An Approach to Detect Smartphone Addiction through Activity Recognition and App Usage Behaviour

by

Nur Uz Zaman
19301052
Afroza Akther
19301076
Nowshin Tabassum
19301251
Md. Khaliduzzaman Khan Samrat
19301114
Swad Mustasin Khan
19101599

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science And Engineering
School of Data and Sciences
Brac University
May 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

NUR UZ ZAMAN

Nur Uz Zaman
19301052

Afroza

Afroza Akther
19301076

Nowshin Tabassum

Nowshin Tabassum
19301251

Samrat Khan

Md. Khaliduzzaman Khan Samrat
19301114

Swad Mustasin Khan

Swad Mustasin Khan
19101599

Approval

The thesis titled "An Approach to Detect Smartphone Addiction through Activity Recognition and App Usage Behaviour" submitted by

1. Nur Uz Zaman(19301052)
2. Afroza Akther(19301076)
3. Nowshin Tabassum(19301251)
4. Md. Khaliduzzaman Khan Samrat(19301114)
5. Swad Mustasin Khan(19101599)

Of Spring, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 22, 2023.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

RRahman

Rafeed Rahman
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The widespread use of smartphones has raised concerns about problematic smartphone use or addiction, which has become a significant issue in today's society. Despite the recognition of this research area, detecting smartphone addiction remains a challenge. Therefore, it is crucial to identify the primary causes of smartphone addiction and understand how individuals' lifestyles contribute to this behavior. Most of the methods in research area are self assessment based and detected via different addiction scales. Moreover, in previous studies daily human activities was never considered as a factor in problematic smartphone use. This study aims to explore a new approach in detecting excessive smartphone usage by considering the impact of sensor based daily activities and smartphone app usage. By examining addictive characteristics of smartphone usage and clustering them based on various independent variables, we sought to determine smartphone addiction and investigate the influence of daily activities. To collect reliable and accurate data, we utilized apps for seven days to capture information on the participants' smartphone usage. Leveraging sensor data and LSTM models, we identified participants' activities and correlated them with daily app usage duration to detect smartphone addiction using clustering methods such as K-Means and K-Medoids. Our analysis revealed that around 28% participants showed addicted behaviour. To validate these findings, we compared our result with survey results using diverse evaluation metrics (RI,FMI), which exhibited 87% accuracy.

Keywords: **Smartphone; Addiction; App; Usage; Activity; Sensor; Cluster;**

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our thesis supervisor Mr. Golam Robiul Alam sir for his kind support and advice in our work. He helped us whenever we needed help.
And finally to our parents without their throughout sup-port it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

BnA Behavioural And Activity

DBI Davies-Bouldin Index

FMI Fowlkes-Mallows Index

$LSTM$ Long short-term memory

PCA Principal Component Analysis

RI Rand Index

RNN Recurrent Neural Network

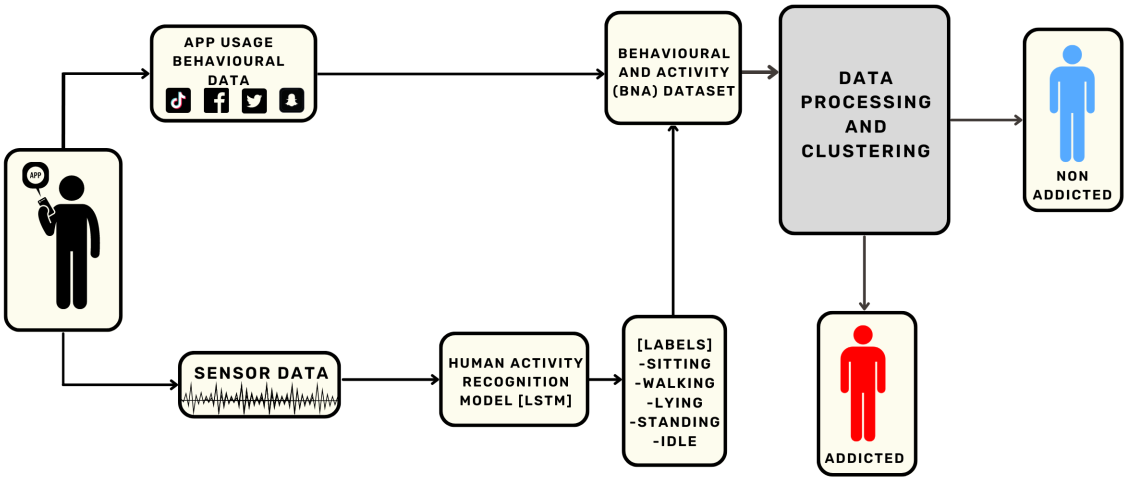
SAS Smartphone Addiction Scale

$SAS - SV$ Smartphone addiction scale Short Version

$WCSS$ Within Cluster Sum of Square

Chapter 1

Introduction



1.1 Introduction

Smartphone addiction, often referred to as mobile phone addiction or nomophobia (fear of being without a mobile phone), is a disorder in which people demonstrate compulsive and excessive usage of their cellphones, having detrimental effects on many parts of their lives. In the current digital era, smartphone addiction is a widespread problem. A growing number of people are becoming overly dependent on their devices due to the quick development of technology and the ubiquitous availability of cellphones. Nowadays, people can get smartphones with all kinds of features that make them more convenient than ever before: from video calling to text messaging and more.According to statistics, there are 6.648 billion smartphone users globally at the moment, which equals an ownership percentage of 83.40 percent[20] and this number is increasing rapidly over the years. Most people find smartphones to be fantastic tools for staying up to date with information, communicating instantly, connecting with others, and learning for themselves and finding enjoyment. It is essential for academics to identify the probable warning signs for mobile phone addiction.

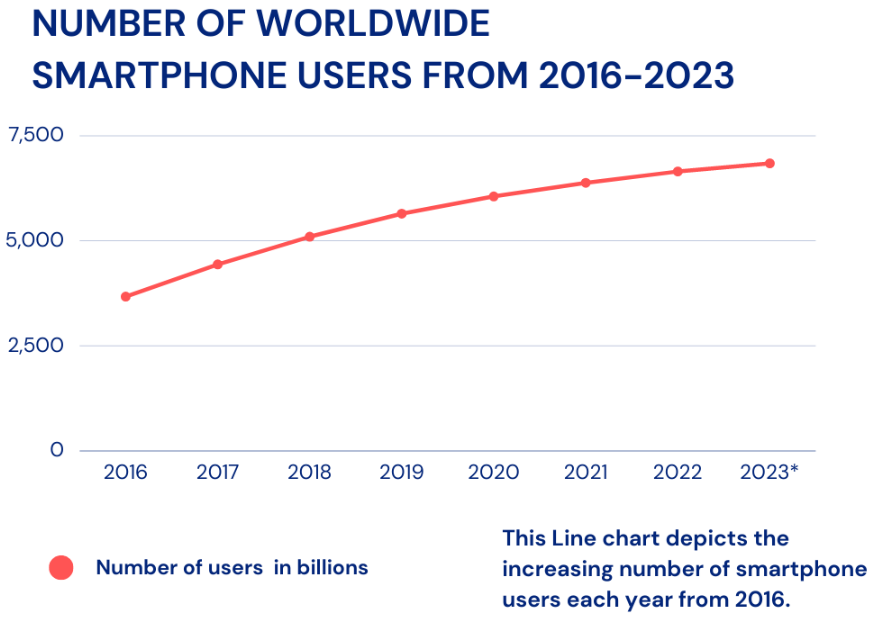


Figure 1.1: Increasing No. of smartphone users from 2016 to 2023

Most people find smartphones to be fantastic tools for staying up to date with information, communicating instantly, connecting with others, and learning for themselves and finding enjoyment. Smartphone addiction is a growing concern in today's fast-paced society. Many experts believe that the prevalence of smartphone use is on the rise, which has led to many questioning the impact these devices have on users' mental and physical health.Therefore, it is essential for academics to identify the probable warning signs for mobile phone addiction. Main key predictors of mobile phone addiction among a variety of other factors by analyzing the usage pattern of users should be identified. For instance, sensation seeking, different substance use [27] , trait procrastination these are some risk factors that have been found through previous studies. An analysis of Statista indicates that younger age groups in the United States are those where smartphone ownership is most prevalent. Only 61% of people 65 and older have smartphones, compared to 95% of people between the ages of 18 and 49. [46]. It is important to quantify the extent of smartphone addiction and how often people use them.

Present-day world, it is clear that people of all ages own smartphones, which accounts for both the rising popularity of smartphones and the growing dependence on them among users. The emergence of numerous social media networks in recent years has intensified smartphone usage. Everyone now uses social media, and this trend is here to stay.Today, over fifty percent of the global population uses social media in some capacity. According to the 2022 research overview of global social media data, 4.70 billion people use social media globally now. [41]. However, in today's interconnected society, it sometimes feels like we're all just trying to fit into this new environment. Despite knowing all the consequences of using smartphones uncontrollably, people are keeping a blindfold. Previous studies showed that the use of smartphones excessively can affect physical well-being. For example, it can increase body fat by decreasing the physical activity of the human body [12]. Furthermore, overuse of smartphones can increase the risk factor for eye damage [7]. Nonetheless, overuse of smartphones can also affect mental health. Research shows that over reliance on smartphones could make the human brain lazy [13].

Detecting smartphone addiction has been an area of research for many years, and various methods and processes have been used to study this addiction to smartphones. One of the most preferred methods used is self-report questionnaires, which are designed to evaluate various aspects of smartphone use and addiction. These questionnaires are typically administered to participants and ask about their smartphone use habits, including how long they spend on their phone each day, how frequently they use their phone and how much they feel the need of using their smartphone [44]. The majority of the study on identifying smartphone users has been done with a focus on college or university students [17], [29]. Recently, studies have also used wearable sensors to detect smartphone addiction. For example, one study used a wristband to measure the user's skin conductance level, which is an indicator of emotional arousal, to determine the user's level of addiction to their smartphone [45].

So, detecting smartphone addiction has been a multi-faceted process that has involved self-report questionnaires, wearable sensors etc.These techniques have been utilized to gain a better understanding of the scope of smartphone addiction and its effects on people's lives. Researchers can create interventions and therapies to assist people in overcoming their addictions and enhancing their general well being by understanding the symptoms of addiction. Still there are different methods to identify smartphone addiction which hasn't been explored yet. However, the problems related to smartphone addiction haven't been solved, instead it is increasing. These kinds of issues related to smartphones are the reason which led us to do this research on the overuse of smartphones in our daily lives. A sincere intention to raise awareness among people, assist parents, teachers, youngsters, and senior folks in creating countermeasures against smartphone addiction and overuse of mobile apps as well .

1.2 Problem Statement

Excessive smartphone use has become a significant issue among people of all ages in the current century. With the growing popularity of social media and mobile applications, smartphone usage has increased significantly in recent years. As many people rely on their smartphones for work, education, and socializing, this problem of excessive use of smartphones has worsened since the COVID-19 pandemic. This can also be called Smartphone Addiction which falls under Problematic Mobile Phone Uses (PMPU). From apparent observations we can notice the huge influx of smartphone users every year. This also reflects on the statistical data. According to a survey conducted by Pew Research Center, 81% of Americans now own a smartphone, and they spend an average of 3.5 hours per day on their phones (Perrin, 2021).

As we can see every year more and more people are using smartphones and it's predicted that it will still grow in upcoming years. Although adaptation to new technology isn't a bad thing, in the case of smartphones it is becoming an eminent part of our everyday life. This over dependence on smartphones can be attributed to behavioral addiction. The excessive use of smartphones has been linked to several problems, including physical health problems such as neck pain, eye strain, and sleep disorders. Furthermore, smartphone addiction can lead to psychological problems such as anxiety, depression, and social isolation.

Behavioral addiction is a form of addiction that doesn't involve substance uses but creates a compulsion to engage in repetitive use of an element – despite any negative consequences to the person's physical, mental, social or financial well-being [3]. This behavioral addiction can easily be caused by smartphones as the whole design of almost every smartphone app nowadays is to keep the users engaged as long as possible. This eventually leads to issues like Procrastination, most importantly academic and bedtime procrastination and social isolation .

Nowadays mobile phones are not limited to phone calls, rather they have become so smart that it is called a "smartphone" as it can be used for different purposes like social media, video games, browsing, study, reading books, watching movies and shows and many more. To avoid negative thoughts, depression, anxiety and other emotional regulation people tend to rely on these recreational elements causing procrastination. Therefore, procrastination refers to prioritizing tasks that are more pleasurable or less unpleasant on your to-do list over those that are actually important or necessary. According to a study by Rabiullah Behzad [33], to find out the connection between procrastination and smartphone addiction, 20% of the respondents firmly agree that using their smartphones has caused them to miss their planned task. About 30% of respondents are only somewhat in agreement with the statement "Missing planned work due to smartphone use" while about 27.50% of respondents more strongly agree that they missed their scheduled tasks. The responses that indicate how "green" the respondents are added up to more than 77%, which amply demonstrates the link between procrastination and smartphone use.

Academic procrastination is a major issue causing damage to students of different fields in their academic performance and their overall study and career. A study conducted in China [37] says that Chinese medical students are subject to strong academic pressure and strict standards, as well as professional and environmental influences. They are hence vulnerable to social anxiety and smartphone addiction. Many researches demonstrate that there is a positive association between smartphone addiction and poor sleep quality. According to another study in China [32] on this issue, their path analysis results confirmed the idea that smartphone addiction was linked to poorer self-regulation, which led to more procrastination at bedtime and less decent sleep.

Day by day people are becoming more concerned about their reel life than their real life which makes them more comfortable in screen-based interaction than face-to-face interaction. Many study's findings showed that people who used their phones more frequently reported feeling more lonely and isolated than their less-device-dependent colleagues. A study conducted in Japan [25] is the first to examine the link between smartphone and internet addiction and the risk of hikikomori in young Japanese individuals. Hikikomori is a condition in which a person without psychosis stays at home for over six months and does not engage in society, such as by attending school or working. Other research also shows that users of smartphones preferred to speak and connect with others behind a screen, which eventually increased their level of social isolation. Social ties were replaced as the primary means of communication by smartphones [34].

To stop individuals from suffering more harm, it is essential to identify smartphone addiction and treat it. Addiction to smartphones can have negative effects on social interactions, academic performance, and work productivity in addition to its negative effects on physical and psychological health. In a study from the University of Derby, Rosen et al. (2013) found a strong correlation between smartphone addiction and poorer academic and professional performance. In order to identify this problem of addiction to smartphones, many researchers have conducted research for a long time. Many have found some incredible results on it as well. Starting from the development of smartphone addiction scale SAS [4] , finding correlation between problematic smartphone usage and other substances [31] or factors [38]. However, most of the research was conducted using a survey as the main source of data which has been used to study this addiction to smartphones. One of the most preferred methods used is self-report questionnaires, which are designed to evaluate various aspects of smartphone use and addiction. These questionnaires are typically administered to participants and ask about their smartphone use habits, including how long they spend on their phone each day, how frequently they use their phone and how much they feel the need of using their smartphone [44]. The majority of the study on identifying smartphone users has been done with a focus on college or university students [17], [29]. But there are some inconsistencies in the scales that are used to measure the smartphone addiction as they focus more on trait personality than actual phone usage of the users. Also it's not reliable to believe in self report for accuracy as people might not want to provide actual information about their smartphone usage and addiction level. Moreover, lack of cross validation also limits the generalizability of findings.

