IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 4, July-August 2024

Impact Factor: 7.394

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

Emotional Prediction from Social Media Patterns

Dr. R. Jayanthi, Shradha S Gadad

Associate Professor, Department of MCA, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

P.G. Student, Department of MCA, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

ABSTRACT: This research aims at undertaking an evaluation of a data set that contains multiple features that describe social media use and the effects on the users' emotions. The collected features are user demographics, age, and gender, as well as the activity measures on social networks,, namely daily active use time, the number of posts per day, likes and comments per post, as well as messages sent per day. Also, it tracks the primary emotions that are felt by each user, which can serve as a way to investigate with which moods people are associated with the usage of social networks. To forecast users' major emotions, we employ a RandomForestClassifier , considering the amount of their social media activity and their demographic characteristics. In addition, feature importance analysis is conducted to understand which factors are the most important in order to make these predictions. The study offers information pertaining to which elements of social media consumption are most likely to impact the levels of happiness. It is the purpose of this study to identify the complex relationship between social media use and emotional health so as to gain insights that could help design the next step in more beneficial use of social media for ultimate improvement of customers' emotional well-being.

KEYWORDS: Social Media Usage, Emotional Well-Being, RandomForestClassifier, Feature Importance, Social Media Impact.

I. INTRODUCTION

Social networking has become a universal necessity in people's everyday life in the framework of the developed information society. In this context, sites popular within Facebook, Instagram, Twitter, and Snapchat enable people to communicate and interact with numerous other people. Due to the increased use of these platforms they have brought about an added increase in social interaction, this cuts across the globe, this is a social interaction. Yet, on the flip side, as the plus of connection and sharing information, questions have arisen touching on the welfare of the customers using social media platforms with regards to their psychological state of being. From a more general standpoint, the present research proposals seek to provide a systematic investigation of the existing connections between overall engagement with social networks and subjective well-being, utilizing rich and diverse data that would include a range of parameters of users' activity and characteristics.

The chosen dataset covers user's demographics dimensions including age and gender, and key social media performance figures. Such measures consist of the daily active time, the daily active post counts, average likes and comments per day, and the daily messages. Also, the dataset provides information about the main affect experienced by the user; thus, the data can be used to analyze the relationship between social media usage and mental health. Finite from these variables we shall look at basic patterns and relationship that indicate how the various facets of social media use pattern emotional states. This will be done using a RandomForestClassifier supplemented by various features derived from users' activity on social media and their demographic characteristics. The RandomForestClassifier is a typical machine learning model which, among other values, is distinguished by its ability to work with large amounts of data. Through tuning the model on the dataset, we expect to achieve feature importance which reflects the features that are relevant in the prediction of dominant emotions; this gives an insight on which factor of social media usage affects the emotional well-being of a user most.

Consequently, a deeper analysis and knowledge of all the effects of social networks on the emotional state of a person is crucial for today's society. Although social media indeed has its benefits, including the maintenance of social relationships and the expression of one's personality and creativity, it also has its drawbacks; it can be a cause of stress, anxiety, and an unfavourable perception of oneself. That way, specifics of social media usage are refined and connected to increased levels of EWB which can then guide better approaches to usage. Such findings can more or less inform strategies meant to cushion the impacts of negative use of social media and foster good psychological wellbeing.

The analysis will be conducted in several steps. Firstly, data preparation and cleaning step will be applied to optimize the dataset for model development. This shall entail dealing with the missing data, transforming the numerical attributes and

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

categorizing the nominal attributes. In the next step, it will be necessary to train RandomForestClassifier and assess it based on the metrics including accuracy, precision, and recall. Next, in order to find out which features are the most potent in predicting dominant emotions we will proceed to feature importance analysis about. Last, we will analyze the results and conclude with an elaboration of their implications to the effect that , or the lack thereof, social network sites have on self-reported measures of affective functioning.

