International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

The Psychology behind Users Mental Health in Social Media

D. Sridhar[1*], V. Kathiresan[2]

[1*]Department of Computer Science, Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, India
[2]Department of Computer Applications(PG), Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, India

[*]Corresponding Author: ursridhardeva@gmail.com

Available online at: www.ijcseonline.org

Accepted: 19/May/2018, Published: 31/May/2018

Abstract— New research uncovers how social network information can be utilized to predict users' psychological and physical wellbeing, adding to a developing number of researchers utilizing social media to make startlingly precise forecasts from the most basic data. The same number of users think what they share on the web, the discoveries give occasion to feel qualms about customary thoughts of "safe" surfing. In spite of the fact that rates of diagnosing mental illness have enhanced in the course of recent decades, numerous cases stay undetected. Side effects related with mental illness are noticeable on Twitter, Facebook, and web forums, and automated techniques are progressively ready to identify misery and other psychological instabilities. In this paper, late investigations that expected to anticipate psychological sickness utilizing web-based social networking are explored. . Mentally ill users have been observed utilizing screening reviews, their open sharing of a determination on Twitter, or by their participation in an online discussion, and they were discernable from control users by designs in their language and online activity. Automated detection techniques may recognize discouraged or generally in danger people through the vast scale passive monitoring of social media, and in the future may complement existing screening methodology.

Keywords— Social Media, Prediction, Classification, Image, Probabilities, Decisiontree

1. INTRODUCTION

Web-based social networking can be a fun and sound action if users take advantage of the site to remain associated with family and old companions, friends and to share interesting and vital parts of their life incidents. The far reaching utilization of social media may give chances to help lessen undiscovered mental illness. A developing number of studies inspect mental health inside web-based social media settings, connecting social media use and behavioural examples with stress, tension, depression, sociality and other psychological instabilities. The best number of investigations of this kind spotlight on depression. Depression keeps on being under-diagnosed, with generally a large portion of the cases distinguished by essential care doctors and just 13% - 49% getting negligibly satisfactory treatment. Automated investigation of online networking conceivably gives strategies to early recognition. If that a robotized procedure could distinguish detect elevated depression scores in a user, that individual could be focused for a more careful evaluation, and gave further resources, support, and treatment. Studies to date have either inspected how the utilization of online social media destinations relates with psychological illness in users or endeavoured to observe dysfunctional behaviour through investigation of the content made by users. This study centres on the last mentioned: examines went for anticipating psychological maladjustment utilizing online social media. We initially consider techniques used to anticipate depression, and after that consider four methodologies that have been utilized as a part of the study. We look at the different methodologies, give guidance for future investigations, and consider moral issues.

2. PREDICTION PARAMETERS

Automated analysis of social media is built by building predictive models, which use 'features,' or variables that have been extracted from social media data. For example, commonly used features include users' language encoded as frequencies of each word, time of posts, and other variables. Features are then treated as independent variables in an algorithm (Linear Regression with built in variable selection, or Support Vector Machines (SVM) to predict the dependent variable of an outcome of interest (users' mental health). Predictive models are trained using an algorithm, on part of the data and then are evaluated on the other part to avoid over fittinga process called cross-validation. The prediction performances are then reported as one of several possible metrics. The following metrics are also used to measure the mental health of a social media user.

- Type of posts
- Type of profile pictures
- Number of likes

- Check-ins
- Posting frequency
- Colours used.

3. ASSESSMENT CRITERIA

Utilizing Social media data from 100 people, we connected machine learning apparatuses to effectively distinguish markers of depression. Measurable highlights were computationally extricated from 100 member Facebook photos, utilizing colour analysis, metadata segments, and algorithmic face identification **(Christopher T. Barry, Chloe L, 2017).**Coming about models outperformed general experts' normal unassisted analytic achievement rate for depression. These outcomes held notwithstanding when the examination was confined to posts made before depressed people were first analysed. Human appraisals of photograph attributes (happy, sad, etc.) were weaker indicators of despondency, and were uncorrelated with computationally-created features. These outcomes propose new roads for early screening and detection of mental illness **(Mike Conway, Daniel O'Connor, 2016)**.

Photos presented on social media offer a huge swath of features that may be broke down for psychological understanding. The substance of photos can be coded for any number of attributes: Are there individuals exhibit? Is the setting in nature or inside? Is it night or day? Picture measurable properties can likewise be assessed at a for every pixel level, including values for normal colour and brightness. Social media metadata offers extra data: Did the photograph get any remarks? What number of 'likes' did it get? At last platform activity measures, for example, utilization and posting frequency, may likewise yield pieces of information as to a social media user's mental state **(Sandra L. Frits, KidsAnxiety, 2018)**. We fused just a restricted subset of conceivable highlights into our prescient models, spurred to some degree by earlier research into the connection amongst state of mind and visual inclinations.

