2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397062

Effectuating Supervised Machine Learning Techniques for Multiclass Classification of Problematic Internet and Mobile Usage

Sneha Sarkar
Dept. of CSE
PESIT-BSC
Bangalore, India
sneha391998@gmail.com

Samanyu Bhandary
Dept. of CSE
PESIT-BSC
Bangalore, India
samanyu973@gmail.com

Arti Arya
Dept. of CSE PES University
Bangalore, India
artiarya@pes.edu

***Abstract*—The internet has slowly become an inevitable part of every facet of our lives. With the power of the world wide web available at the touch of our fingertips, anything seems possible. But mental health disorders due to prolonged usage of the ever-evolving internet and mobile are also on the rise. Studies show there is a strong correlation between excessive internet usage and depression, lower self-esteem, Attention-Deficit Disorder (ADHD), impulsivity, hyperactivity and so on. In this paper, a system is proposed that classifies a persons' internet/mobile usage into four classes (multi class) which are- Normal, Borderline, Critical and Severe. In collaboration with our institutions' Counsellor and considering previous studies, a non-invasive questionnaire was developed to collect the data. The collected data was used to train some efficient and state-of-the-art machine learning models such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), Xtreme Gradient Boosting (XGBoost), Random Forests and Light Gradient Boosting (LightGBM). The model with the highest accuracy was taken forward to deliver the best possible classification of a user into one of four categories. With thorough training and testing linear SVM with radial basis kernel returned the best accuracy and thus it was chosen to moveforward with.**

Index Terms*—*problematic internet usage, problematic mobile usage, machine learning, multiclass classification

I. Introduction

As technology evolves, the internet is becoming not only easier but also cheaper to access. The mass general population is now able to get "online". While this has brought its share of advantages some of which include better access to education, ease of payment transactions and so on, it is not without its own set of issues. Problematic Internet Usage (PIU) or Problematic Mobile Us- age (PMU) arises due to excessive and uncontrolled usage of the internet or your mobile phone respectively. This makes for an internet/phone addiction that is left undiagnosed and unchecked most of the time. However, these issues are not new and have been in existence for a very long time with very little research explaining how to classify and further help users based on this classification. Problematic Internet/Mobile Usage refers to repetitive impairing internet usage in various forms such as video gaming, social media usage, gambling, online shopping etc., which often leads to addiction.

Previous studies have shown that addictive usage of the inter- net leads to the development of mental disorders like depression, lower self-esteem, Attention-Deficit Disorder (ADHD),impulsivity, hyperactivity [1][2][7][8] etc. as well as overallpoor physical well-being, lower self-esteem and lower self-control [3]. Seo D.G. et. al [1] have spoken with respect to impact excessive usage of mobile phones in adolescents. These are disorders that strongly impact a person's quality of life. Low self-esteem leads to poor decision making, working jobs below potential and in worse cases these people become victims of abuse and toxic relationships. Depression can manifest physically and cause weight fluctuations, insomnia, fatigue andincrease risk of heart attacks. It can lead to suicidal thoughtsand overall deteriorating quality of life.

Previous research [13] further shows that problematic social media usage leads to social anxiety and academic procrastination. There has been very little examination of the subjective urges or cravings of behavioral addictions compared to substance addictions. Alcohol problems, mood swings and obsessive problems in many can be traced back to Problematic Internet Usage as well. The relation between sleep quality and internet addiction has been studied in [14], wherein authors have considered 984 students from different universities in Nepal and established the correlation between internet us-age and sleep quality and depression. They also discussed the relation between internet usage and academic grades. Newer research explores the idea of using prevention programs for the

ISBN: 978-1-7281-8529-3/21/$31.00 ©2021 IEEE

purpose of curtailing Problematic Internet Usage [17]. This is a novel approach that can be implemented after classification. Models based on decision trees have been developed to match users to the right program designed for them.

While previous studies have presented a binary classification i.e., addicted or not addicted, in this paper a four-category classification based on the collected data is proposed. Multiple state-of-the-art machine learning models have been evaluated and after a holistic comparison, the one with the best accuracy has been taken forward.

