

# Optimizing the Quality of Predicting the Ill Effects of Intensive Human Exposure to Social Networks Using Ensemble Method

Manjunath Gadiparthi<sup>1\*</sup>, E Srinivasa Reddy<sup>2</sup>

E-mail: gmanjunathc2000@gmail.com, esreddy67@gmail.com

<sup>1</sup>Department of Computer Science & Engineering, ANU College of Engineering & Technology, Andhra Pradesh, India.

<sup>2</sup>Department of Computer Science & Engineering, ANU College of Engineering & Technology, Andhra Pradesh, India.

**Keywords:** Social network, Support Vector Machine, Neural Networks, Random Forests, Ensemble model

**Received:** May 31, 2022

*People may quickly obtain information through a variety of channels including social media, blogs, websites, and other online resources. These platforms have made it possible for information to be shared more easily. As a result, the amount of time that individuals spend on various social networking applications has increased. This research predicts the effects of human exposure to social networks in the near future. In this work a competent model for predicting the ill-effects is provided, that is both accurate and efficient. This model represents a combination of the independent models that have been operating independently so far. Thus each of these models makes a forecast and ultimate selection is decided based on whether or not there exists a majority of convergence of results from the operation of various models. By employing the majority voting system, this strategy attempts to take the benefits of the predictions produced by all the models while also to reduce the inaccuracies generated by each model. Theoretically, this model should outperform the use of individual models in terms of performance. Important features are extracted from the datasets using the proposed model, and the extracted features are then classified using an ensemble model that consists of four popular machine learning models: support vector machines (SVMs), logistic regression (logistic regression), random forest (random forest classification), and neural networks (NN). We have analyzed our prediction performance with the existing methods for number of times by changing the train set and test set data. In all the cases our novel method has been predicting with 3% to 4% improved performance in accuracy, precision, F-1 score, and specificity. From the dataset it has been achievable to attain the highest training and testing accuracy from among the existing models.*

*Povzetek: Optimizacija kakovosti napovedovanja slabih učinkov intenzivne izpostavljenosti človeka družbenim omrežjem z uporabo ansambelske metode.*

## 1 Introduction

As a consequence of the proliferation of social networking sites such as Facebook, LinkedIn, and Twitter [15], people's communication habits have been significantly altered [1]. As a consequence of this transition, predicting concerns that may develop among a large number of people has become more challenging. In accordance with studies, persons who spend a significant amount of time on social media are more likely to suffer from disorders such as social anxiety or depression, as well as being exposed to inappropriate content. User behavior predictions based on data gathered from social network use time are becoming more popular as a method of accurately projecting user behavior and are

becoming more popular. The data received by SN is typically imbalanced, in contrast to conventional data, which may make it harder to foresee user issues based on this data.

Machine Learning algorithms will aid us in learning about algorithms that learn from large data sets, generalize from them, and predict from them. In the future, when we need to find out exactly how to employ these algorithms, this will be really helpful. Calculative statistics and decision-making are inextricably linked in that machine learning is a component of both fields of research. Robotics techniques, such as machine learning, are used in a number of applications such as predicting

product sales, calculating the possibility of rainfall happening in a certain region, and so on [2]. Systems analysis paired with machine learning algorithms will assist in the development of prediction models for specialized situations involving social network time spent, for example, it involves the tracking of adverse occurrences in users while they are on their trial run, as well as the selection of the best forecast for each user. To foresee the issue in this research, we used machine learning methodologies and offered an integration framework for a decision support system to accomplish this. In our previous work we have applied the existing models Support Vector Machines, logistic regression, random forests, and neural networks in order to build the decision support system for predicting the issue, and the system's performance was evaluated [3].

The current research entails conducting experiments on a survey dataset that has been acquired from users via the use of a series of questionnaires. The data is cleansed and changed in accordance with the needs of the machine learning model in order to forecast difficulties that may arise as a result of the widespread usage of social networking apps. Using k-fold cross validation, 70 percent of the dataset is utilized for training, and the remaining 30 percent is used to test the classification model as part of the suggested technique. The collected characteristics are then identified using an ensemble machine learning model that consists of a Decision Tree (DT) classifier, a Random Forest (RF) algorithm, and logistic regression to further classify the data. The following are the most significant contributions made by this paper:

1. Preparation dataset for training machine learning model.
2. Selection of an Ensemble model to achieve optimized accuracy in prediction.
3. Reduction of training time of the ensemble learning model.
4. Compare the results with existing model.