Individuals with low self-esteem are more likely to share on Facebook than in person, but because their status updates tend to express more negative thoughts, they are perceived as less likable.People with low self-esteem are also more likely to feel insecure in their romantic relationships, and consequently are more likely to post about their partners as a way to boost their self-worth and refute others' impressions that their relationship is poor **(Simon M. Rice, Rosemary Purcell, 2018)**.

People have different reasons for using their mobile device to check-in to a location. For some, their reason can be as simple as just wanting to have a record of places they've visited on their vacation **(Gillian Fergie, Kate Hunt, 2016)**. For others, their reason could be so they can earn rewards such as badges, mayorship, or to redeem some special offer or discount from participating businesses. Finally for some, it could be the sense of achievement or accomplishment for being where they say they've checked-in to. (After all, checking-in at the peak of Mt. Everest is sure to result in many Likes, comments, and bragging rights!).

The potential to get responses from our friends is a major motivator of social media activity, and most of us provide plenty of feedback to our online friends. A survey of Facebook users found that "liking" friends' posts was a common activity, with 44% of users saying they liked friends' content on a daily basis. And these likes and comments may be a major reason why people post content on sites, such as Facebook **(Eugene Brusilovskiy, Greg Townley, 2016).** That same survey found that 16% of men and 29% of women felt that receiving support from others was a major reason they used Facebook, and approximately 16% of all users agreed that getting feedback on postings was a primary motivator of their use of the site.

In studies associating mood, colour and mental wellness, healthy people distinguished darker, dim colours with antagonistic temperament, and by and large preferred brighter, more vivid colours. By differentiate, depressed people were found to incline toward darker, dim colours. Also, Barrick, Taylor, and Correa found a positive connection between self-identification with depression and an inclination to see one's surroundings as dim or grey or lacking in colour (**Ang Li, Dongdong Jiao, Tingshao Zhu, 2018)**.These discoveries inspired us to incorporate measures of hue, immersion, and shine in our investigation. We likewise followed the utilization of social media channels, which enable users to modify the colour and tint of a photo.

Depression is firmly connected with decreased social action. As Facebook is utilized to share personal experiences, it is sensible to derive that posted photographs with individuals in them may catch parts of a user's social life. On this commence, we utilized a face detection algorithm to dissect social media posts for the presence and number of human faces in each photo. We likewise tallied the quantity of remarks and likes each post got as measures of group engagement, and utilized posting recurrence as a metric for user engagement **(Renwen Zhang, 2017)**.

4. PREDICTION METHODS

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, we construct a frequency table for each predictor against the target. It has been shown that OneR produces rules only slightly less accurate than state-of-the-art classification algorithms while producing rules that are simple for humans to interpret.

Algorithm 1: OneR Algorithm
For each predictor, *For each value of that predictor, make a rule as follows;* *Count how often each value of target (class) appears* *Find the most frequent class* *Make the rule assign that class to this value of the predictor* *Calculate the total error of the rules of each predictor* *Choose the predictor with the smallest total error.*

Table 1: OneR algorithm's Classification based on Images/Post

Colour	Types of Posts	Posting frequency	Check-ins	Score	Mental status	Problem
Bright	Violence	High	High	2	Depression	Yes
Bright	Events	Normal	Normal	6	Happy	No
Shadow	Sad	Low	High	3	Sad	Yes
Mild	Conceptual	Normal	Normal	4	Neutral	No
Dark	Sad	Low	Low	2.5	Depression	Yes
Dark	Violence	High	Low	1.8	Depression	Yes
Bright	Events	Normal	Normal	4.5	Happy	No
Bright	Foods	High	High	5	Happy	No
Bright	Events	High	Normal	7	Happy	No
Shadow	Conceptual	Normal	Normal	6	Happy	No
Shadow	Violence	High	High	4	Depression	Yes
Mild	Conceptual	Low	Normal	5	Neutral	No
Dark	Violence	High	Low	3.2	Depression	Yes
Bright	Events	High	High	6	Happy	No
Shadow	Sad	Low	Low	3.6	Sad	Yes
Bright	Events	High	Normal	5.6	Happy	No

Table 2: Frequency Tables of image classification probabilities (A to E)