Therefore, in this analysis, it is necessary to consider and explain all key aspects related to the correlation between the regular use of social networks and changes to the individual's emotional state. Using a large number of features and state of the art machine learning algorithms, it is our intention to identify meaningful patterns that would assist in the healthier use of social media and the enhancement of the users' psychological wellbeing. Thus, through this study, we aim to present the Social media – Emotional health relationship in a balanced and accurate manner and use the findings for improving the approach to the problem of social media usage in the modern society.

II. TOOLS AND TECHNOLOGIES USED

For an extensive study on the prediction of emotions from patterns recognized in Social Media, several techniques & technologies are used each of which has a specific use in analysis & model building.

Python and Jupyter Notebook
Python is chosen as the key language because it is tremendously effective in encouraging interaction with data and has a kaleidoscope of libraries and frameworks for machine learning, data analysis, visualization, etc. Jupyter Notebooks are used for its interactivity and easiness in sharing files that comprise of code, data analysis and visualization in one single instrument. This environment is very important for sequentially passing through the model building and model assessing steps.

Machine Learning Libraries
Scikit-Learn: High-level libraries of Python that cover aspects of machine learning application, containing classes, which are necessary for applying machine learning models; the RandomForestClassifier for predicting emotions; functions for estimating and improving the quality of the model.

Pandas: Aids the cleaning and transformation of data which prepares it for analysis, as well as makes it easier to handle datasets.

NumPy: Handles numbers and array computations, useful when working through mathematical computations that define most machine learning models and the outcomes derived from them.

Data Visualization
Matplotlib: Used to generate complex graphics that allow to enhance the interferential results such as the feature importance plots and the model performance plots.

Seaborn: Improves the presentation of aesthetic designs and makes the handling and transmission of complicated graphics more manageable, in effect assisting in the presentation of data and its possibly present patterns.

Machine Learning Evaluation Metrics
The performance of the emotion prediction model can be evaluated by means of **accuracy**, which points at the global level of right platforms, and by the means of **precision**, the type of which meets the evaluate of positive platforms. **Recall** quantifies potential oversights on the part of the model regarding situations that should have been taken into consideration and the **F1-score** gives a measure of the average of the precision and recall rates. The **ROC-AUC** represents the model's classification capability between the different classes, while the confusion matrix provides the comparison of the assigned classes and the actual classes.

Classification Algorithm Employed
The **RandomForestClassifier** is the classification algorithm used in this study to predict the dominant emotions from the social media patterns. This algorithm is chosen because it is good for dealing with large data set and can model relation between features and targets even if those relations are non-linear in nature. They also note that RandomForestClassifier makes many decision trees, each used to classify examples that are constructed from different subsets of features and data, which counteracts overfitting and increases the model's stability. Using ensemble learning

strategies makes it possible to reduce the exposure of segmenting the emotions based on the social media usage patterns and demographical variables.

III. PROPOSED SYSTEM

The current proposed system is to be expected to forecast the major moods that the user is likely to portray by his or her behaviour on social media. The system would employ a RandomForestClassifier; therefore, the features include daily active time, posts per day, daily likes, daily comments, and daily messages, and several user's characteristics like age and gender. The method is to determine what major emotion a user is feeling, which would offer a finer-grained idea of precisely how various elements of SM interaction are connected to subjective well-being.

The process of analysis includes several important activities. It is first necessary to clean the data as well as scale the variables if there are any missing values or if the ranges are sensitive. Subsequently, we perform another training for the RandomForestClassifier with a part of the data set for fine-tuning the parameters in order to improve the surface prediction model. After the model is trained with the data, the performance is calculated with the remaining data with the purpose of testing the model's performance on yet unseen data. Another element of the architecture of our system is feature importance analysis, which consists in the identification of features that are most important for the prediction of dominant emotions. It serves the purpose of identifying how each of the social media activity metrics relates to users' emotional experiences.

In this way, it is possible to provide further understanding about the correlation between the level of activity on social platforms and the overall emotions/vitality state of the users, having them join this proposed system. As such, our study would help to present a set of factors that affect users' emotional states, which may for the better understanding of efforts to create healthier attitudes toward social media usage and guide interventions meant to enhance general well-being among users. Hence, this system provides an opportunity for developing further studies and practical applications for improving the population's e-motional health in the context of modern digital technologies.