II. Literature Survey

In [1], Seo and Park et al. examined and provided evidence in relation to the effects of mobile phone dependency both in the interpersonal and intrapersonal spheres of adolescents in South Korea. It considered effects on the social and academic capacities of students. It further showed a decline in the academic achievements of students more severely affected. Mental health was also considered and higher rates of mobile phone dependency negatively predicted attention spans and depression. Sariyska et al. [2] have established a connection between problematic/excessive internet use and Attention-Deficit Dis- order (ADHD) as well as depression. The study uses a quiz, which is divided into 3 parts - First part tests the internet usage, second part tests for ADHD and the third part tests for depression. The three parts are checked for a statistical relationship among each other to establish a connection. The study uses the Pearson correlation metric to establish a relationship between different variables. The correlation between Internet usage score and ADHD score is 0.34, and the correlationbetween Internet usage and Depression is 0.24. Thus, both ADHD and Depression scores show a positive correlation withInternet usage.

Mei et al. [3] have focused on Problematic Internet use among the youth in China, based on health, gender, social-status, income and other socio-demographic features. It also establishes the relationship between well-being, self-control, self- esteem and internet usage patterns, where internet addiction is severe for people who perform poorly in the three mentioned domains. The study uses a questionnaire (Young's Diagnostic Questionnaire for Internet Addiction) to collect input from user. Using Multivariate regression analysis, specific attributeslike gender, grade, expenses, income, having a single parent has shown to have a significant impact on internet use patterns and addiction. This study gives an insight on all the factorsthat affect internet usage among the youth and based on socio-demographic and above-mentioned measures. The study concludes that early onset of internet addiction can be found among younger individuals and early mentoring might curb severe addiction in adults. The study conducted by Lopez et al [4] for PIU/PMU among teenagers in the UK and Spain, using socio-demographic measures as a predictor. Teenagers were subjected to a questionnaire and the following attributes were considered for prediction - Age, Center, Drugs, Parents' studies and Gender. The model (Binary Logistic Regression) predicts the student as - excessive user or non-excessive user. The model is saidto have an accuracy between 72%-79%. Study [5] by Shuai et al. focuses primarily on social media use such as online-relationship addiction, internet compulsion and information overload. Data mining is done to detect early onset of mental disorders caused by excessive social media usage. The paper proposes a machine learning framework which uses social interaction features and personal features, which is fed to a learning model. A regularized logistic regression model gives an accuracy of 77.9% and Transductive Support vector machine (TSVM) gives an accuracy of 83%. Gundogar et al. [6] conducted a study with high school students and provides evidence that Internet Addiction is a common problem among such an age group and that there isa strong correlation between ADHD and Internet Addiction. Students with higher internet addiction scores also showed higher ADHD feature subscale scores. Yen et al. [7] had evaluated the association between ADHD, hyperactivity, inattention, impulsivity and internet addiction. It also provides insight into whether gender factors into this association. Results of this study show deficit of attentionis the most common symptom in association with Internet addiction closely followed by impulsivity. Further it states that this association is stronger in females.

In [18], Tomczyk et.al had used a modified version of Young's questionnaire (consisted of 10 questions) and based their studyon a dataset of 3569 teenagers in Poland. The paper explores data collected and segregated them based on their internet usages. Our work provides a larger dataset, higher dimensions of data, clear classification of users based on questionnaire data and extensive use of predictive and classification models to help classify unseen data. Lutz Wartberg et al. [19] have used machine learning models to find indicators that predict the spontaneous remission of PIU in adolescents. The paper used bivariate and multivariate logistic regression to point out specific factors, chief among which was emotional regulation, that can be further promoted to regulate and prevent PIU.

During the COVID-19 pandemic, internet usage spiked all over the world and the issue of PIU/PMU came to the forefront in many fields of study. Orsolya Király et al. [20] have focused on preventive measures to be taken specifically during the pandemic. The use of psychoactive substances and reinforcing behaviors (gambling, video gaming, etc.) is also cited and general and specific steps that can be taken to consciously curb problematic internet usage before it becomes a glaring issue. The study in this paper outlines a multi-class classification of users based on their internet/mobile usage patterns. This categorization was established based upon characteristics of the data collected and in consultation with the school's student counsellor. Classes in descending order of problematic usage are presented as Severe, Critical, Borderline and Normal.

III. Proposed Methodology

Data for the study was collected via a questionnaire, where users answered 18 questions which evaluated their personal behaviour with the internet/ mobile devices. Most questions were based on the Likert scale [8]. Likert scale is usually meant for scaling the responses in questionnaires. Any questionnaire based on Likert concept has responses to the questions over a range like strongly agrees/agrees/weekly agree/disagree. The options for each question ranged from 0 (does not apply), 1 (rarely) all the way to 5 (always) for the questionnaire given in Table I.