(A)		Problem	
		Yes	No
Types of Posts	Violence	4	0
	Conceptual	0	2
	Sad	2	0
	Events	0	5

(B)		Problem	
		Yes	No
Types of Colour	Bright	1	6
	Dark	3	0
	Shadow	1	2

(C)		Problem	
		Yes	No
Posting frequency	High	4	4
	Low	3	1
	Normal	3	3

(D)		Problem	
		Yes	No
Check-ins	High	0	3
	Low	3	1
	Normal	2	3

(E)		Problem	
		Yes	No
Score	High (7 to 10)	0	7
	Low (1 to 4)	4	0
	Normal (4 to 7)	0	3

		Problem	
		Yes	No
Types of Posts	Violence	4	0
	Conceptual	0	2
	Sad	2	0
	Events	0	5

Table: Illustration of the best predictor

IF Type of Posts = Violence THEN Problem = Yes

IF Type of Posts = Conceptual THEN Problem = No

IF Type of Posts = Sadness THEN Problem = Yes

IF Type of Posts = Events THEN Problem = No

Predictors Contribution: Simply, the total error calculated from the frequency tables is the measure of each predictor contribution. A low total error means a higher contribution to the predictability of the model.

The following confusion matrix shows significant predictability power. OneR does not generate score or probability, which means evaluation charts (Gain, Lift, K-S and ROC) are not applicable.

Table 3: Confusion Matrix

Confusion Matrix		Problem			
		Yes	No		
One R	Yes	5	2	Positive Predictive Value	0.71
	No	3	6	Negative Predictive Value	0.75
		Sensitivity	Specificity	Accuracy = 0.79	
		0.63	0.75		

5. C4.5 ALGORITHM

In our previous paper we used Naïve Bayes algorithm for predicting the mental health status of social media users. In this paper we use C4.5 algorithm for predicting mental health status of social media users. C4.5 constructs a classifier in the form of a decision tree. In order to do this, C4.5 is given a set of data representing things that are already classified. Our data set contains bunch of online users and we know we know various things about each users like age, occupation, gender, frequency of posts, type of posts, check-ins which is called as attributes.Given these attributes need to predict mental health status. The user can fall into 1 to 10 health score will get the mental health status. C4.5 is told the class for each user.

Using a set of user's attributes and the user's corresponding class, C4.5 constructs a decision tree that can predict the class for new user's based on their attributes.C4.5 algorithm builds tree based on the information (information gain) obtained from the training instances and then uses the same to classify the test data. C4.5 algorithm generally uses nominal attributes for classification with no missing values.The pseudo code of this algorithm is very simple. Given a set of attributes not target C1, C2, ...,Cn, C the target attribute, and a set S of recording learning.

Algorithm 2: Pseudocode of C4.5 algorithm

InputsR: a set of non- target attributes, C: the target

AttributeS: training data

Output: returns a decision tree

Start:

Initialize to empty tree;
If S is empty **then**
Return a single node failure value
End If
If S is made only for the values of the same target **then**
Return a single node of this value
End if
If R is empty **then**
Return a single node with value as the most common value of the target attribute values found in S
End if
D ← the attribute that has the largest Gain (D, S) among allthe attributes of R
{dj
j = 1, 2, ..., m} ← Attribute values of D
{Sj with j = 1, 2, ..., m} ←The subsets of S respectively constituted of dj records attribute value D
Return a tree whose root is D and the arcs are
labeled by d1, d2, ..., dm and going to sub-trees ID3 (R-{D},
C, S1), ID3 (R-{D} C, S2), .., ID3 (R-{D}, C, Sm)
End

In our study the classification of the target is "Is social media user mental health in problem?" which can be Yes or No. Weather attributes colors, type of posts, posting frequency, check-ins and mental state. They can take the following values: Colours = {bright, mild, shadow, dark} Type of posts = {Violence, conceptual, sadness, events} Posting frequency = {High, low, normal} Check-ins = {High, low, normal} Mental State = {Happy, Sadness, neutral, depression} Examples of the set S are:
We need to find the attribute that will be the root node inour decision tree. The gain is calculated for the four attributes.