Section 2 gives background information and a suggested model. Section 3 gives a detailed review of the most important works in this field of study, which is followed by Section 4 which shows the results and discussion, and Section 5 brings the conclusion of our paper.

## 2 Related work

In the last few years, there has been a big rise in the amount of literature about depression. Choudhury and other researchers say that depression is a good way to check one's own and one's own family's well-being [4]. Every year, a lot of people get hurt because of their despair, but only a small number of them get the help they need. They also looked into the idea of using the internet to look for and look for signs of a serious depression problem in people who might be having it.

These people used web-based social networking sites like Facebook [16] and MySpace to keep track of their behavior credits for social engagement, emotion, dialect and semantic styles, sense of self-system, antidepressant medication notes, and how they felt.

As a possible tool for public health, a team of academics [5] looked to social media sites like Twitter to construct predictive models for how childbirth affects new mothers' behavior and disposition in the weeks and months that follow. They used Twitter tweets to gauge social engagement, mood, informal community, and phonetic style, among other things, in 376 mothers who had recently given birth. According to O'Dea and colleagues [6], the general public's psychological well-being status, including sadness and suicidal thoughts, is increasingly being studied via Twitter. They observed that their technique can determine the level of worry exhibited in suicide-related tweets using both human coders and a programmed machine classifier.

Individuals at high risk of suicide may be recognized via online networking sites such as micro blogs, according to research conducted by Zhang and colleagues [7]. A dynamic intervention system can be implemented to assist them in preserving their lives, according to Zhang and colleagues.

According to the findings of several studies, the effective usage of user-generated content (UGC) can assist in assessing an individual's level of psychological well-being (psychological well-being). In the case of social networking sites (SNS), Aldarwish and Ahmad [8] observed that the usage of SNSs is rising these days, particularly among those of a younger generation. With the accessibility of social media, customers are able to discuss their interests, sentiments, and daily routines, which assists in the development of trust and loyalty among them.

Nguyen and colleagues [9] employed machine learning and statistical methodologies to eliminate references to substances from the postings written by participants from the depression and control groups. This allowed them to distinguish between the two groups' online communications. They accomplished this by deleting topics such as temperament, psycholinguistic processes, and substance abuse from the posts that were produced by members of these groups.

Park et al. [10] conducted an evaluation of people's attitudes and actions toward online web-based social networking in order to identify whether or not they are discouraged from participating in the activity. A semi-

organized meet-up and personal interaction with 14 dynamic Twitter users was planned, with half of them being discouraged by the experience and the other half not being discouraged by it, according to the results. Additionally, they looked at a few plan implications for future social networks that may better suit users suffering from depression and supplied bits of information to aid discouraged individuals in dealing with their worries via online web-based social networking sites, among other things.

Ensemble learning is a method of learning that involves combining the opinions of several algorithms in order to get a final judgment that is considered to be the most intelligent [11]. It is not necessary to be concerned about selecting a single learning algorithm that does not perform well when using ensemble learning. An alternative approach is to bring together numerous distinct algorithms to build a smart group that can increase the performance of a single algorithm. When it comes to a wide range of scenarios, ensemble learning has been proved to be quite successful. It has been shown that ensemble learning may be utilized to come up with alternative strategies to foresee economic crises and make better financial decisions [13]. Ensemble learning has been utilized for land-cover mapping, as well as picture classification and other applications. Several studies have demonstrated that ensemble learning and deep learning are effective when used to speech-based emotion detection and face expression recognition. A similar approach to ensemble learning has been used to detect medical diagnoses in large, high-dimensional medical datasets [12]. The battery management system may also be used to determine how long a battery cell will survive. If you do not have a battery management system, you can use this tool to determine how long a battery cell will last. Despite this, the application of ensemble learning to forecast the capacity of a battery is still in its early stages of development [14]. The application of ensemble learning for estimating the capacity of Li-ion batteries while they are charging is the first time, to our knowledge, that this has been done.