A. *Formulation of the questionnaire and Data Collection*

After thorough examination of papers detailing instruments to assess Problematic Internet Usage [9][10][11], the Young's Diagnostic Questionnaire [11] was taken and altered slightly. The formulated questionnaire was further examined by the student counsellor for a better psychological perspective of the data that would be collected for the study. The questions were phrased with the help of student counsellor in such a way that the subject consider the question in a positive way and answer appropriately. The questionnaire consists of 18 questions and is targeted to obtain data based on the internet and mobile usage of the user as shown in Table I. It includes personal and social behavior questions in relation with their internet usage. The process of data collection was spread over a period of three months, October to December 2019.

The questionnaire was made via Google Forms and circulated online. Further, face to face data collection was done all over Bangalore, India to allow a larger set of subjects.

B. *Exploratory Data Analysis*

The form with the questionnaire received a total of 8000 responses from various users. The data received had no outliers since user input was sanitized on entry and thus did not require cleaning. The heat map in Fig.1 shows the Pearson Correlation coefficient among the attributes with each other. Heatmaps help visualizing the similarity between two entities, with each color signifying a specific value of similarity.

TABLE I
Data attributes and questionnnaire

S.No.	*Attributes*
Q0	How old are you?
Q1	I stay online/on my phone longer than I intended.
Q2	I try to cut down the amount of time I spend online/on my phone but fail.
Q3	I neglect household chores / school work / job tasks to spend more time online/on my phone.
Q4	I form new relationships with fellow online users.
Q5	People complain about the amount of time I spend online/on my phone.
Q6	I check my social media/phone before something else that I need to do.
Q7	I block out disturbing thoughts about my life with soothing thoughts of the internet.
Q8	I anticipate when I will go online/use my phone again.
Q9	I fear that life without the internet or my mobile would be boring, empty and joyless.
Q10	I snap, yell or act annoyed if someone bothers me while I'm online or on my phone.
Q11	I say "just a few more minutes" when I'm online/on my phone.
Q12	I try to hide how long I've been online/on my phone.
Q13	I feel depressed/moody/nervous when I'm offline which goes away when I'm online/on my phone.
Q14	I choose the internet/phone over being with my partner/family/friends.
Q15	I think about asking for help in relation to my internet/phone usage.
Q16	I use my phone for work/school purposes more than anything else.
Q17	Gender (M,F or O)

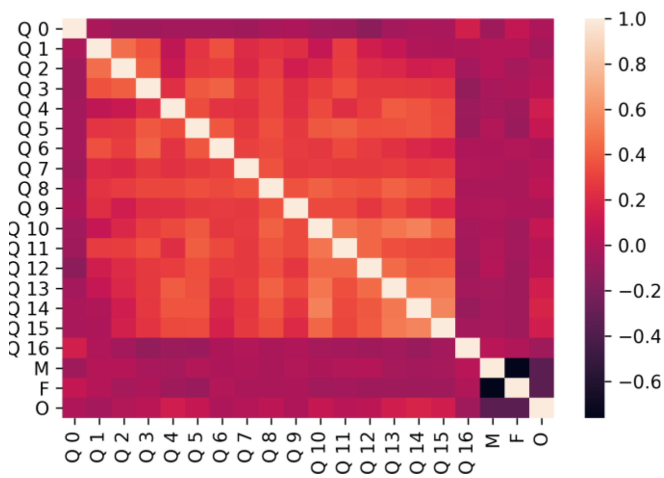

Fig. 1. Correlation heatmap

In Fig.1, each attribute is compared with each other in the form of a matrix. The white box running diagonally (showing perfect correlation) is since the attribute is being compared to itself, hence resulting in a white box. The Gender attributes, M and F, have a black box marked against each other, showing total negative correlation, or in other words are completely opposite. Zero relation (or correlation) is shown by the boxes marked with a mildly dark red color, which is the case for most of the heatmap. The heatmap visualization shows us the relationship among all attributes. Correlation depicts how each attribute relates with/ depends on each other as shown in Fig.1. As the figure shows, most questions have a slight negative correlation to the age attribute. This shows that people who are older tend to have lower scores, which in turn tells us that older people are less prone to internet addiction as compared to younger people.