The entropy of the set S:

*Entropy (S) =-9/14*log2 (9/14)-5/14*log2 (5/14) = 0.94*

Calculation for the first attribute
Here we take only three types of posts to decrease the complexity
Gain(S, Type of Posts) = Entropy (SConceptual)-5/14*Entropy (Events)
*-4/14*Entropy (Ssad)*
-5/14 Entropy (SVoilance)*
*=0.94 – 5/14*0.9710-4/14*0 – 5/14*0.9710*
Gain(S, Type of Posts)= 0 .246

Calculation of entropies:

*Entropy (SColor) = -2/5*log2 (2/5)-3/5* log2 (3/5) = 0.9710*
*Entropy (SFrequency)) = -4/4*log2 (4/4)-0* log2 (0) =0*
Entropy (SCheckins) = -3/5 log2 (3/5) -/5* log2 (2/5) =0.9710*

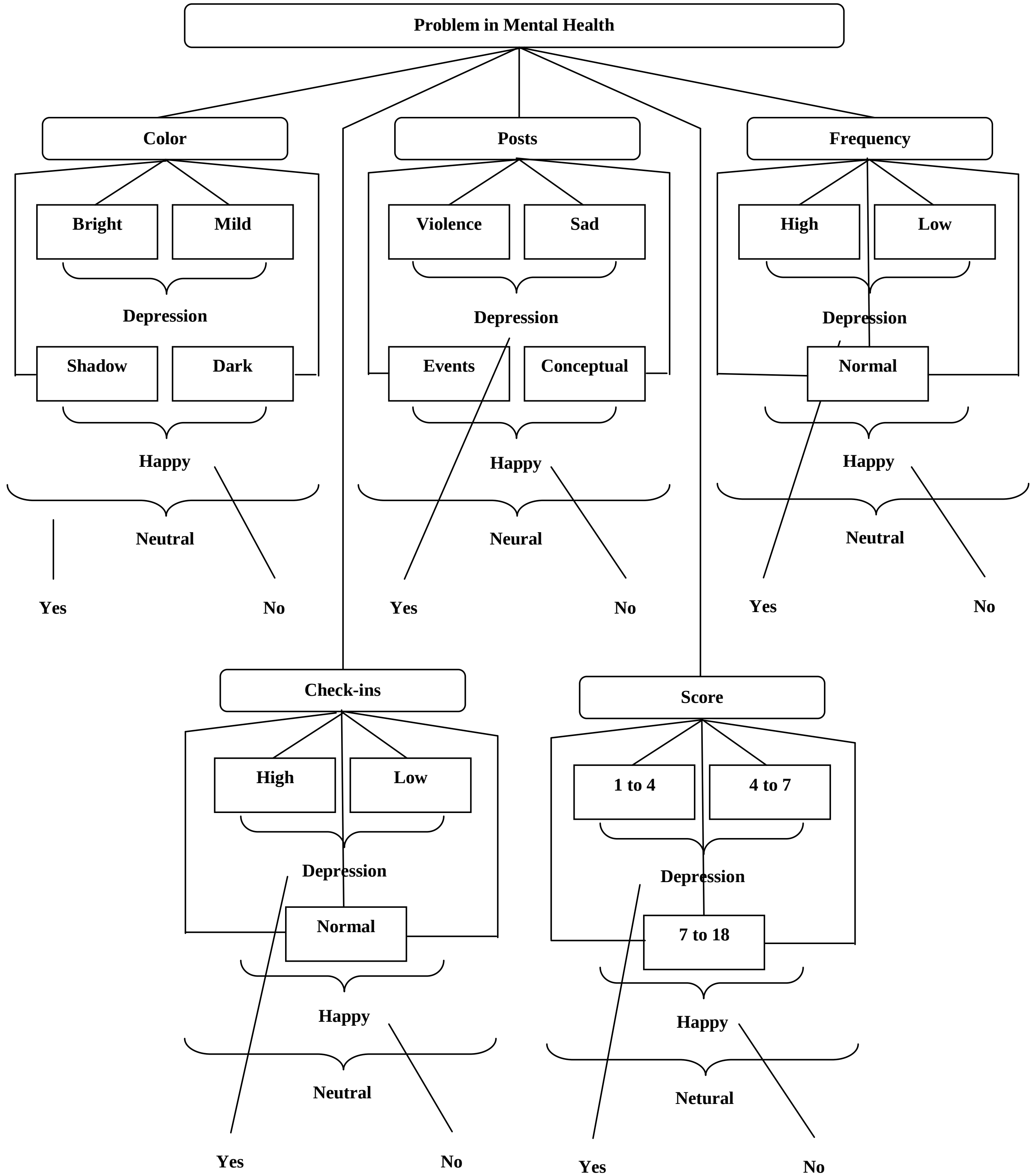

Figure 1: ID3 final tree

As well we find for the other variables:

Gain(S, Score) = 0.1515

Outlook attribute has the highest gain, so it is used as a decision attribute in the root node of the tree (Fig 1).

