

Predicting Psychosomatic Disorders Arising from Intensive Exposure to Social Networks - Using Machine Learning Techniques



Manjunath Gadiparthi^{*}, Edara Srinivasa Reddy^{*}

Department of Computer Science & Engineering, ANU College of Engineering & Technology, Andhra Pradesh 522510, India

Corresponding Author Email: manjunath.gadiparthi@aau.edu.et

<https://doi.org/10.18280/ria.370105>

ABSTRACT

Received: 12 November 2022

Accepted: 7 January 2023

Keywords:

social network apps, machine learning, logistic regression, support vector machine, artificial neural network, Random Forest

This study forecasts a variety of problems that people would suffer as a result of the amount of time they spend on various social networking apps. The information is gathered from 1092 members with the purpose of anticipating the occurrence of certain problems. We have taken current machine learning paradigms and applied them to the data from our survey to see what results we get. For problem prediction, we used techniques like as Support Vector Machines, logistic regression, Random Forests, and neural networks (NN). On the basis of the data set obtained from the survey, we evaluated the performance of these four strategies. In light of the findings of the comparison, we propose that the test set and training set be changed in order to obtain the most efficient model for prediction. Using social networks to predict the problem, our analysis show that the performance of artificial neural networks is improved by 4% when social networks are used. In some circumstances, logistic regression is more efficient than nonparametric regression. Finally, we recommend the most effective prediction strategy, as well as how to train the model to get better outcomes.

1. INTRODUCTION

People's communication patterns have been considerably transformed as a result of the growth of social networking (SN) sites like as Facebook, LinkedIn, and Twitter. Predicting issues that will arise among a big number of users has become quite a difficulty as a result of this transformation. According to research, people who spend a momentous quantity of time on social media are added probable to suffer from issues such as social anxiety, depression, and the exposure to improper material. User behavior predictions based on data obtained from social network usage time are gaining popularity as a means of precisely forecasting user behavior. In contrast to traditional data, data acquired through SN is frequently unbalanced, which might make the prediction of user problems based on such data difficult to achieve.

Machine Learning algorithms will assist us learn about algorithms that learn, generalise, and predict from big data sets. This will be useful in the future, when we need to figure out how to use these algorithms. Calculative statistics and decision-making are closely connected in that machine learning is a component of both. Machine learning paradigms are employed in a variety of applications, like as forecasting the sales of a product, determining the likelihood of rainfall occurring in a specific location, and so on. Systematic analysis combined with machine learning algorithms will aid in the construction of prediction models for customized problems related to social network time spent, the monitoring of adverse events in users during their trial run, and the identification of the best prediction for each user. Machine learning approaches were worn to forecast the problem in this study, and we presented an integration framework for a decision support system to do this. The SVM kernel approach was used to

construct the decision support system for forecasting the problem, and the performance of the system was analyzed.

The article is prepared as follows: Section 1 give the introduction, Section 2 covers relevant work in problem prediction, Section 3 deals with the CNN model for social network problem prediction, and Section 4 deals with the implementation of a prediction model. In Section 5, the findings were discussed, and Section 6 ends with a discussion of the findings of this study effort.

These days, many people's daily lives revolve on their use of social media and messaging apps like Facebook and WhatsApp. The majority of social network mining research focuses on finding ways to improve people's lives by uncovering the wisdom in the data. Even while online social networks (OSNs) seem to increase users' capacity to build social relationships, they may actually reduce face-to-face encounters in the real world. Because of the increasing prevalence of this illness, new words such as Phubbing (Phone Snubbing) and Nomophobia (No Mobile Phone Phobia) have been coined to characterize persons who are unable to refrain from using mobile social networking applications.

It has only lately been recognized that people might suffer from social network mental diseases (SNMDs), such as Information Overload and Net Compulsion. 1. In accordance with statistics, one out of every eight Americans has difficulties while accessing the Internet. 2. Overuse, melancholy, social disengagement, and other negative consequences have been reported in journals such as the American Journal of Psychiatry [1]. A diagnosis of SNMDs is based on these symptoms, which include excessive use of social networking applications, which can lead to a loss of time or a lack of attention to essential needs, and withdrawal, which can include emotions of irritation and/or despair while the

computer/apps are unavailable. SNMDs are social in nature, and they are more likely to occur in those who often engage in online social media interactions. In order to compensate for the absence of face-to-face connection, those with SNMDs often turn to online relationships. Children's mental health is frequently a passive responsibility of their caregivers (such as teachers or parents). Because of the lack of physical risk factors, patients seldom seek out medical or psychological care. Consequently Thus, people will only seek medical help if their illnesses worsen to an unbearable degree. Suicide attempts have been found to be strongly linked to SNMDs in a recent research.

Teenagers who are addicted to social media have a substantially higher risk of suicide thoughts than those who aren't. Emotional status can be significantly impacted by social network addiction, which can lead to increased aggression, depression, and obsessive behavior. An even more disturbing fact is that people's social functioning can be severely harmed if early intervention is delayed. In a nutshell, early detection of SNMD users on OSNs is a desirable capability. SNMDs have been linked to a number of mental characteristics that have been studied in the past, however they are only used as diagnostic criteria in survey questionnaires. The extraction of these characteristics to measure the online mental states of OSN users is quite difficult. OSN users' loneliness and social inhibitions, for example, are difficult to see. 3 As a result, new methods for recognizing OSN users with SNMD are needed. We believe that utilizing an individual's social network data in addition to more traditional methods of psychological assessment offers a fantastic potential to catch these situations early on. In this study, we provide Social Network Mental Disorder Detection, a machine learning method for identifying SNMDs (SNMDD). Problems we considered are FOMO, anxiety and depression, obesity. Out of which obesity is physical health related issue and remaining all are psychological problems.

