However, we implemented a pre-padding method to add zeros to the beginnings of the vector representation.

Understanding deviation of ratings: -

- The ratings or reviews which areshowing a trend of continuous growth but suddenly shows negativity is simply displaying a deviation from the normal ratings.

Sentiment analysis of the product review: -

It is necessary for the system to understand whether the review is positive or negative, which further helps to understand the deviation from either the positivity or the negativity in the reviews. The analysis willhelp us to understand the overall aspect of the products so that few spam reviews doesn't affect the overall statistics ofproducts.

- The posted reviews will undergothe process of sentiment analysis, IP address track, and its deviation from overall reviews. In-case of miscalculations, reviews will be analyzed and detected.

Web Scripting is an automatic method to obtain large amounts of data fromwebsites. Most of this data is unstructured data in an HTML format which is thenconverted into structured data in a spreadsheet or a database so that it can be used in various applications. This large amounts of datafrom a website are used to train analgorithm. Web scraping requires two partsnamely the crawler and the scraper. The crawler is an artificial intelligence algorithm that browses the web to search the data required by following the linksacross the internet. The scraper, on the otherhand, is a specific tool created to extract thedata from the website. The design of the scraper can vary greatly according to the complexity and scope of the project so that it can quickly and accurately extract the data. When a web scraper needs to scrape a site, first it is provided the URLs of the required sites. Then it loads all the HTML code from those sites and a more advancedscraper might even extract all the CSS and JavaScript elements as well. Then the scraper obtains the required data from this HTML code and outputs this data in the format specified by the user. Initially, awebsite is created which contains featured products of famous brands. Users have to login to the website for entering reviews. Once the reviews have been entered,machine learning algorithms will be used for classifying them into fake or real. Fake or spam reviews will be removed thereafterfrom the website. Only thereviews which remain truthful gets published in thisprocess. Thus, the product review website is an efficient and effective way for users to know about the actual information of theproduct.

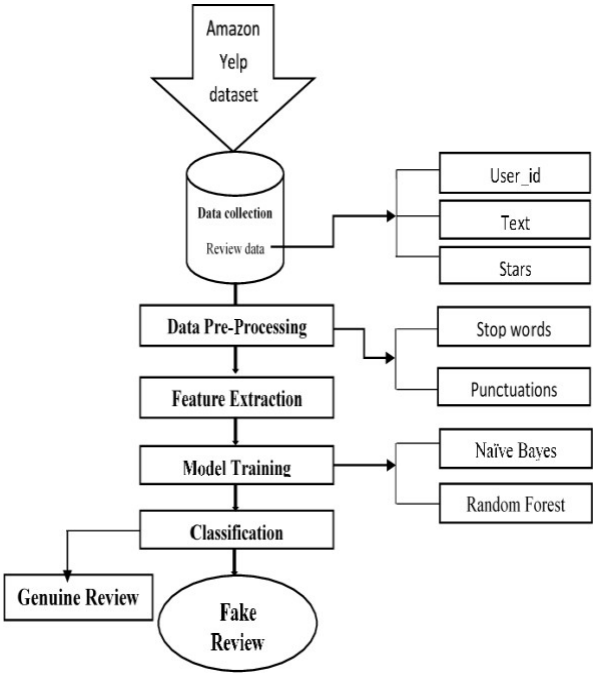
B. *We Are Using Two Machine Learning Algorithms*

1) TF-IDF Vectorizer: TF-IDF Vectorizer (Term Frequency-Inverter Document Frequency): TF-IDF which stands for Term Frequency– Inverse Document Frequency is a statistical method of evaluating the significance of word in given documents. This is very common algorithm totransform text into a meaningful representation of numbers which is used tofit machine algorithm for prediction. TF- IDF vectorizer is defined with parameter (stop words= 'English') which eliminates all the common English words.
2) *Naïve Bayes Classifier:* Naïve Bayes Classifier is one of the simpleand most effective Classification algorithms which helps in

building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts based onthe probability of an object. It is calledBayes because it depends on the principle of Bayes theorem, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability. Naïve Bayes Classifier works on the following steps:

- Convert the given dataset into frequency tables. Generate Likelihood table by finding the probabilities of given features. Now, use Bayes theorem to calculate the posterior probability. Formula: P (c|x) = P(x|c) P(c) / P(x) Referred from Bayes's theorem, in probability theory, a means for revising predictions considering relevant evidence, also known as conditional probability or inverse probability.
- Passive Aggressive Classifier Passive-Aggressive algorithms are calledso because Passive- If the prediction is correct, keep the representation and do not make any interchanges. i.e., the data in the example is not enough to cause anychanges in the representation. Aggressive- If the prediction is incorrect, make interchanges to the representation. i.e., some interchange to the representation may correct it. Understanding the mathematics supportingthis algorithm is not very simple and issupporting the scope of a single article. This section provides just an overview of the algorithm and a simple implementation of it. To learnmore about the mathematics supporting this algorithm.

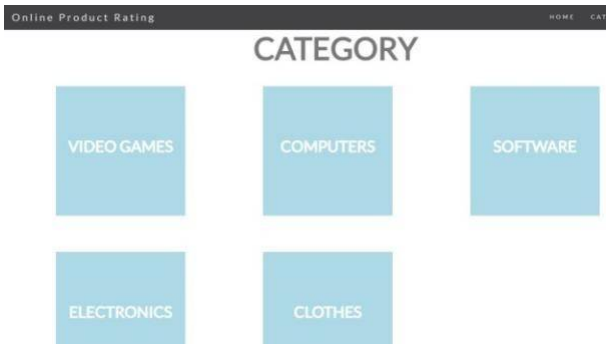
IV. SYSTEM ARCHITECTURE



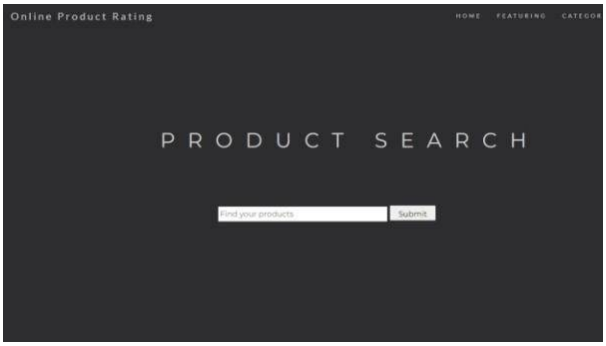
V. RESULTS



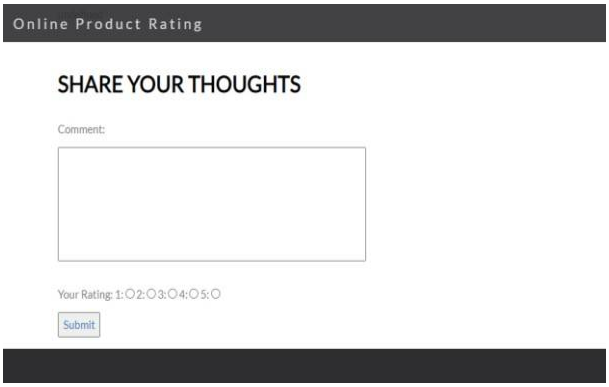
Products Featuring Page



Products Category Page



Products Search Page



Product Review and Rating Page

VI. FUTURE DEVELOPMENTS

For future developments, a web application can be designed which makes theprocess of finding out fake reviews easier. Every user will be given an account through which they can write reviews for various products. The app would automatically filterout fake reviews based on the proposed Machine Learning algorithm. Eventually, customer will get rid of fake reviews presentin online shopping websites.

