iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.39027

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 9 Issue XI Nov 2021- Available at www.ijraset.com

Emotion Detection using Social Media Data

Ansh Mehta[1], Shubham Patil[2], Soham Dave[3], Prachi Tawde[4]
[1, 2, 3]Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
[4]Assistant Professor, Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India,

Abstract: Previous research on emotion recognition of Twitter users centered on the use of lexicons and basic classifiers on pack of words models, despite the recent accomplishments of deep learning in many disciplines of natural language processing. The study's main question is if deep learning can help them improve their performance. Because of the scant contextual information that most posts offer, emotion analysis is still difficult. The suggested method can capture more emotion sematic than existing models by projecting emoticons and words into emoticon space, which improves the performance of emotion analysis. In a microblog setting, this aids in the detection of subjectivity, polarity, and emotion. It accomplishes this by utilizing hash tags to create three large emotion-labeled data sets that can be compared to various emotional orders. Then compare the results of a few words and character-based repetitive and convolutional neural networks to the results of a pack of words and latent semantic indexing models. Furthermore, the specifics examine the transferability of the most recent hidden state representations across distinct emotional classes and whether it is possible to construct a unified model for predicting each of them using a common representation. It's been shown that repetitive neural systems, especially character-based ones, outperform pack-of-words and latent semantic indexing models. The semantics of the token must be considered while classifying the tweet emotion. The semantics of the tokens recorded in the hash map may be simply searched. Despite these models' low exchange capacities, the recently presented training heuristic produces a unity model with execution comparable to the three solo models.
Keywords: Hashtags, Sentiment Analysis, Facial Recognition, Emotions.

I. INTRODUCTION

People now have a new way to generate and consume a large amount of information on the internet thanks to the advent of social network platforms. People used to acquire information from portal websites in the past. Many websites offer a diverse range of topics, from politics to entertainment. Traditional internet information sources are useful, but they are inefficient since they frequently contain redundant data. People, on the other hand, have tended to obtain information via online social network platforms since its introduction because of their quick and efficient characteristics. Users can pick and choose whatever information sources they want to use on these sites. And many social network platforms such as Twitter, Google+, and Facebook provide information for users. Users need individualized services now more than ever. To benefit users, more and more tailored services are now available. People require this individualized service to live more efficiently in their fast-paced life. Users on the Twitter network submit a significant volume of information every day. Many research studies focus on Twitter and this data collecting since these data are related to user behavior. User modelling is one of the study works in the world of Twitter. Researchers began investigating the ranking and recommendations of web pages cited from Twitter to give a tailored service. A lot of research is focused on predicting users' interests based on their publicly available tweets. [5] Regardless of the content of tweets or the possible utility of Twitter, researchers discovered that tweets frequently transmit relevant information about users' emotional states. As a result, emotion analysis on Twitter has become a hot topic in the microblogging world. Most of the emotion research relies on Twitter sentiment classification. Several features and approaches for training sentiment classifiers on the Twitter network have been investigated in recent years, with mixed results. There are also several additional research projects on Twitter that deal with emotion analysis. One of the experiments in this field focuses on obtaining product feedback by analyzing client sentiment on the Twitter network. Previous research has also focused on analyzing public sentiments by extracting emotions from Twitter tweets.

II. LITERATURE SURVEY

"People's emotional states, also known as moods, are fundamental to their thoughts, ideas, and views, which influence their attitudes and conduct." In this research, we present a method for detecting the emotion or mood of a tweet and categorizing it into the proper emotional category. Our method is referred to as a two-step method because it employs two methods for classification: one is based on rules, and the other is based on machine learning. The first approach is the Rule Based Approach (RBA), to which we contributed pre-processing, tagging, feature selection, and knowledge base construction as modest contributions. Tags are used to select features. The classifier in our second approach, Machine Learning Approach (MLA), is based on a supervised machine learning algorithm called Nave Bayes, which requires labelled data.