2. RELATED WORK

The use of various machine learning algorithms in clinical data analytics research is being done by numerous researchers in order to make predictions about various diseases. A variety of health care analytics research, chronic renal disease, diabetes, and cardiovascular disease, as well as numerous machine learning methods for analyzing clinical data, are covered in this area. According to the study [2], data analytics can have a positive impact on health care because of its ability to identify patterns and trends. Clinical trial design may be improved by the development of statistical and algorithmic models and algorithms, as detailed by the authors. Unwanted and noisy data will be removed from the dataset during data preparation.

Authors [3] discussed the challenges of preparing critical care data. The authors of this paper divided the difficulties in vital care paying attention in three categories: 1) data classification, 2) preprocessing, and 3) prediction. The central repository has all of the patient data that is maintained by the various hospitals. To improve prediction, writers Cooke and Iwashyna and Black and Payne suggested integrating databases with hospitals. The quality of the data is critical when building a predictive model. Outliers in the dataset can be removed using a variety of methods that have been described in Noura and Trabelsi, Imhoff [4].

Recently, there has been a lot of attention paid to online social networks and mental problems. For sentiment analysis and topic recognition, textual characteristics based on user-generated material (such as blogs and social media) are among the features that may be retrieved. This study uses an NLP-based technique to detect borderline personality disorder and bipolar disorder patients by analysing online social media for linguistic and content-based characteristics. The thematic and linguistic aspects of online social media are extracted by van der Velden [5] for the analysis of depression patients' patterns. Emotional and linguistic styles of social media information for Major Depressive Disorder (MDD) are examined by Choudhury. There has been a be short of of awareness paid to the structure of social networks and their potential Psychological aspects in prior study. The created approaches also do not have the capability of dealing with the sparse data that comes from numerous OSNs. A multi-source machine learning (STM) method is proposed to extract proxy characteristics in Psychology from OSN topologies such as Cyber-Relationship dependence and Net Compulsion, which need thorough investigation of the OSN topology.

In the majority of studies on the impacts of social media, researchers focus on how social media affects a person's behavior and character before and after extensive use [6-10]. To recently, researchers have simply observed and analyzed the problem of social media addiction without offering a cure. Using machine learning, this study intends to address the issue of social media addiction by providing beneficial activities and practices that can assist an individual avoid excessive use of social media. Since public media platforms like WhatsApp, Facebook, Instagram, and YouTube have grown in popularity over the past year, so has the study's methodology. In the last two years, the percentage of people who use social media has risen dramatically [11]. For this, we can thank Facebook and its peers for coming up with fresh approaches to boosting user engagement with the platform's software [12]. Naive Bayes and decision tree data mining techniques were utilized by Banu and Gomathy [13] to forecast various illnesses. Predictions of heart disease, diabetes, and breast cancer were the major focus. Confusion measures were used to generate the findings.

3. DATA SET AND ML METHODS

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 [14]. The following is a list of the survey questions that were used to gather the information for this report.

How much time do you spend on public media every day?
how much time you spend on viewing Facebook?
how much time you spend on viewing YouTube?
how much time you spend on viewing WhatsApp?
how much time you spend on viewing Telegram?
how much time you spend on viewing Tik Tok?
Problems faced if any due to use of application?

In response to above questions users responded with how much time they are spending on specific social network application and if any problems they are suffering. The collected data is fine tuned for machine learning technique to accept for train and test. Figure 1 below is a sample screen shot

of collected data for training and testing on different prediction models. The following table provides information regarding the hours spent using various social networks, as well as challenges encountered as a result of the use of social network apps. A score of '0' indicates that there are no health risks associated with using apps, while a score of '1' indicates that there are health risks associated with using apps.

User No.	SM time	Facebook	Youtube	Whatsapp	Telegram	anxiety and depression	FOMO	obesity
1	3	1	4	2	1	1	0	1
2	4	1	1	1	1	0	0	0
3	4	1	1	1	1	0	0	0
4	4	1	1	1	3	0	0	0
5	4	1	1	2	2	0	0	0
6	4	1	1	1	2	0	0	0
7	4	1	2	1	1	0	0	0
8	4	1	4	1	1	0	0	1
9	2	1	1	1	3	0	0	0
10	3	1	2	3	1	0	1	0
11	2	1	1	1	3	0	0	0
12	2	1	1	1	1	0	0	0
13	2	1	1	1	1	0	0	0
14	2	1	1	1	1	0	0	0
15	4	1	5	4	1	1	1	1
16	5	1	5	1	4	1	0	1
17	4	1	4	1	3	1	0	1
18	3	1	5	1	2	1	0	1

Figure 1. Sample picture of collected data set received from users for training the model