S.No	Authors	Problem with identified
1	Seebohm, P, Chaudhary	Serious depression problem in people who used web-based social networking sites like Facebook and MySpace
2	Hagg, E., Dahinten, V. S., & Currie	Twitter to construct predictive models for how childbirth affects new mothers' behavior and disposition in the weeks and months that follow

3	Seabrook, E. M., Kern, M. L	Determine the level of worry exhibited in suicide-related tweets using both human coders and a programmed machine classifier
4	Huang, X., Li, X., Liu, T	Suicide may be recognized via online networking sites such as micro blogs
5	Latif, A. A., Cob, Z. C	With the accessibility of social media, customers are able to discuss their interests, sentiments, and daily routines , which assists in the development of trust and loyalty among them
6	Islam, M., Kabir, M. A.,	Temperament, psycholinguistic processes, and substance abuse from the posts that were produced by members
7	Park, M., McDonald, D.,	Evaluation of people's attitudes and actions toward online web-based social networking

### 3 Proposed model

The Training algorithm takes in labeled data as an input and outputs ensemble classifier. Each classifier is trained independently (i.e., line 4) with the input data (line 5-7). After training the models the Ensemble voting algorithm takes in the trained models (H) and a test data point. Each of the test data is passed to the ensemble model (i.e., line 12). And the class with the majority vote is chosen as prediction (i.e., showed as > 2 in line 12)

Algorithm:

```

Algorithm Training
Input: training data D = {xi, yi}
Output: ensemble classifier H
# train each classifier (Logistic regression,
SVM, NN, Random Forest)
For m=1 to T do
Train hm on D
End for
Algorithm Ensemble Voting
Input: test data D = {xj}, ensemble model
H
Output: Majority voted class
For j=1 to n do
H(xj) > 2
End for
    
```

Algorithm 1: Ensemble classifier algorithm

The new method works by ensambling the models that were previously working as independent models.

Each of the models make prediction and the final decision is made by seeing if the majority of the models predicted a certain class. This method tries to take advantage of the predictions across the models and tries to lessen the errors made by each model by using a majority vote. Theoretically, this method should perform better than the individual models. Everything else in the code is exactly the same as the previous trainings.

The prediction is managed by the formula

$$\sum_i y_i - 2 \geq 0$$

Where  $y_i$  is the class prediction from model  $i$

### 4 Dataset and training model

Survey forms from 1092 people from around the world have been used to get this information. Data is gathered by posing questions on how much time users spend on various social networking sites and the difficulties they encounter when using such sites.

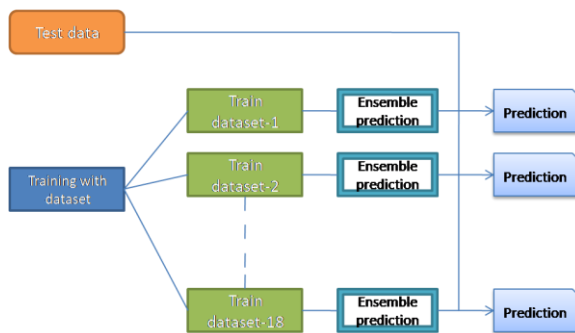


Figure 1: Pipeline schematic.

### 5 Results discussion

In this study, we employed an ensemble classifier, which is a mixture of four distinct prediction models: Logistic Regression, SVM, Random Forest, and FNN (Functional Neural Network) (Functional Neural Network). Python was selected as the programming language of choice for this particular project. After gathering all of the necessary information, we began by filtering it to remove any extraneous fields so that the model could use the information for training and testing reasons. To facilitate prediction analysis, we divided the dataset into two parts: the train dataset and the test dataset. The train dataset contains the training data, while the test dataset contains the testing data. As a result, we updated the test data set and training data set a total of 18 times in order to acquire accurate results, with average

performance serving as the final performance number for the perdition research.

Below table shows the improvement of our noble model to previous models

		with existing methods	Ensemble method	with existing methods	Ensemble method	with existing methods	Ensemble method
S.NO	Performan ce metrics	AD-1	AD-H	FOMO-1	FOMO-H	OBESITY -1	OBESITY -H
0	Accuracy	0.959817	0.96621	0.961644	0.968037	0.976256	0.974429
1	Precision	0.957629	0.980591	0.963561	0.977939	1	1
2	Recall	0.871343	0.888934	0.741563	0.826313	0.843449	0.850691
3	F1	0.91211	0.931818	0.836635	0.895713	0.911371	0.918871
4	Specificity	0.987976	0.993826	0.995855	0.996777	1	1

Table 1: Prediction performance table for ensemble to previous methods.

Figure 2 shows the comparison of accuracy between our previous work which predicts with existing model and ensemble prediction technique. There is clear improvement in proposed method which prediction between 96% to 100% for predicting problems with extensive social network use.