IV. IMPLEMENTATION

To implement the proposed system for predicting dominant emotions based on social media patterns, we follow these steps:

Data Collection

Gather the emotion prediction from social media pattern dataset(from Kaggle), ensuring the detailed information on users' social media activity, including metrics like daily usage time, posts, likes, comments, and messages.
Additionally, user demographics such as age and gender are recorded. The dominant emotion experienced by each user is also captured to analyze the impact of social media on emotional well-being.

Data Cleaning and Preprocessing

Step 1: Import the Libraries

By importing necessary Python libraries including Data handling library (Pandas), Numerical computation library (NumPy), and Data visualization Library (Matplotlib). These libraries are used for loading cleaning and transformation of data required for pattern analysis of social media users and for accurate identification of the emotional state.

Step 2 : Reading the Dataset

In the process of reading the dataset, a Pandas library is employed to load the dataset of social media activity metrics as well as demographical information of users. This includes such parameters as using time per day, daily posts, likes, comments, messages, age, and gender. It is used in the subsequent preprocessing steps that are used to structure, clean, modify and transform the raw dataset ready for prediction and analysis.

Step 3: Analysing data

In order to analyze the data and come up with a model that provides accurate results through the Random Forest algorithm, one need to initialize the RandomForestClassifier from scikit-learn, then fit it to the preprocessed data for the results to be obtained and fine tune the hyperparameters such as the number of trees and the maximum depth of the trees for the best results. Ensemble learning is used in the algorithm that identifies dominant emotions of the users according to the analysis of their activity and demographic characteristics in order to reveal the factors affecting emotions in the social network to a greater extent. This work in turn makes it easier to gauge the prompt emotional reactions from social media patterns with robust and accurate analysis.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

- Handle missing values and incorrect entries in the dataset.
- Convert categorical variables into numerical values using one-hot encoding.
- Scale the features to ensure consistency in the data.

```
[6]: df_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1001 entries, 0 to 1000
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   User_ID                     1001 non-null   object
 1   Age                         1001 non-null   object
 2   Gender                      1000 non-null   object
 3   Platform                    1000 non-null   object
 4   Daily_Usage_Time (minutes)  1000 non-null   float64
 5   Posts_Per_Day               1000 non-null   float64
 6   Likes_Received_Per_Day      1000 non-null   float64
 7   Comments_Received_Per_Day   1000 non-null   float64
 8   Messages_Sent_Per_Day       1000 non-null   float64
 9   Dominant_Emotion            1000 non-null   object
dtypes: float64(5), object(5)
memory usage: 78.3+ KB
```

Fig.1. Describes the data Information

Model Training

- Split the data into training and testing sets.
- Train a Random Forest Classifier on the training data.

Feature Importance Determination

- Compute the feature importances from the trained Random Forest model.
- Identify the top 10 most important features that contribute to the prediction of dominant emotions.

Focused Analysis on top features

- Retrain the model using only the top 10 features.
- Evaluate the performance of the model with the reduced feature set.
- Analyze the results to understand the key factors influencing emotional well-being.

Implicitly, this approach will help us better understand how various aspects of social media use relate to the overall well-being of the user by disentangling the positive and negative associations between patterns of social media use and users' affective experiences. These insights can be applied to recommendations for safer social media use and extend to the development of therapeutic approaches for users' emotional well-being.

V. RESULTS AND DISCUSSION

Model Evaluation

Utilize various evaluation metrics, such as accuracy, confusion matrix, precision, f1-score, and classification report to assess the model performance.