Still there are different methods to identify smartphone addiction which hasn't been explored yet. However, the problems related to smartphone addiction haven't been solved, instead it is increasing. These kinds of issues related to smartphones are the reason which led us to do this research on the overuse of smartphones in our daily lives. In our situation, we considered employing the daily smartphone usage time as well as the daily activities of smartphone users as sensor values to obtain a more accurate result on the identification of smartphone addiction. Because it's possible that some questions won't get an accurate response from the participants due to the likely difficulties in remembering the information. While investigating other works on this topic we haven't come along with researchers considering daily activities of users in the detection process previously. So, we have considered using daily human activities like sitting, walking, standing, lying as some features to detect smartphone addiction to observe if it works better. The question that this research is trying to answer is:

How effective is the use of observing sensor based human activity in detecting smartphone addiction along with individuals daily smartphone usage?

As mentioned above, as previous works were mostly survey oriented this research tried to come up with new methodologies beside the smartphone addiction scale in detection addiction to smartphones more accurately considering some new and unused features. It is indeed needed to find a more optimum and comprehensive way address this issue.

1.3 Research Contribution

This study uses innovative data collecting techniques to examine smartphone users' addiction. Analyzing smartphone addiction, its relation to daily human activities. Main objective of our research is to introduce a new method of detecting smartphone addiction beside the empirical SAS-SV scale. Also, analyzing the effect of human daily activities on this detection procedure. After fulfilling the our objectives, the contributions that we have made through this research are listed below:

1. We have collected all possible smartphone data, specifically, app usage data and smartphone sensor data along with SAS survey data for achieving more precise outcome from our research. Because through our findings from studying several previous works, it was coherent that only self questionnaire based survey Data might not be consistent for detecting smartphone addiction.
2. We have extracted five types of human activity from the sensor data that we have acquired by training the LSTM model. We have contemplated these human activities in our research as the prior works in this field never considered human activities as a factor that can cause people to be addicted to smartphone.
3. We have constructed a new 'Behavioural and Activity Fusion Dataset' by combining app behavioural dataset and activity dataset which can aid us to gain more rigorous result as this dataset holds all accurate information about smartphone.
4. We have Analyzed and Clustered those acquired features by applying some unsupervised clustering techniques as our 'Behavioural and Activity Fusion Dataset' did not have any classes.
5. Finally, we evaluated the performance of the clusters by comparing those with the SAS-SV score that was computed from the SAS-SV survey response dataset using some extrinsic and intrinsic evaluation metrics.

1.4 Document Outline

Chapter 1 gives a brief background history about smartphone and smartphone addiction, the research problems that motivated our thesis, the contribution we have made through our research and overall thesis outline. **Chapter 2** holds all the summaries of background study that was done by reading previous work articles or journals. **Chapter 3** describes broadly about our working methodology. Starting with a workflow diagram, how the datasets were acquired, all pre-processing phases along with a short description of all datasets and finally broad explanation of all the models that was used for the entire research work has been portrayed here. **Chapter 4** narrates how each model worked on the datasets and what was the result from those models by showing some illustrations of those results. The evaluation metrics that was used are also explained here. Finally, **Chapter 5** concludes our thesis along with some short report on the limitations that was found during the work and what should be done in the future.

Chapter 2

Literature Review

Smartphones have revolutionized how we interact with the digital world, communicate, and obtain information in modern life. The rising use of smartphones has, however, led to worries about addiction and its possible effects on both individuals and society. In the process of this research it is noticeable how cellphones have changed over time, from their invention to the present. It examines the technological developments, traits, and prowess that have helped them become so widely used. We can contextualize the impact of smartphones on society and human behavior by comprehending the development of smartphones. As every good research builds upon the previous works in a particular domain, we have also read many of the previous works in the subject of our research. In this section we tried to enlist some of the papers that are relevant to our domain. In one form or another these studies provide light on problematic smartphone use detection and its effects on a range of demographics, including kids worldwide and students in high school and colleges. They made clear the detrimental effects of smartphone addiction, including worse academic achievement, greater loneliness, mental health problems, and ruined social relationships. These researches also look at underlying processes, smartphone use, social media use, and behavioral patterns as well as other aspects of smartphone addiction. Overall, these studies contribute to our understanding of the intricacies surrounding problematic smartphone use and provide useful data for the detection process which are related with our research goals.

2.1 Related Works

The concept of smartphone addiction, also known as problematic or compulsive smartphones can be understood by looking at the theoretical frameworks, including behavioral addiction models, that are utilized to understand addiction. Also by examining the scales and diagnostic criteria used to measure smartphone addiction, as well as the prevalence rates found in various research. The well known developed smartphone addiction scale which contains 33 items are generally used mostly in PSU (Problematic Smartphone use). The Smartphone Addiction Scale (SAS), which was created and validated by Kwon et al., is a novel scale that uses self-reporting to assess smartphone addiction. It has 33 questions and 6 items. This SAS scale, which covers six characteristics including daily-life disruption, positive anticipation, withdrawal, cyberspace-oriented connection, overuse, and tolerance, was designed in order to address the limitations of the previous measures.The internal-consistency test result (Cronbach's alpha) throughout its development phases was 0.967. The SAS internal-consistency test result in this study (Cronbach's alpha) was 0.966. In [15] a newly reformed short version of this smartphone addiction scale. Proposed SAS-SV contains 10 questions which were considered enough to detect smartphone addiction of adults according to Kwon, Kim, Cho and Yang (2013). The detection was mainly done using the ROC analysis where the cut-off value was set 31 and 33 for girls and boys accordingly.The SAS-SV showed good validity and reliability for the assessment of smartphone addiction . This brief version of the smartphone addiction scale, which was created and validated, definitely gave us an insight in our work of research in effectively utilizing to assess smartphone addiction.

Through analysis we came across a lot of research where empirical scales like SAS helped enough in getting good observations on it. This is necessary because smartphones have become a crucial part of our daily life with the influx of different types of apps. These apps are designed in a way that makes a person use it repetitively. This becomes a problem when the individual is addicted. Even though it appears to be similar to other substance-based addictions, the researchers assert that smartphone addiction is distinct from gambling, online auction, and gaming addictions. Instead, it is a pattern of individuals' uncontrollable urges to check their smartphones repeatedly. This inability to control an overwhelming urge to check their phones has negative effects on all aspects of their lives. For instance, in [15] it is mentioned of trying to get a comprehensive understanding of problematic smartphone uses through empirical research. Here, Wang, Chuang and Lee (2016) primarily identified the characteristics of smartphone addiction. Which is divided into two groups depending on Underlying mechanisms, rewarding based and Compensation based. Both of them can be divided into further categories based on characteristics of smartphone addiction. They have collected data through questionnaires (SAS-SV) which is a subset of the Smartphone Addiction Scale [4]. We have seen that multiple research work has used this scale to determine the addictive behavior of smartphone users.

Smartphone addiction can vary from person to person. Some show extreme addictive behavior and some say they are neutral about using smartphones. We came along solely focused on categorizing the smartphone addicts based on their behaviors depending on two major emerging factors: user-related characteristics and technology-related characteristics in [6]. They also divided the effects of smartphone use into three key categories: impacts on productivity, impacts on social life and impacts on well-being. They used a grounded theory technique to analyze the data acquired through in person interviews and questionnaires. Lapointe and Liette (2013) aimed to investigate the many forms of smartphone addiction-related behaviors, the factors that contributed to the establishment of smartphone addiction, and the consequences of such usage. And the survey was distributed to smartphone users enrolled in an undergraduate program at a large university in North America. Four distinct profiles of smartphone users could be distinguished from their data study, which revealed some trends. The "Addicts" user features in the first profile are those of typical addicts, and they would probably qualify for an addict in clinical diagnosis. Most of them failed to reduce their smartphone usage. They also describe themselves as shy, dependent, conformist, and introverted in their profiles.

The ”Copycats” user profile comes in second place with 25.73% of the total user. Despite the fact that they engage in a number of addictive activities, they do not generally share the traits of normal addicts. The ”Regulars” user profile stands in third place with 38.46% of the users. Much less addictive and problematic behaviors are displayed by these users. The final group of smartphone users represents the ”Moderates.” They have very few addictive behaviors and use smartphones less frequently. The fact that this research was limited on behaviors of smartphone addicts, not the reason and factors behind this. Now-a-days use of the internet has increased tremendously, sometimes people get the idea that both internet usage and smartphone usage might be considered the same but that may not be the case which we have seen in [11]. In [11] they have divided users of internet addiction and smartphone addiction into two categories: those at high risk and those who may be at risk. According to the classification of internet addiction and smartphone addiction by the Korea Information Society Agency, there are three groups based on the degree of internet addiction and smartphone addiction: high-risk group, potential-risk group, and general user group.The conclusion given by Jun, Woochun(2015) during the presentation found that smartphone addiction is becoming more prevalent than internet addiction. In the near future, it is anticipated that smartphone addiction will become independent from internet addiction. Second, both dependencies must be subdivided into more specialized dependencies. So, it is noticeable that smartphone addiction is increasing rapidly and the internet is not the only factor that is considered as a mediator.

Although the use of these scales has been going on for a long time, over numerous researches we came across a paper which showed the limitation of smartphone addiction scales. In [26] a systematic review of 78 validated scales developed over the past 13 years to measure problematic mobile phone and smartphone use. The study evaluates the theoretical foundations and psychometric properties of these scales. The analysis revealed that numerous scales published in studies lack satisfactory internal consistency and test-retest reliability. Despite the existence of several self-report scales investigating the phenomenon, there remains a need for more dependable and valid measures to assess problematic mobile phone use. The research paper highlights the growing concern regarding excessive smartphone usage and its potential adverse effects on mental well-being. Pienaar and Malekian (2019) emphasized that a limitation of the reviewed scales is their exclusive reliance on self-reported data, which cannot reliably capture actual phone usage. This limitation underscores the necessity to address this issue in the field of research. Additionally, during the development of these new scales, many researchers hypothesized that problematic phone use would correlate not with actual usage, but rather with related personality traits such as self-esteem and impulsivity. So, this gave us the idea not to depend wholly on self-questionnaires either.

In [24] Podo, Luca, Taccardi, Benito and Colonna stated empirical research on problematic smartphone addiction through smartphone app usage is necessary as some apps have become a crucial part of our day life. For example, Facebook, Instagram are some popular apps that are used daily by millions of users. A paper about youth’s social media addiction [9] Arora, Shivani, Okunbor and Daniel aimed to study the social media, mainly Facebook usage pattern of two countries; Western

Asia (India), North America (United States) and compared their addictive behavior. This paper also used questionnaires and distributed those to the undergraduate students of Delhi University and Fayetteville state university. This study employs a mixed research methodology for its research design. Content analysis was chosen for this research.The purpose of the paper is to investigate how different social media activities affect respondents from the USA and India. They analyzed various symptoms of the respondents and compared those activities. They discovered that users in both countries spend more than an hour every day on social media. Users also reported that since all popular social media sites have an app version for their smartphone, it has become easier for them to frequently access social media. According to the analysis, only 25% of Indian respondents check social media websites first thing in the morning, compared to more than half (54%) of American respondents. Also, the majority of U.S. respondents (76%) check their social networking accounts before bed, although only 67% of Indian respondents do so, which may be detrimental to the user's mental health. Although fewer than half of users in both nations think that social media is a way of life, more than half accept that social media is addictive. According to studies, the "need to belong" motivates the use of social media platforms.

As Social media is a driving factor for smartphone addiction, it is necessary to analyze the app usage pattern to measure people's addiction level. In the article [42] researchers suggested a paradigm for identifying everyday trends in online behavior across mobile app usage. It is only feasible to define the patterns of one or a small number of individuals with the limited data and controlled research available.Large groups of people often behave in similar ways, which may reveal disruptive tendencies in our society as well as their demographics and lifestyles. To detect the activity aspects of app usage windows acquired via a time-based segmentation methodology, they develop a probabilistic topic model-based method.Here they investigate the semantic data of app categories and previous knowledge of activity temporal patterns to annotate retrieved cyber actions with semantic words. For each retrieved activity, calculate an average feature vector for the app category. Then, in an action known as internal ranking, rank app categories based on the density in the computed vector. Additionally, assign an external score to each app category's actions.The consistency of users' online behavior by gauging the average intra-distance of their patterns across several days. 2,000 well-known mobile applications were used by 653,092 individuals throughout a city-scale dataset comprising 971,818,946 use records, which was made available by a major Chinese ISP. According to their study, people often follow the activity patterns from yesterday, but as time passes, the patterns start to diverge.There are five typical everyday patterns of online behavior: commuting, pervasive socializing, nighttime amusement, afternoon reading, and nightly socializing. Mentioned research has significant consequences for understanding user demographics, lifestyles, habits, and service needs. It also helps to spot other disruptive trends, such working overtime and gaming and social media addiction. So using app usage information may help us to measure smartphone usage pattern in our research.

Mostly, there are some driving factors that lead humans to smartphone addiction. Like in the [14] authors tried to stand Mere exposure effect as a factor behind smartphone addiction. Mere exposure effect denotes positive emotions that may appear naturally after being exposed to repeated stimuli. In this study, they translate MEE into four factors—convenience, habit, enjoyment and concentration—to investigate it. Here convenience means, smartphones are more convenient than other electronic devices; habit of using smartphone means people use it frequently; enjoyment denotes users experience positive feelings while using smartphone; and concentration shows the degree to which attention is absorbed while using a smartphone. Then they proposed some hypotheses to prove the influence of MEE on smartphone addiction using these factors. To collect data they gathered participants for an online survey from two significant Chinese institutions and used structural equation modeling (SEM) to analyze this data. Their research demonstrates that individuals develop smartphones using habits out of convenience. Convenience and habit might encourage users to enjoy themselves by encouraging them to concentrate on their smartphone activities. Additionally, concentration and enjoyment can encourage people to develop addictions. Overall, our research suggests that the unusual feature of extended exposure might pose a ”danger” to smartphone users. Firstly, again this research is limited only to undergraduate students as well as focus on four driving factors or MEE. Second, they concentrated on the current state of smartphone addiction in its particular setting. Third, there might be some other significant elements lacking given the model's ability to only explain relatively minor differences. Fourth, as a preliminary study, they showed the risks of elements that were previously thought to be favorable in IS research. In another research paper on Kids' Smartphone addiction [35] has done a systematic literature review to identify the contributing variables and negative impacts of preschool kid's smartphone addiction. All the elements were determined after doing a thorough quality analysis on the selected 20 papers, and they were then divided into ten independent variables, one dependent variable, and four negative impacts. Ten independent factors, including parental use, gaming, peer influence, accessibility, learning applications, video and photo viewing, gadget use, loneliness, and single-family living. While having four negative impacts, including aggression, withdrawal, silence, and anxiety. A relationship view of the independent and dependent variables is presented by their conceptual model and they developed some hypotheses. These papers helped us to brainstorm about any factor that we can study further to detect smartphone addiction.