VII. CONCLUSION

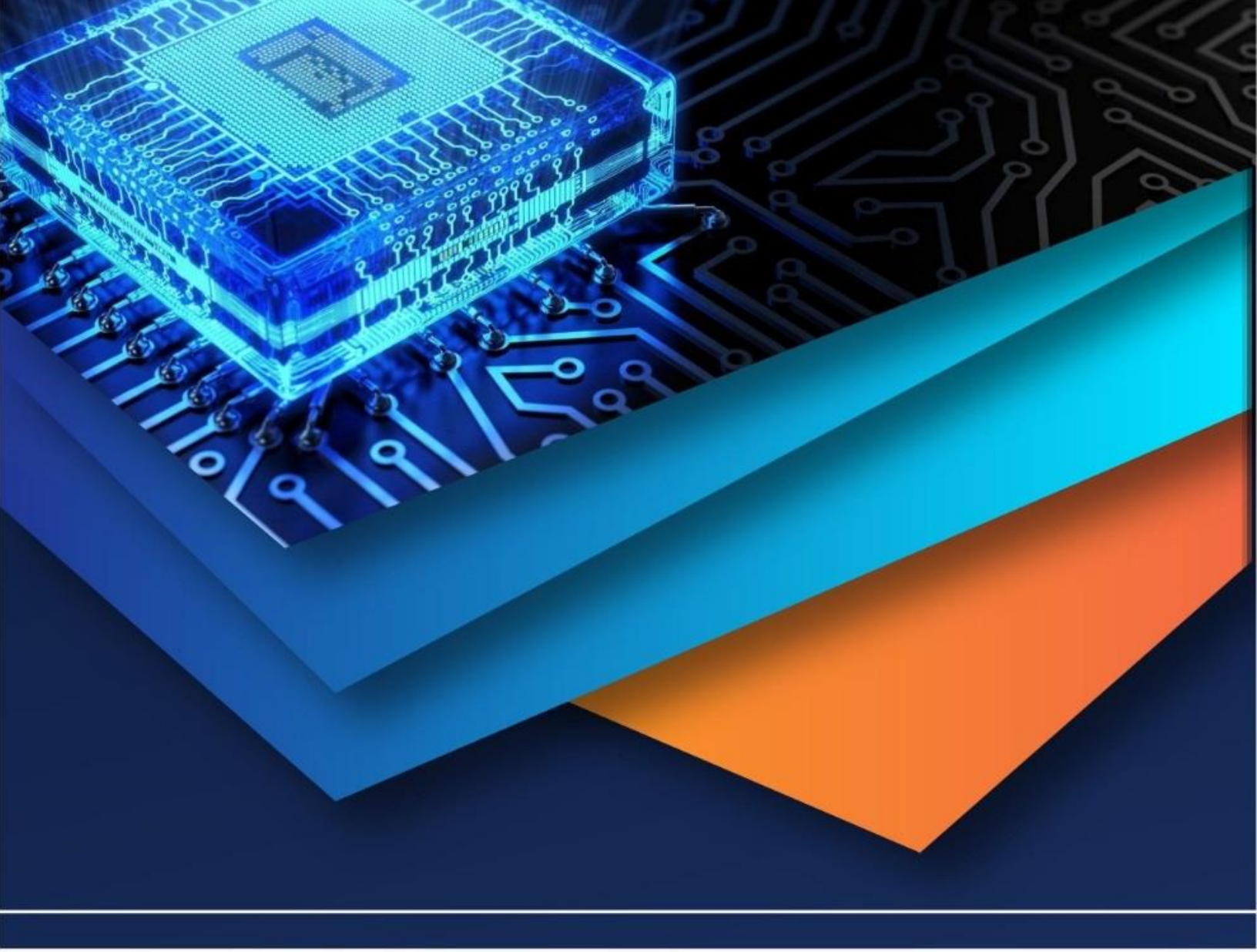
Determining and classifying a review into fake or truthful one is an important and challenging problem. As part of future work, we can incorporate review spammerdetection into the review detection and viceversa. Exploring ways to learn behaviorpatterns related to spamming to improve the accuracy of the current regression model. To evaluate ourproposed methods, that conducts user evaluation on an Amazon dataset containing reviews of different manufactured products.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Gyandeep Dowari, Dibya jyoti Bora, "Fake Product Review Monitoring and Removal using Opinion Mining, IEEE conference publication,2020.
[2] Eka Dyar Wahyuni, Arif Djunaidy, "Fake Review Detection from a Product Review Using Modified Method ofIterativeComputation Framework", MATEC Web of conferences, 2016.
[3] Abishek Pund, Ramteke Sanchit, Shinde Shailesh, "Fake product review monitoring & removal and sentiment analysis of genuine reviews", International Journal of Engineering and Management Research (IJEMR), 2019, Volume 9:Issued
[4] Long- Sheng Chen, Jui-Yu Lin, "A study on Review Manipulation Classification using Decision Tree", Kuala Lumpur, Malaysia, pp 3-5, IEEE conference publication, 2013.
[5] Ivan Tetovo, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization "Ivan Department of Computer Science University of Illinois at Urbana, 2011
[6] N. Jindal and B. Liu, "Opinion spam and analysis," International Conference on Web Search and Data Mining, 2008, pp. 219-230.
[7] R.Narayan,J. Rout and S. Jena, "Review Spam Detection Using Semisupervised Technique", Progress in Intelligent Computing Techniques: Theory,Practice, and Applications, pp. 281-286, 2018.
[8] W. Etaiwi,G. Naymat, "The impact of applying pre-processing steps on review spam detection", The 8th international conference on emerging ubiquitous system and pervasion networks, Elsevier, pp. 273- 279, 2017.
[9] A. Rastogi, M. Mehrotra, "Opinion spamDetection in Online Reviews", Journal of information and Knowledge Management, vol. 16, no. 04, pp. 1-38, 2017.
[10] N. Jindal and B. Liu, "Review spam detection", Proceedings of the 16th internationalconference on World Wide Web - WWW 07 (2007), ACM, pp. 1189–1190, 2007.