3.1 Correlation between the problems and social network apps

The first step is to look into the correlations that exist between social network apps and the components that serve as predictors. This is done with the intention of selecting which variables ought to be incorporated into our data set model. It has been demonstrated that a significant relationship exists between the prevalence of obesity and the use of YouTube. The relationship between the problems and their applications is illustrated in Figure 2. It is possible for obesity to result in a wide range of complications, some of which are serious in nature and have an impact on an individual's physical health. However, a correlation factor ranging from 0.43 to 0.52 indicates that the amount of time spent on public media platforms like Facebook, Telegram, and YouTube is a major risk factor in the prediction of anxiety and depression. When compared to other social networking programmes, Whatsapp has a lesser relationship with any ailment than any other social networking software, which means it poses a reduced risk to one's health. If the value of the correlation is close to one, this indicates that there is a high risk of developing health problems, whereas a value that is closer to zero indicates that the risk is lower.

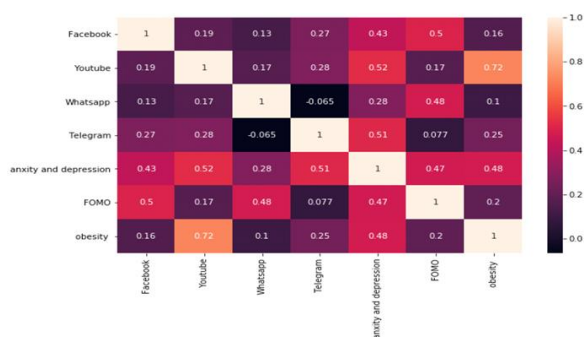


Figure 2. Correlation between the problems and social network applications

There is a correlation between one issue and the amount of time spent using other social networking applications (SN apps). It's possible that the common information shared throughout the many Social Network apps and groups of people connected through various Social Network apps is the aspect that has an effect. When compared with Telegram, for instance, which has a correlation value of 0.27, Facebook's influence is significantly weaker on YouTube, as seen by the fact that the correlation coefficient is only 0.19. In contrast, the correlation coefficient for Facebook's impact on Telegram is 0.27.

3.2 SVM model

Support vector machines (SVMs) are supervised machine learning techniques that are both powerful and versatile [15]. They are used in both classification and regression applications. However, in the vast majority of cases, they are used to resolve categorization difficulties. Statistical learning machines (SVMs) were initially presented in the 1960s, but they were modified even more in 1990. As opposed to other machine learning algorithms, SVMs feature a unique implementation method that sets them apart from the competition. It is because of their ability to handle enormous numbers of both continuous and categorical data that they have gained popularity in recent years. Based on support vector machines, a support vector machine (SVM) model represents multiple classes on an orthogonal hyper plane in multidimensional space. The SVM will construct the hyper plane on a number of occasions in order to limit the amount of error created to an absolute minimum. When using SVM, the goal is to divide datasets into classes in order to discover the hyper plane with the largest marginality.

Support Vectors are the data points that are closest to the hyper plane and are used to determine the distance between two points on a graph. Support Vectors are used to compute the distance between two points on a graph [16]. They will be divided by the data points that have been collected between the two of them. A margin is the distance between two lines drawn on the nearest data point of two different categories that are drawn on the same line. A simple way to compute it is to divide the perpendicular distance between the line and the support vectors by the number of support vectors there are. An increase in profit margin is regarded a positive margin in business; on the other hand, a decrease in profit margin is considered a negative margin.

The fundamental purpose of SVM is to division of datasets into classes in organize to locate the marginal hyperplane with the greatest amount of information. This may be accomplished by following the two steps mentioned in the following section:

- i. First, SVM will create hyper planes that best split the classes into subcategories repeatedly.
- ii. The hyperplane that best isolates the classes from one another will be selected by the algorithm.

SVM in Python: The standard libraries that will be loaded in order to implement SVM in Python are as follows: SVM Kernels are a split of machine learning kernels. In practice, the SVM algorithm is accomplished through the use of a kernel that turns an input data space into the format required by the method's needs. SVM is trained using a method known as the kernel trick, which involves converting a low-dimensional input space into a higher-dimensional space using a low-dimensional input space [17]. To put it another way, kernel transforms non-separable issues into separable problems by

introducing more dimensions to the problem domain. As a result of this improvement, SVM becomes more powerful, adaptive, and accurate. SVM uses a variety of kernels, some of which are listed below.

Linear Kernel refers to a kernel that is linear in nature.

It may be used to compute the dot product of any two observations by multiplying them together. The mathematical procedure for the linear kernel is as follow:

$$Kn(V, Vi) = sum(V, Vi) \tag{1}$$

As illustrated in the formula 1, the product of two vectors, v and v_i , may be expressed as the amount of the multiplications of every pair of contribution values.

Polynomial coefficients in a kernel.

It is a more sophisticated linear kernel that can discriminate between input areas that are curved and those that are nonlinear. The polynomial kernel formula is presented in the following example.

$$Kn(V, Vi) = 1 + sum(V * Vi)^{de} \tag{2}$$

This polynomial has a degree of d_e , which must be explicitly stated during the learning process in order for it to be able to function properly.