Besides, using scales and app usage to measure level of addiction there was a different approach which was mentioned in [16]. In [16] researchers used a different method of data collection rather than self-report method. They came up with the concept of using smartphone sensors to get more unbiased data regarding usage habits. Because Zhou, Rong and Shi, Dianxi (2016) discovered that reporting a user's daily consumption is subjective and erroneous, thus 16 people's log data has been gathered for more than 6 months. Over the course of the trial, they discovered the monthly usage frequency and duration of more than 300 APPs. After reviewing this research work we also got intuition to use direct smartphone usage data and sensor data from participants' mobile phones. For instance, in [21] Alruban, Alobaidi, Clarke and Li tried to recognize human activity through smartphone sensor data. They collected the sensory data using an app called 'Androsensor' for a total of 6 activities like, normal, fast, with bag, downstairs, upstairs walks, and sitting. They recorded each activity for a 3 minute period that logged data of different sensors like Accelerometer, gyroscope, gravity etc at 3 axes . Then they had done feature extraction on the raw gyroscope and accelerometer data and further preprocessing and feature selection to run three supervised models, NN, SVM and XGB. This methodology reached almost 98% accuracy in detecting these activities from smartphone sensors. From these we got to know about the Androsensor app which can log smartphone's sensor data.

In [43] paper authors tried to recognize an activity of a human specifically eating through accelerometer values obtained from smartwatch sensors. They were successful in developing a LSTM-ANN architecture that can correctly identify eating bites by comparing other activities such as puffing or jogging 90% of the time. The study focused on eating, smoking, using medications, and jogging. Smoking and medication use were added because of how difficult it is to detect eating when these behaviors are present. While jogging data was gathered from an available public database, hand-to-mouth events for eating, smoking, and taking medications were recorded using accelerometer and gyroscope sensors included in smartwatches. This paper gave an idea of using LSTM for better detection of human activity. In another paper [23] chalk Wilhelm Pienaar and Reza Malekian again approached human activity detection using the LSTM algorithm rather than only using RNN architecture. They used LSTM instead of RNN because LTSM overcomes the limitation of RNN called vanishing gradient problem. They used raw sensor data provided by WISDM that had data of six activities, namely sitting, jogging, standing, walking, upstairs and downstairs. These sensory data were also three-axial in nature only for the accelerometer sensor. After preprocessing and train-test splitting they trained LSTM with a part of the dataset for train and they predicted activity from the other test dataset. From the result they found out that their model gets confused between upstairs and downstairs walking. But other than that LSTM-RNN deep neural architecture reached about 90% accuracy while predicting the activities. Thus these papers motivated us to use LSTM architecture to human activity from smartphone's sensor for our phone addiction research.

2.2 Background Study

These research investigations provide light on a variety of smartphone addiction-related topics, such as its categorization, contributing causes, effects on various age groups, and suggested remedies. They give insights on prospective therapies and tactics for controlling excessive smartphone usage, which helps us better comprehend the intricacies of smartphone addiction.Although these study articles may not directly relate to our research's goals, they have a lot to say about smartphone addiction and may offer some helpful insights into the overall situation.

In [10], proposed an integrative pathway model intended to serve as a theoretical framework for future research in the field of PMPU (Problematic Mobile Phone Use). According to Billieux et al. (2015), despite the fact that many scholars consider PMPU to be a behavioral addiction, there is still a dearth of evidence that either confirms or refutes this notion. In fact, only research methods based on self-report data collected through convenience samples support the hypothesis that this condition can be categorized as an addiction to date. Two main arguments were presented in [10]. First, the lack of evidence supporting the behavioral addiction model's applicability to PMPU was highlighted. The second discovery was that PMPU is a complex and heterogeneous condition. This model proposed by Billieux et al. (2015) served as a guide for our own research.

Because excessive smartphone use has a harmful influence on our daily lives. One of the essential components of digital medications for smartphone addiction, the app limitation feature, is the topic of [39]. Here, Yasudomi et al. (2021) examine how the feature is used and confirm its efficacy. The outcome reveals a clear distinction between individuals who utilized the app limitation option and those who did not in terms of psychological and behavioral traits. A growing field of study called "digital therapeutics" (DTx) uses software like smartphone applications to track and treat medical conditions. They installed a home app on the individuals' cell phones that has a function that limits app use, and they evaluated the function app's efficacy as a digital medication for smartphone addiction.Data from the individuals' smartphone usage logs and questionnaire responses were used in the research. Data from the individuals' smartphone usage logs and questionnaire responses were used in the research. Here, smartphone usage logs were gathered using Macromill's U-Logger for data collecting.Only the logs associated with the four actions—launching applications, hitting the home button, turning off the screen, and shutting off the phone—were retrieved.They used a home app created by KDDI Research to assess the effects of the app limitation features common to these applications.Here they found users of this app limitation feature spent an average of 2.56 hours on their smartphones each day, compared to 3.76 hours for non-users.

The app limitation feature may be useful for individuals who use their cellphones seldom, i.e., those who have not yet shown signs of smartphone addiction and those who are still in the early stages of addiction. However, if a person's smartphone reliance is severe, further intervention strategies are needed. So, to make sure of one's stage of smartphone addiction we need to categorize the severity of this. Which can be also helpful in terms of age structural smartphone addiction analysis.

Additionally in [31], about trait procrastination and mobile phone addiction, the authors tried to find a relation between excessive use of smartphones and procrastination characteristics and if gender plays a role as a mediating factor not. They took samples from 1004 Chinese college students. The samples contained information about mobile phone usage, stress and procrastination characteristics, and demographics. They mentioned that, Empirical studies have shown there is a firm association between stress and procrastination characteristics [1], [2], [8]. With the support of these findings, they tried to show that stress should mediate linking mobile addiction and procrastination. Yang et al.(2020) found that the relationship between trait procrastination and smartphone addiction is partially mediated by stress. Moreover, they tried to show Gender also plays a role in moderating the association between smartphone addiction and trait procrastination, and that was also found. They conclude that stress and trait procrastination are more vital indicators of mobile phone addiction for males than female college students.

In [36], the authors sought to discover smartphone use behaviors that are connected with excessive smartphone use among preschoolers. They discovered virtually little study on PSU habits among younger children, despite the fact that the average age of first smartphone use is declining. In this study, researchers analyzed data from a 2017 countrywide survey performed by the South Korean Ministry of Science and ICT and the National Information Society Agency regarding smartphone addiction. They applied binomial logistic regression on 1,378 preschool kids to analyze. Their research found that children are more attracted to consuming media over the smartphone, which causes problematic smartphone use. However, in contrast to adults, they are not very engaged in games and social media. They discovered that seventeen percent of the collected data meets the PSU's criterion. There are children who use smartphones more than two hours a day, and their frequency of using a smartphone is higher than others. According to Park et al. (2021), one in five preschoolers who use a smartphone are susceptible to PSU. Their finding says the caregiver of children is more responsible for this. They suggested that caregivers should control how a child will use a smartphone, and they need to monitor how much time a child is spending on a smartphone and what they are consuming while using smartphones. The caregivers should engage the children in other activities rather than using smartphones. They conclude by saying that smartphones used by children should be monitored because frequent use of smartphones for a short period may be a symptom of problematic use of smartphones.

Moreover, we came across some papers which were dedicated to find out the factors behind the increase of this problematic issue. In [38], it was stated that problematic smartphone use has been increased by stress levels. The Authors scrutinize the association between stress, problematic smartphone use, and FOMO (Fear of missing out). Their goal was to examine the effect of smartphone usage frequency and FOMO as mediators between stress and problematic smartphone usage. In July 2019 they surveyed a large sample of 2,276 Chinese undergraduates. Yang et al. (2021) discovered that stress was connected with PSU based on the outcome. Moreover, the outcome indicated that FOMO was positively linked with SUF and PSU severity. Their research suggested that FOMO may play a crucial role in understanding why some individuals with elevated stress levels may engage in excessive smartphone use. Further, they argued that FOMO is a novel and essential psychological construct that is strongly connected with the severity of PSU and SUF. Multiple studies have demonstrated a correlation between negative emotions and FOMO, which is highly correlated. In the paper's discussion, it was stated that FOMO was a determinant of SUF and PSU intensity. Moreover, in their research, they also got similar findings about FOMO.

Besides the factors that correlates with smartphone addiction there are numerous physical and mental negative impacts of it. From the very beginning of this problem researchers are working on finding the impacts of it to be aware and save people from the bad consequences of it. The [40]is about how excessive use of smartphones can affect works that require mental effort and creativity. The authors argue that the use of a smartphone can be habitually disruptive, with the most major unfavorable consequence being the incapacity to concentrate for lengthy durations. This restriction may hinder innovation in the actual world and the development of domain-specific knowledge. Other studies have indicated that the association involving mobile use and cognitive characteristics such as attention and working memory is minimal, indicating that digital technology does not have a detrimental effect [19]. However, they argued that disruptive smartphone use had a detrimental influence on people's ability for prolonged cognitive effort while doing activities not related to smartphones. Aru et al. (2022) stated that the human brain performs a cost-benefit analysis before choosing from multiple tasks and then chooses the task which has a better ratio of cost-benefit. In addition, they asserted that cognitive effort is intrinsically more expensive than smartphone usage [18]. That is why smartphone use over other tasks is more likely to be chosen by the human brain. They advised that individuals could increase the cost-benefit value of other activities or reduce the ratio of smartphone usage, or both.

During searching for work on this topic we came across a lot of work of the researchers where they tried to find a pattern of this issue. A research was done on the students of a Peruvian college to find a common behavioral pattern of college-going students [28]. Through the analysis Diaz et al. (2020) found that problematic mobile phone users were more likely than healthy mobile phone users to be single, in their early years of their careers, and to partake in dangerous activities. 921 college students in Peru were polled using questionnaires. The first section of the poll included questions about socio demographics, smartphone usage habits, and drug use. They came to the conclusion that one of the indicators of problematic smartphone use is sensation seeking.

Also there are other psychological problems like phubbing. It is an act of neglecting someone when in a social setting by focusing on the phone (such as to reply to messages or see notifications in their smartphone) [22]. The purpose of this study was to elucidate the psychological rationale for phubbing behavior which can be also helpful to analyze smartphone addiction. According to the findings of the hypothesis test, R Square = 0.496, which means that 49.6% of the variation of phubbing behavior is explained by all independent variables, while the remaining 50% is impacted by other factors. According to a preliminary study, 85.1% of social media users have intentionally engaged in phubbing activity, with friends being the biggest percentage of targets (88.1%), followed by partners and family members. Self control has also been connected to issues with smartphone usage, according to research, which also demonstrates how closely these behaviors are tied to addiction. Students at colleges who own cell phones and are from big cities make up the study's demographic. 246 individuals that ranged in age from 17 to 20 years old were the samples for this study.Manual and online questionnaires are used to collect data.Using the phubbing scale developed by Karadag and translated into Indonesian, phubbing behavior is measured.The Perceived Social Norms of Phubbing Scale is used to quantify norms in society. In this study, researchers employed four different answer options—extremely appropriate (SS), appropriate (S), not suitable (TS), and very not suitable—across five different Likert scale models (STS).Using the SPSS 24 program, researchers conducted a regression analysis.A regression coefficient for each IV is examined by the researcher. To determine what proportion (%) of the variation of DV is explained by IV(norms, disruption of daily life, withdrawal, tolerance, perspective taking, fantasy, empathic concern, personal distress, self-control, and gender), the researcher first examines the amount of R square .The findings revealed an R-Square of 0.496, or 49.6%, indicating that differences in total IV may account for 49.6% of the variance in phubbing behavior, whereas the remaining 50% is impacted by factors outside the scope of this study.Significant regression coefficient values may be found for the following four variables: (1)norm; (2) withdrawal; (3) tolerance; and (4) self-control. Norm, self-withdrawal, tolerance has a favorable effect on phubbing behavior, which implies those who exhibit these IV characteristics more passionately are likely to engage in phubbing. On the other side, it was shown that people tend to phubbing more the less self control they had. It also shows how smart phone addiction is related with our behavioral and mental life as well as has a great influential relation with our social life.

Many of the researches were focused on university students. Another research [30] was done to find the disorders of smartphones among university going students. Studies have shown that smartphones have always had an effect on a person's mental and physical health. Every age group of users has some effect on them. Subramaniam et al.(2020) published the findings of a quantitative study that was carried out to assess smartphone addiction issues among college students in 2020. Online surveys with 26 SPAI questions were conducted using Google Forms. The study's findings demonstrated that when they are unable to use their smartphones, college students experience mental distress, anxiety, and tension. When they temporarily cease using their smartphones, they experience anxiety. Sixty-one percent (61%) of students said that spending too much time on their smartphones has negatively impacted their health and put them in danger, and 58% said that it has had a negative impact on their academic performance. 56% of students also agreed that they are sleeping fewer hours as a result of their excessive smartphone use.

As smartphones are becoming an essential part of human's daily life and consuming a large amount of time, some researchers also tried to propose solutions to this problem as well. One paper to face the distraction issue in case of smartphone usage [24] has proposed an app called '9Seconds', which is like using a smartphone to fight smartphone addiction. This app is based on challenges and rewards. If anyone wants to reduce their smartphone usage, they can start a challenge by searching an opponent through the matchmaking system built in the app. Interested opponents will enter himself into the game room. Users will get points for avoiding smartphone distractions for every nine seconds during the task. This application can detect if the user is using a distracting application or productive or important application. If a user breaks any rule points will be deducted. When the game ends, their earned points will be shown in a weekly leader-board along with the winner who will be rewarded.

Finally, after reading and reviewing all these papers in our research domain we got to know about some wonderful research works which can help us in our own work. Especially, the Smartphone Addiction Scale that can help us to get a clear understanding of a person's state of smartphone addiction through questionnaires. We also noticed some of the shortcomings of the papers, some notables points are, none of the previous works tried to collect in depth smartphone usage data or the data from the sensors, most of the research only focus on users answers to the questionnaires but this method can easily fail if a user lies or simply can not answer the question accurately in the moment of enquiry. In our research, we will try to mitigate both of the problems by directly collecting data from the smartphone and cross checking the questionnaire with the collected data. This will make our work less biased and depend on the actual uses rather than the users' opinion.

Chapter 3

Methodology

The following diagram shows the overview of all the work done during this research.

