iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5472

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: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Data Analysis in Social Media

Vishal Khedkar[1], Harshil T. Kanakia[2]
[1, 2]*Dept. of Master of Computer Application, Sardar Patel Institute of Technology*

Abstract: ***Over the last few years, the use of social media in our everyday lives has skyrocketed. The availability of good hardware devices at affordable prices and an increase in connectivity with lower data rates at most public places has made people increasingly use social media sites.***
Social networking sites like Twitter, LinkedIn, Facebook, Instagram, etc. are continually developing their products to make them addictive, engaging and convenient to use for people of any age or linguistic group. All of this results in a huge chunk of data that people around the world are generating at any moment. The huge size data is one of these top-scale corporations' most valuable assets.
About 2.5 quintillion bytes of data was generated every day from the usage of social media sites in 2018, and it is increasing exponentially as a result of the Internet of Things.
Also, 90 percent of this data was produced only in the last 3 years. Organizations make use of this data by analysing it using different methods and generating insights into context from the same. The observations arising from this are often used for different purposes.
Keywords: Data Analytics, Social Media, User Behaviour, Prediction, Big Data

I. INTRODUCTION

The rise in use of social media has enabled people to reach out their voices on various social media platforms. These opinions can be anything from comments, textual posts, photos, post sharing etc. This leaves companies with many opportunities to take better decisions and make use of those decisions to create better marketing and selling strategies, considering the amount of data generated by individuals.

When Google launched Google Nexus 4 in mid-2012, many users expressed their views stating how they like this phone's idea at a reasonable price. Such kind of positive response helps the marketing team identify potential customers willing to purchase their products.

Even people from other departments such as engineering and manufacturing may recognize what flaws or drawbacks, they find about their proposed product, are. Using this knowledge, they can create better and better versions of existing goods, thereby making the company a benefit. Analysing the opinions of consumers on social media can provide good support when making such business decisions.

It's generally been found that there are three main types of social media responses. The first is the Positive group; their input to some individual is positive. The second is the Negative group, who have a negative outlook to the entity and the last is the Neutral group, which has a neutral approach to the entity.

II. EXISTING METHODOLOGIES

There are usually two major types of methods for the analysis: 1. Cognitive method. 2. Lexical–based method [1]. The cognitive process model is provided with an input of known properties, extracted from labelled data. These data are analysed, and unlabelled new data are forecast. The relation between the features of the text segment is extracted when the data type is simply text. Naïve Base Classifier [2], Support Vector Machines, Maximum Entropy Classifier, and Extreme Learning Machine are some examples of the cognitive method. To produce accurate classification, the learning-based methodologies require a large dataset of labelled data. They look for an emotional meaning in the text while dealing with Lexical –- based approaches. All the measured sets of emotions are aggregated to establish the highest-dominant polarity.

One of the benefits of lexical-based approaches is that they are applicable to various types of datasets; it is also not a prerequisite to mark i.e. label the tasks. Even though lexical–based approaches tend to be easy to implement, they have their own limitations. It is not possible to construct a single lexical dictionary and extend it to various applications. Because of this, each new usage needs clean samples, which need to be manually generated. Another drawback of lexical approaches is that they cannot have a detailed emotional description.

©IJRASET: All Rights are Reserved

III. CONDUCTED SURVEYS

Over the years, many independently employed individuals and organisations have performed surveys and research. The findings of these surveys have been used in various fields, such as education, public safety, to study the effects of demonetization, etc.

A. Twitter Survey on Public Safety

This survey was conducted by Dennis Thom. According to Dennis, in the modern world, social media is used not for public protection but for commercial benefits. The research was carried out on Twitter during the year 2013, when flooding impacted Germany. A system was developed, called the 'Scatter Blogs System', to apply numerous other techniques. The Scatter Blogs System evaluated the visual representation of the data from social media. Visual analytics was used along with situation evaluation. There was a wide variety of methods and techniques available for the researchers to use their ongoing study. Natural language processing along with LDA (Latent Dirichlet Allocation) subject modelling was made use of during the visualization process.

B. Demonetization

A study by an independent organisation showed that Indian citizens had a varied response to the Indian government's demonetization of Rs. 500 and Rs. 1000 in 2016 [3]. Shortly after the decision was announced, people took to the social media to express their views about the demonetization decision. The responses were of a mixed nature and there was no strong positive or negative response which could be said to be dominant. The demonetization hashtag was the trendiest for both the same day and the next day.