©IJRASET: All Rights are Reserved

The emotion of a tweet is detected and classified using the Naive Bayes algorithm. Because MLA requires labelled data, which we have already prepared using RBA, the result of RBA is given to MLA as input. We examined the accuracies of both approaches and found that the rule-based approach can categorize tweets with an accuracy of approximately 85 percent, while the machine learning approach can classify tweets with an accuracy of around 88 percent. The performance of the machine learning strategy is better than the rule-based approach, and the performance has increased because we deleted the incorrect data while training the model. For the construction of the system, the approaches involve the ideas of Natural Language Processing, Artificial Intelligence, and Machine Learning. The identification of emotion for non-hashtagged data and the labelled data construction for machine learning approaches without manual creation are the two main contributions in this study. [1]

Individual emotional states, often known as moods, are important in the expression of thoughts, ideas, and views, which influences attitudes and conduct. Individuals are increasingly using social media platforms like Twitter to broadcast their daily activities or to report on an external event of interest; comprehending the rich 'landscape' of emotions will help us better interpret millions of people. This paper proposes a Rule-Based technique for detecting the emotion or mood of a tweet and categorizing it into the relevant emotional category. The technology has an accuracy of 85 percent. It is feasible to understand the deeper layers of emotions (finer grained) using the proposed system rather than sentiment (coarse grained). Sentiment indicates if a tweet is favorable or negative, but the proposed approach provides additional information about the tweet, which has negative implications in the fields of psychology, intelligence, and social and economic trends. [2]

We examine if common machine learning techniques known to perform well in coarse-grained emotion and sentiment classification can also be applied successfully on a set of fine-grained emotion categories. We first describe the grounded theory approach used to develop a corpus of 5,553 tweets manually annotated with 28 emotion categories. We identified two machine learning algorithms that perform well in this emotion classification task based on our preliminary experiments, demonstrating that it is possible to train classifiers to detect 28 emotion categories without a significant performance drop when compared to coarser-grained classification schemes. Automatically detecting fine-grained emotions is a difficult task, but we have shown that a classifier can be trained to do rather well in identifying as many as 28 emotion types. Our 28 emotion categories are a supplement to the six to eight emotion categories widely employed in current research (Alm et al., 2005; Aman & Szpakowicz, 2007). Some of the 28 emotion categories are like those found in previous emotion theories like Plutchik's (1962) 24 categories on the emotion wheel and Shaver et al (2001)'s tree-structured list of emotions. Emotion theories in psychology haven't been created specifically based on emotions portrayed in text. Therefore, our emotion categories offer a more fitting framework for the study of emotion in text. [3]

Individuals are increasingly using social media and microblog platforms to communicate their views and ideas in the form of brief text messages. Detecting emotions in text has a wide range of uses, including recognizing individual anxiety or sadness and gauging a community's well-being or mood. In this research, we present a novel method for automatically classifying persons' text messages to infer their emotional states. We use the well-known Circumplex model to model emotional states, which characterizes effective experience along two dimensions: valence and arousal. We chose Twitter messages as our input data set because they provide a big, diversified, and openly accessible set of emotions. Our method uses hashtags as labels to train supervised classifiers to detect many kinds of emotion on potentially large data sets with no operator intervention. Unigrams, emoticons, negations, and punctuation are among the traits investigated for their utility in detecting emotion. We use a vocabulary of emotions to solve the problem of sparse and high-dimensional feature vectors in messages. For identifying Twitter messages, we compared the accuracy of numerous machine learning methods, including SVM, KNN, Decision Tree, and Naive Bayes. Our technique has an accuracy of over 90%, while demonstrating robustness across learning algorithms. [4]

The properties of byte-level recurrent language models are investigated by the authors. These models learn representations that include disentangled features corresponding to high-level ideas when given enough capacity, training data, and computation time. It looks for a single unit that can execute sentiment analysis. These unsupervised representations exhibit state-of-the-art performance on the binary portion of the Stanford Sentiment Treebank. They are also incredibly data-savvy. This method matches the performance of strong baselines trained on entire datasets when only a few labelled instances are used. Authors also demonstrate the sentiment unit has a direct influence on the generative process of the model. Simply fixing its value to be positive or negative generates samples with the corresponding positive or negative sentiment. [5]

Discourse Parsing and Sentiment Analysis are two crucial tasks in Natural Language Processing that have been proved to work together. The authors of this paper create and compare two neural models for learning both tasks simultaneously. The inventors of this method start by creating a vector representation of all the text segments in the input sentence. It then uses three different Recursive Neural Net models for discourse structure prediction, discourse relation prediction, and sentiment analysis. Finally, the authors put these Neural Nets together in two joint models: Multi-tasking and Pre-training.