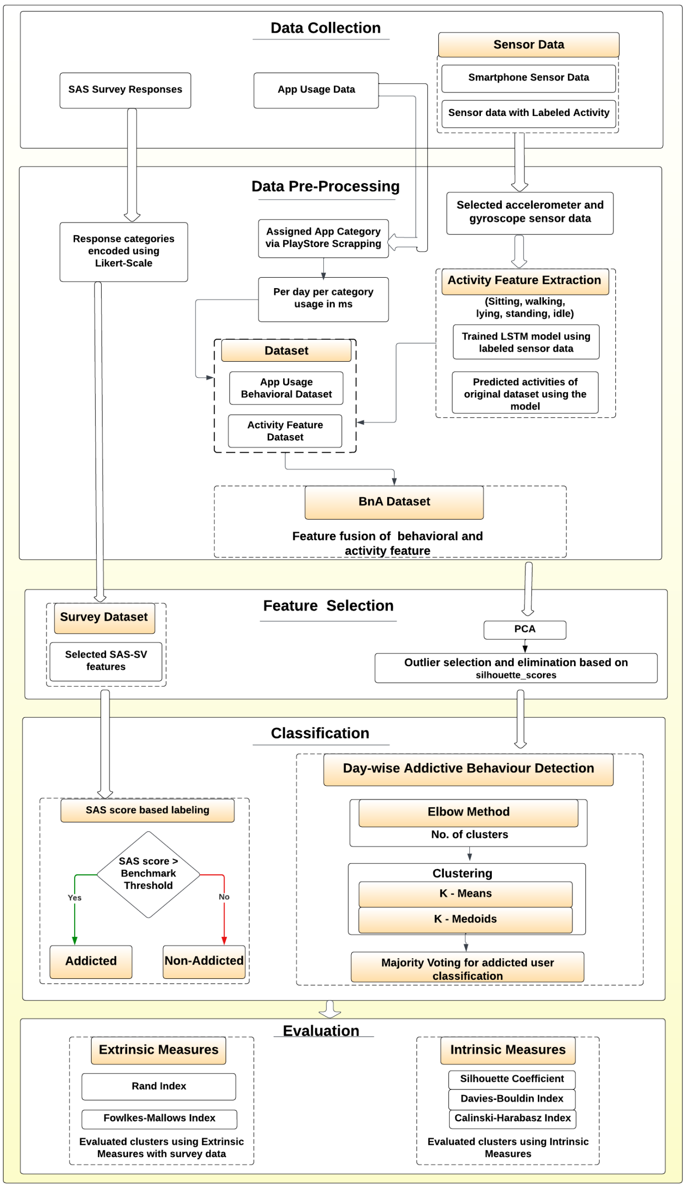


Figure 3.1: Top Level Overview Of The Proposed Method

At first as shown in figure 3.1 our research started with data collection. We have used primary dataset which was gathered using various methods from a fixed group of participants. We collected 4 types of dataset in total which are - SAS survey dataset, app usage behavioural dataset, sensor dataset, sensor data with labelled activity dataset. After doing necessary pre-processing on the sensor data we created a dataset containing activities which was extracted based on a LSTM model. Moreover, app dataset was pre-processed into per day per category usage time. Then, these two datasets feature were fusioned to create a new dataset named BnA Dataset. In the later phase, essential features were selected as shown, SAS-SV questions were selected from survey responses and through PCA features from BnA dataset were selected for further classification. For the classification part clustering algorithms were used to classify the addicted and non-addicted participants from BnA dataset. On the other hand, SAS-SV survey was labeled based on SAS scores of the parcipants response. Lastly, result of the BnA dataset clusters were evaluated using different intrinsic and extrinsic measures.

3.1 Data Acquisition

This study employed a research design using data collected in various types and ways. In order to conduct our research we have collected three datasets from different sources. These are:

1. SAS Survey Data,
2. App Usage Data,
3. Sensor Data.

Here, we have used a survey procedure as a data collection method and two different mobile app data to analyze. A self- administered questionnaire was made with the SAS scale to do the survey. For the app's data collection it was done on a 30 number of participants for a time period of 7 days.

Survey questions which were used as the data collection method to know about participants' smartphone and social media usage habits were conducted within a period of 4 weeks starting from Dec 7 2022 to Jan 12 2023. This questionnaire is divided into two sections. Firstly, demographic information for example age, gender, occupation etc. Then we have 33 questions from the SAS (Smartphone Addiction Scale) [5]. A six-point Likert scale, from strongly disagree to strongly agree, was used to evaluate these questions.

After creating the questions, the next step was to disseminate them. To collect the participants' responses, we constructed a google form that had all of the questions. We used a variety of social networking platforms, including Facebook, Whatsapp, Slack, Discord, and others to distribute the survey form to selected 30 participants. So that's how we exclusively gathered the survey's data. We also provided detailed guide to install the apps to them.

Socio-demographic Characteristics Of Survey			
Variables		**N**	%
Gender	Male	12	42.86%
	Female	16	57.14%
Age	below 22	11	39.29%
	above 22	17	60.71%

Table 3.1: Socio-demographic characteristics of survey respondents

Here, as we can see in the table 3.1, out of the total survey respondents, 12 participants (42.86%) were male, and 16 participants (57.14%) were female. Similarly, 11 respondents (39.29%) were below the age of 22, while 17 respondents (60.71%) were above the age of 22.

The next method of data collecting involved use of a monitoring app called "My phone time" and 'AndroSensor' for collecting sensor data which are available in

google play store. This was used to compile information about the previously mentioned group of participants' smartphone usage.
'My phone time' mostly logs which apps are used and for how long by a user. The data included the app name, app package name, start time, and duration of use for each app.Each app's start time was broken out into individual columns for the day, the year, the month, the hour, the minute, and the second. This allowed for the calculation of daily smartphone usage, which was defined as the period from 00:00 hours to 23:59:59 hours.

App Uses Dataset Summary		
Total Features	**Time Frame**	**Participants**
4	7 days	30

Table 3.2: Summary of Data Collected from "My Phone Time" app

All the sensor data gets stored as a csv file in a file in the user's phone. These csv files contain 31 columns in which all of them contain sensor values except two. Sensor values such as Accelerometer, gyroscope, gravity, magnetic field etc, were recorded for all three axes.

Finally, in order to train our used model to determine human activity we recorded some sets of activities, namely walking , sitting, standing, lying, idle phone facing down and phone facing up from 15 participants. In this study, we utilized the selected sensor data sampled at a frequency of 0.2Hz. These behaviors were recorded from few participants for around 10 minutes to use as a labeled dataset in order to detect human activities.

Smartphone Sensor Dataset Summary			
	Total Features	**Time Frame**	**Participants**
With Activity Label	31	7 days	30
Without Activity Label	32	10 min	15

Table 3.3: Summary of Data Collected from "AndroSensor" app

We used both online and offline methods to handle the participants in our study. Additionally, any data pertaining to survey or app activity by participants was collected with their agreement. The survey and tracking app data, as well as the participant's personal information, have all been protected in accordance with legal requirements ensuring confidentiality.

3.2 Data Pre-Processing

After acquiring total four types of dataset, these were pre-processed in various methods to make it usable for our model. Table 3.4 contains all the datasets used in our research. This section contains all the pre-process phases in a descriptive manner.

Dataset Name	Description	Acquiring Process
SAS Survey Responses	33 SAS scale questions, 3 socio-demographic questions	By distributing Google forms online to control group
App Usage Behavioral Dataset	App package, App name (in object), start time (in ms), duration	Installing an app available in play store which keep tracks of the daily smartphone usage of users
Activity Feature Dataset	Acceleration x, y, z (m/s^2), 'Linear Acceleration x, y, z (m/s^2)', 'Gyroscope x, y, z (rad/s)', Gravity x, y, z (m/s^2) , activity label	Sensor dataset was collected using an app available in play store . All the data's were collected in 5 second interval from control group participants. It was collected in with and without label form. After extracting activity Feature vector from LSTM model this dataset was created.
BnA (Behavioral And Activity) Dataset	Date, ID, time of 6 activities (in seconds), per day time duration of 33 categories(in seconds),	This final dataset was obtained by merging the App usage behavioral dataset and Activity feature dataset

Table 3.4: Dataset Overview

3.2.1 Survey Data

Survey Data were collected through google forms from our selected 30 participants. Initially the csv file of this dataset had columns and 30 rows. The columns or feature contained 3 demographic questions like, Age, Gender and Employment status along with 33 questions from SAS-scale.

First of all, unnecessary columns such as 'timestamp' and 'I permit the use of my data for research purposes' were dropped from the dataset. Then null values and duplicate values were checked in case any participant made an error while submitting the form. Thankfully, there were none so we didn't have to drop any rows or columns. A short version SAS scale with only 10 questions from the previous 33 questions was proposed by [5], where it was stated that only these 10 questions are enough to evaluate smartphone addiction. That's why we, at first, selected 10 questions or features that was filtered out by [5] out of those 33 questions along with other 3 demographic questions. We also added another feature called 'ID' to uniquely identify each participant for our further research work without disclosing their personal identification. Now the dataset holds total 14 features. The answers of the questions were all categorical data with 6 types of values from 'Strongly Disagree' to 'Strongly Agree'. So we had to encode these categorical data from 'Strongly Disagree' to 'Strongly Agree' into 0 to 5 integer values. Then this 'SAS-SV Survey Response dataset' were used to calculate SAS-SV Score to label each ID Addicted or Non Addicted.

Items	
1	Missing planned work due to smartphone use
2	Having a hard time concentrating in class, while doing assignments, or while working due to smartphone use
3	Feeling pain in the wrists or at the back of the neck while using a smartphone
4	Won't be able to stand not having a smartphone
5	Feeling impatient and fretful when I am not holding my smartphone
6	Having my smartphone in my mind even when I am not using it
7	I will never give up using my smartphone even when my daily life is already greatly affected by it.
8	Constantly checking my smartphone so as not to miss conversations between other people on Twitter or Facebook
9	Using my smartphone longer than I had intended
10	The people around me tell me that I use my smartphone too much.

Figure 3.2: SAS-SV Questionnaires

3.2.1.1 SAS-SV Score Calculation

To classify the users we calculated the participant's total score and then used the cut-off value mentioned in the SAS-SV scale from our "SAS-SV Survey Response Dataset". Which is 33 for Females and 31 for Males. Finally we've set a label to the participants based on their score in the dataset.

3.2.2 App Usage Data

Through the "My Phone Time" app, we have also We collected App Usage information from our 30 participants for our research. This data was collected for 7 days or below from each participants. This application logs the following information: App package, App name (in object form), Start time (in milliseconds), and duration (for each app usage). For the purpose of examining various analogies, this data must be preprocessed.

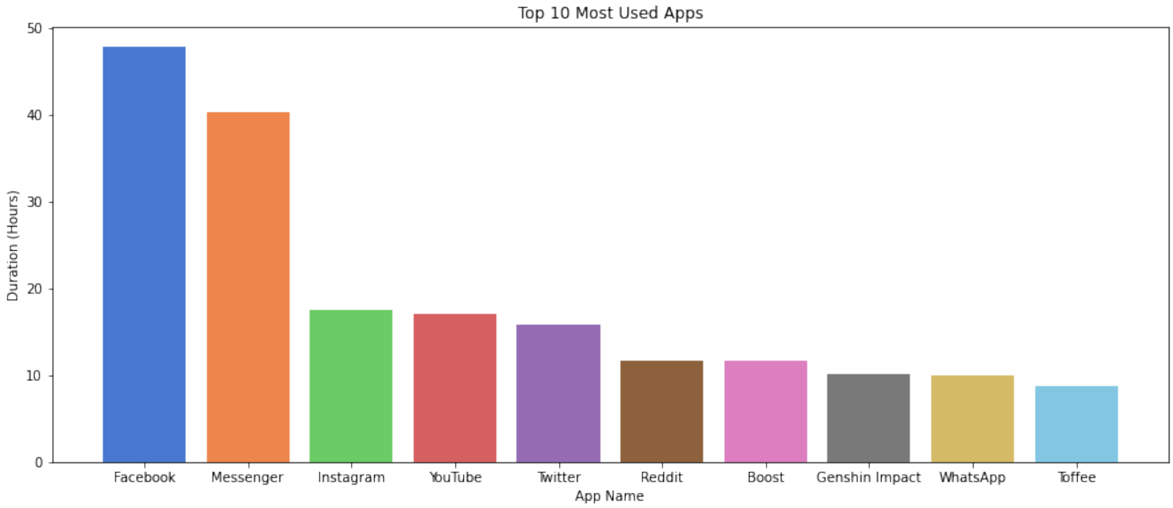


Figure 3.3: Most Used Apps by the Participants

In order to do that, we must first read the csv files that we have gathered from various people. The Exported CSV file from each participant had total 4 features with inconsistent number of rows for each participant. We had to merge the csv files because there total 30 of them. Therefore, we imported glob and listed every path to each of those csv files. Then, we constructed dataframes by creating a loop around that list and read each csv file. In order to uniquely identify each dataframe and its contents, we simultaneously assigned a unique id to each dataframe. After that, we combined every data frame into a single dataframe. In this step we assigned same ID that was assigned to the survey data to each person to maintain the harmony between survey responses and App usage data of each participants for future evaluation of our model.

We then created a column called ”Category” that contains the category to which each app belongs. The ’google play scraper’ module makes this simple to implement, and we used app() to scrape the category of each app from the Google Play Store. We developed a function that receives an App package string and returns the app’s category. Then, we simply attached the id and category columns after mapping those categories to our merged dataset.

Finally, we made another data frame by grouping dates of each participants. In this step we mainly calculated total time spent in milliseconds in each day by each participants at each category type of app. So, this new dataframe contains 35 features which are, Date, ID and app category names and had 222 rows with 30 participant’s 7 or below days data. From this dataset we found that, most used category was Communication and Social whereas health and fitness, travel and local etc. This dataset is called ’App Usage Behavioural Dataset’ which will be used to make our main dataset by combining with preprocessed sensor data.