Fig.2 shows the variance of each attribute as a bar chart. This gives an idea about which attribute has the most effect on the classification. Attributes with higher variance have more influence on the classification. In this case the first attribute, i.e., the age has the most influence on the dataset, while the last three attributes (Gender) have the least influence.

C. Data Preprocessing

- Removing redundant attributes - As Fig.1 shows, most of the attributes do not show high correlation, which implies that each attribute is unique and contributes information that is required by the model. The gender attributes (M and F) show a strong negative correlation, which is expected, and hence one of the attributes is dropped. Either male or female attribute is sufficient rather than having both.

- Data Normalization - Fig.2 shows the variances for each question and indicates that the age attribute, as expected, has a very high variance. This is solely since age ranges from 0-100 (or more) whereasthe questions range from 0-5 and gender is binarized into three digit 0-1 variables (The gender attribute can only be one of the following, Male, Female or Others. Hence, this is binarized (one-hot encoded) into three separate columns, where if the user belongs to the specific gender, that gender is set to 1 and the othertwo are set to 0. E.g., Male - 100, Female - 010 and Others -001). Therefore, any statistical model would give the age attribute the most importance and the gender attribute the least. The questions would have a much lower importance which is undesirable. To redistribute the importance, normalization is done for all the data (for each attribute) so all data values range between 0 and 1.

- Dimensionality Reduction - PCA (Principal Component Analysis) is used to reduce the number of attributes in the dataset. Fig.3 shows the relationship between percentage of information retained and the number of PCA components chosen. Ideally the dataset should have high information retention and a low number of PCA components. Choosing 16 components gives us almost a 100% retention in information, thus reducing the dimensions of the dataset by 2. This has shown to significantly improve the compute speed in the training and prediction phase of the models with the collected dataset.

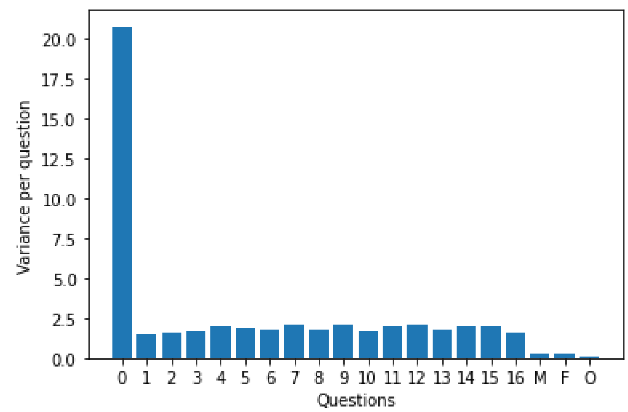

Fig. 2. Attribute variances

- Train-Test Split - The preprocessed data is split into train and test data. Training set consists of 70% of the total number of records, i.e., 5600 records, and the Test set consists of the remaining 30% of the records, i.e., 2400 records. However, when K-fold cross validation tests are performed, the train and test sets are combined as the folds selected act as test/ validation sets. K-fold cross validation test chooses random sets of data as test cases and does this k times (k is the number of folds taken), where for each k, a different set of data is chosen.

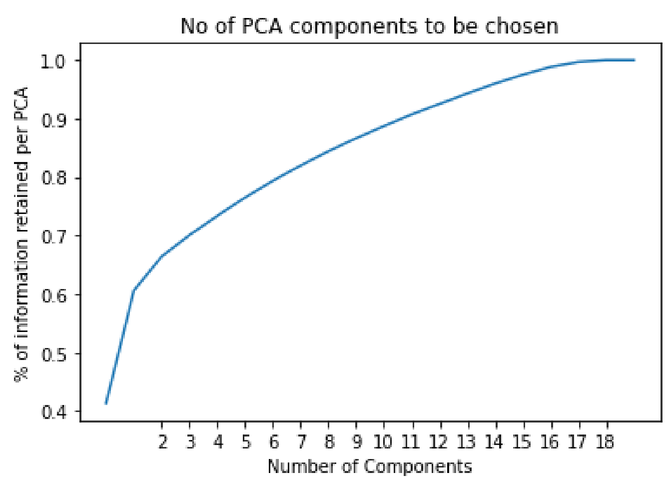

Fig. 3. PCA - Information retention graph

This ensures that our model being evaluated is not overfitting for a specific split, hence multiple splits are taken, and the final accuracy taken is the mean of all k results (accuracies for the specific k split) obtained. In our case, we have taken 5-10 splits for cross validation.

