

MACHINE LEARNING MODEL FOR PREDICTION OF SMARTPHONE ADDICTION

Arpita Mazumdar¹, Gahina Karak², Srishti Sharma³

¹Department of Computer Science & Engineering, Haldia Institute of Technology, India

²Department of Computer Science & Engineering, Haldia Institute of Technology, India

³Department of Computer Science & Engineering, Haldia Institute of Technology, India

ABSTRACT

Purpose: The primary objective of the present study is to predict the levels of smart-phone addiction and also to find the correlation between different smart phone activities, and their relationship across male and female users.

Methodology: The survey was conducted using a well-designed questionnaire which enquires about the usage of smartphone of an individual. College undergraduates (N = 115) participated in the survey and completed the questionnaire as part of their class requirements. The data thus collected is trained to form a machine learning model based on clustering.

Results: The findings significantly shows that males tend to use smartphones more than females to access books and e-books. that female has the largest count for possession of phones for more than 12 hours, whereas, male have the largest count for possession of their phones for less than 6 hrs. The results show that most of the male have their phone's battery last for a day, whereas for females the count of "yes" and "no" are almost equal. The whole population is categorized in 3 clusters such as Highly addicted group, moderately addicted group, non addicted group.

Conclusions: This prediction model certainly be highly useful for understanding the phone usage level and eventually predicting certain possible threats prevalent amongst addictive smartphone users.

Keywords -addiction, clustering, machine learning, smart-phones

1. INTRODUCTION

There has been tremendous growth in the use of the mobile phones in India. It is reported that the India's telecommunication market is the second largest in the world. The mobile phones are available to the people right from the age of 12 years. The mobile phone technology has brought the world closer. It provided great convenience in communication among people by way of either calling or texting. Now, the mobile phones are coming up with variety of features like internet access, sending e-mails, games, access to social networking sites like face-book, listening to music, playing radio, reading books, dictionary and so on. The mobile phones are also used to overcome the feeling of loneliness. The majority of the users are in the age group of 15 to 25 years. The current-day fascination with the smart phones highlights the latest technology that, for better or worse, appears to be encouraging people to spend relatively more time with technology and less with fellow humans[1]. Nowhere is this fascination with technology more intense than in young adults – college students in particular[2] [3]. College students commonly view their cell-phone as an integral part of who they are, and/or as an important extension of themselves [4]. Present-day cell-phones are seen as critical in maintaining social relationships and conducting the more mundane exigencies of everyday life [5] [6].The contacts are established instantly with the help of mobile phones which was not possible earlier. However, though the mobile phone provided many advantages, it has also caused some problems also. Some people are using the mobile phones so excessively that it assumes the form of addiction. The use of mobile phones has reduced the face-to-face communication. It is observed that the people sending text messages while talking to others. Even the visitors, guests are kept waiting till the mobile phone use is

over. The students are using mobile phones for playing games, sending messages, calling even when the class is in progress. The mobile phones are used at places like hospitals, judicial courts, petrol pumps where their use is banned. The mobile phone use during driving is commonly observed which may increase the chances of involving in accidents. The use of mobile phone while driving can distract attention of the driver visually, physically and cognitively. Earlier studies have shown that various personality traits like neuroticism, extraversion, psychoticism, etc. and mobile use have some relation to each other. The gender may also play some role in high mobile phone use. The provision of the additional features like internet, music, radio, etc. may be resulting in excessive use of mobile phones. It is, therefore, thought to study the various aspects of mobile phone use including gender role.

2. Methodology

The steps in the proposed work flow involves data collection, pre-processing of data, training and testing with specific machine learning models.

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. Types of Machine Learning: Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output parameters. Unsupervised learning is when the model is getting trained on an unlabeled dataset. As in this study all are unlabeled and unclassified data, hence unsupervised machine learning techniques are adapted. Unsupervised Machine Learning is a branch of machine learning that learns from test data that has not been labelled, classified or categorized. Instead of responding to feedback, unsupervised learning identifies commonalities in the data and reacts based on the presence or absence of such commonalities in each new piece of data.