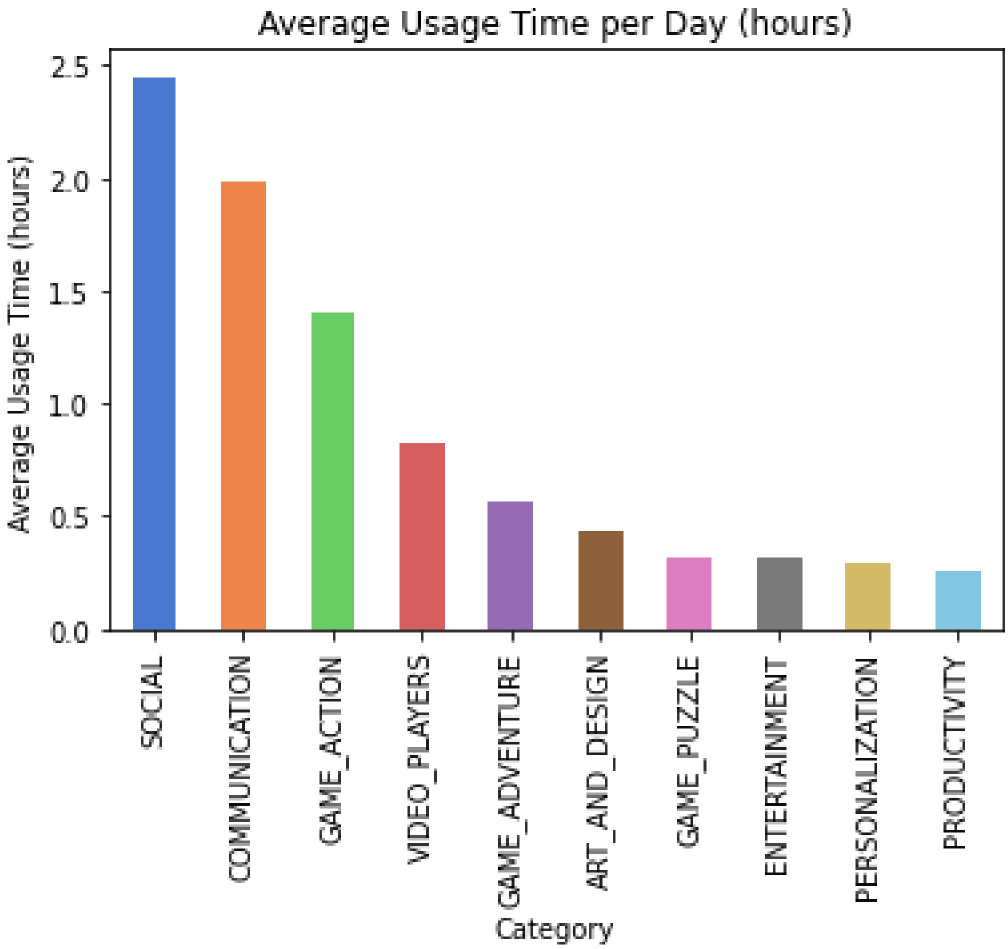


Figure 3.4: Average Daily App Uses Based on Category

3.2.3 Smartphone Sensor Data

We have collected the smartphone sensor data through the app called 'AndroSensor' which includes the 'Acceleration x, y, z (m/s^2)', 'Gravity x, y, z (m/s^2)', 'Linear Acceleration x, y, z (m/s^2)','Gyroscope x, y, z (rad/s)', 'Orietation x, y, z', etc. sensor features from the participants' smartphones during the 7-day data collection period. We also collected labeled sensor data from 15 persons of our previous 30 participants. We divided the activity labels into five categories: **idle phone facing down**, **idle phone facing up**,**walking**, **standing**, **sitting**, and **lying**.

The smartphone sensor information that was collected from 30 people for 7 days were initially separated into multiple CSV files. So, we had to merge those CSV files for each participants at first and then merged every participants data file into one. While merging each persons CSV file we had to assign unique ID which is again same as the survey dataset or 'App Usage Behavioural Dataset' to maintain the consistency between every data files.

To train the LSTM model which will be used to label our sensor dataset later, we had to collect sensor data for six specific activity, that led us to total 90 CSV files in our hand from 15 participants for all types of activity. Each activity's CSV file contains sensor information of 10 minutes duration for each activity from every participant. So, we had to merge these CSV files using glob and had added another column as label called 'Activity'.

So we had two different dataset holding sensor data, one with labelled activity and another without any label. Then We only selected Acceleration x, y, z (m/s^2)', 'Linear Acceleration x, y, z (m/s^2)','Gyroscope x, y, z (rad/s)', 'Gravity x, y, z (m/s^2)' in these two datasets to train the LSTM model and predict activities. In order to verify that the raw sensor data is in a consistent time-series format, the data was further pre-processed by applying a low-pass filter to eliminate high-frequency noise.

After pre processing, the sensor data was segmented into fixed-length time windows of 15 seconds with 50% overlap. This ensures that the activity recognition model can capture the dynamic changes in the sensor data over time.

3.2.3.1 Activity Detection Using LSTM

We used the Long Short-Term Memory (LSTM) algorithm proposed by [23] to train the model for activity detection. LSTM is an enhanced version of Recurrent Neural Network(RNN) Algorithm. This LSTM was introduced because of the long term dependency problem or vanishing gradient problem that RNN faces practically, though it was theoretically proven that RNN can hold long term memory. A simple RNN structure is basically a chain of some neural network modules with only a simple tanh layer. But in the LSTM structure shown in the figure - 3.5, each repeating module has four layers with sigmoid and tanh functions that helps to forget unnecessary information and pass through important information for a long time to solve the vanishing gradient problem in RNN.

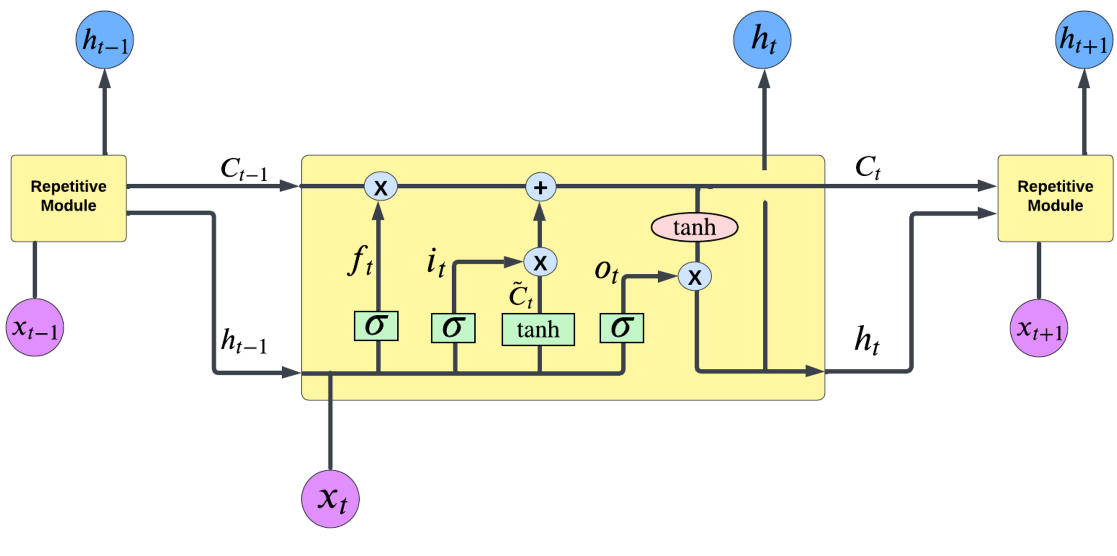


Figure 3.5: The Repeating Modules in an LSTM Structure

This figure - 3.6 describes the notations used to depict the LSTM structure. Other than that there is something called a gate which holds a Sigmoid function and a point-wise multiplication operation also shown this figure. As the Sigmoid function can give values from 0 to 1, these gates control how much information will pass through the gate. Moreover each repeating module in LSTM is called a cell and the cell state from the previous module is passed to the next through the horizontal line running through the top of each cell. Additionally, in the LSTM, there are a total of three gates structure named: Input Gate, Forget Gate and Output Gate.

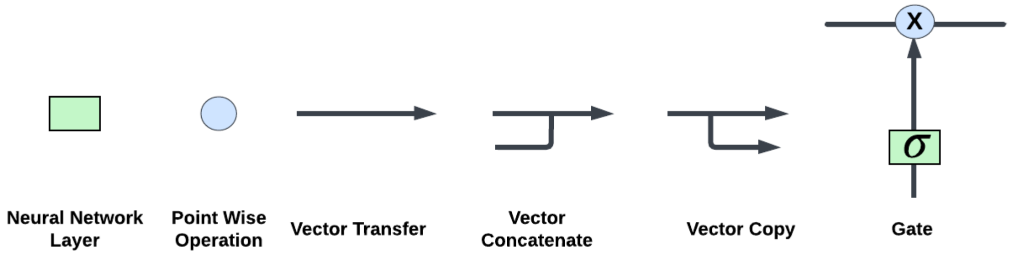


Figure 3.6: Notations And Gate of LSTM Structure

In the beginning each cell decides how much information is going to be thrown away from the previous hypothesis and it is done by the Forget gate. It concatenates previous hypothesis vector h_{t-1} and current input vector x_t, which then is multiplied by the weight matrix of the forget gate. Then it adds a bias b_f and passes through the sigmoid function to calculate f_t as shown in the figure - 3.7.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.1}$$

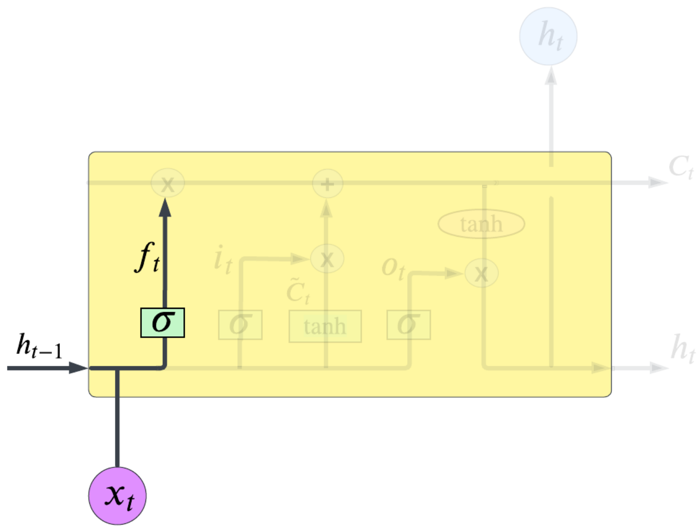


Figure 3.7: Working Procedure of Forget Gate Layer

The amount of fresh data to be stored in the current cell state is decided in the next phase. In order to determine which value from the input and prior hypothesis will be used for this, the input gate layer first concatenates these two vectors and multiplies an input weight matrix W_i.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.2}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3.3}$$

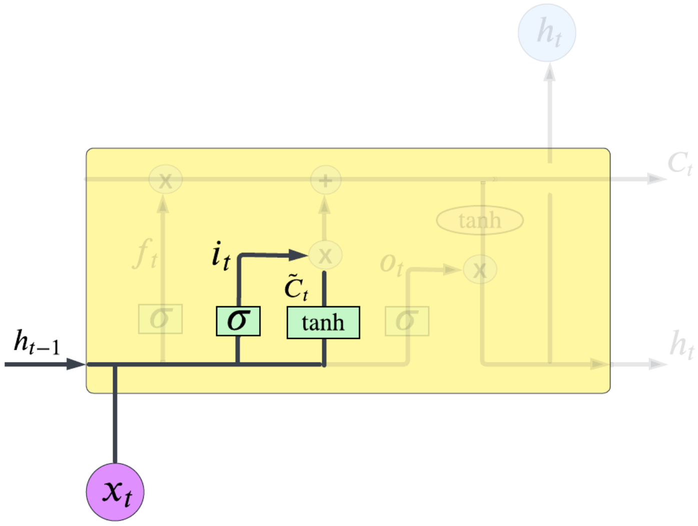


Figure 3.8: Working Procedure of Input Gate Layer

As like the previous forget gate it also adds a bias b_i and passes these into the sigmoid layer for input gate and calculates i_t. Another *tanh* function enhances or shrinks the information from the h_{t-1} and x_t by multiplying these with W_c. It also adds a bias b_c before passing through the *tanh* function and calculates $\tilde{C}_t$ as shown in the figure 3.8.

Then it updates previous cell state C_{t-1} by multiplying with f_t and adding another multiplication of i_t and $\tilde{C}_t$. Basically in this step, the current state is enhanced while forgetting the earlier things that were made a decision to forget and by adding $i_t \odot \tilde{C}_t$ which are the updated candidate values, scaled according to how much each state's value will be updated. This gives Current cell state C_t as shown in the figure - 3.9.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3.4}$$

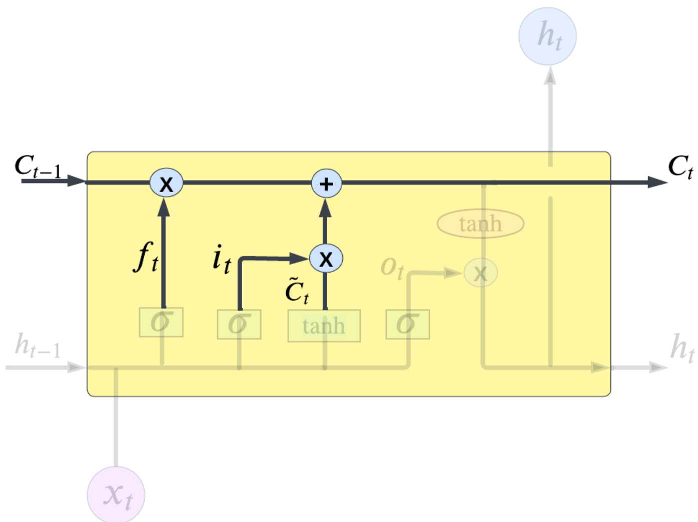


Figure 3.9: Current Cell State Calculation

The output is then determined by running the current data through the output gate layer's sigmoid function. The current cell hypothesis, denoted by the symbol h_t, is then obtained by multiplying the cell state, C_t, by the output of the output gate, o_t. The calculation of o_t is done in a manner similar to how f_t and i_t were done. This hypothesis then has been passed through two lines, one of them is for the next cell's calculation and another is the output of the current cell which is sometimes passed though as output after applying softmax.

$$h_t = o_t \odot \tanh(C_t) \tag{3.5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3.6}$$

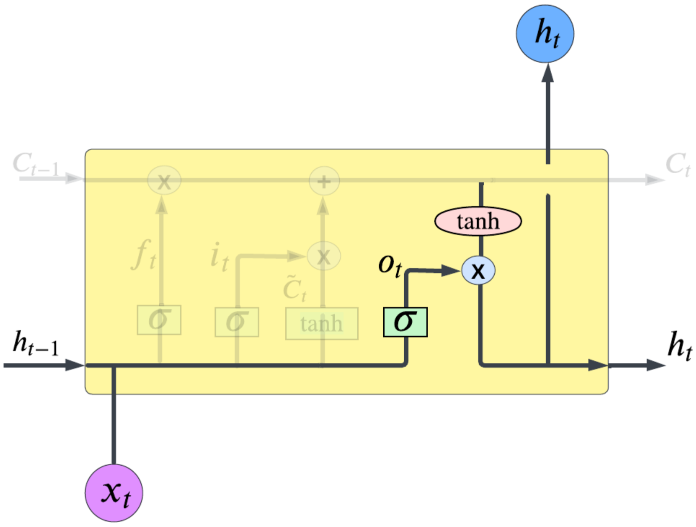


Figure 3.10: Output and Current Cell Hypothesis Calculation

In our research, The LSTM model takes a sequence of feature vectors (i.e., the pre-processed sensor data for each time window) as input and outputs the predicted activity label for each window. To train the LSTM model, we used a supervised learning approach where we provided the model with the labeled activity data. We randomly split the labeled data into training and testing sets with a ratio of 80:20. We used cross validation score to tune the model hyperparameters and prevent over-fitting. We trained the LSTM model using the Adam optimizer with a learning rate of 0.000001 and a batch size of 32. The model was trained for 100 epochs, and the best model was selected based on the validation set accuracy. The chosen model's total test set accuracy of 97.44 % showed the viability of the suggested approach for activity identification while using a smartphone.