D. Clustering the Collected Preprocessed Data

The users taking the quiz were to be classified into one of the four categories - Normal, Borderline, Severe and Critical internet and mobile usage. In order to make collected data, a supervised dataset, k-means clustering is applied. In this case, the dataset is clustered into four clusters because the whole dataset is expected to get assigned with four distinct class labels, a cluster for each class.

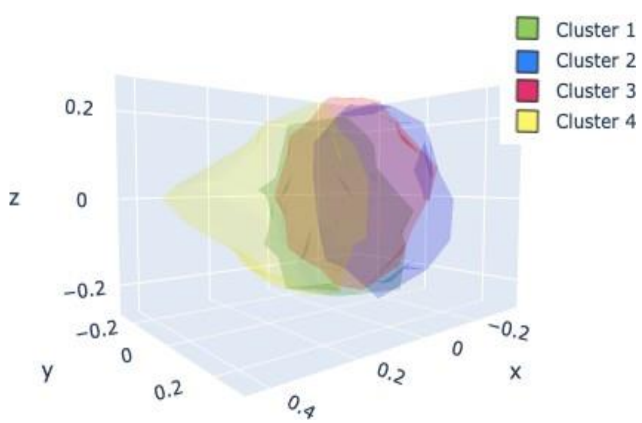

Fig. 4. K-means cluster visualization in the dataset

The results are as shown in Fig. 4. Further analysis is required to determine which cluster associates to which class. Fig.4 shows four distinct clusters of different colors, each color pertaining to one of the classes. As per the cluster visualization graph, the four clusters are not very distinct Yellow cluster lies on a different plane which is not clearly visible in three dimensions.

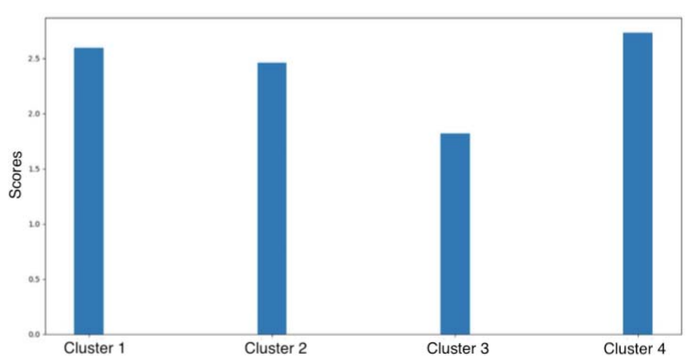

Fig. 6. Bar graph showing averaged scores per cluster

E. Analysis of Clustered Data

Upon obtaining four unlabeled clusters from the process of unsupervised learning, the data within each cluster was analyzed for assignment of classes to the clusters. A bar graph detailing average scores for each question/feature from each cluster, each with its own designated color is shown in Fig.5. Further averaging total scores per cluster in Fig.6 makes labeling clear. The clusters with scores in descending order are labeled as severe, critical, borderline and normal, respectively.

F. Training Different Machine Learning Models using Clustered Data

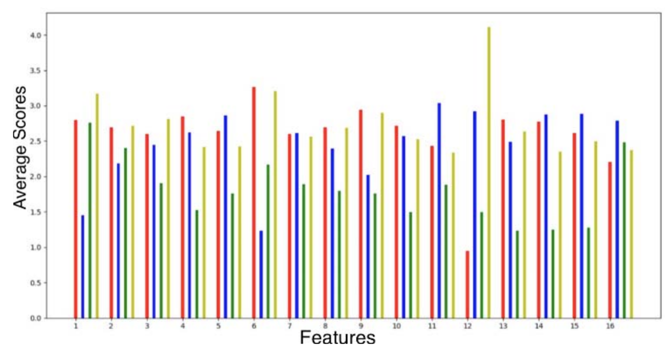

Fig. 5. Bar graph showing data within clusters

For classifying new users who take the quiz, training asupervised learning model to predict their internet usage isthe next step. The Machine Learning models trained for this were - Multinomial Logistic Regression, Support Vector Machine (SVM), XGBoost, LightGBM, Decision Trees (ID3), Random Forest (Ensemble ID3). All the models were tuned for hyperparameters using the GridSearchCV function (a function present in the scikit library), that picked the most optimal hyperparameter combination out of a range of parameters. A k-fold cross validated test (5-10 folds chosen) was performed for each model to determine model accuracy, for all the models that have been used in this research. None of the models showed overfitting behavior throughout all our tests with k- fold cross validation.