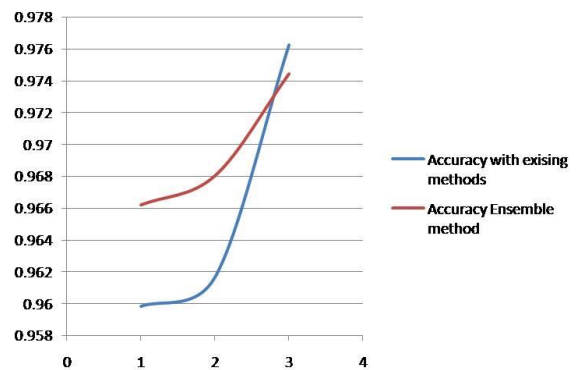


Figure 2: Accuracy comparison graph.

Figure 3 and 4 shows the comparison of precision and recall between our previous work which predicts with existing model and ensemble prediction technique. There is clear improvement in prediction. In proposed method precision varies between 98% to 100% for predicting problems with extensive social network use.

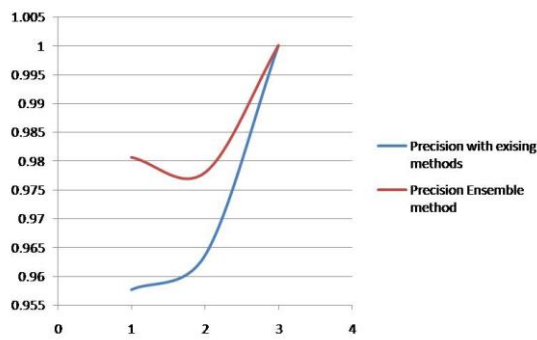


Figure 3: Precision comparison graph.

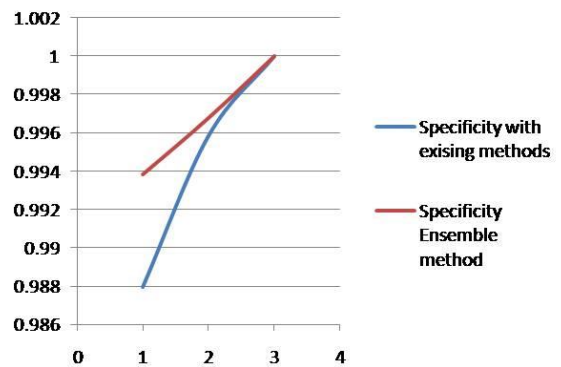


Figure 6: Specificity comparison graph.

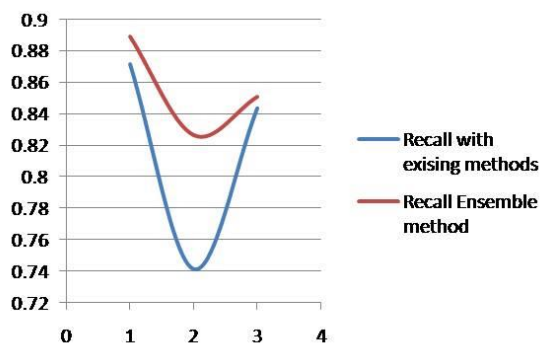


Figure 4: Recall comparison graph.

Figure 5 and 6 shows the comparison of F1-Score and Specificity between our previous work which predicts with existing model and ensemble prediction technique. There is clear improvement in Specificity of proposed method which prediction varies between 99% to 100% for predicting problems with extensive social network use.

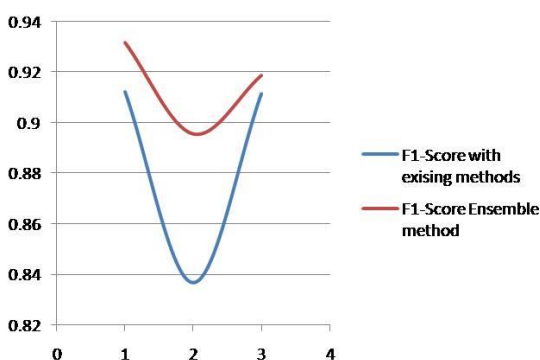


Figure 5: F1-Score comparison graph.