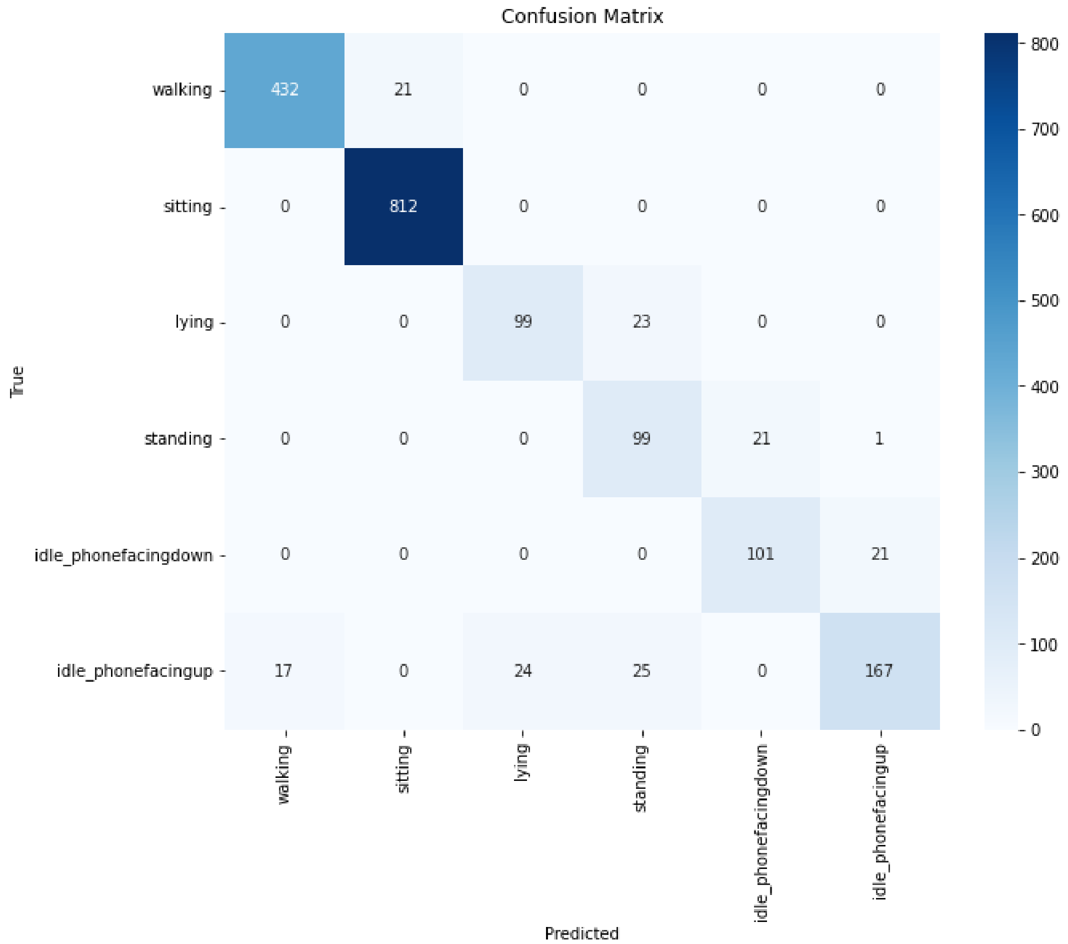


Figure 3.11: LSTM Activity Recognition Model's Confusion Matrix

After predicting the activity from smartphone we labelled the sensor dataset (without label) with the predicted values. Then we created another dataset called 'Activity Feature Dataset' by calculated total spending time for each category at each day for every ID. So this 'Activity Feature Dataset' has 8 columns, i.e, ID, Date, idle phone facing down, idle phone facing up, walking, standing, sitting, and lying along with 222 rows for 30 participants 3-7 days activity information data.

3.2.4 Behavioural and Acitivity (BnA) Fusion Dataset

After preprocessing 4 different raw datasets, we are now left with the creation of BnA dataset combining 'App Usage Behavioural Dataset' and 'Activity Feature Dataset'. For that we just fusioned these two datafreames by matching the ID and timestamp column and got the 'Behavioural and Acitivity (BnA) Dataset'. This Data set holds 222 rows and 42 columns carrying date, ID, Activities and app categories. We found out that some ID has only 3 days data which may act as outliers while clustering these dataset. so we dropped those IDs that has only 3 days data and left with a dataset with 216 rows. Then we saw that we can combine idle phone facing up and idle phone facing down activity as it both means phone were entirely idle that time. So, we combined these two feature by summing the values and finally got 41 feature for further clustering purpose.

BnA Dataset Summary		
Total Features	**Time Frame**	**Participants**
41	7 days	28

Table 3.5: Summary of BnA Dataset

3.2.4.1 Dimentionality Reduction Using Principal Component Analysis (PCA)

Dimensionality reduction is a crucial technique in data science that involves reducing the number of features or variables in a dataset while preserving the essential information. When working with high-dimensional data, it is especially helpful since dimensionality reduction can make the analysis simpler, increase computational efficiency, and lessen the negative effects of dimensionality. There are two main approaches in dimensionality reduction. Those are feature selection and feature extraction. In feature selection, the features which are more relevant to the analysis are taken and other features are not taken. Some of the common methods of feature selection are filter methods, wrapper methods and embedded methods. Feature extraction is different from feature selection. In feature extraction, it creates new features containing the information from the original features. The new generated features are known as “latent variables” or “components” are a combination of the original features but typically the number of newer features is less than the original features. One of the widely used feature extraction methods is Principal Components Analysis (PCA).

Principal Components Analysis (PCA) helps to extract the features from higher dimensional data. Initially, the data should be standardized as PCA is sensitive to the scale of the variables. Therefore it is recommended to perform the standardization before PCA. Standardization can be done by subtracting the mean from the dataset and naming the new dataset RowDataAdjust. In PCA algorithm, it starts with calculating the covariance matrix of the features. If the data is of N dimension then the covariance matrix would be of NxN size. Covariance is a measure of how much two dimensions vary from the mean with respect to each other. In the next step, It requires to calculate the eigenvectors and eigenvalues for the derived covariance matrix. The high information axis’ directions are represented by eigenvectors or the covariance matrix.

Additionally, the eigenvalues are defined as the coefficients of these eigenvectors. Next, it needs to sort the eigenvalues in descending order. Along with that, the eigenvectors need to be sorted accordingly. The eigenvectors with the highest eigenvalues capture the most significant variability in the data. Therefore it is needed to select the top k eigenvectors to form a lower-dimensional subspace and here k is the desired number of dimensions for reduced data. The optimum number for k can be found in various methods. One of the methods is to follow the Kaiser’s rule. Now there will be N number of eigenvalues, if k number of eigenvalues are taken then the dimension would be changed from N to k. Now form the RowFeatureVector matrix by the k number of sorted eigenvectors. To find the transformed reduced data, take the transpose of the vector and multiply it on the left of the original dataset. Finally,

after transposing the derived matrix it will get the reduced dimensional data.
Covariance-

$$cov(X,Y) = \frac{\sum_{i=1}^{\mathbf{n}} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{Y}_i - \bar{\mathbf{Y}})}{(\mathbf{n}-\mathbf{1})} \tag{3.7}$$

Eigenvector and Eigenvalue- If v is a vector that is not zero, then it is an eigenvector of a square matrix A if Av is a scaler multiple of v , where λ is the eigenvalue . This can be written as

$$Av = \lambda v \tag{3.8}$$

Find the optimum number of k components in PCA- According to Kaiser's rule, it is recommended to keep all the components with eigenvalues greater than 1. For instance, in Figure 3.12, the first 3 components are having eigenvalues greater than 1. Hence here needs to take 3 components in PCA.

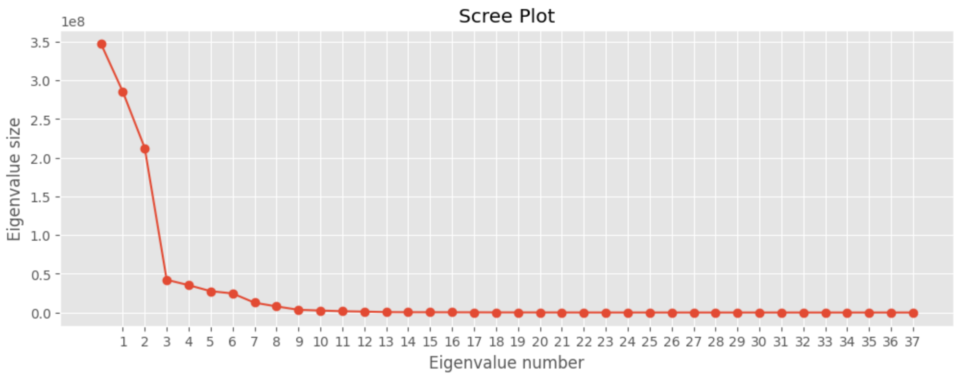


Figure 3.12: Eigenvalue number x Eigenvalue size

Initially, the BnA dataset had 41 features, making the analysis computationally expensive and may lead to overfitting. Hence dimensionality reduction technique was needed. Principal Component Analysis (PCA) is one of the most used dimensionality reduction methods. PCA can reduce the number of features in a dataset while maintaining the most information. This can be helpful in data visualization and analysis. PCA can be used to identify the underlying structure of the dataset and it can be used to find the patterns in the dataset which can be difficult to see with the original features. Along with that, applying PCA increase the performance of the machine learning algorithms. Data can be pre-processed using PCA before being fed to a machine learning system. This may enable the algorithm to learn more efficiently and perform better.

To start with the Principal Component Analysis (PCA), we needed to determine the optimal number of components to take from the features. For that, we plot the eigenvalue numbers against the eigenvalue size in Figure -3.12. From Figure -3.12, we can see three components are having values greater than 1. According to Kaiser's rule, it is recommended to keep all the components with eigenvalues greater than 1. In Figure -3.12, it is showing that three components are having values greater than 1. Besides that, three components of the PCA is giving a variance of more than 80 per cent and that can be seen the Figure -3.13. Hence we took three components in our dataset. After that, the dimension reduced version of BnA dataset has 216 rows and 3 columns, which is a huge reduction in dimensionality compared to the original dataset.

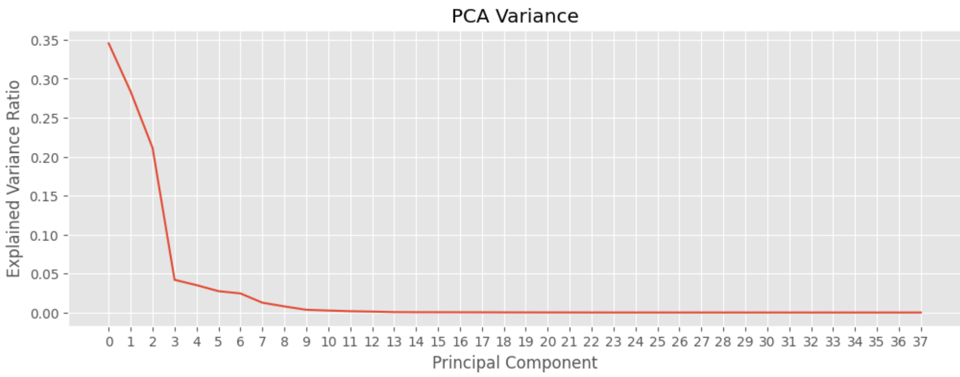


Figure 3.13: Explained variance for PCA

After the dimensionality reduction, BnA dataset Data points are plotted in the Figure -3.14.

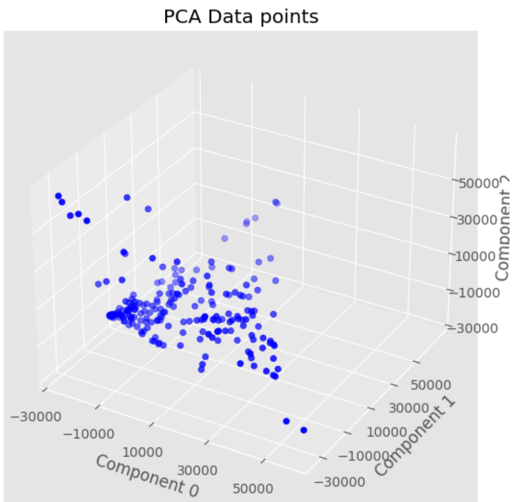


Figure 3.14: Data Points after dimentionality reduction

3.3 Working Procedure Of The Models

3.3.1 K-Means Algorithm

Unsupervised learning approaches are used by several machine learning algorithms to handle unlabelled datasets, which require a distinct approach. Clustering is one of them, and it compares similarities in traits and patterns between several clusters of the unlabeled data. Clustering techniques come in a variety of forms, including partitioned clustering, fuzzy clustering, hierarchical clustering, etc. K-means clustering is a type of partitioning clustering algorithm that, on its own, mostly splits the dataset into k groups without the need for any training.

Let's imagine we want to find clusters in a data set that looks like the figure-3.15, where the x and y axes represent the two different features. Now that the data set has been provided to us, we are unsure of what to look for because we don't have any information about the target variables. All we're trying to do is identify some structure into it and one way of looking into this is these two clusters like in the figure-3.15. We can see this data set has two clusters just by looking at it and k means will help us to find these clusters.

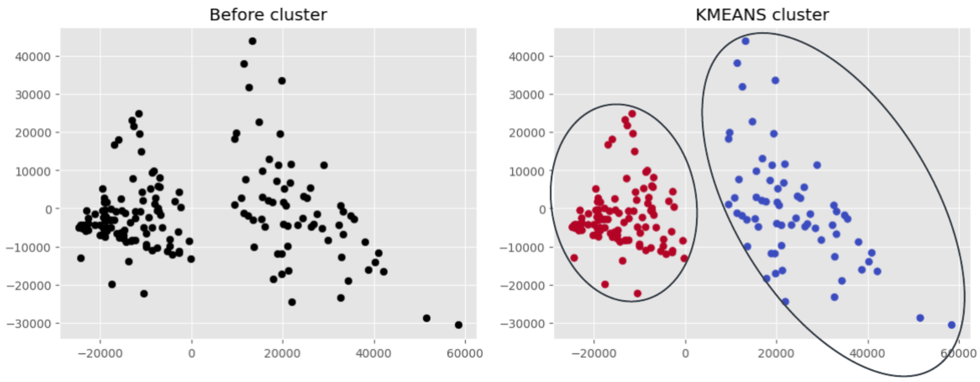


Figure 3.15: Data points and cluster

Finding random points that can be used as the cluster centers is the first stage in the k-means method. We also refer to them as centroids. The next step is to calculate the separation between each data point and the centroids. If a data point is closer to the red centroid, for instance, we will say that it belongs to the red cluster, whereas a different data point is closer to the blue centroid, in which case it will belong to the blue cluster. Using any distance technique, such as the manhattan distance, euclidean distance, etc., we can easily determine the distance between the centroids and data points.