2 a. Dataset preparation

The survey was conducted amongst under graduate students. The survey was conducted using a questionnaire which enquires about the usage of smartphone of an individual. The questionnaire was prepared on a way such that, each question is independent of other. This was done keeping in mind the fact that, in unsupervised learning algorithm all the features should be independent to avoid any conflicts and thus, avoiding errors in the model. The 20 questions included in the survey are:

1. Do you use your phone to click pictures of class notes?
2. Do you buy books/access books from your mobile?
3. How often is phone in your hand?
4. Does your phone's battery last a day?
5. When your phone's battery dies out, do you run for the charger?
6. Do you worry about losing your cell phone?
7. Do you take your phone to the bathroom?
8. Do you use your phone in any social gathering (parties)?
9. How often do you check your phone without any notification?
10. Do you check your phone just before going to sleep/just after waking up?
11. Do you keep your phone right next to you while sleeping?
12. Do you check emails, missed calls, texts during class time?
13. Do you find yourself relying on your phone when things get awkward?
14. Are you on your phone while watching TV or eating food?
15. Do you have a panic attack if you leave your phone elsewhere?
16. How often do you click selfies/photos?
17. You don't mind responding to messages or checking your phone while on date?
18. For how long do you use your phone for playing games?

19. Do you prefer an extended phone time or a vacation?

20. For how long do you talk over the phone in a day?

The answers for these questions were scaled on a 3 point scale (1-always, 2- sometimes, 3- never).

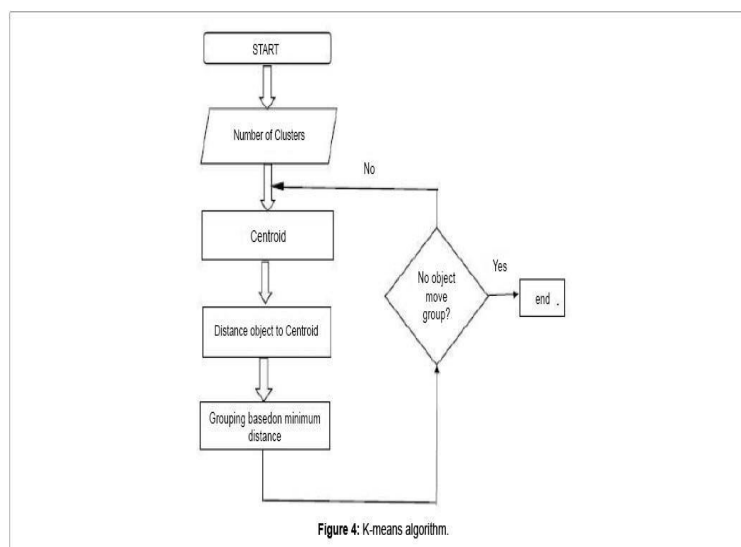
2 b. ALGORITHM

Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields. In Data Science, clustering may be applied while analysis to gain some valuable insights from the data by seeing what groups the data points fall into when we apply a clustering algorithm. There are different types of clustering methods.

- **Hard Clustering:** In hard clustering, each data point either belongs to a cluster completely or not. For example, in the above example each customer is put into one group out of the 10 groups.
- **Soft Clustering:** In soft clustering, instead of putting each data point into a separate cluster, a probability or likelihood of that data point to be in those clusters is assigned. For example, from the above scenario each customer is assigned a probability to be in either of 10 clusters of the retail store.
- **Hierarchical clustering:** as the name suggests is an algorithm that builds hierarchy of clusters. This algorithm starts with all the data points assigned to a cluster of their own. Then two nearest clusters are merged into the same cluster. In the end, this algorithm terminates when there is only a single cluster left.

Four of the most used clustering algorithms: • K-means • Fuzzy K-means • Hierarchical clustering • Mixture of Gaussians