## 6 Conclusion

Specifically, under this research, an efficient prediction machine learning model, referred to as ‘ensemble model has been developed. This method comes handy in predicting the psychosomatic disorders associated with the length of time individuals spend on the social networking applications. For the purpose of the dataset a combination of models has been chosen to analyze, including support vector machines (SVMs), logistic regression (logistic regression), random forest (random forest classification), and neural networks (NN). The process entailed several iterations for ensuring accurate results by adjusting the data drawn from the train set and test set which required a long time. After carrying out substantial testing and verifying of the hypotheses, a comparison of the results of the ensemble approach with those results from the four unique performance measures that had been used earlier has been made, it is concluded that the ensemble model is showing an improvement of 96% to 98% in accuracy, 94% to 99% in precision, 98% to 100% in F1-score, and 94% to 100% in specificity. So ensemble method is the most effective one in anticipating the problems arising out of intensive exposure to social networks on our data set. These findings stand as the conclusive evidence to convict that this ensemble approach is the most effective in predicting the hazards facing the individuals as they get exposed to social networks for longer durations.

## References

- [1] Ali Taha, V., Pencarelli, T., Škerháčková, V., Fedorko, R., & Košíková, M. (2021). The use of social media and its impact on shopping behavior of Slovak and Italian consumers during COVID-19 pandemic. *Sustainability*, 13(4), 1710. <https://doi.org/10.3390/su13041710>
- [2] Thakur, R., & Rane, D. (2021). Machine Learning and Deep Learning for Intelligent and Smart Applications. In *Future Trends in 5G and 6G* (pp. 95-113). CRC

- Press. 6(1), 1-12. <https://doi.org/10.1007%2Fs13755-018-0046-0>
- [3] Manjunath, E Sreenivasa Reddy., (2021). Predicting Psychosomatic Disorders Arising from Intensive Exposure to Social Networks - Using Machine Learning Techniques. Unpublished manuscript. Acharya Nagarjuna University, Dr. YSR ANUCET.
- [4] Seebohm, P., Chaudhary, S., Boyce, M., Elkan, R., Avis, M., & Munn-Giddings, C. (2013). The contribution of self-help/mutual aid groups to mental well-being. *Health & social care in the community*, 21(4), 391-401. <https://doi.org/10.1111/hsc.12021>
- [5] Hagg, E., Dahinten, V. S., & Currie, L. M. (2018). The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review. *International journal of medical informatics*, 115, 92-105. <https://doi.org/10.1016/j.ijmedinf.2018.04.010>
- [6] Seabrook, E. M., Kern, M. L., & Rickard, N. S. (2016). Social networking sites, depression, and anxiety: a systematic review. *JMIR mental health*, 3(4), e5842. <https://doi.org/10.2196/mental.5842>
- [7] Huang, X., Li, X., Liu, T., Chiu, D., Zhu, T., & Zhang, L. (2015, October). Topic model for identifying suicidal ideation in chinese microblog. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation* (pp. 553-562). <https://aclanthology.org/Y15-1064>
- [8] Latif, A. A., Cob, Z. C., Drus, S. M., Anwar, R. M., & Radzi, H. M. (2021, December). Understanding Depression Detection Using Social Media. In *2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (Vol. 6, pp. 1-6). IEEE.
- [9] Islam, M., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health information science and systems*, 6(1), 1-12. <https://doi.org/10.1007%2Fs13755-018-0046-0>
- [10] Park, M., McDonald, D., & Cha, M. (2021). Perception Differences between the Depressed and Non-Depressed Users in Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 476-485. Retrieved from. <https://ojs.aaai.org/index.php/ICWSM/article/view/14425>
- [11] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- [12] Chaibub Neto, E., Pratap, A., Perumal, T. M., Tummalacherla, M., Snyder, P., Bot, B. M., ... & Omberg, L. (2019). Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ digital medicine*, 2(1), 1-6. <https://doi.org/10.1038/s41746-019-0178-x>
- [13] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. <https://doi.org/10.3390/risks7010029>
- [14] Aykol, M., Gopal, C. B., Anapolsky, A., Herring, P. K., van Vlijmen, B., Berliner, M. D., ... & Storey, B. D. (2021). Perspective—combining physics and machine learning to predict battery lifetime. *Journal of The Electrochemical Society*, 168(3), 030525. <https://doi.org/10.1149/1945-7111/abec55>
- [15] Kakar, S., Dhaka, D., & Mehrotra, M. (2021). Value-Based Retweet Prediction on Twitter. *Informatica*, 45(2). <https://doi.org/10.31449/inf.v45i2.3465>
- [16] Albayati, M. B., & Altamimi, A. M. (2019). An empirical study for detecting fake Facebook profiles using supervised mining techniques. *Informatica*, 43(1). <https://doi.org/10.31449/inf.v43i1.2319>