[11] Rajashree S. Jadhav, Prof. Deipali V. Gore, "A New Approach for Identifying Manipulated Online Reviews usingDecisionTree ". (IJCSIT) International Journal ofComputer Science and Information Technologies, Vol. 5 (2), pp 1447-1450,2014
[12] Jiawei Zhang, Bowen Dong, Philip S. Yu, "FAKE DETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network" published in August 2019.
[13] Steni Mol T S and Shreeja P Sin, "Fake News Detection on Social Media-A Review" published in April 2020.
[14] Monther Aldwairi and Ali Alwahedin, "Detecting Fake News in Social Media Networks" published in 2018.
[15] Natali Ruchansky, Sungyong Seo and YanLiu, "CSI: A Hybrid Deep Model for FakeNews Detection".
[16] Ray Oshikawa, Jing Qian, William Yang Wang, "A Survey on Natural Language Processing for Fake News Detection" publishedin March 2020.
[17] Shekhar Pandey, Supriya M, Abhilash Shrivastava,"Data Classification using machine learning approach"publishedin January 2018.
[18] Sang-Woon Kim, Joon-Min Gill,"Classification Systems based on TF-IDF and LDA schemes"publishedin August 2019.
[19] Shakib Hakak, Mamoun Alazab, Suleman Khan," An ensemble machine learning approachthrough effective feature extraction to classify fake news"publishedin April 2021.
[20] Azizur Rahman, "Statistics-Based Data Preprocessing Methods and Machine Learning Algorithms for Big Data Analysis" , International Journal of Artificial Intelligence, vol. 17, no. 2, pp. 44-65, 2019.
[21] Ms. Reema Anne Roy, Dr. Sunita R Patil,"Fake Product Monitoring System using Artificial Intelligence"publishedin May 2021.
[22] Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel," Creating and detecting fake reviews of online products"publishedin September 2021.
[23] Jyoti Bist, Neha Hulsurkar, Deepali Narkhede, Shraddha Bhalerao," CommentSentiment Analysis and Fake Product Review Detection" International Research Journal OfEngineering and Technology (IRJET) Volume:07 Issue: 05 May 2020.
[24] C. Reddineelima, V. Haritha, U. Dinesh, B. Kalpana," Spotting and Removing Fake Product Review in Consumer Rating Reviews" International Research Journal Of Engineering and Technology (IRJET) Volume: 06 Issue: 03March 2019.
[25] Ching-Lung Fan, " Evaluation of Classification for Project Features with MachineLearning Algorithms" publishedin February 2022.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)