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.2022.42569

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

Sentimental Analysis on Live Twitter Data

Tanai Ranpise[1], Divya Mavkar[2], Sharang Sarda[3], Rushikesh Pawar[4]
[1, 2, 3, 4]*Tssms Bhivrabai Sawant College of Engineering and Research, Pune*

Abstract: ***The entire world is transforming quickly under the present innovations. The Internet has become a basic requirement for everybody with the Web being utilized in every field. With the rapid increase in social network applications, people are using these platforms to voice them their opinions with regard to daily issues. Gathering and analyzing peoples' reactions toward buying a product, public services, and so on are vital. Sentiment analysis (or opinion mining) is a common dialogue preparing task that aims to discover the sentiments behind opinions in texts on varying subjects.***
In recent years, researchers in the field of sentiment analysis have been concerned with analyzing opinions on different topics such as movies, commercial products, and daily societal issues. Twitter is an enormously popular microblog on which clients may voice their opinions. Opinion investigation of Twitter data is a field that has been given much attention over the last decade and involves dissecting "tweets" (comments) and the content of these expressions. As such, this paper explores the various sentiment analysis applied to Twitter data and their outcomes.

I. INTRODUCTION

In the past few years, there has been a huge growth in the use of microblogging platforms such as Twitter. Spurred by that growth, companies and media organizations are increasingly seeking ways to mine Twitter for information about what people think and feel about their products and services.

While there has been a fair amount of research on how sentiments are expressed in genres such as online reviews and news articles, how sentiments are expressed given the informal language and message-length constraints of microblogging has been much less studied. Features such as automatic part-of-speech tags and resources such as sentiment lexicons have proved useful for sentiment analysis in other domains, but will they also prove useful for sentiment analysis in Twitter?

In this paper, we begin to investigate this question.

Another challenge of microblogging is the incredible breadth of topic that is covered. It is not an exaggeration to say that people tweet about anything and everything. Therefore, to be able to build systems to mine Twitter sentiment about any given topic, we need a method for quickly identifying data that can be used for training. In this paper, we explore one method for building such data: using Twitter hashtags (e.g., #bestfeeling, #epicfail, #news) to identify positive, negative, and neutral tweets to use for training three-way sentiment classifiers.

The online medium has become a significant way for people to express their opinions and with social media, there is an abundance of opinion information available.

Using sentiment analysis, the polarity of opinions can be found, such as positive, negative, or neutral by analyzing the text of the opinion. Sentiment analysis has been useful for companies to get their customer's opinions on their products predicting outcomes of elections , and getting opinions from movie reviews. The information gained from sentiment analysis is useful for companies making future decisions. Many traditional approaches in sentiment analysis uses the bag of words method. The bag of words technique does not consider language morphology, and it could incorrectly classify two phrases of having the same meaning because it could have the same bag of words .

The relationship between the collection of words is considered instead of the relationship between individual words . When determining the overall sentiment, the sentiment of each word is determined and combined using a function . Bag of words also ignores word order, which leads to phrases with negation in them to be incorrectly classified. Other techniques discussed in sentiment analysis include Naive Bayes, Maximum Entropy, and Support Vector Machines. In the Literature Survey section, approaches used for sentiment analysis and text classification are summarized.

Sentiment analysis refers to the broad area of natural language processing which deals with the computational study of opinions, sentiments and emotions expressed in text. Sentiment Analysis (SA) or Opinion Mining (OM) aims at learning people's opinions, attitudes and emotions towards an entity. The entity can represent individuals, events or topics.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

A. Sentiment Analysis of Twitter Data

Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau

In the work carried out by the author's Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Ram bow Rebecca Passonneau, they examined sentiment analysis on Twitter data. The contributions of this paper are:

1) To introduce POS-specific prior polarity features.
2) To explore the use of a tree kernel to obviate the need for tedious feature engineering. The new features (in conjunction with previously proposed features) and the tree kernel perform approximately at the same level, both outperforming the state-of-the-art baseline.