Result

The following figures describe the results found by the model.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

IJARETY

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

```
Accuracy with Top 10 Features: 0.95
Classification Report with Top 10 Features:
              precision    recall  f1-score   support

       Anger       0.90      1.00      0.95         9
     Anxiety       1.00      0.95      0.98        21
     Boredom       1.00      0.88      0.93        16
   Happiness       0.86      1.00      0.92        12
     Neutral       0.96      1.00      0.98        26
     Sadness       0.92      0.86      0.89        14

    accuracy                           0.95        98
   macro avg       0.94      0.95      0.94        98
weighted avg       0.95      0.95      0.95        98

Confusion Matrix with Top 10 Features:
[[ 9  0  0  0  0  0]
 [ 0 20  0  1  0  0]
 [ 0  0 14  0  1  1]
 [ 0  0  0 12  0  0]
 [ 0  0  0  0 26  0]
 [ 1  0  0  1  0 12]]
```

Fig.2. Random Forest

Visualization

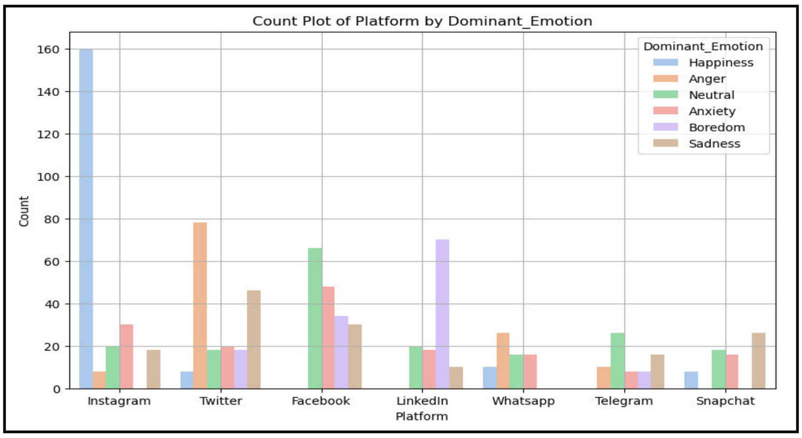

Fig.3. Count Plot of Platform by Dominant_Emotion

The provided data visualization displays a count plot showing distribution of the dominant emotive states across several social media accounts. Instagram, Twitter, Facebook, LinkedIn, WhatsApp, Telegram, and Snapchat are analyzed, tracking six emotions: positive emotions which included joy, negative emotions including rage, middle emotions including neutrality, and complex emotions which included both, concern and lethargy and lastly the negative emotion which is sadness. Instagram has by far the highest count of happiness: it significantly surpasses the number of all the other emotions and platforms. In turn, there is a defined increase in the expression of anger on Twitter, accompanied by sadness. Facebook analysis reflects rather a moderate and balanced emotional background, and therefore, the most popular of all types of sentiments identified is the neutral one. It is important to mention that the count of boredom has the highest value in the case of LinkedIn. In general, there is a considerable variation in the averages of emotional counts for different platforms; these new platforms such as; WhatsApp, Telegram and snapchat registers low counts compared to the most frequently used platforms. This interpretation describes how the phenomenon of valence may reveal the affective characteristics of social media environments and, by extension, people's behaviors in those settings. It implies that perhaps, the various social media services may trigger or invite a specific affect in its users as well as a specific affect being more dominant in particular service.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