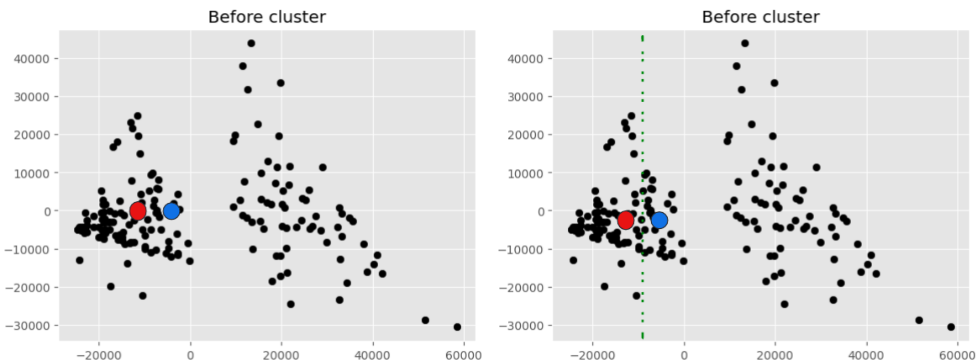


Figure 3.16: Centroids and Clustering

We then encounter imperfect, chunky clusters and work to make them better. At each level, we'll strive to tweak the centroids to make the clusters better and better. For these two clusters, for instance, we will try to locate the center of gravity and place the red center there for the red cluster with four data points, and we will do the same for the blue cluster. The distance between each of these points and these centroids is now calculated once more by using the same procedure. We place a point in a red cluster if it is closer to red than blue; otherwise, we place it in a blue cluster. When none of these data points shift positions after trying to recalculate everything, we are done.

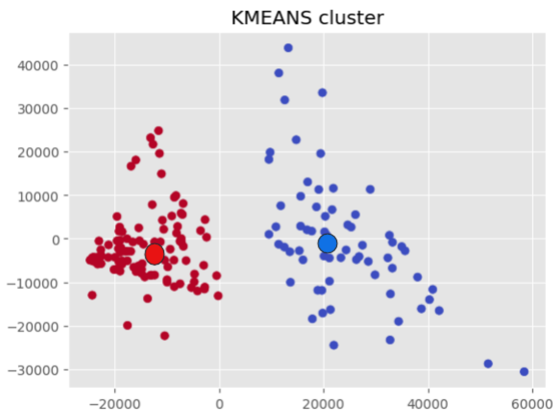


Figure 3.17: After Adjusting Centroids and Recomputing Clusters

The equation for k-means clustering is typically represented using the following notation:

$$\min_{\mathbf{c}1,\ldots,\mathbf{c}k} \sum_{i=1}^{n} \sum_{j=1}^{k} ||x_i - \mathbf{c}_j||^2 \tag{3.9}$$

where $\mathbf{c}_j$ is the j-th cluster centroid, x_i is the i-th data point, $||x_i - \mathbf{c}_j||^2$, is the squared Euclidean distance between x_i and $\mathbf{c}_j$. The overall goal is to minimize the sum of the distances between each data point and its closest cluster centroid. It can also be represented as

$$\min_{\mathbf{c}1,\ldots,\mathbf{c}k} \sum_{j=1}^{k} \sum_{x_i \in S_j} ||x_i - \mathbf{c}_j||^2 \tag{3.10}$$

where S_j is the set of all data points assigned to cluster j.

3.3.1.1 Elbow Method

The most crucial aspect of K-Means and K-Medoids Algorithms is that we must provide k to these algorithms and determine a suitable value for k since, in practice, we will have a large number of features and it will be challenging to represent that data on a scatter plot. The elbow method is a technique that can be used in this situation. Therefore, the task is to determine the best k number using the elbow method.

With this approach, we begin with some k and attempt to calculate the sum of square error. This suggests that we try to compute the average distance between each data point and the centroid for each cluster, square that distance, and then sum the findings. As a result, we obtained the sum of square errors for each cluster, and at the end, we obtained the entire sum of squared errors. After that, we only square to handle negative values. We repeat the entire procedure for various k values and create a plot similar to this figure 3.18.

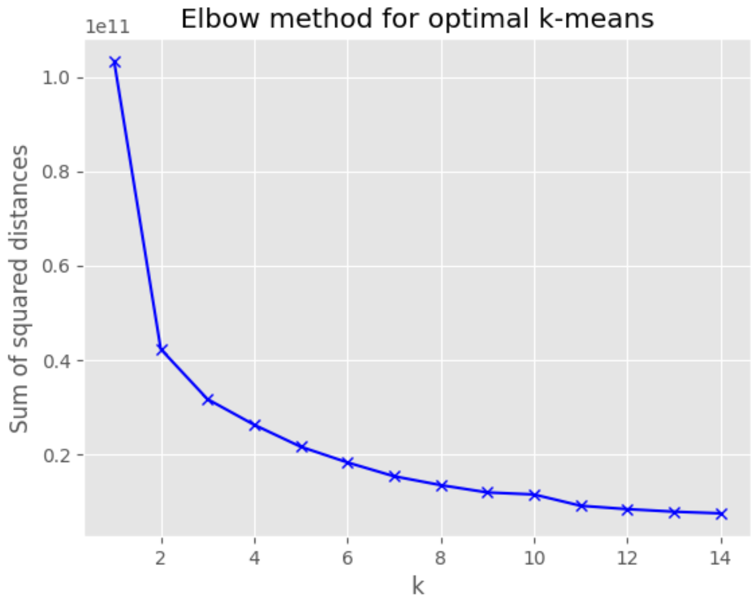


Figure 3.18: Sum of Square Error for different k values

Here, k ranges from 1 to 14, and the squared error sum is plotted on the y axis. We can see from the plot that the error will reduce as the number of clusters rises. It makes sense that the sum of square error will eventually approach zero if we can think of each of our data points as belonging to different clusters. We need to identify the elbow on the graph, which is denoted by the point after which the sum of square error changes the slowly.
commonly used metrics for sum of square include:
Within-cluster sum of squares (WCSS):

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \tag{3.11}$$

Where C is the set of all clusters, μ_i is the mean of points in cluster i, and k is the number of clusters.

3.3.2 K-Medoids Algorithm

K-medoids clustering is a variant of the k-means algorithm. K-medoids is also a type of partitioning clustering algorithm that, on its own, mostly splits the dataset into k groups without the need for any training. K-medoids typically perform better when there are outliers as k-means is sensitive to outliers. In k-means the centroid is taken by the mean value of the object in a cluster but in k-medoids the centroid is taken from the object which is the most centrally located. Medoid is a point in the cluster which has a minimum sum of distances to other data points in the cluster.

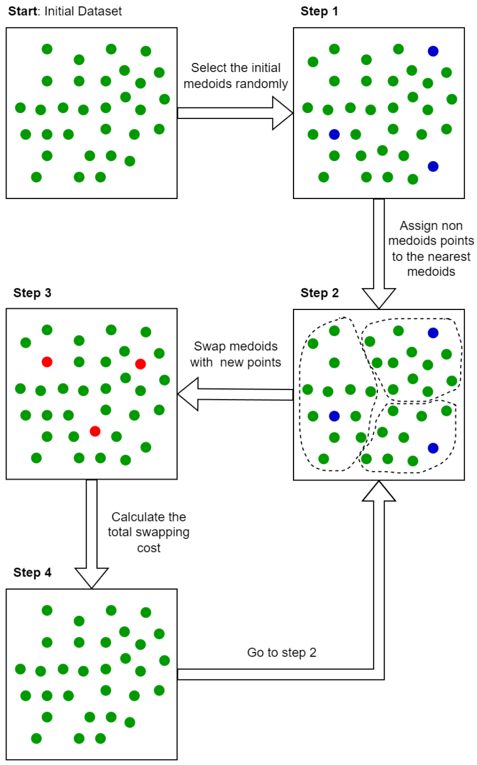


Figure 3.19: K-Medoids algorithm process

The initial step of the k-medoids method is to choose k random medoids from the data points. Here k, represents the number of clusters. For all the other data points, calculate the dissimilarity between the points and the medoids. There are various dissimilarity methods to calculate the dissimilarity between the non-medoid points and medoid points. Some of them are euclidean distance, manhattan distance, cosine distance etc. By using any of the common distance metrics the dissimilarity can be calculated. In the next step, each point is assigned to that cluster whose dissimilarity is less than the other one. After that, the cost is calculated. Cost is calculated by adding the minimum dissimilarity of the non-medoids points. After that, one of the medoids gets swapped with any of the non-medoids points. Later on, the procedure of finding dissimilarities and assigning the data points to the medoids cluster gets repeated. Then the new cost is calculated. After that, the swap cost is calculated by subtracting the present cost from the previous cost. And if the swap cost is not less than zero then the swap got undone and the previous medoids are taken. The process gets repeated until no change is encountered with new medoids to classify data points. These steps are shown in the Figure -3.19
The cost in the k-Medoids algorithm is given as

$$\mathbf{c} = \sum_{\mathbf{Ci}} \sum_{\mathbf{Pi}\mathcal{E}\mathbf{Ci}} |\mathbf{Pi} - \mathbf{Ci}| \tag{3.12}$$

The dissimilarity of the medoid(Ci) and object (Pi)is calculated by using $E = |Pi - Ci|$

Chapter 4

Result and Discussion

4.1 Implementation

4.1.1 K-Means And K-Medoids Clustering Results

We applied Silhouette Analysis to identify outliers in the data. Low silhouette scores for each data point can be used to identify probable outliers, while Silhouette Analysis is best known for assessing the quality of the clusters. In this procedure, we used clustering with the k-means algorithm on the dataset. Then silhouette scores for each data point were calculated. The score range from -1 to 1, data points with a score less than 0 indicate overlapping points. We set a threshold of 0.2, where data points less than 0.2 are overlapping and potential outliers. After that, those points with lower scores were omitted in order to reduce the noise. The data we are using in this are not labelled data. Hence, we need to apply unsupervised learning on the data like clustering. Later on, for clustering it is important to find out the optimal number for clusters. The Elbow Method is a widely used technique for determining the optimum K value. In Elbow Method, the sum of squared distance is calculated for the k range from 1 to 15. Then the values of the sum of squared distance is being plotted against each k value. The value of the sum of squared distance is highest when the value of k is 1 and the value gradually decreased with k values. In the plot, there is a drastic downward change in the sum of squared distance and that point of change indicates the elbow and that point is considered the optimum number of clusters. The elbow method was calculated for the k-medoids algorithm and the k-means algorithm. Figure -4.1 is showing the elbow method for the k-medoids and Figure -4.2 is showing the elbow method for the k-means algorithm.

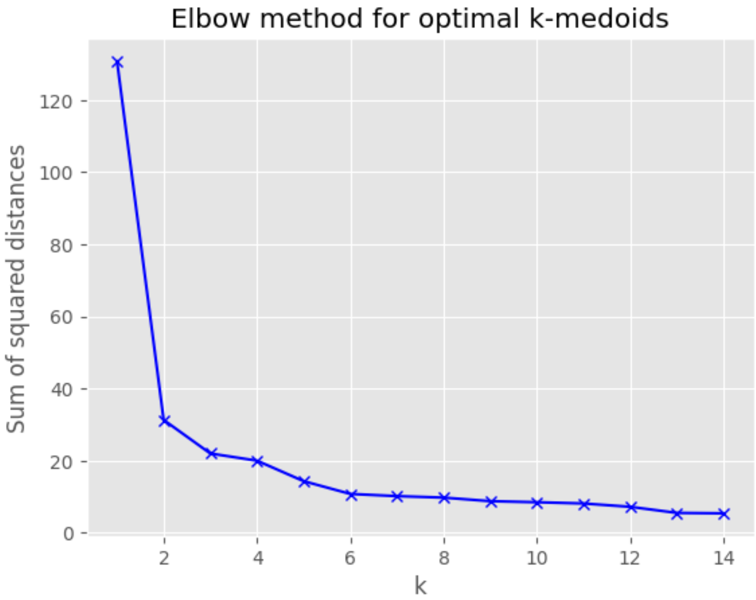


Figure 4.1: Sum of Square Error for different k values for k-Medoids

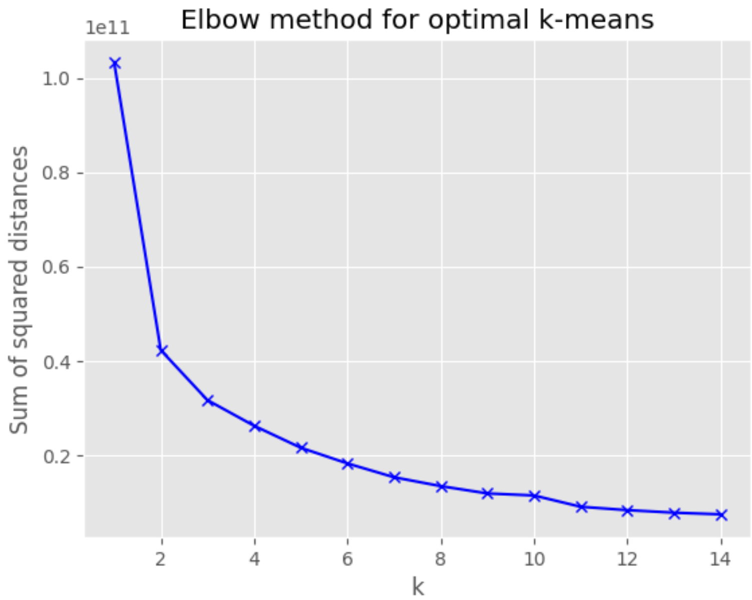


Figure 4.2: Sum of Square Error for different k values for k-Means

In both the elbow methods plots it is showing that the elbow is on point 2, which indicates the optimum number of clusters is 2. This means we can get two optimum clusters from the data points in the dataset. Along with the elbow method we implemented the Silhouette method to find the optimum cluster number.