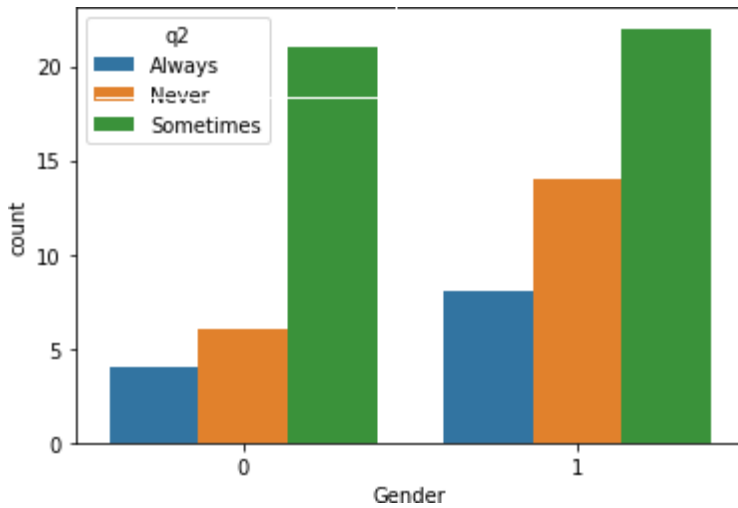
In this study, K-means algorithm is considered for simplicity. K-means is one of the most popular “clustering” algorithms. K-means stores k centroids that it uses to define clusters. A point is considered to be in a particular cluster if it is closer to that cluster’s centroid than any other centroid. K-Means finds the best centroids by alternating between (1) assigning data points to clusters based on the current centroids (2) choosing centroids (points which are the center of a cluster) based on the current assignment of data points to clusters. K-means clustering algorithm computes the centroids and iterates until we find its optimal centroid. It assumes that the number of clusters are already known. It is also called flat clustering algorithm. The number of clusters identified from data by algorithm is represented by ‘ K ’ in K-means. In this algorithm, the data points are assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroid would be minimum. It is to be understood that less variation within the clusters will lead to more similar data points within same cluster. Figure 4. shows the flowchart of the algorithm.



3. Results and discussion

A primary objective of the present study is to find the degree of smart phone addiction, for which 20 independent questions are considered for building the machine learning based prediction model. We initially investigated if there is any difference across male and female cell-phone users in terms of the cell-phone activities used. The following are the few significant findings.

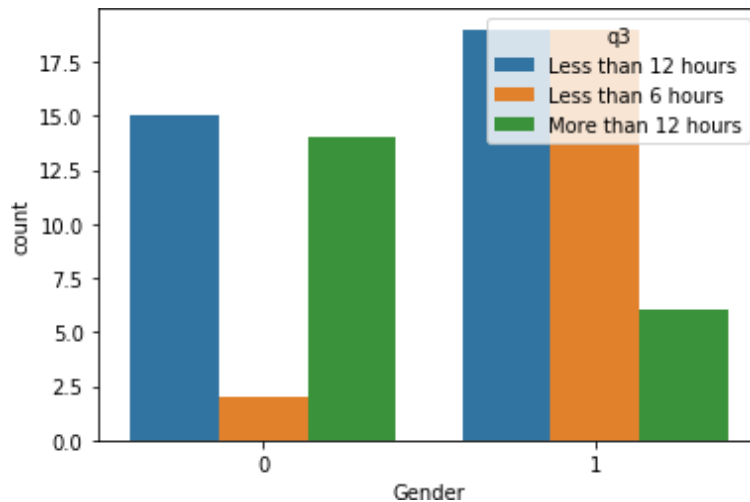
A. RELATION BETWEEN GENDER AND USAGE OF SMART PHONE TO ACCESS BOOKS



Here, the field “1” refers to “Male” and field “0” refers to “female”. The plot significantly shows that male tend to use smartphones more than females to access books and e-books.

B. RELATION BETWEEN GENDER AND POSSESSION OF SMARTPHONES IN HAND

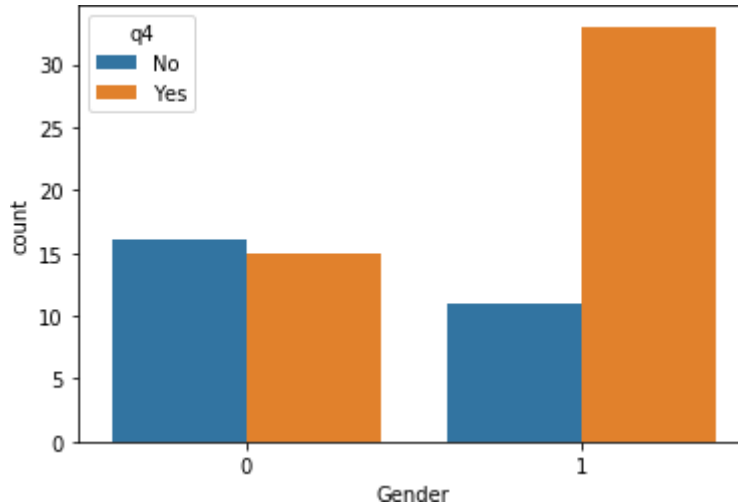
This plot shows the correlation of time of smartphone possession amongst male and female. It is clear from the



plot that female have the largest count for possession of phones for more than 12 hrs, whereas, male have the

largest count for possession of their phones for less than 6 hrs.