They presented results for sentiment analysis on Twitter by use of previously proposed state-of-the-art unigram model as their baseline and report an overall gain of over 4% for two classification tasks: a binary, positive versus negative and a 3-way positive versus negative versus neutral. Here they presented a comprehensive set of experiments for both these tasks on manually annotated data that is a random sample of stream of tweets. They also investigated two kinds of models: tree kernel and feature based models and demonstrated that both these models outperform the unigram baseline. feature-based approach, they did feature analysis which reveals that the most important features are those that combine the prior polarity of words and their parts-of-speech tags. They tentatively concluded that sentiment analysis for Twitter data is not that different from sentiment analysis for other genres.

B. Sentiment Analysis of Twitter Data: A Survey of Techniques

Vishal A. Kharde S.S. Sonawane

This survey focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases. In this paper, they provided a survey and a comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms like Naive Bayes, Max Entropy, and Support Vector Machine, they provided research on twitter data streams. They also have discussed general challenges and applications of Sentiment Analysis on Twitter.

In this paper, they provided a survey and comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics. Research results showed that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases, which require few effort in human-labeled document .They also have studied the effects of various features on classifier. Here they concluded that more the cleaner data, more accurate results can be obtained. Use of bigram model provides better sentiment International Journal of Computer Applications accuracy as compared to other models. they focused on the study of combining machine learning method into opinion lexicon method in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages.

C. Naïve Bayes Algorithm for Sentiment Analysis on Twitter

Agarshana K , Karpan P ,Leo Celestine ,S .Vasantha ,V. Kumar

In this study, a sentiment analysis application for twitter analysis was conducted on 2019 Republic of Indonesia presidential candidates, using the python programming language. There are several steps taken to conduct this sentiment analysis, which is to collect data using libraries in python, text processing, testing training data, and text classification using the Naïve Bayes method. The Naïve Bayes method is used to help classify classes or the level of sentiments of society.

The results of this study found that the value of the positive sentiment polarity of the Jokowi-Ma'ruf Amin pair was 45.45% and a negative value of 54.55%, while the Prabowo-Sandiaga pair received a positive sentiment score of 44.32% and negative 55.68%. Then the combined data was tested from the training data used for each presidential candidate and get an accuracy of 80.90% ≈ 80.1%.

In this study a comparison was carried out using the naïve bayes, svm and K-Nearest Neighbor (K-NN) methods which were tested using RapidMiner by producing a naïve bayes accuracy value of 75.58%, svm accuracy value of 63.99% and K-NN accuracy value of 73.34%. In this study it was found out that out of the algorithms that are K-Nearest Neighbors (K-NN) , SVM and the Naïve Bayes the most useful , accurate and feasible was the Naïve Bayes algorithm

©IJRASET: All Rights are Reserved

D. Analysis of Behavior Extraction on Social Life Issues Using Sentiment Analysis:

1) Review

2) Adnan Beg

The authors in this paper , stated that "At present time most of the world can be found on internet. The popularity of the social networking sites has been growing fast, parallel to emerging technologies, with increase in the number of users to the social networking sites that actively express their opinions on these sites. " .

To forecast the sentiment analysis they have used the data stored on social sites stockpiling. their goal was to retrieve data from social sites and analyze the emotion of a particular person on certain topic. For analysis they have used the data available on Twitter, as Twitter has a large amount of data that people post, which gives the output beyond the polarity but those polarities can be used in product profiling, trend analysis and forecasting. This paper presents an overview of past and current research on twitter sentiment analysis and presents better ideas for future work.

The authors work , concludes that as social media website like twitter is so vast that getting all information is almost infeasible but possible steps can be taken to get most of the useful information from it. This can be achieved through sentiment analysis as there are so many areas in sentiment analysis field like data refinement, topic modeling of sentiments and summarization of sentiment tweets which are still unrecognized and untouched. These areas build strong chances for innovation of new outcomes, techniques and methods if we look forward to explorer these areas.

III. PROBLEM DEFINATION

Sentiment analysis of in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews , documents, web blogs/articles and general phrase level sentiment analysis . These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text. The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised and semi-supervised approaches, and there is a lot of room of improvement. Various researchers testing new features and classification techniques often just compare their results to base-line performance. There is a need of proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications.