Since Visibility has four possible values, the root node has

Four branches (High, low, normal)

So by using the three new sets, the information gain is calculated for the temperature, humidity, until we obtain subsets Sample containing (almost) all belonging examples to the same class (Fig 1).

C4.5 uses "Information gain," This computation does not, in itself, produce anything new. However, it allows to measure a gain ratio.

Gain ratio, is defined as follows:

$$GainRatio\ (p,T) = \frac{Gain\ (p,T)}{SplitInfo\ (p,T)}$$

where SplitInfo is:

$$SplitInfo(p, test) = -\sum_{j=1}^{n} p'\left(\frac{j}{p}\right) * log\ (P'\left(\frac{j}{p}\right))$$

P' (j/p) is the proportion of elements present at the position p, taking the value of j-th test. Note that, unlike the entropy, the foregoing definition is independent of the distribution of examples inside the different classes.

Decision trees are built in C4.5 by using a set of training data or data sets as in ID3. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision.

6. COMPARISON BETWEEN SEVERAL ALGORITHMS

Naïve Bayer Vs C4.5: Naïve Bayer algorithm selects the best attribute based on the concept of entropy and information gain for developing the tree.

C4.5 algorithm acts similar to Naïve Bayer but improves a few of Naïve Bayer behaviors:

- A possibility to use continuous data
- Using unknown (missing) values
- Ability to use attributes with different weights
- Pruning the tree after being created
- Pessimistic prediction error
- sub-tree Raising

Performance Parameters:

Accuracy: The measurements of a quantity to that quantity's factual value to the degree of familiarity are known as accuracy. The Table 4 presents a comparison of Naïve Bayer and C4.5 accuracy with different data set size, this comparison is presented graphically in Fig 2. This comparison study was done used by wavelet tool.

Table 4: Accuracy comparison between Naïve Bayer AND C4.5 algorithms

Size of data Set	Algorithm	
	Naïve Bayer	C4.5
15	94.15	96.2
25	79.68	83.65
36	85.5	89.9

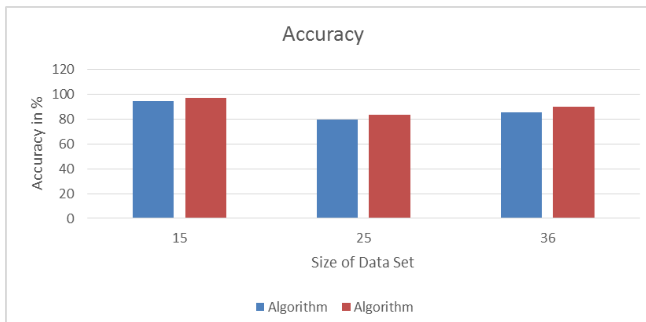

Figure 2: Comparison of accuracy for Naïve Bayer& C4.5 Algorithm

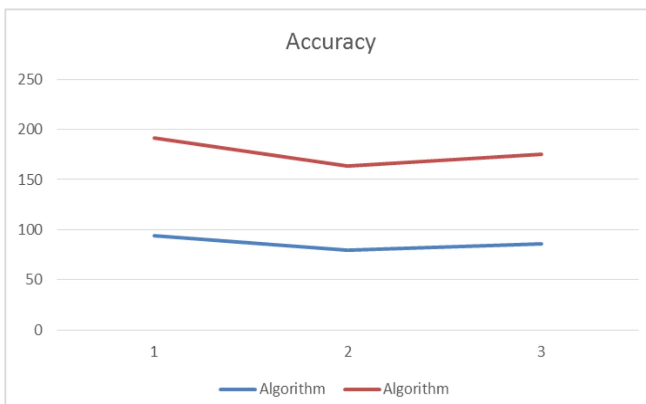

Figure 3: Comparison of accuracy for Naïve Bayer& C4.5 Algorithm

The 2nd parameter compared between Naïve Bayer and C4.5 is the execution time, Table 5 present the comparison. This comparison is presented graphically in Fig 4.