IV. THE PROCESS OF ANALYSING DATA

Analytics on social media involves a three-stage process: capture, understand and present [4] (See Fig. 1). The capture stage includes the collection of relevant social media data by tracking or "listening" to various social media outlets, archiving relevant data and extracting relevant information. This method can be done either by a organization itself, or by a third party vendor. Not all reported data would be useful. The understand stage selects appropriate data for modelling, eliminates noisy, low quality data and uses various advanced data analytics approaches to analyse and obtain insights from the data retained

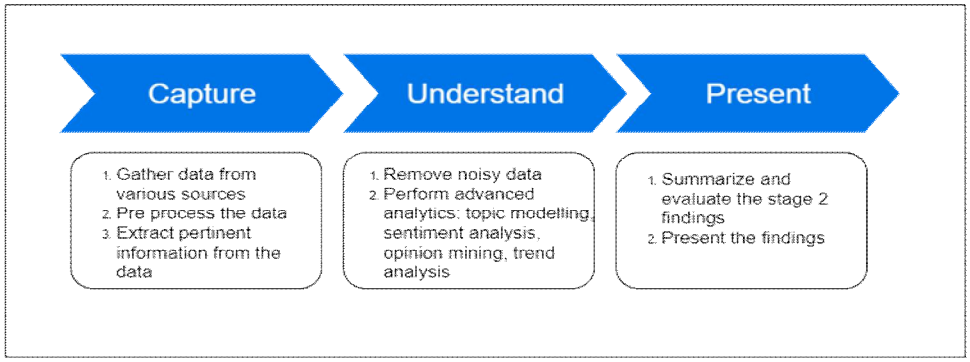


Fig. 1 Analytics Process.

A. Capture

The capture stage allows a business engaged in social media analytics to identify conversations related to its activities and interests on social media platforms. This is achieved by gathering large quantities of relevant data using news feeds, APIs, or crawling through hundreds or thousands of social media outlets. The capture process covers mainstream networks such as Facebook, Twitter, LinkedIn, YouTube, Pinterest, Google+, Tumblr, Foursquare and so on, as well as smaller, more specialized outlets such as Internet forums, blogs and microblogs, wikis, news sites, photo sharing sites, podcasts, and social bookmarking sites

B. Understand

If a company collects the conversations related to its goods and activities, it must then analyse their context and create valuable indicators for decision taking. This is the 'understand' stage. Since the capture stage gathers data from multiple users and sources, a significant amount may be noisy and will need to be extracted before any substantive analysis is performed. For this cleaning function, plain, rule-based text classifiers or more sophisticated classifiers trained on labelled data can be used. Significance evaluation from the cleaned data may include various statistical methods and other techniques derived from text and data mining, natural language processing, machine translation and network analysis.

©IJRASET: All Rights are Reserved

C. Present

The present stage is the final stage in the cycle of social media analytics. The findings from different analytics will be compiled, analysed and presented in an easy-to-understand format to the users. Different visualization techniques can be employed to provide valuable details. The visual dashboard is one of the most widely used GUI styles, which aggregates and shows information from various sources.

V. THE TECHNIQUES OF ANALYSING DATA

Social media analytics is a growing domain encompassing a number of specific areas of modelling and analytical techniques. Below, we highlight the most influential in identifying, evaluating, and presenting vast volumes of social media data. These techniques can support social media analytics at various stages. Evaluating the feelings and evaluating the pattern mainly help the understand stage. Topic modelling and analysis of social networks primarily have applications in the understand stage but can also help the capture and present stages. Visual analytics helps the understand and the present stages [4].

A. Sentiment Analysis (Opinion Mining)

It is the central methodology behind many applications of trend analysis and monitoring of social media networks. It leverages computational linguistics, natural language processing, and other text analytics approaches to automatically extract user feelings or thoughts from text sources at any level of granularity (words or phrases up to entire documents). Such subjective information obtained about individuals, products and services or other entities supports various tasks, including predicting stock market movements, determining market trends, analysing defects in or probable improvements to products and managing emergencies.

B. Topic Modelling

It is used to sift the dominant themes (topics) across broad bodies of captured text. The discovered themes can be used to provide clear labels to further explore the selection of text, or to create successful navigational interfaces. Themes uncovered by topic modelling can also be used to find out user desires, detecting emerging themes in forums or social media posts, or summing up sections (or all) of a set of text. Recent developments in topic modelling also allow for the use of these algorithms with streaming data. Twitter and other continuous feeds of the data, making the technique an ever more effective analytical method. Theme modelling uses a number of sophisticated machine learning and statistics techniques.