IV. METHODOLOGY

A. Naïve Bayes Classification Algorithm

Many language processing tasks are tasks of classification, although luckily our classes are much easier to define than those of Borges. In this classification we present the naive Bayes algorithms classification, demonstrated on an important classification problem: text categorization, the task of classifying an entire text by assigning it a text categorization label drawn from some set of labels. We focus on one common text categorization task, sentiment analysis, the ex-sentiment analysis traction of sentiment, the positive or negative orientation that a writer expresses toward some object. Are view of a movie, book, or product on the web expresses the author's sentiment toward the product, while an editorial or political text expresses sentiment toward a candidate or political action. Automatically extracting consumer sentiment is important for marketing of any sort of product, while measuring public sentiment is important for politics and also for market prediction.

Naive Bayes is a probabilistic classifier, meaning that for a document d, out of all classes c ∈C the classifier returns the class ˆ c which has the maximum posterior probability given the document. We use the hat notation to mean "our estimate of the correct class".

c = erg max P(c |d) where c ∈C

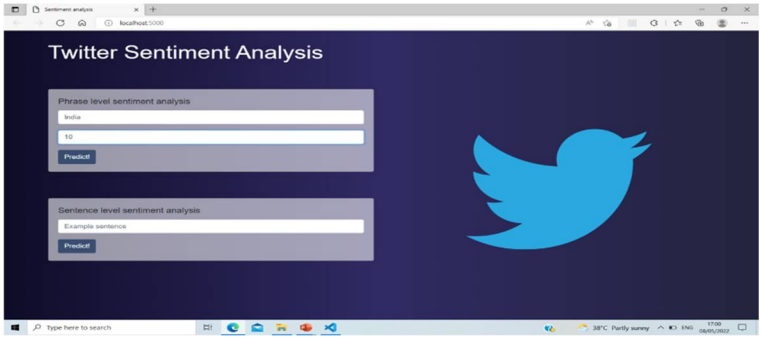
The idea of Bayesian inference has been known since the work of Bayes (1763) Bayesian inference and was first applied to text classification by Most Eller and Wallace(1964).The intuition of Bayesian classification is to use Bayes' rule to transform in to other probabilities that have some useful properties. It gives us a way to break down any conditional probability P(x |y) into three other probabilities:

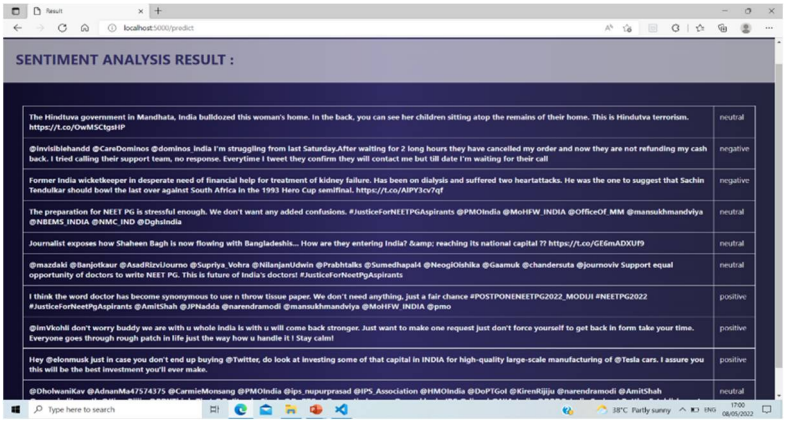
P(x |y) =P(y |x)P(x) /P(y)