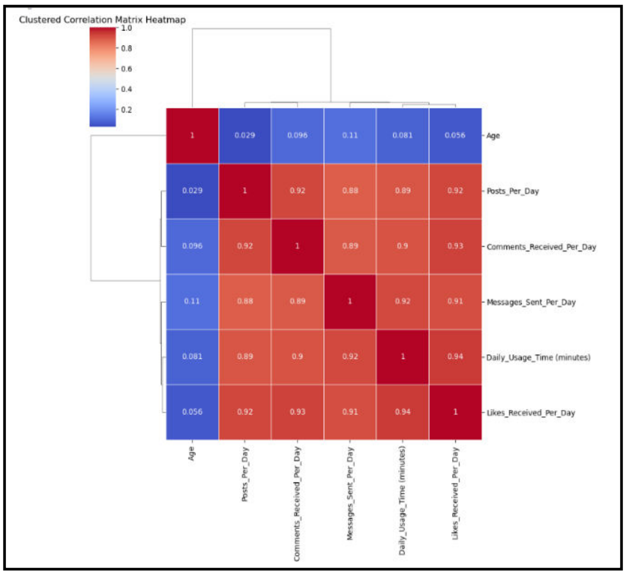

Fig.4. Clustered Correlation Matrix Heatmap

The data visualization depicts a clustered correlation matrix heatmap to represent the associations between multiple characteristics of social networking sites. The features are Age, Posts per day, Comments received per day, Messages sent per day, Daily usage time in minutes and Likes received per day. A color scale is used on the heatmap, ranging from blue, for it stands for low or negative correlation coefficient, to red for high positive correlation coefficient. The degree of rising of the color corresponds with the degree of the correlation between the two, which can be from negative one to one. The heatmap has to be created with some custom features, such as clustering, which emphasizes similar features regarding their correlation with samples. The high density of the connection between the features among a, b, and c as depicted by their clustering shows that they are strongly interconnected with the aspect of the social media activity apart from the Age. In all, this heatmap offers an efficient way of showing how various aspects of social media usage relate concerning the frequency with which they are engaged by the users; this is because it outlines where various behaviours are related.

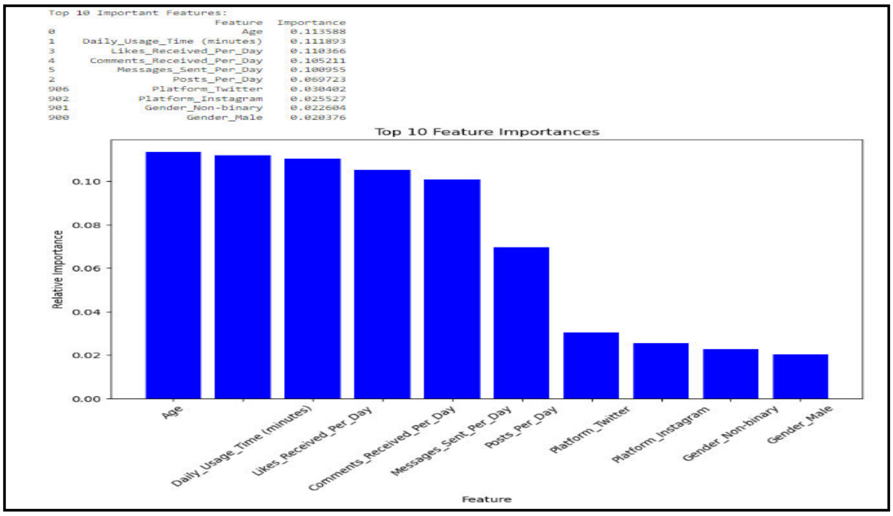

Fig.5. Feature Importances

The bar plot shows the details of the ten most important features for a machine learning model and their potential in determining the outcome regarding the use of social media. The horizontal axis is labeled features, and the vertical axis is labeled by the importance of the features. Age is established as the most important factor followed by Daily Usage Time (minutes) and Likes Received per Day both with importance scores slightly above 0.The importance of Comments Received Per Day and Messages Sent Per Day is also nearly equal, both being roughly 0. Posts Per Day can be

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

categorized as having a moderate importance of roughly 0. undefined Twitter with importance value close to 0 and Instagram – are less important but still significant platforms. undefined undefined. Non-binary and Male gender categories have the least importance among the top features with values nearly equal to 0.This plot also helps to emphasize the key features that define the model's predictions, with age and numerous indicators related to the activity on social networks.

In the current research work, we conducted a comprehensive analysis of emotion prediction from social media patterns using a RandomForestClassifier, highlighted essential aspects of social media consumption that affected the users' emotional status. Categorisation of features also maintained that more hours spent in the social media every day elicited negative emotional responses while the number of likes and comments one receives in a day elicited positive emotions pointing towards the need for social acceptance derived from the app. Likely, the frequency of posts also had a moderate effect, which indicates that either the type of post or the reason behind it plays a role. In sum, the current research shows the multifaceted relationship between social media engagement and emotions; while moderate use and pleasant experiences might have beneficial effects on people's well-being, excessive usage and seeking approval can do harm.