C. Trend Analysis

Trend analysis is used on the basis of empirical data gathered over time to define and forecast potential events and behaviours. Trend analysis applications include forecasting company or revenue growth, predicting ad campaign effectiveness, staying ahead of changes in customer preferences, forecasting stock market movements, etc. Trend analysis is based on long-standing statistical methods such as analysis of time series or regression and other more recent modelling techniques such as neural networks and vector machines.

D. Visual Analytics

Visual analytics is "analytical thinking science enabled by immersive digital interfaces" Initially spurred on by U.S. defence needs, visualization works across multiple application areas to enable integration, exploration, discovery, and validation of data insights usually voluminous and distributed through different sources. Visual analytics includes a variety of tasks, from data collection to decision taking that is assisted by data. Although visual analytics are underpinned by many statistical approaches (such as reducing high-dimensional data to smaller and very small dimensions), the capacity of humans to interpret patterns and draw conclusions are also key factors. Indeed, when there is a flood of knowledge that needs to be done rapidly, this combination of computer and human power is crucial, both in decision making and in being able to articulate and justify it.

E. Social Network Analysis

Analysis of social networks is used to evaluate a social network graph to understand its underlying structure, interactions and theoretical properties as well as to define the relative importance of various nodes within the network. A graph of a social network consists of nodes (users) and associated relationships (by edges). User actions that directly connect two people (such as accepting another user as a "friend") typically detect the relationships, although they may be inferred from indirect behaviours that create relationships, such as voting, tagging, or commenting. Study of the social network uses a number of techniques applicable to the interpretation of graph mathematical structures. These range from simpler methods (such as counting the number of edges a node has or calculating path lengths) to more sophisticated methods that calculate individual vectors (similar to Google's PageRank algorithm) to classify main nodes within a network.

©IJRASET: All Rights are Reserved

VI. BUSINESS VALUE OF ANALYTICS

As we discussed the different strategies that help social media analytics, they can be put to a number of business uses. Here, we look in more detail at certain uses.

A. Product Design and Development

Trend analysis will help bring any shifts in preferences, attitudes and other emotions that may impact product design and development to the forefront. These tools can allow adding or modifying features and can help to establish ample lead time for product development.

Social media analytics can also promote product innovation by capturing conversations involving either of two groups and understanding them. At the one hand, it can expose valuable insights from the most “fanatical” customers of a company. Conversations with "ordinary" consumers on the other hand, can also contribute to product changes.

B. Product Manufacturing

A company can also use social media analytics to discover that another company it competes with (or maybe doesn't) is having problems with a supplier, which can be useful in helping it predict and prevent the same problem, even though it doesn't yet encounter that. Social media may provide advance notice when circumstances are less predictable, like political tensions abroad that could interfere with the flow of metals, minerals or other critical supplies for manufacturing [4].

C. Product Disposal

A customer can face decisions about how to dispose of a product near the end of life, and what to substitute it with. For a variety of people, being able to dispose of a product (possibly a computer) in an environmentally friendly manner may affect their overall opinion of a business and its goods. Therefore, it is important to make this convenient and to make sure customers are conscious that it is convenient.\\ Social media analytics can monitor and companies themselves can participate in dialogue on disposal. Companies monitoring these social media interactions will of course also deduce the disposal can be followed by a purchase of a new item and using it in their ads.

VII. FAKE NEWS AND FACT CHECKING

In recent years, with the exponential growth of social media, digital content and internet traffic, a new issue of fake news has arisen. From Wikipedia, “Fake news is a news that contains deliberate lies or hoaxes or incorrect information that gets distributed via conventional news media (print and broadcast) or social media online to fool people into believing it to be real news.” Fake news is typically written and published with the intention of misleading in order to harm an organization, institution, or person and/or benefit financially or politically, often using sensationalist, deceptive, or outright fabricated headlines to increase readership. Similarly, the click-bait reports and headlines from this operation help publishers gain advertisement revenue.

One of the main applications of social media analytics has been to fact checking viral messages, news or posts. Given the state of social media use in India, a majority of India's population is actively involved on social media platforms. While there are users verifying the truth behind most of the posts found on these sites, there is also a community of people interacting with posts that are in reality not right, and are spread intentionally to mislead people into believing certain claims that might politically or economically benefit a particular organization or an entity.