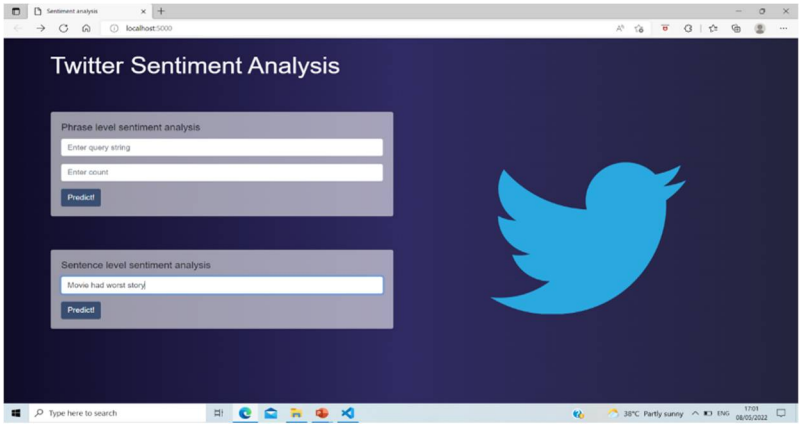
Naive Bayes is a generative model that make the bag of words assumption (position doesn't matter) and the conditional independence assumption (words are conditionally independent of each other given the class).

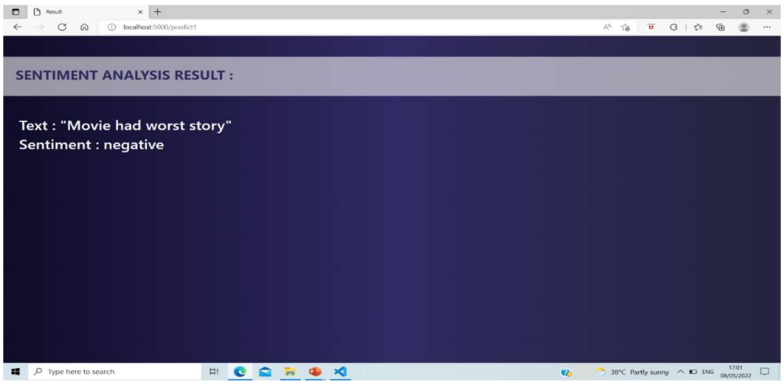
Naive Bayes with binarized features seems to work better for many text classification tasks.

V. IMPLEMENTATION AND RESULTS









VI. CONCLUSION

The task of sentiment analysis, especially in the domain of micro-blogging, is still in the developing stage and far from complete. So we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance. As the sentiment analysis is the hot topic in machine learning right now, we are still far to detect the sentiments of corpus of texts very accurately because of complexity in English language and even more if we consider Devnagri or our Indian Languages like Hindi , Marathi etc.

VII. ACKNOWLEDGMENT

We would like to show gratitude to our project guide prof. Dr. S.N. Gujar who motivated and guided us to make this project successful. Our project guide enabled us to continue with him professional guidance , kind advice, suggestions and timely inspiration which has helped us decide on our project. We would also extend our deepest gratitude to our HOD Dr.S.N.Gujar. We would like to thank him for simulating suggestions and encouragement, also with areas for improvement which helped us for implementation and writing of this dissertation.

REFERENCES

[1] S. K. Khare , V. Bajaj, and U. R. Acharya, ”PDCNNet : An Automatic Framework for the Detection of Parkinson’s Disease Using EEG Signals,” in IEEE Sensors Journal, vol. 21, no. 15, pp. 17017-17024, 1 Aug.1, 2021, DOI: 10.1109/JSEN.2021.3080135.

[2] M. Rumman, A. N. Tostem, S. Farzana, M. I. Pavel, and M. A. Alam, ”Early detection of Parkinson’s disease using image processing and artificial neural network,” 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2018, pp. 256-261, DOI: 10.1109/ICIEV.2018.8641081.