Radial Basis Function (RBF) is a sort of basis function (RBF) that converts input space into an indefinite dimensional space, which is why it is most usually employed for SVM classification. A mathematical explanation is provided by the following formula:

$$K_e(V, Vi) = e^{-Ga * sum(V - Vi)^2} \tag{3}$$

Ga is Gamm between 0 and 1 in this example. It must be expressly defined in the learning technique, which takes time. Gamma is set to 0.1 by default, which is an ideal starting point.

SVM may be built in Python for data with the intention of is not linearly discrete in the same way that it was built for linearly separable data. It is feasible to achieve this by using kernels.

3.3 Logistic regression model

The logistic function, as its name indicates, is the function that is at the heart of the logistic regression technique's operation. Invented by scientists in order to define the features of population increase in ecology, the logistic function, also known as the sigmoid function, includes characteristics such as population expansion that is outpaces the environment's carrying capacity [16]. By using an S-shaped arch, every real-valued number may be turned into a value between zero and one, but never into a value that is exactly within those two limitations.

$$\frac{1}{(1 + e_{val})} \tag{4}$$

In this equation, $eval$ denotes the natural logarithm base (Euler's number or the EXP() function in your spreadsheet), and $value$ is the numerical value that will be transformed to a natural logarithm. A representation of the values between -5 and 5 that have been changed to the range of 0 and 1 using the logistic function is shown in the following example: Please have a look at the picture below. The chance of being assigned

to the defaulting class is estimated via the use of a logistic regression model (e.g. the first class).

Isn't it true that we're generating probability forecasts? When we first heard about logistic regression, we believed it was a categorization procedure. Recall that in order to apply a probability forecast; it must first be translated hooked on a binary value, which can be found in the following table: (0 or 1). More on this later in the chapter, when we discuss how to make educated guesses.

When it comes to linear techniques, logistic regression is considered to be one of the most efficient. The logistic function is used to translate the forecasts into something more useful. A consequence of this is that predictions cannot be thought of as a linear grouping of the data inputs, as we might do in the case of linear regression. Continuing on from the previous paragraph, the model might be described as follows:

$$P_b(X) = \frac{e^{(C_0 + C_1 * V)}}{1 + e^{(C_0 + C_1 * V)}} \tag{5}$$

For simplicity, we may invert the previous equation as follows (remember that we can delete the e from one side of the equation by adding an exponential logarithm (\ln) to the other side):

$P(X)$ is half the square root of the square root of the square root. Because the input on the left is a log of the probability of the default class, the output on the right is also a log (which is beneficial for classifying data).

The ratio on the right shows the likelihood of the default class (we have been using odds for a long time; for example, odds are used in horse racing rather than probabilities since it is more historical). The probabilities of an event occurring are computed as follows: $0.8 / (1 - 0.8) = 4$. Alternatively, we might type:

$$O_{ds} = C_0 + C_1 * V_{ln}(odds) \tag{6}$$

Because the chances are logarithmic, the left side is called the log-odds, while the right is called the probability. Since various functions might be used for the transform (beyond the scope of this article), the transform that links the linear regression equation to the probabilities is called the link function. If the exponent is moved to the right, it becomes:

$$O_{ds} = e^{(C_0 + C_1 * X)} \tag{7}$$

where, C_0 and C_1 are the initial values.

As a result of this, we can see that the approach remains a linear grouping of the data inputs, but that the linear blend is tied to the default group's log-odds.

3.4 Random Forest

In machine learning, a supervised learning strategy known as a Random Forest, which is also known as an artificial neural network (ANN), is used for classification and regression. But it is widely utilised in the solution of categorization issues, which is why it is so popular [18]. To put it simply: a forest is made up entirely of trees, and a forest comprised entirely of more trees is a more vigorous forest, as we all know. The Random Forest algorithm also performs other tasks, such as creating decision trees from a collection of data samples, extracting the prediction from each of these decision trees, and eventually selecting the best answer using a procedure known

as vote selection [19]. It is decided to employ an ensemble technique rather than a single decision tree. Instead of a single decision tree, averaging the results prevents over-fitting and reduces false positives, making it a more desirable method than one that uses just one decision tree. The Random Forest Method is a sort of optimization algorithm that uses a Random Forest to find the best solution.

It is possible to obtain a knowledge of the Random Forest algorithm's functioning by following the steps outlined below. Choosing random samples from a given dataset and organising them in a separate data table is the first stage in the process. Afterwards, this algorithm will construct a decision tree for each sample that is received in the second stage of the process. Step 3: Make a list of all of the things you want to do. In the next stage, it will get the prediction results from each decision tree in the system. This stage will comprise voting for each and every projected result, which will be carried out in the third phase once the vote has concluded. Finally, in the fourth phase, choose the forecast result that received the most number of votes as the final prediction result.

3.5 Feed forward neural network

The artificial neural network's connections do not create loops. A Feed Forward Neural Network. Like a feed forward neural network, it has certain pathways that recur. This is a recurrent neural network. The feed forward approach simply looks at one item at a time, making it the simplest to grasp. Because data may pass via numerous hidden nodes, it always flows one way.