LightGBM [15] is an extension of the gradient boosting decision tree (GBDT) algorithm developed by Microsoft in 2017. It is said to be almost 20 times faster than the traditional GBDT algorithms, due to its greedy nature of choosing gradients with higher Information Gain. It utilizes two techniques, Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), which is responsible for estimating large gradients for high information gain and reducing the number of features used. Smaller gradients were ignored as they made very

TABLE II
ML MODELS ACCURACIES OVER THE PIU/PMU DATASET

ML Model	*Data Size*	*Accuracy*	Author's Model	*Data Size*	*Accuracy*
Support Vector Machine	8000	98.4%	Transductive Support vector machine (TSVM) [5]	3126	83%
Logistic Regression	8000	97.5%	Binary Logistic Regression [4]	2228	72-79%
Logistic Regression	8000	97.5%	Regularized Logistic Regression [5]	3126	77.9%
Decision Trees	8000	72%	J48 Decision Tree [5]	3126	74.4%
XGBoost	8000	93%	Not Mentioned	Not Mentioned	Not Mentioned
Random Forest	8000	92%	Not Mentioned	Not Mentioned	Not Mentioned
LightGBM	8000	94%	Not Mentioned	Not Mentioned	Not Mentioned

little contribution to the result and this reduction in computation made the algorithm much faster. Table II provides a detailed accuracy metric for each model used.

The concept of fuzziness was applied when a probability measure (degree of membership) of classification among all 4 classes was produced, i.e., a user may have a probabilitymeasure associated with more than one class; the highest measured probability class was the taken as final predictionof the user's class. The probabilities were calculated using the Platt Scaling and Recalibration technique. The experiments were carried out using the Scikit Learn library for Python, which has all the models and probability calculations (predict_proba()) available as in-built functions and APIs. It gives a more accurate and precise classification, which makes it significantly easier to diagnose the issues.

IV. RESULTS AND DISCUSSION

The experiments were conducted on a system having configuration as Intel i5-3427U, 1.8Ghz processor with dual-core. System had 8GB RAM with DDR3 at 1600MHz and Intel HD 6000 1.5GB VRAM graphics card.

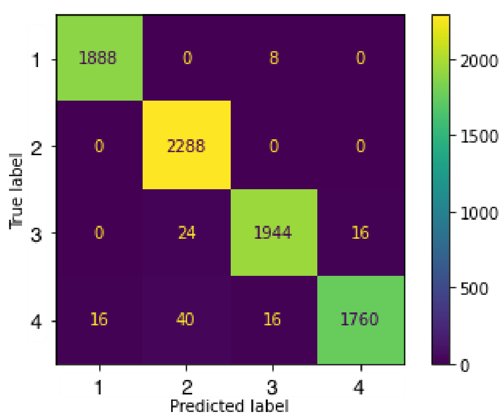

Fig. 7. Multi-Class Confusion Matrix for the dataset

The goal to develop a novel method of classifying users according to their answers to our mobile/internet usage questionnaire was competently achieved.Given the size of our dataset, all the models considered forthis study have given excellent results with accuracies well over 90% which is also significantly higher than results that are available in literature, refer Table II. The Supervised Learning models trained with the data and their respective model accuracies are mentioned in Table II. The accuracies forall the models are calculated by using the cross_val_score() function from the scikit-learn python library.

For multiclass classifiers, the various evaluation metrics were accuracy, micro precision, micro recall, micro f1 score, macro f1 score etc. For evaluating the proposed system, accuracy, micro precision, micro recall and micro F1 score was used.

TABLE III
PRECISION AND RECALL TABLE

Class	Precision	Recall
1	0.9915	0.9957
2	0.9727	1.00
3	0.9898	0.9798
4	0.9865	0.9607

For multiclass classification,

$$micro - F_1 = 2\frac{P_{micro} - R_{micro}}{P_{micro} + R_{micro}}$$

$$P_{micro} = \frac{\sum_{i=1}^{4} TP_i}{\sum_{i=1}^{4} TP_i + FP_i}$$

$$= \frac{7880}{7880 + 128} = \frac{7880}{8008} = 0.984$$

$$R_{micro} = \frac{7880}{7880 + 120} = \frac{7880}{8000} = 0.985$$

$$micro - F_1 = 2\frac{0.984 \times 0.985}{0.984 + 0.985} = \frac{1.938}{1.969} = 98.4\%$$

Clearly Accuracy, micro precision (also known as averaged precision), micro recall and micro-F1 score are all same. The above computations are with respect to Support vector machine and same computations were done for other models, but Support Vector Machines gave the highest accuracy (on Cross Validated Tests) of 98.4% and hence s used as the primary classifier for the users.