In the years 2018-2019, social media giants like Facebook, Twitter and YouTube gave top priority to fact checking. Political parties and their staff underlined the use of social media to support and promote their respective political parties during the 2014 Indian general elections. However, there have also been cases where posts with false data have been uploaded (fake news), resulting in the spread of inaccurate information to the public and their manipulation for political advantage.

In 2019, Facebook agreed to extend its fact-checking system, typically achieved with the assistance of third-party organizations [5]. Facebook collaborated with Factly, Today Group and Fact Crescendo to tackle the problem of spreading false news on their website. Facebook has equipped tools to review posts that will help them identify and take action against misinformation of any sort. In total, Facebook has collaborated with seven different partners, covering six languages, and will check the posts for accuracy.

YouTube, the largest video streaming site, has also prepared to battle the war with fake, misguiding, incorrect videos that users watch. YouTube built a new tool in India, 'fact-check,' a pop-up that will appear whenever a search result is shown on the screen which could offer misinformation [6]. Disclaimers will be created whenever the platform queries sensitive topics. This tool also helped combat fake information spread during the COVID-19 pandemic. The figure 'Fig. 2' shows how this tool works. A dialog box is shown which states that ibuprofen has no adverse effects on patients being treated with it for COVID-19, as told in fake videos and posts.

©IJRASET: All Rights are Reserved

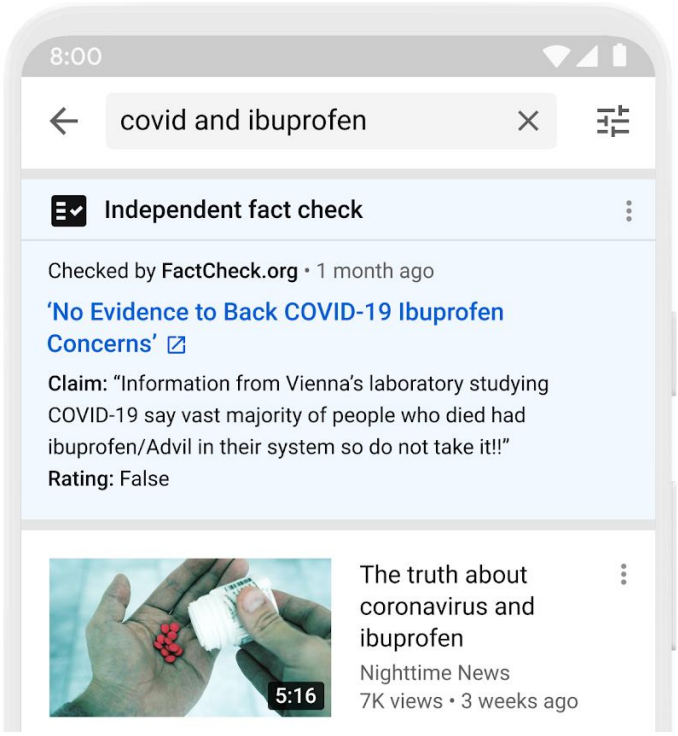


Fig. 2 A screenshot of the YouTube app showing fact check alert for the related search term.