©IJRASET: All Rights are Reserved

The findings of two conventional corpora show that both techniques boost performance in each test, however multitasking outperforms Pre-training. Shows advances in the prediction on the set of contrastive relations, specifically for Discourse Parsing. [6] The authors take it a step further in this research, creating a new method for sentiment learning in the MapReduce framework. Their method uses the hash tags and emoticons in a tweet as sentiment labels, then uses a parallel and distributed classification procedure to classify different sentiment kinds. Furthermore, Bloom filters are used to reduce the storage capacity of intermediate data and improve the algorithm's performance. It establishes that this approach is efficient, robust, and scalable, as well as confirming the quality of sentiment identification, through a comprehensive experimental evaluation. [7]

III. RESEARCH AND METHODOLOGY

Social platforms have become increasingly important in people's daily lives as networks have grown. Twitter, as the most popular microblogging network, contains a wealth of information in the form of tweets exchanged by millions of users. It's challenging to extract meaningful information for users from this data stream because it's always growing. More and more users wish to use these data to profit from Twitter's customized service. People can get a tailored service on Twitter by extracting the semantic meaning of Twitter and modelling their interests. Meanwhile, studies demonstrate that people use Twitter to convey their emotions. When contrasted to other tweets, these emotional tweets usually clearly show the users' preferences. [4] [7] Social platforms have become increasingly important in people's daily lives as networks have grown. Twitter, as the most popular microblogging network, contains a wealth of information in the form of tweets exchanged by millions of users. It's challenging to extract meaningful information for users from this data stream because it's always growing. More and more users wish to use these data to profit from Twitter's customized service. People can get a tailored service on Twitter by extracting the semantic meaning of Twitter and modelling their interests. Meanwhile, studies demonstrate that people use Twitter to convey their emotions. When contrasted to other tweets, these emotional tweets usually clearly show the users' preferences [3].

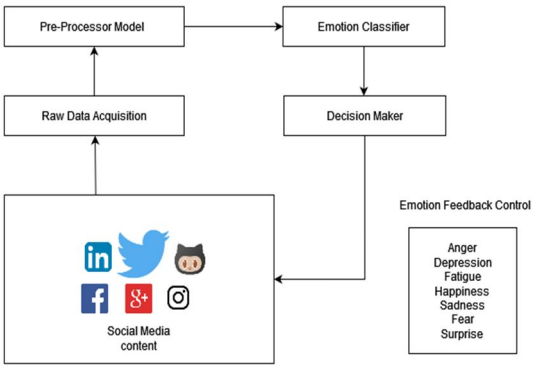


Fig.1 Architecture Diagram

In recent years, the importance of social media in our daily lives has grown dramatically. It is currently used for more than just social connection; it is also a vital platform for exchanging information and news. Twitter, a microblogging website owned by Facebook, connects millions of users around the world and enables for the dissemination of information and news in real time. In these social media platforms, fake news has become a serious issue. In these social media platforms, fake news has become a serious issue. Fake news has a huge impact in today's culture. The detection of fake news is a crucial step. This research uses the Support Vector Machine (SVM) method to detect fake news using machine learning approaches. Before applying the machine learning approach to classify data, it is necessary to cleanse the data using the normalization method. This approach considers the content's trustworthiness as well as the user's reputation. By learning to predict accuracy ratings in a credibility-focused Twitter dataset, this research develops a way for automating fake news detection on social media.

As a result, the purpose of this research is to develop some emotion-based user modelling methodologies that take advantage of emotional data. The methods for detecting emotion on Twitter are introduced and analyzed in this paper. First, it assesses and compares the performance of several emotion detecting techniques. Then, for the purpose of user modelling, apply these emotion detection algorithms to a Twitter sample dataset. For the basis of the observed emotional data, a set of emotion-based user modelling tactics on the Twitter platform was also proposed. It also assesses emotion-based user modelling methodologies and examines their effects on typical user profiles. Proposed system results show that emotion-based user profiles enhance the quality of user profiles and have a better performance. [1]

©IJRASET: All Rights are Reserved

The Profile of Mood States [6] is a psychological tool for determining a person's mood state. It defines 65 adjectives that are rated on a five-point scale by the subject. Each adjective belongs to one of the six groups. Feeling irritated, for example, will positively add to the anger category. Except for relaxed and efficient, whose contributions to their respective categories are negative, the greater the adjective's score, the more it contributes to the overall score for its category. POMS groups these scores into six categories: anger, despair, exhaustion, vigour, tension, and confusion. We adopted the structure from Norcross et al. [7], which is known to closely match POMS's categories, because POMS isn't publicly available. We added to it with data from the Brian Mac Sports Coach website. We removed the adjective blue from the original structure since it only seldom refers to an emotion rather than a color, and word-sense disambiguation techniques were unable to discern between the two meanings.