C. RELATION BETWEEN GENDER AND DOES THEIR PHONE'S BATTERY LAST FOR A DAY



This plot shows that most of the male have their phone's battery last for a day, whereas for females the count of "yes" and "no" are almost equal.

In cluster analysis, the elbow method is a heuristic used in **determining the number of clusters in a data set**. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use. Elbow is one of the most famous methods by which we can select the right value of k and boost the model performance. We also have performed the hyperparameter tuning to choose the best value of k as 3. After following the K-means algorithm the prediction model is developed. The entire population is thus clustered in 3 groups labelled as highly addicted, moderately addicted, not addicted. Here clusters are evaluated based on some similarity or dissimilarity measure such as the distance between cluster points. If the clustering algorithm separates dissimilar observations apart and similar observations together, then it has performed well. The two most popular metrics evaluation metrics for clustering algorithms are the Silhouette coefficient and Dunn's Index [7].

- **Silhouette Coefficient**

$$s = \frac{b - a}{\max(a, b)}$$

The Silhouette Coefficient is defined for each sample and is composed of two scores:

- a: The mean distance between a sample and all other points in the same cluster.
- b: The mean distance between a sample and all other points in the next nearest cluster.

The Silhouette Coefficient for a set of samples is given as the mean of the Silhouette Coefficient for each sample. The score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters. The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster. The mean Silhouette coefficient measured on the dataset has the highest value for k=3 which confirms the result of the elbow method.

Conclusion

This study aimed to reveal the level of smartphone addiction of under graduate students, and to identify the predictors of smartphone addiction. Therefore, we consider demographic characteristics, daily usage duration of a smartphone, commonly used content and game usage pattern etc.

This prediction model certainly be highly useful for understanding the phone usage level and eventually predicting certain possible threats prevalent amongst addictive smartphone users.

References

1. Griffiths, M. D. (2000). Does Internet and computer “addiction” exist? Some case study evidence. *CyberPsychology & Behavior*, 3(2), 211–218.
2. Massimini, M. & Peterson, M. (2009). Information and communication technology: Affects on U.S. college students. *CyberPsychology: Journal of Psychosocial Research on Cyberspace*, 3(1), 1–12.
3. Shambare, R., Rugimbana, R. & Zhoua, T. (2012). Are mobile phones the 21st century addiction? *African Journal of Business Management*, 62(2), 573–577.
4. Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15(2), 139–168.
5. Junco, R. & Cole-Avent, G. A. (2008). An introduction to technologies commonly used by college students. *New Directions for Student Services*, 124(Winter), 3–17.
6. Junco, R. & Cotton, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers & Education*, 59, 505–514.
7. [Evaluation Metrics For Machine Learning For Data Scientists \(analyticsvidhya.com\)](http://analyticsvidhya.com)
8. Bian, M., Leung, L.: Linking loneliness, shyness, smartphone addiction symptoms, and patterns of smartphone use to social capital. *Social Science Computer Review* 33(1), 61–79(2015)
9. Blehm, C., Vishnu, S., Khattak, A., Mitra, S., Yee, R.W.: Computer vision syndrome: a review. *Survey of ophthalmology* 50(3), 253–262(2005)
10. Cohen, S., Kamarck, T., Mermelstein, R.: A global measure of perceived stress. *Journal of health and social behavior* 33, 385–396(1983)
11. van Deursen, A.J., Bolle, C.L., Hegner, S.M., Kommers, P.A.: Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender. *Computers in human behavior* 45, 411–420(2015)
12. Fengou, M.A., Mantas, G., Lymberopoulos, D., Komninos, N., Fengos, S., Lazarou, N.: A new framework architecture for next generation e-health services. *IEEE journal of biomedical and health informatics* 17(1), 9–18(2013)
13. Group, E., et al.: Open service access (osa); application programming interface (api); part 3: Framework (parlay5); 203 915-3 v1. 2.1. ETSI/The Parlay Group(2007)