For new users, a web interface using Flask, HTML, CSS and JavaScript has been constructed, to enable the user to take the quiz and see the results. The results displayed also provide useful tips and article links to help the user curb Internet/ Mobile addiction. It also features graphs to show the user their scores per class which indicates that user may belong to more than one class at one point of time with different degree of membership. Thus, users can see a fuzzy classificationof results. Fig.8 shows for one new instance that it belongsto each of the four classes by some degree of membership. This pie chart clearly depicts that for a particular user the concept of fuzziness provides more flexibility and thus instead of providing an immutable diagnosis, users are given a clearerpicture of their results in terms of four different classes and how that user is assigned to four class labels. This graph sets forth the probabilities with which said user belongs to each of the four classes.

Further, the in-depth results provided to the user notifies them of the results from each implemented model. This is done to provide a more holistic insight into the inner workings of the system. Table II compares the test set accuracy obtained by the proposed model with the other existing works available in the same domain for the same algorithms.

V. Conclusion and Future Scope

The proposed approach takes a fresh look at problematic internet/mobile usage and attempts to develop a novel classification of users whose usage falls inside the problematic threshold. It incorporates the findings of studies done in this field in the past and proposes a new methodology. Data collected via the internet was cleaned and normalized in preparation.

After this clustering and subsequent classification was done. Various state of the art models were tested for the best results. Among the diverse classification algorithms that were tested Support Vector Machines gave the highest accuracy and was thus used as the primary classifier.

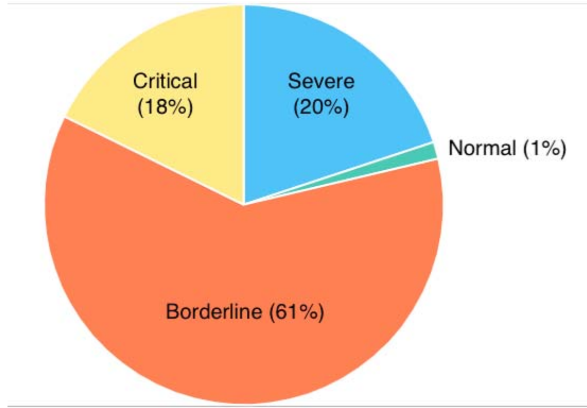

Fig. 8. Pie chart showing scores across multiple classes for a single instance

The proposed proposed approach in this paper gives higher accuracy readings compared to models proposed in past studies. The current model provides a fuzzy multiclass classification which gives an accurate insight on the user's internet and mobile usage. However, the reports generated don't make much sense to a user and requires a thorough analysis by an expert to give the user an intuition of the problem being faced and the steps to take to curb this problem.

Instead of continuous external intervention of experts, if the model could evaluate and provide explainable solutions to the user, it would be a much better solution. The next iteration of the project would be to add an Explainable AI module, which gathers the results to generate an intuitive and meaningfulresult for the user and generate personalized solutions to curb their internet addiction. This would require further data collection and expert intervention for various use cases and results.

References

[1] Seo, Dong Gi, Yujeong Park, Min Kyung Kim, and Jaekook Park. "Mobile phone dependency and its impacts on adolescents' social and academic behaviors." Computers in human behavior 63 (2016): 282-292.

[2] Sariyska, Rayna, Martin Reuter, Bernd Lachmann, and Christian Montag."Attention deficit/hyperactivity disorder is a better predictor for problem- atic Internet use than depression: Evidence from Germany." Journal of Addiction Research Therapy 6, no. 209 (2015): 1-6.

[3] Mei, Songli, Yvonne HC Yau, Jingxin Chai, Jinhua Guo, and Marc N. Potenza. "Problematic Internet use, well-being, self-esteem and self-control: Data from a high-school survey in China." Addictive behaviors 61(2016): 74-79.