A Feed Forward Neural Network is most often seen as a one layer preceptor, the most basic kind of neural network. To reach the ultimate outcome, people may input a model several items and then multiply the weights. To get the sum of the weighted input values, add them all together. The output value is normally one if the total of the values is greater than a threshold, which is usually zero [20]. If the total is below the threshold, the result is normally zero. The single layer perceptron is a popular form of feed forward neural networks used for classification. Single layer perceptrons may also utilise machine learning components. The delta rule allows the neural network to compare node outputs to predetermined values. This enables the network to adjust its weights to improve accuracy. So a gradient descent pattern forms. The process for adjusting multi-layered perceptrons' weights is almost same. Back-propagation is the term used here. Because each hidden layer in the network changes when the previous layer's output changes.

Because of its simplicity, Feed Forward Neural Networks may be suitable for certain machine learning applications. A moderate intermediate layer may exist between two or more feed forward neural networks that function independently. Like the human brain, this process requires more neurons to manage and perform more. Because each network operates independently, the findings may be integrated to form a single, cohesive product.

The data obtained is used to train four dissimilar machine learning paradigms: logistic regression, support vector machine, Random Forest, and fuzzy neural networks. The data collected is used to train four different machine learning models. Figure 3 clearly demonstrates the flow of our model for prediction. The data set has been separated into two groups: the train set and the test set. Eighty percent of the complete data is provided as train set data, while the remaining twenty

percent is provided as test data set. We modified the train set data and test set data for a total of 18 times in order to acquire better performance results. Figure shows our model of prediction from the data set we predict whether a person suffers from some problems or not. The output is either suffering from any problem (pos) or healthy (Neg).

There are only four possible scenarios in which person may find themselves. The following will serve as a guideline for our discussion.

- TP-True positive: Prediction is positive and person is facing problem
- TN-True negative: Prediction is negative and person is healthy
- FP-False positive: Prediction is positive and person is healthy, *counterfeit alarm, terrible*
- FN-False negative: Prediction is negative and person is diabetic, *the most horrible*

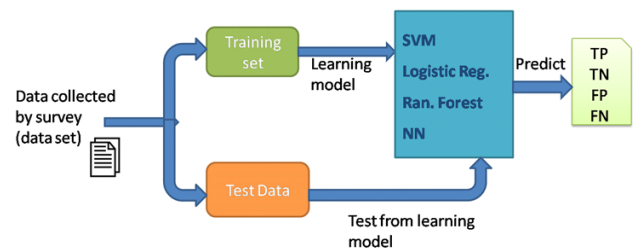


Figure 3. Flow of prediction model for finding different problems

4. RESULTS

In our studies, we used four different prediction models: Logistic Regression, SVM, Random Forest, and FNN (Functional Neural Network). Python was chosen as the programming language for this project. We started by filtering the acquired data to remove any superfluous fields so that the model could use the information for training and testing purposes. For the purpose of prediction analysis, we have separated the dataset into two parts: one is the train dataset and the other is the test dataset. In order to obtain correct findings, we altered the test data set and training data set a total of 15 times, with the average performance being used as the final performance value for the prediction study. Every iteration has resulted in a different calculation of the recall, precision, accuracy F1-score, and specificity of all models.

Accuracy: The percentage of correctly labelled subjects in relation to the total number of subjects is defined as the accuracy of the data. It is the one that comes to mind first. Accuracy provides a response to questions such as, "How many people did we accurately categorise out of all the people in the dataset?". Figure 4 and Table 1 show the accuracy performance of different techniques.

$$\text{Accuracy} = (TN + TP) / (TP + FP + FN + TN) \quad (8)$$

Precision: Precision is defined as the ratio of the correctly +ve labelled by our algorithm to the total number of correctly +ve labelled. Answers to questions such as How many of people who we labelled as troublesome are genuinely suffering from a problem are provided with pinpoint accuracy.

$$\text{Precision} = TP / (TP + FP) \quad (9)$$

Numerator: +ve people who have a problem are in this group.

Denominator: +ve people who have been labelled as such by our programme, whether or not they're actually in pain or not. Figure 5 and Table 2 shows the precision performance of dissimilar techniques.

Table 1. Accuracy of LR, SVM, RF, and NN vs AD, FOMO, OBESITY

	Model	AD	FOMO	OBESITY
0	Logistic Regression	0.959817	0.961644	0.975342
1	Support Vector Machine	0.946119	0.952511	0.972603
2	Random Forest	0.860274	0.919635	0.976256
3	Neural Network	0.959817	0.961644	0.964384

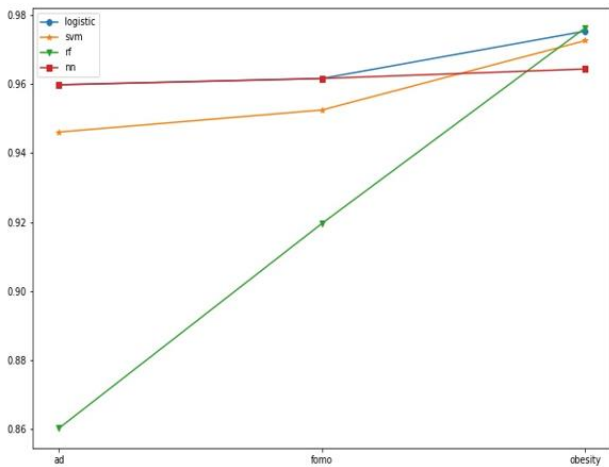


Figure 4. Accuracy of different prediction models with respect to problems

Table 2. Precision of LR, SVM, RF, and NN vs AD, FOMO, OBESITY

	Model	AD	FOMO	OBESITY
0	Logistic Regression	0.957629	0.963561	0.993103
1	Support Vector Machine	0.923241	0.913217	1.000000
2	Random Forest	0.828590	0.830065	1.000000
3	Neural Network	0.957629	0.963561	1.000000

Recall: When our programme correctly labels someone as positive, this is called "recall." This is the percentage of people who are actually having a problem that our programme correctly labels as positive. It tells us how many people who have a problem, out of all of them, did we correctly guess. Figure 6 and Table 3 show the recall presentation of different techniques.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) / \text{TP} \quad (10)$$

Numerator: True positive anyone with any problem.