VI. CONCLUSION

In conclusion, it can be stated that the analysis of the possibility of emotion prediction from social media patterns based on the application of RandomForestClassifier model has given much understanding of the interconnection of the given aspect. The dominantly extracted emotion from the employees can be predicted with classifier accuracy in terms of daily usage time, number of Likes, number of comments, and number of posts per day. Other insights suggest that long time spent on certain types of sites daily is likely to lead to negative feelings, which stress the possible drawbacks of the intensive use of social networks. On the other hand, gaining more likes and comments positively affects the target's mood, indicating the value of social recognition and interaction for increasing the level of well-being.

Also, the study showed that frequency of posts per day also moderately influenced the subjects' emotional health, and that content and purpose of posting represents an essential factor to focus on. These results corroborate the suggestions made in the literature regarding responsible and purposeful usage of the social media platforms which, once used for meaningful interactions and sharing of the meaningful content, can act as potential source of positive emotions.

From the analysis drawn above, it is possible to foster practical suggestions for positive changes regarding the use of social media accounts. Some of them may include; providing more guidance to users on the amount of time they spend on social media, and the fostering of positive and healthy online communities rather than dwelling on the addictive part of the social media interactions and the promotion of digital health. It is possible to prevent the negative outcomes by knowing the subtle shifts in people's mood that social media brings when formulating the corrective measures to emphasize the benefits of online interactions. Future studies could expand these findings within different age, gender and ethnic populations as well as other platforms, enhancing the knowledge of how technology can be used to promote general emotional health. We have implemented an automatic text detection technique from an image for Inpainting. Our algorithm successfully detects the text region from the image which consists of mixed text-picture-graphic regions. We have applied our algorithm on many images and found that it successfully detect the text region.

REFERENCES

[1] Sivanantham, K. (2021). Sentiment analysis on social media for emotional prediction during COVID-19 pandemic using efficient machine learning approach. Computational intelligence and healthcare informatics, 215-233.
[2] He, S., Zheng, X., Zeng, D., Luo, C., & Zhang, Z. (2016). Exploring entrainment patterns of human emotion in social media. PloS one, 11(3), e0150630.
[3] M. Roshanaei, R. Han, S. Mishra, "Features for mood prediction in social media", In the proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1580-1581, 2015.
[4] P. Chiranjeevi, V. Gopalakrishnan, P. Moogi, "Neutral face classification using personalized appearance models for fast and robust emotion detection", IEEE Transactions on Image Processing ,vol. 24, no. 9, pp.2701-2711, 2015.
[5] G. T. Giancristofaro, A. Panangadan, "Predicting Sentiment towards Transportation in Social Media using Visual and Textual Features", In the proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp. 2113-2118, 2016.
[6] M. Tasviri, S. A. H. Golpayegani, H. Ghavamipoor, "Presenting a Model Based on Social Network Analysis in Order to Offer a Diet to Users Proper to Their Mood", In the proceedings of the 3th International Conference on Web Research (ICWR), pp. 133-139, 2017.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104033

[7]] A. Cernian, A. Olteanu, D. Carstoiu, C. Mares "Mood Detector – On Using Machine Learning to Identify moods and Emotions", In the proceedings of the 21st International Conference on Control Systems and Computer Science, pp. 213-216, 2017.

[8] S.Taylor, E.Nosakhare, A. Sano, R. Picard, "Personalized Multitask Learning for Predicting Tomorrow"s Mood, Stress, and Health", IEEE Transactions on Affective Computing, vol. 14, no. 8, 2017.

[9] Tang, J., Zhang, Y., Sun, J., Rao, J., Yu, W., Chen, Y., & Fong, A. C. M. (2011). Quantitative study of individual emotional states in social networks. IEEE Transactions on Affective Computing, 3(2), 132-144.

[10] Roshanaei, M., Han, R., & Mishra, S. (2015, August). Features for mood prediction in social media. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (pp. 1580-1581).

IJARETY

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर
NISCAIR

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com