VIII. CONCLUSIONS

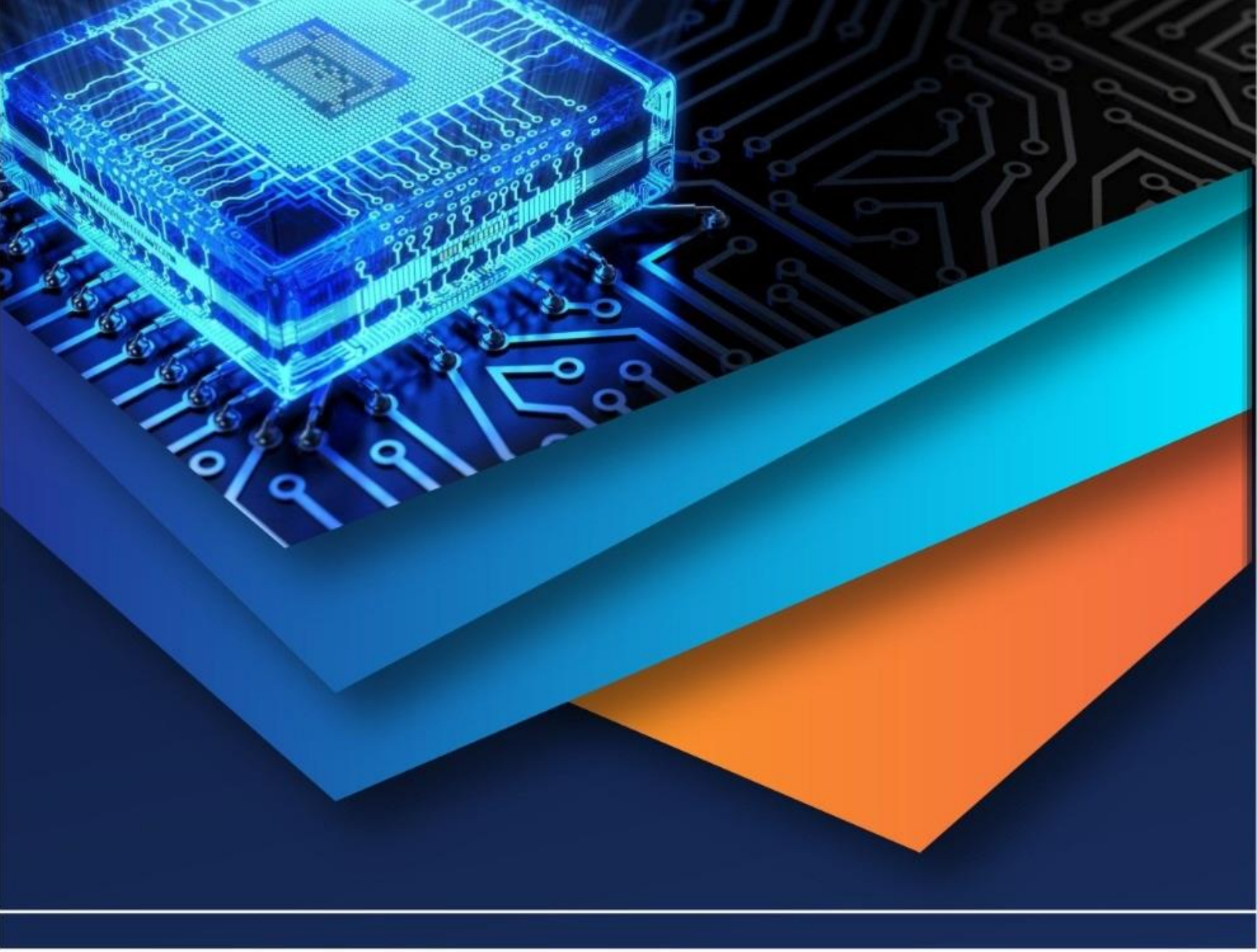
For any organization social media can be powerful as a tool. Using social media can increase an organization's visibility, strengthen partnerships, strengthen customer engagement and provide a forum for users to provide their input. However, companies should develop an engaging plan to make the best use of it and adopt the best practices for the plan. Social networking provides new opportunities, but it's also important to devise the strategy and metrics of success. Organizations that have been able to use social media successfully have started experiments with new technologies to better understand their consumers.

It is also necessary to develop systems that combat the negative aspects of the growing data generated by social media usage, like we discussed about fact checking tools. If not kept in check, it could lead to mass hysteria, political discomforts or even economic instability.

REFERENCES

[1] Zhang, Lei & Ghosh, Riddhiman & Dekhil, Mohamed & Hsu, Meichun & Liu, Bing. (2011). "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis".
[2] L. Zhang, H. Yuan and R. Y. K. Lau, "Predicting and Visualizing Consumer Sentiments in Online Social Media", 2016 IEEE 13th International Conference on e-Business Engineering (ICEBE), Macau, 2016, pp. 92-99, doi: 10.1109/ICEBE.2016.025.
[3] P. V. Paul, K. Monica and M. Trishanka, "A survey on big data analytics using social media data," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, 2017, pp. 1-4, doi: 10.1109/IPACT.2017.8245092.
[4] Fan, Weiguo & Gordon, Michael. (2014). "The Power of Social Media Analytics". Communications of the ACM. 57. 74-81. 10.1145/2602574.
[5] https://economictimes.indiatimes.com/tech/internet/facebook-expanding-fact-checking-in-india-before-election/articleshow/67940960.cms
[6] https://youtube.googleblog.com/2020/04/expanding-fact-checks-on-youtube-to-united-states.html.

©IJRASET: All Rights are Reserved



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)