[4] Lopez-Fernandez, Olatz, Jose´ Luis Losada-Lopez, and Mª Luisa Honrubia-Serrano. ”Predictors of problematic Internet and mobile phone usage in adolescents.” Aloma: revista de psicologia, cie`ncies de l'educacio´ i de l'esport Blanquerna 33, no. 2 (2015): 49-58

[5] Shuai, Hong-Han, Chih-Ya Shen, De-Nian Yang, Yi-Feng Lan, Wang-Chien Lee, Philip S. Yu, and Ming-Syan Chen. ”Mining online social data for detecting social network mental disorders.” In Proceedings of the 25th International Conference on World Wide Web, pp. 275-285. 2016.

[6] Gundogar, A., B. Bakim, O. A. Ozer, and O. Karamustafalioglu. ”P-32- The association between internet addiction, depression and ADHD among high school students.” European Psychiatry 27, no. S1 (2012): 1-1.

[7] Yen, Ju-Yu, Cheng-Fang Yen, Cheng-Sheng Chen, Tze-Chun Tang, and Chih-Hung Ko. ”The association between adult ADHD symptoms and internet addiction among college students: the gender difference.” Cy- berpsychology Behavior 12, no. 2 (2009): 187-191.

[8] Nemoto, Tomoko, and David Beglar. ”Likert-scale questionnaires.” In JALT 2013 Conference Proceedings, pp. 1-8. 2014. Nemoto, Tomoko, and David Beglar. ”Likert-scale questionnaires.” In JALT 2013 Conference Proceedings, pp. 1-8. 2014.

[9] Young, Kimberly S. Caught in the net: How to recognize the signs of internet addiction–and a winning strategy for recovery. John Wiley Sons, 1998.

[10] Chen, Sue-Huei, Li-Jen Weng, Yi-Jen Su, Ho-Mao Wu, and Pin- Feng Yang. ”Development of a Chinese Internet addiction scale and its psychometric study.” Chinese Journal of Psychology (2003).

[11] Young, Kimberly S. ”Internet addiction: The emergence of a new clinical disorder.” Cyberpsychology behavior 1, no. 3 (1998): 237-244.

[12] Durak, Hatice Yıldız. ”Modeling of variables related to problematic internet usage and problematic social media usage in adolescents.” Current Psychology (2018): 1-13.

[13] Moretta, Tania, and Giulia Buodo. ”Autonomic stress reactivity and craving in individuals with problematic Internet use.” PloS one 13, no. 1 (2018): e0190951.

[14] Bhandari, Parash Mani, Dipika Neupane, Shristi Rijal, Kiran Thapa, Shiva Raj Mishra, and Amod Kumar Poudyal. 2017. "Sleep Quality, Inter- net Addiction and Depressive Symptoms among Undergraduate Students in Nepal." BMC Psychiatry 17 (1). BioMed Central Ltd. doi:10.1186/s12888- 017-1275-5.

[15] Guolin Ke1, Qi Meng2, Thomas Finley3, Taifeng Wang1, Wei Chen1, Weidong Ma1, Qiwei Ye1, Tie-Yan Liu1.2017.”LightGBM: A Highly Efficient Gradient Boosting Decision Tree”

[16] Ioannidis, Konstantinos, Samuel R. Chamberlain, Matthias S. Treder, Franz Kiraly, Eric W. Leppink, Sarah A. Redden, Dan J. Stein, Christine Lochner, and Jon E. Grant. ”Problematic internet use (PIU): associations with the impulsive-compulsive spectrum. An application of machine learn- ing in psychiatry.” Journal of Psychiatric Research 83 (2016): 94-102.

[17] Probierz, Eryka, and Adam Gałuszka. ”Problematic Use of the Internet- Using Machine Learning in a Prevention Programme.” In International Conference on Human-Computer Interaction, pp. 231-238. Springer, Cham, 2020.

[18] Tomczyk, Łukasz Szyszka, Michał Stosic, Lazar. (2020). Problematic Internet Use among Youths. 10. 161. 10.3390/educsci10060161.

[19] Wartberg, Lutz, and Katajun Lindenberg. ”Predictors of Spontaneous Remission of Problematic Internet Use in Adolescence: A One-Year Follow-Up Study.” International Journal of Environmental Research and Public Health 17, no. 2 (2020): 448.

[20] Kira´ly, Orsolya, Marc N. Potenza, Dan J. Stein, Daniel L. King, David C. Hodgins, John B. Saunders, Mark D. Griffiths et al. ”Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance.” Comprehensive Psychiatry (2020): 152180.