For each category, we used the following adjectives:

1) *Anger:* Angry, peeved, grouchy, spiteful, annoyed, resentful, bitter, ready to fight, deceived, furious, bad tempered, rebellious,
2) *Depression:* Sorry for things done, unworthy, guilty, worthless, desperate, hope- less, helpless, lonely, terrified, discouraged, miserable, gloomy, sad, unhappy,
3) *Fatigue:* Fatigued, exhausted, bushed, sluggish, worn out, weary, listless,
4) *Vigor:* Active, energetic, full of pep, lively, vigorous, cheerful, carefree, alert,
5) *Tension:* Tense, panicky, anxious, shaky, on edge, uneasy, restless, nervous,
6) *Confusion:* Forgetful, unable to concentrate, muddled, confused, bewildered, uncertain about things.

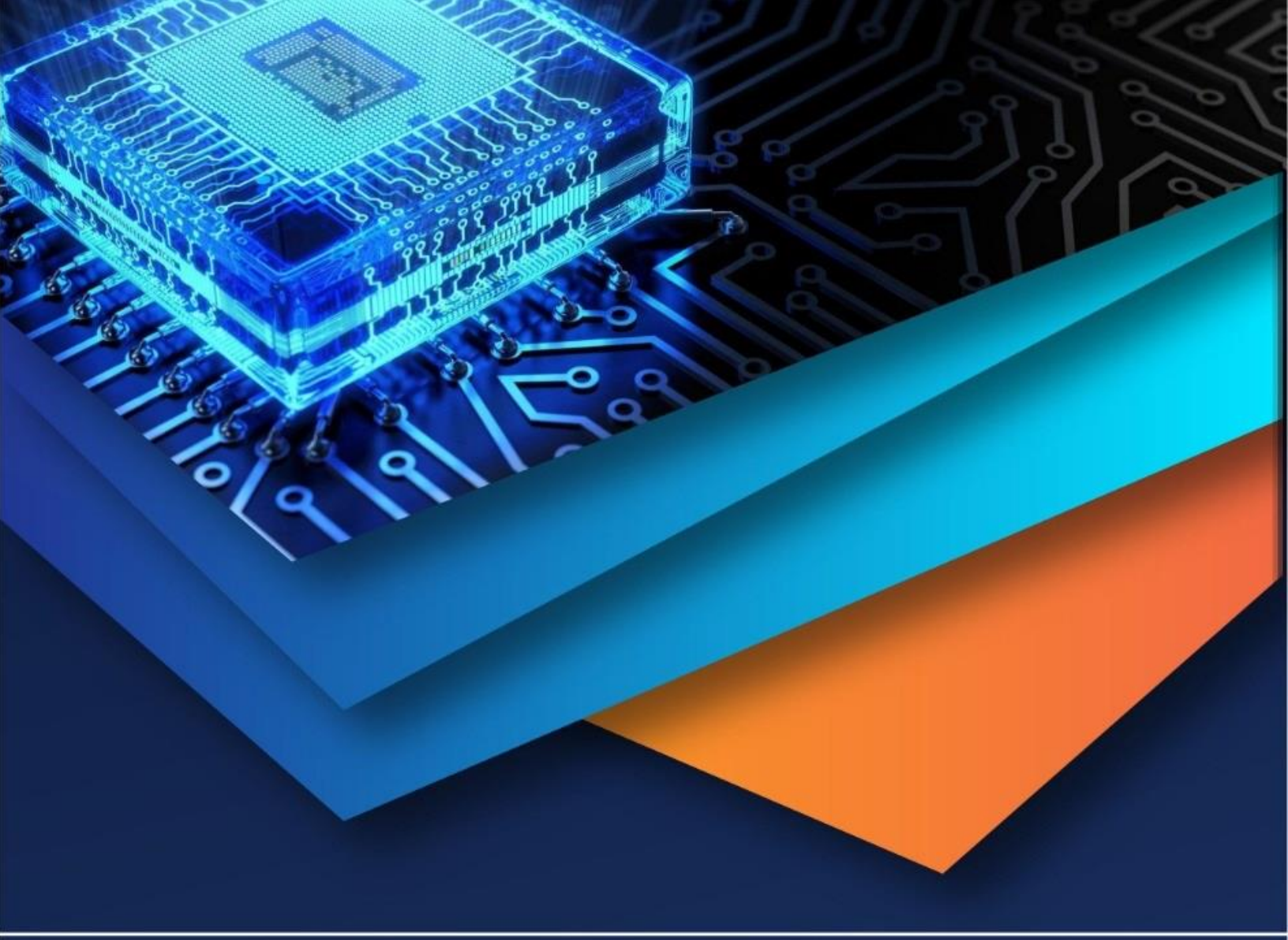
IV. CONCLUSION

The paper's major goal is to increase the use of deep learning and data mining approaches in twitter emotion recognition. To obtain the twitter stream, the proposed system is directly connected to the tweeter's public account. The results of the experiment are displayed using three algorithms: SVM, Random Forest, and CNN. To obtain a better result, it is necessary to tokenize the tweets and obtain the semantic of the token. The proposed semantic with pre-processing system has a lot higher accuracy than the present system. Instead of using letters as input, the proposed system uses words. This suggested system uses years of tweets to analyze the largest data collection for emotion prediction. We examined the system's utility for several classes of emotions to design a universal emotion recognition algorithm. Because the training data was automatically annotated, our method is language agnostic and adaptable to different languages. This research will help with user modelling on the Twitter network, and it aims to integrate two hotspots: emotion and user modelling. Because information is easily transmitted around the internet community by unsupported sources, recognizing disinformation in online social media platforms is crucial in this endeavor. It can be important in evaluating activist movements to be able to automatically recognize bogus news and stop it from spreading. User accounts that tweet a lot of URLs, @username mentions, and hashtags. This research also addresses the issue of determining the reliability of information on Twitter. The topic of information reliability has gotten a lot of attention lately, especially in social media, which is now being used as a primary source of information. The system's future scope is to recognize emotion based on user-uploaded multimedia (i.e., photos).