Denominator: All people who are having any kind of pain (whether detected by our programme or not).

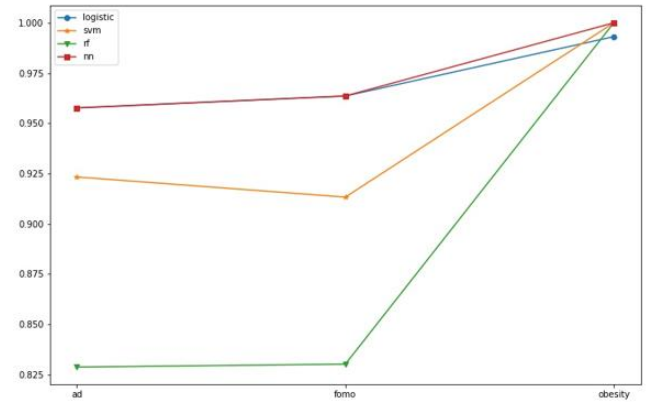


Figure 5. Precision of different prediction models with respect to problems

Table 3. Recall of LR, SVM, RF, and NN vs AD, FOMO, OBESITY

	Model	AD	FOMO	OBESITY
0	Logistic Regression	0.871343	0.741563	0.843449
1	Support Vector Machine	0.848107	0.714718	0.819374
2	Random Forest	0.530305	0.501241	0.843449
3	Neural Network	0.871343	0.741563	0.766800

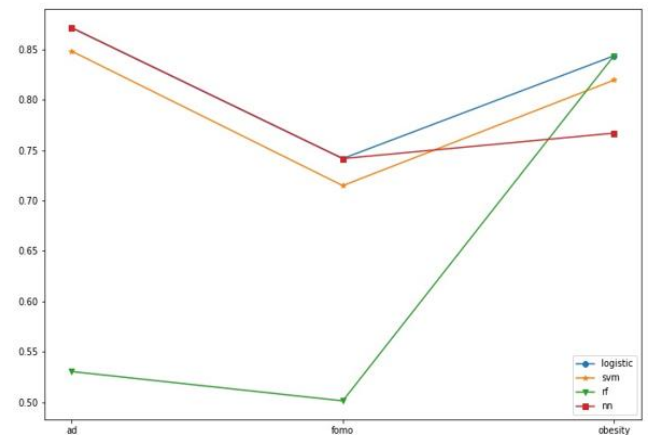


Figure 6. Recall of different prediction models with respect to different problems

F1-score: The F1 Score takes into account both accuracy and recall. It is the harmonic mean (mean average) of the accuracy and recall measurements. Precision (p) and recall (r) should be balanced in the system in order for the F1 Score to be optimally calculated. If one metric is enhanced at the cost of the other, the F1 Score is not as high as it would otherwise be. Example: If P is 1 and R is 0, the F1 score equals 1. Figure 7 and Table 4 show the F1-Score of different techniques.

$$F1^{Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (11)$$

Table 4. F1-Score of LR, SVM, RF, and NN vs AD, FOMO, OBESITY

	Model	AD	FOMO	OBESITY
0	Logistic Regression	0.912110	0.836635	0.911371
1	Support Vector Machine	0.883819	0.799611	0.900362
2	Random Forest	0.645562	0.623713	0.914431
3	Neural Network	0.912110	0.836635	0.865494

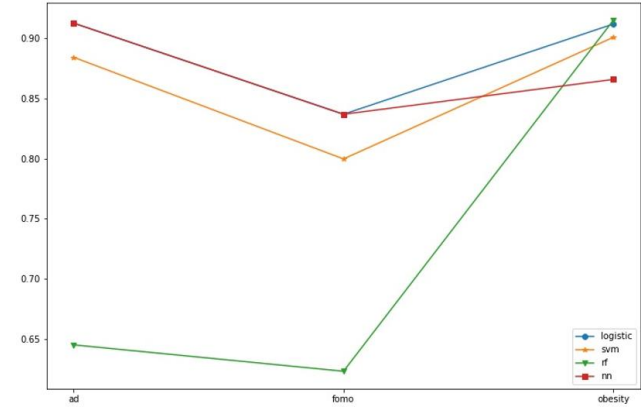


Figure 7. F-1 Score of different prediction models with respect to problem faced due to social network apps

Specificity: Everyone who is healthy in real life gets labelled with specificity by the computer, and this is exactly how it should be done. Fill up the blanks with your responses to the following question. People who are in good health include: How many of these did we believe we were familiar with? Specificity is defined as the ratio of the numerator to the denominator (TN + FP): The healthy folks I've referred to as "truly healthy people" are those who aren't labelled as either "plus" or "minus" in any way. Figure 8 and Table 5 shows the effectiveness of specificity for different techniques.