Table 5: Comparison of execution time for Naive Bayer& C4.5 algorithm

Size of data Set	Algorithm	
	Naïve Bayer	C4.5
15	0.315	0.0218
25	0.42	0.28
36	0.28	0.15

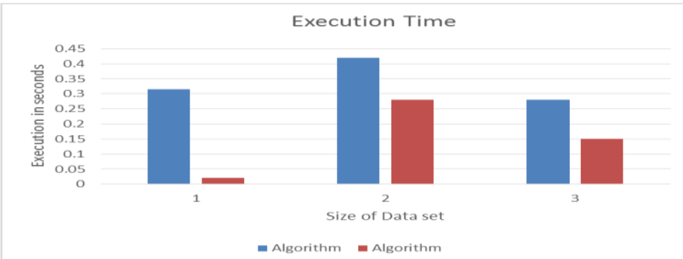

Figure 4: Comparison of Execution Time for Naïve Bayer& C4.5 Algorithm

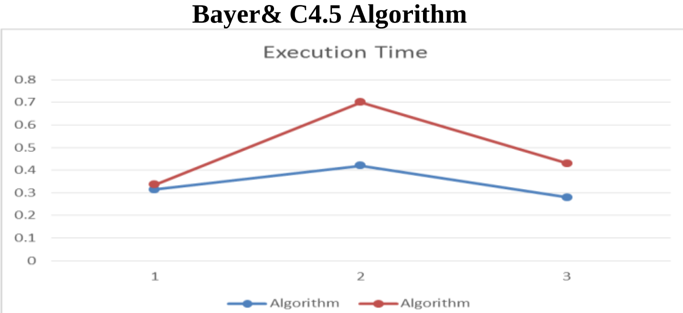

Figure 5: Comparison of Execution Time for Naïve Bayer& C4.5 Algorithm

7. CONCLUSION

Decision trees are simply responding to a problem of discrimination is one of the few methods that can be presented quickly enough to a non-specialist audience data processing without getting lost in difficult to understand mathematical formulations. In this article, we wanted to focus on the key elements of their construction from a set of data, then we presented the algorithm Navies Bayer and C4.5 that respond to these specifications. And we did compare Navies Bayer and C4.5, which led us to confirm that the most powerful and preferred method in machine learning is certainly C4.5.

References

[1] Ang Li, Dongdong Jiao, Tingshao Zhu, Detecting depression stigma on social media: A linguistic analysis, Journal of Affective Disorders, May 2018.

[2] Christopher T. Barry, Chloe L. Sidoti, Shanelle M. Briggs, Shari R. Reiter, Rebecca A. Adolescent social media use and mental health from adolescent and parent perspectives, Journal of Adolescence, December 2017.

[3] Eugene Brusilovskiy, Greg Townley, Gretchen Snethen, Mark S. Salzer, Social media use, community participation and psychological well-being among individuals with serious mental illnesses, Computers in Human Behavior, December 2016.

[4] Gillian Fergie, Kate Hunt, Shona Hilton, Social media as a space for support: Young adults' perspectives on producing and consuming user-generated content about diabetes and mental health, Social Science & Medicine, December 2016

[5] Jenna Glover, Sandra L. Frits, KidsAnxiety and Social Media: A Review, Child and Adolescent Psychiatric Clinics of North America, April 2018.

[6] M. Krausz, Social media and e-mental health, European Psychiatry, April 2017.

[7] Mike Conway, Daniel O'Connor, Social media, big data, and mental health: current advances and ethical implications, Current Opinion in Psychology, June 2016

[8] Renwen Zhang, The stress-buffering effect of self-disclosure on Facebook: An examination of stressful life events, socialsupport, and mental health among college students, Computers in Human Behavior, October 2017.

[9] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, Johannes C Eichstaed, Detecting depression and mental illness on social media: an integrative review, Current Opinion in Behavioral Sciences, December 2017.

[10] Simon M. Rice, Rosemary Purcell, Patrick D. McGorry, Adolescent and Young Adult Male Mental Health: Transforming System Failures Into Proactive Models of Engagement, Journal of Adolescent Health, March, 2018.

Authors Profile

Mr.D.Sridhar is Research Scholar and Assistant Professor, Department of Computer Science at Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore. He received his B.Sc., in 2005 and MCA in 2009 from Bharathiar University, Coimbatore. He obtained his M.Phil. in the area of Data mining from Bharathiar University, in 2013. His research interest lies in the area of Data mining.

Dr.V.Kathiresan is an Associate Professor and Head, Department of Computer Applications (PG) at Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore. He received his B.Sc., in 2003 and MCA in 2006 from Bharathiar University, Coimbatore. He obtained his M.Phil. in the area of Data mining from Periyar University, Salem in 2007. He has completed his doctoral degree from Bharathiar University in 2015. His research interest lies in the area of Data mining.