REFERENCES

[1] X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, “Users participation and social influence during information spreading on twitter,” PloS one, vol. 12, no. 9, p. e0183290, 2017.
[2] Niko Colneric and Janez Demsar, “Emotion Recognition on Twitter: Com- parative Study and Training a Unison Model”, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, FEBRUARY 2018.
[3] J. Guo, W. Che, H. Wang, and T. Liu, “A Universal Framework for Inductive Transfer Parsing across Multi-typed Treebanks”, Proc. of the 26th Int. Conf. on Computational Linguistics (COLING-16), pp. 12–22, 2016.
[4] A. Radford, R. Jozefowicz, and I. Sutskever, “Learning to Generate Reviews and Discovering Sentiment”, 2017.
[5] B. Nejat, G. Carenini, and R. Ng, “Exploring Joint Neural Model for Sen- tence Level Discourse Parsing and Sentiment Analysis”, Proc. of the SIG- DIAL 2017.
[6] N. Nodarakis, S. Sioutas, A. Tsakalidis, and G. Tzimas, “Using Hadoop for Large Scale Analysis on Twitter: A Technical Report”, arXiv preprint arXiv: 1602.01248, 2016.
[7] S. M. Mohammad and S. Kiritchenko, “Using Hashtags to Capture Fine Emo- tion Categories from Tweets”, Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015
[8] Pawar, Miss Priya, and Pankaj Agarkar. "TWITTER SENTIMENT ANALYSIS USING TEXTUAL INFORMATION AND DIFFUSION PATTERNS." *INTERNATIONAL JOURNAL* 5.7 (2020).
[9] Thakre, Miss Preeti, and Pankaj Agarkar. "Customer emotions recognition using facial and textual review." *INTERNATIONAL JOURNAL* 5.5 (2020).
[10] Y. Zhang and B. C. Wallace, “A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification”, ArXiv preprint arXiv:1510.03820v4, 2016.
[11] J. Bollen, H. Mao, and X.-J. Zeng, “Twitter mood predicts the stock market,” J. of Computational Science, vol. 2, no. 1, pp. 1–8, 2011.
[12] B. Plank and D. Hovy, “Personality Traits on Twitter —or— How to Get 1,500 Personality Tests in a Week”, Proc. of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.

©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)