Table 5. Specificity of LR, SVM, RF, and NN vs AD, FOMO, OBESITY

	Model	AD	FOMO	OBESITY
0	Logistic Regression	0.987976	0.995855	0.99893
1	Support Vector Machine	0.976964	0.989527	1.00000
2	Random Forest	0.965073	0.984257	1.00000
3	Neural Network	0.987976	0.995855	1.00000

4.1 Experimental results and analysis

Figures 9 and 10 are simply a comparison of two graphs. The classes are classified in two ways: The number 1 indicates that there are concerns with time spent on social media networks, i.e., if the yellow scatters throughout the graph, it is class 1; otherwise, it is class 2. The second-class designation is for the individual who has no difficulties, as shown by the number 0, i.e., if the black does not scatter on the graph, that individual is in class zero. When viewing the graph, the x-axis

indicates the number of values shown from the test data sampled from the full dataset. On the y-axis, the class labels of the datasets used to generate the graph are displayed. On the left side of the graph, in the left portion, the charting of the class labels of real test data used to validate the trained model is shown. The produced model/projection algorithm's of class labels onto the same data set used to evaluate the generated model is presented in the right section of the graph to the right of the graph.

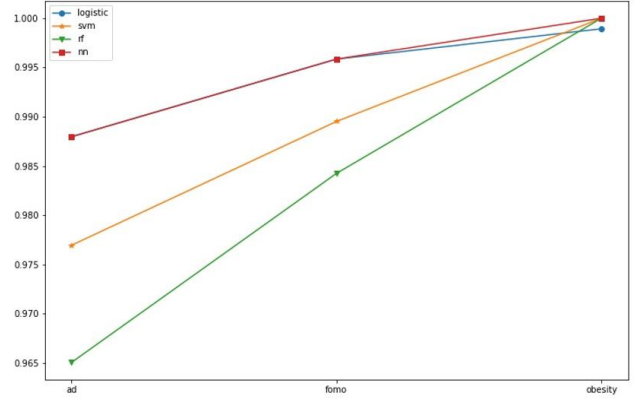


Figure 8. Specificity of different prediction models with respect to social network problems

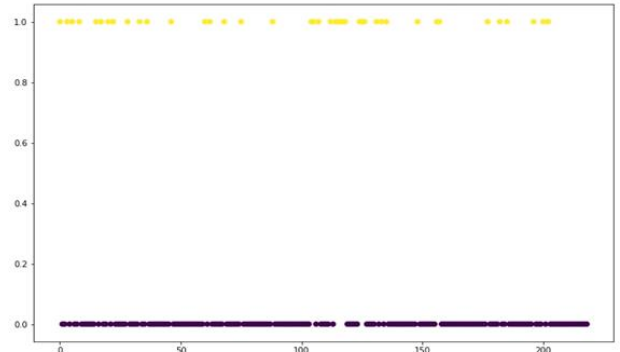


Figure 9. Comparison of prediction to actual test data sample -1

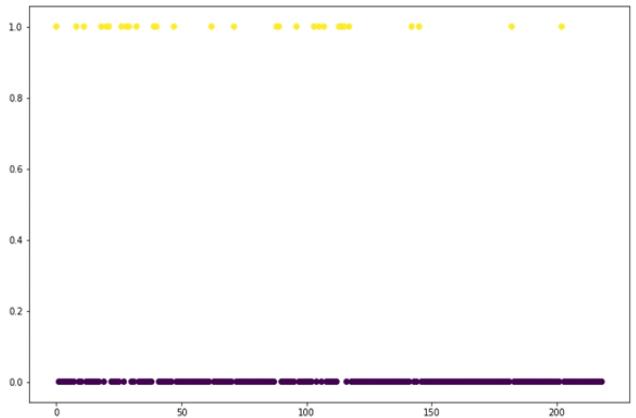


Figure 10. Comparison of prediction to actual test data sample-2

In this study, we offer an effective prediction machine learning model for anticipating difficulties in relation to the quantity of time spent on public networking sites (social

network time). For our data set, we have picked models such as Logistic Regression, SVM, Random Forest, and FNN (Functional Neural Network) to analyse. We have performed a lot of iterations in order to obtain correct findings by altering the data from the train set and test set. We looked at the findings of five distinct performance metrics after doing significant experimentation and validating our hypotheses. As a consequence of the findings, we conclude that a FNN and Logistic Regression are the most effective models for forecasting difficulties in social networks. As a possible extension of this research, it is likely that the results will be used in social media apps, with the user being informed of any potential problems that may arise as a result of spending more time on a given application.

5. CONCLUSION

In this study, we offer an effective prediction machine learning model for anticipating difficulties in relation to the quantity of time spent on public networking sites (social network time). For our data set, we have picked models such as Logistic Regression, SVM, Random Forest, and FNN (Functional Neural Network) to analyse. We have performed a lot of iterations in order to obtain correct findings by altering the data from the train set and test set. We looked at the findings of five distinct performance metrics after doing significant experimentation and validating our hypotheses. As a consequence of the findings, we conclude that a FNN and Logistic Regression are the most effective models for forecasting difficulties in social networks. As a possible extension of this research, it is likely that the results will be used in social media apps, with the user being informed of any potential problems that may arise as a result of spending more time on a given application.