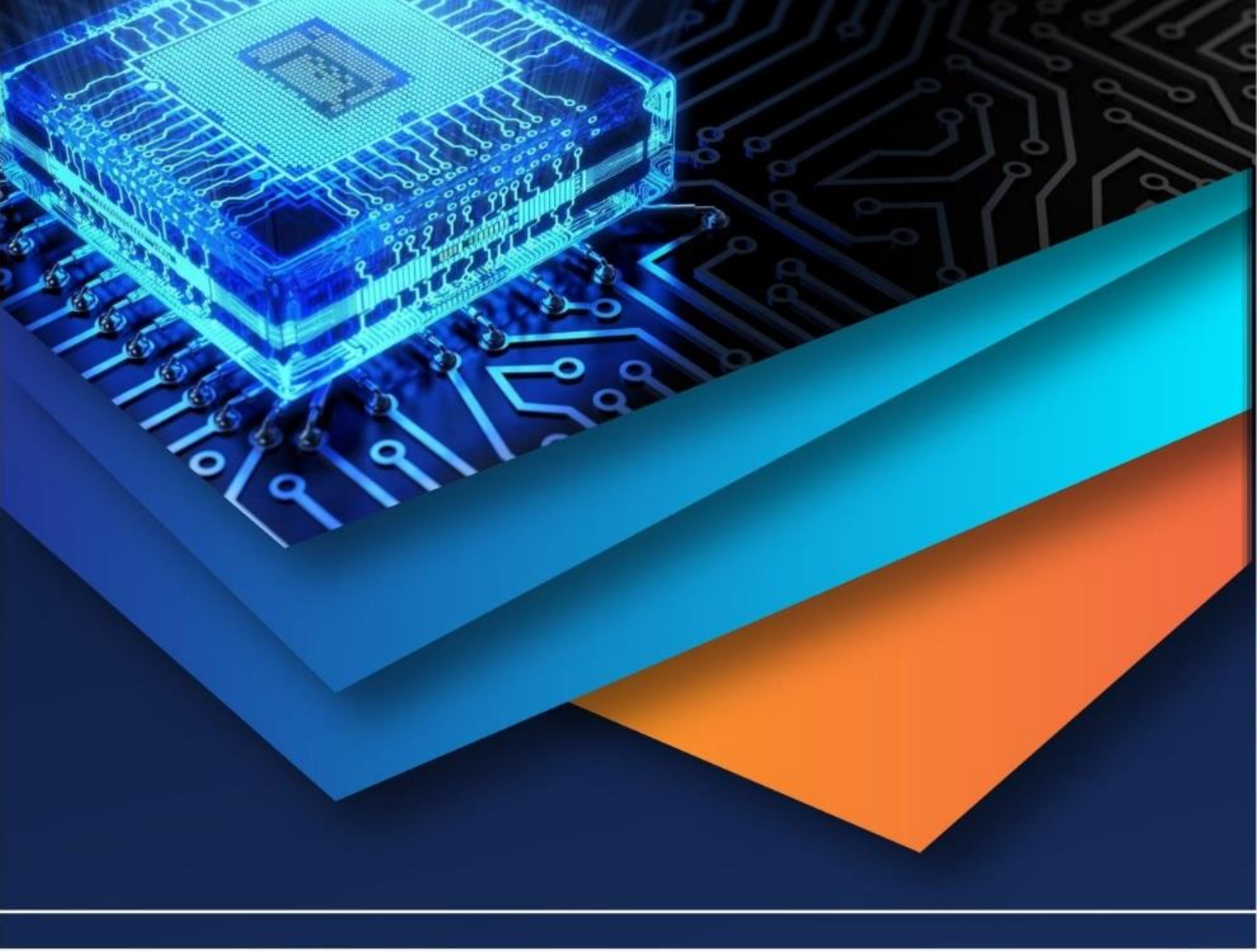
[3] K. H. Leung et al., ”Using deep-learning to predict outcome of patients with Parkinson’s disease,” 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), 2018, pp. 1-4, DOI: 10.1109/NSS- MIC.2018.8824432.

[4] K. Hu et al., ”Graph Sequence Recurrent Neural Network for Vision-Based Freezing of Gait Detection,” in IEEE Transactions on Image Processing, vol. 29, pp. 1890-1901, 2020, DOI: 10.1109/TIP.2019.2946469.

[5] Z. -J. Xu, R. -F. Wang, J. Wang, and D. -H. Yu, ”Parkinson’s Disease De- tection Based on Spectrogram-Deep Convolutional Generative Adversarial Network Sample Augmentation,” in IEEE Access, vol. 8, pp. 206888-206900, 2020, DOI: 10.1109/ACCESS.2020.3037775.

[6] P. M. Shah, A. Zeb, U. Shafi, S. F. A. Zaidi and M. A. Shah, ”Detection of Parkinson Disease in Brain MRI using Convolutional Neural Network,”2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1-6, DOI: 10.23919/IConAC.2018.8749023.

©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)