REFERENCES

- [1] uran, N., Özdemir Aydın, G., Kaya, H., Aksel, G., Yılmaz, A. (2019). Male nursing students' social appearance anxiety and their coping attitudes. *American Journal of Men's Health*, 13(1): 1557988319825922.
- [2] Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J. (2019). Dice loss for data-imbalanced NLP tasks. *arXiv preprint arXiv:1911.02855*. <https://doi.org/10.48550/arXiv.1911.02855>
- [3] Prabhakaran, D., Jeemon, P., Sharma, M. et al. (2018). The changing patterns of cardiovascular diseases and their risk factors in the states of India: The global burden of disease study 1990–2016. *The Lancet Global Health*, 6(12): e1339-e1351. [https://doi.org/10.1016/S2214-109X\(18\)30407-8](https://doi.org/10.1016/S2214-109X(18)30407-8)
- [4] Harimoorthy, K., Thangavelu, M. (2021). Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. *Journal of Ambient Intelligence and Humanized Computing*, 12(3): 3715-3723. <https://doi.org/10.1007/s12652-019-01652-0>
- [5] van der Velden, P.G., Setti, I., van der Meulen, E., Das, M. (2019). Does social networking sites use predict mental health and sleep problems when prior problems and loneliness are taken into account? A population-based prospective study. *Computers in Human Behavior*, 93: 200-209. <https://doi.org/10.1016/j.chb.2018.11.047>
- [6] Talwar, S., Dhir, A., Singh, D., Virk, G.S., Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. *Journal of Retailing and Consumer Services*, 57, 102197.
- [7] Yankah, S., Adams, K.S., Grimes, L., Price, A. (2017). Age and online social media behavior in prediction of social activism orientation. *The Journal of Social Media in Society*, 6(2): 56-89.
- [8] Kuss, D.J., Griffiths, M.D. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health*, 14(3): 311. <https://doi.org/10.3390/ijerph14030311>
- [9] van Rooij, T., Ferguson, C.J., Van de Mheen, D., Schoenmakers, T.M. (2017). Time to abandon internet addiction? Predicting problematic internet, game, and social media use from psychosocial well-being and application use. *Clinical Neuropsychiatry*, 14(1): 113-121.
- [10] Coyne, S.M., Padilla-Walker, L.M., Holmgren, H.G., Stockdale, L.A. (2019). Instagrowth: A longitudinal growth mixture model of social media time use across adolescence. *Journal of Research on Adolescence*, 29(4): 897-907. <https://doi.org/10.1111/jora.12424>
- [11] Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J., Yeo, B.T.T. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. In *NeuroImage*, 222: 117203. <https://doi.org/10.1016/j.neuroimage.2020.117203>
- [12] Khobzi, H., Lau, R.Y., Cheung, T.C. (2019). The outcome of online social interactions on Facebook pages: A study of user engagement behavior. *Internet Research*, 29(1): 2-23.
- [13] Banu, M. N., Gomathy, B. (2014). Disease forecasting system using data mining methods. In *2014 International Conference On Intelligent Computing Applications*, Coimbatore, India, pp. 130-133. <https://doi.org/10.1109/ICICA.2014.36>
- [14] Manjunath, G., Reddy, E.S. (2022). Impact of individuals' engagement in social network-an extensive analysis. *Webology*, 19(1): 2782-2796. <https://doi.org/10.14704/WEB/V19I1/WEB19185>
- [15] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408: 189-215. <https://doi.org/10.1016/j.neucom.2019.10.118>
- [16] Ahmad, I., Basher, M., Iqbal, M.J., Rahim, A. (2018). Performance comparison of support vector machine, Random Forest, and extreme learning machine for intrusion detection. *IEEE Access*, 6: 33789-33795. <https://doi.org/10.1109/ACCESS.2018.2841987>
- [17] Puspaningrum, A.P., Endah, S.N., Sasongko, P.S., Kusumaningrum, R., Ernawan, F. (2020). Waste classification using support vector machine with SIFT-PCA feature extraction. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, pp. 1-6. <https://doi.org/10.1088/1757-899X/288/1/012042>
- [18] Balyan, A.K., Ahuja, S., Lilhore, U.K., Sharma, S.K., Manoharan, P., Algarni, A.D., Elmannai, H., Raahemifar, K. (2022). A hybrid intrusion detection model using egapso and improved Random Forest method. *Sensors*, 22(16): 5986. <https://doi.org/10.3390/s22165986>

- [19] Geetha, R., Sivasubramanian, S., Kaliappan, M., Vimal, S., Annamalai, S. (2019). Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using Random Forest classifier. *Journal of Medical Systems*, 43(9): 1-19. <https://doi.org/10.1007/s10916-019-1402-6>
- [20] Popoola, S.I., Adetiba, E., Atayero, A.A., Faruk, N., Calafate, C.T. (2018). Optimal model for path loss predictions using feed-forward neural networks. *Cogent Engineering*, 5(1): 1444345. <https://doi.org/10.1080/23311916.2018.1444345>