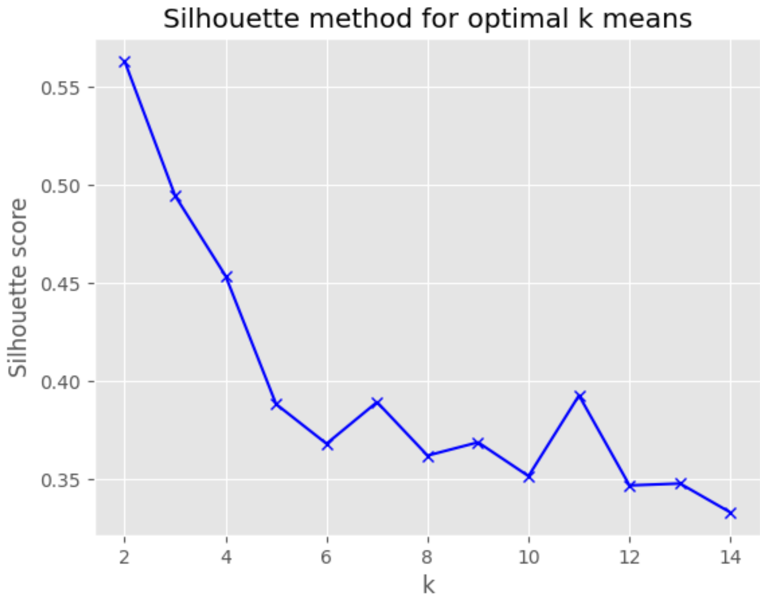


Figure 4.3: Silhouette score for each k numbers of clusters

In Figure -4.3 we can see, it is giving the maximum Silhouette score for k=2. That means the optimum number of clusters should be 2 according to the elbow method and Silhouette method. For clustering the k-medoids and k-means algorithms were applied. According to elbow method the cluster number we found were 2. Hence, we initialize the k-medoids algorithm with cluster number to 2 and the distance method was 'Manhattan' distance method. Then the dimension reduced version of BnA data were fitted to the k-medoids algorithm. After that we got two clusters from the algorithm. Later on, we plotted principal components 1 vs 2 for both clusters to get a visual representation of the clusters. The plot is shown in Figure -4.4

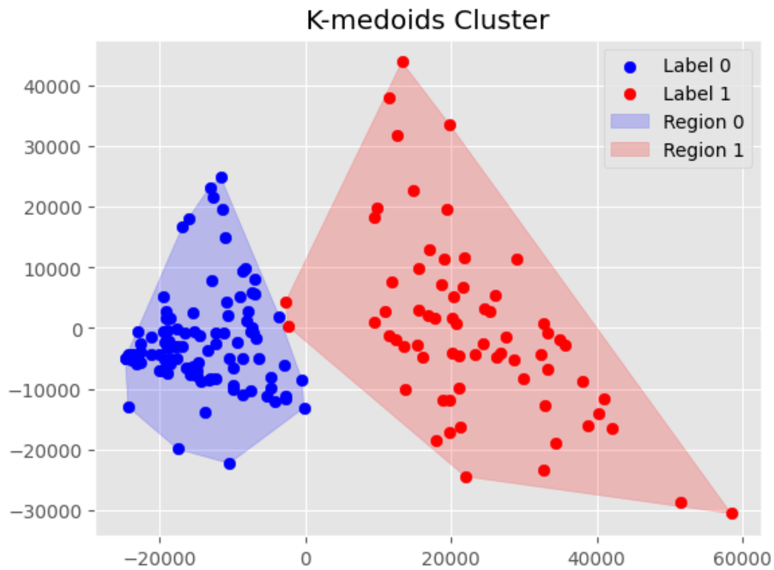


Figure 4.4: Cluster using K-medoids

For K-means cluster, we trained the k-means model with the cluster number k=2. Then we fitted the dimension reduced version of BnA data to the k-means model.

After the clustering, the data got into two clusters according to labels 0 and 1. Later on, we plotted principal components 1 vs 2 from PCA data to get a visual representation of the clusters. The clusters are shown in Figure -4.5.

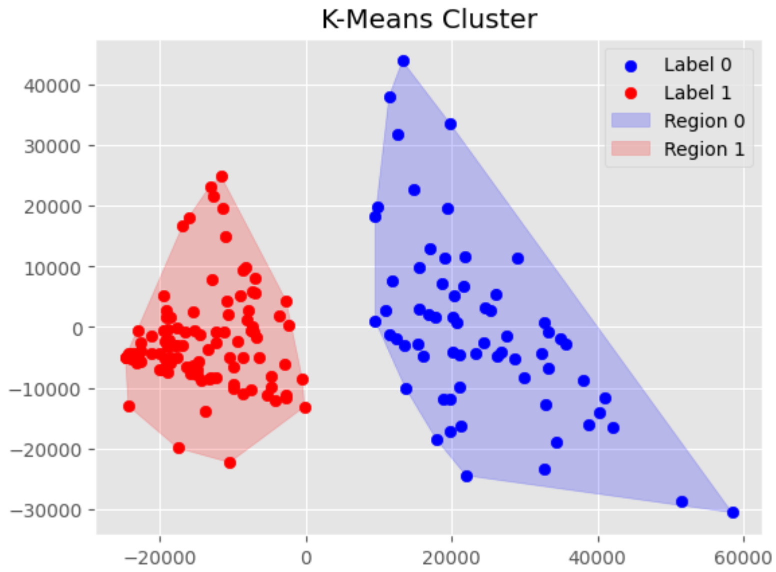


Figure 4.5: Cluster using K-Means

4.2 Performance Evaluation

After we got the clusters from the K-Means and K-Medoids We applied majority Voting to Determine if an ID falls under addicted or Non Addicted category. For example , if one ID has 5 Data points in the Addicted Cluster and 2 in the Non Addicted Cluster, we classify that person as addicted.
Finally after this step we divided our participants into addicted and non addicted from the cluster we Used extrinsic measure to validate smartphone addicted by using the classification from SAS-SV.

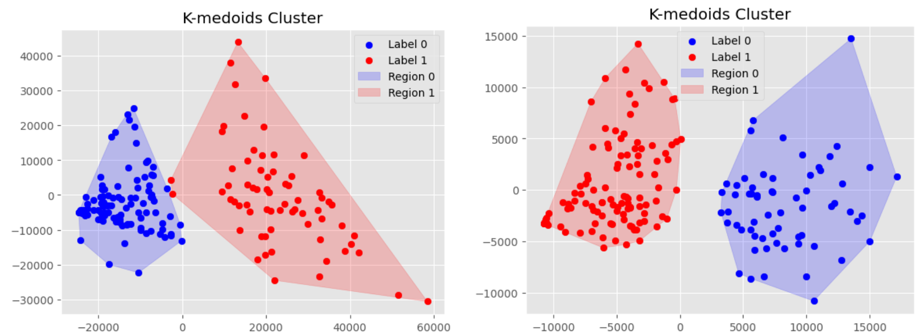
4.2.1 Performance Metrics

The labels in supervised learning are known, and the degree of accuracy may be determined by comparing the predicted values to the labels. However, since there is no ground truth in unsupervised learning, it is challenging to assess the degree of accuracy. A prominent method for unsupervised data analysis is clustering. Cluster validity metrics are employed in data clustering analysis to rate the caliber or efficacy of clustering solutions. The effectiveness of the clusters' capture of the underlying structure or patterns in the data is evaluated quantitatively by these metrics.

A good clustering method seeks to maximize the between-cluster variance, which indicates that clusters are distinct to one another, while minimizing the within-cluster variance, which indicates that data points within a cluster are similar to one another. There are two main categories of assessment measures for clustering:

(a) Cluster With Human Activity Data (b) Cluster Without Human Activity Data

Figure 4.6: K-Means Cluster Comparison



(a) Cluster With Human Activity Data (b) Cluster Without Human Activity Data

Figure 4.7: K-Mediods Clusters Comparison

4.2.1.1 Extrinsic Measures

The data points for these assessment metrics must have known class assignments or ground truth labels. These metrics evaluate how well the clusters match the reality by comparing the clustering outcomes to the actual labels. The Rand Index and the Fowlkes-Mallows Index are a few examples of extrinsic metrics.
By contrasting the results of the clustering algorithms with the real labels or ground truth, the Rand Index and the Fowlkes-Mallows Index are two common extrinsic metrics used to assess the effectiveness of clustering algorithms. These measures gauge how well the algorithm's clusters and actual class assignments match up.

Rand Index (RI): The Rand Index computes the percentage of agreements between the clustering results and the true class assignments. It considers all pairwise data points and checks whether they are in the same cluster (both in the clustering result and the true class assignment) or in different clusters (either in the clustering result or the true class assignment). The formula for Rand Index is as follows:

$$RI = (a + b)/(a + b + c + d) \tag{4.1}$$

where: - a represents the number of pairs that are in the same cluster in both the clustering result and the true class assignment. - b represents the number of

pairs that are in different clusters in both the clustering result and the true class assignment. - c represents the number of pairs that are in the same cluster in the clustering result but in different clusters in the true class assignment. - d represents the number of pairs that are in different clusters in the clustering result but in the same cluster in the true class assignment.
The Rand Index ranges between 0 and 1, where 1 indicates a perfect match between the clustering result and the true class assignment.

Fowlkes-Mallows Index (FMI): Another extrinsic measure that determines the geometric mean of recall and accuracy is the Fowlkes-Mallows Index. When comparing all the pairings allocated to the same cluster in the clustering result, precision is the percentage of true positive pairs (pairs that are in the same cluster in both the clustering result and the real class assignment). Recall measures the proportion of true positive pairs among all the pairs assigned to the same cluster in the true class assignment. The Fowlkes-Mallows Index is given by the following formula:

$$\text{FMI} = \sqrt{(Precision * Recall)} \tag{4.2}$$

where:

$$Precision = a/(a+c)$$

$$Recall = a/(a+b)$$

The Fowlkes-Mallows Index ranges between 0 and 1, with 1 indicating a perfect match between the clustering result and the true class assignment.
Both the Rand Index and the Fowlkes-Mallows Index provide a quantitative measure of the agreement between the clustering outcomes and the actual labels. Higher values indicate better clustering performance, while lower values suggest poorer cluster quality and less agreement with the ground truth.

For our Clusters the Extrinsic Measures are (Table 4.1),

	With Activity Data		**Without Activity Data**	
	K- Means	**K - Medoids**	**K- Means**	**K - Medoids**
Rand Index	0.86243	0.86243	0.52380	0.52380
Fowlkes-Mallows Index	0.87005	0.87005	0.54818	0.54818

Table 4.1: Extrinsic Measures

From the table we can see that when activity data is included, both the K-Means and K-Medoids algorithms achieve high values for both the Rand Index and Fowlkes-Mallows Index. For the Rand Index, both algorithms obtain a score of 0.86243, indicating a high level of agreement between the clustering results and the ground truth labels. Similarly, the Fowlkes-Mallows Index yields a score of 0.87005 for both algorithms, further supporting the consistency between the clustering outcomes and the known class labels.

In contrast, when activity data is not considered, the performance of both algorithms decreases in terms of the Rand Index and Fowlkes-Mallows Index. The Rand Index drops to 0.52380 for K-Means and K-Medoids, indicating a significantly lower level of agreement between the clustering results and the ground truth. Similarly, the Fowlkes-Mallows Index decreases to 0.54818 for both algorithms, reflecting a decreased similarity between the clusters and the known class labels.
From these results, it can be concluded that incorporating activity data significantly improves the clustering performance based on the external criteria. The high values of the Rand Index and Fowlkes-Mallows Index with activity data suggest that the clustering outcomes align well with the known class labels, indicating successful clustering of the data points. On the other hand, excluding activity data leads to poorer agreement and similarity between the clusters and the ground truth labels.

4.2.1.2 Intrinsic Measures

These assessment metrics can be used to evaluate clustering outcomes in unsupervised contexts and do not rely on ground truth labels. These evaluations often focus on internal clustering properties like compactness and cluster separation. The Silhouette Coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index are a few examples of intrinsic measurements.
When ground truth labels are not provided or when it is desired to examine and compare various clustering techniques or parameter settings without taking into account predetermined class memberships, intrinsic measures are more frequently utilized.

Silhouette Coefficient: Evaluating the quality of clustering results is crucial in order to assess the effectiveness and validity of clustering algorithms. One commonly used measure is the Silhouette Coefficient, which provides a quantitative assessment of how well individual data points fit within their assigned clusters. The Silhouette Coefficient takes into account both the compactness of a data point within its own cluster and the separation between the data point and other neighboring clusters. A higher Silhouette Coefficient indicates better clustering results.
The Silhouette Coefficient for a data point i is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4.3}$$

where $a(i)$ is the average dissimilarity of data point i to all other data points within the same cluster, and $b(i)$ is the minimum average dissimilarity of data point i to any other cluster, i.e., the dissimilarity between i and the nearest neighboring cluster.
The Silhouette Coefficient ranges from -1 to 1. A value close to 1 indicates that the data point is well-matched to its own cluster and poorly-matched to neighboring clusters, suggesting a good clustering result.The contrary, when the data point is poorly matched to its own cluster and well matched to nearby clusters, is indicated by a number that is close to -1, which raises the possibility of misclassification. A value near 0 suggests that the data point is on or very close to the decision boundary between two clusters.
To obtain an overall assessment of the clustering quality, the Silhouette Coefficient can be computed for all data points and averaged to calculate the average Silhouette

Coefficient for the entire dataset. A higher average Silhouette Coefficient indicates better overall clustering performance.
The Silhouette Coefficient provides a reliable and intuitive measure for evaluating the quality of clustering results, allowing researchers and practitioners to compare different clustering algorithms and parameter settings to identify the most suitable clustering solution for their specific datasets.

The Calinski-Harabasz Index: The Calinski-Harabasz index is a quantitative indicator of how successfully a clustering solution divides data points into different clusters and is used as a measure of cluster validity or quality in data clustering research. Better overall separation and more clearly defined clusters are indicated by higher index values. The Calinski-Harabasz Index, also known as the Variance Ratio Criterion, compares the total of dispersion inside and across clusters, where dispersion is defined as the sum of the square of the distance.A greater ratio denotes a more distinct cluster that is located far from its neighboring clusters. So, the cluster is well-defined.
The between-cluster dispersion, denoted as B, is calculated as:

$$B = \sum_{q \in k} n_q (c_q - c_E)(c_q - c_E)^T \tag{4.4}$$

The within-cluster dispersion, denoted as W, is calculated as:

$$W = \sum_{q \in k} \sum_{x \in \text{cluster } q} n_q (x - c_E)(x - c_E)^T \tag{4.5}$$

Finally, the Calinski-Harabasz score (s) is calculated as:

$$s = \frac{B}{W} \times \frac{n_E - k}{k - 1} \tag{4.6}$$

Where,
k: Number of clusters
n_q: Number of clusters in q
c_q: Cluster center of cluster q
n_E: Number of data points
c_E: Cluster center of all points

This score is used as a measure of the separation between clusters in a clustering algorithm.
Choosing the right number of clusters for a dataset is one of the difficulties in clustering. This procedure can be aided by the Calinski-Harabasz index, which compares various clustering approaches with various numbers of clusters. For solutions with an appropriate number of clusters, the index tends to be higher, suggesting a better fit to the data. Comparing the Calinski-Harabasz index to some other cluster validity measures, it is computationally effective and quite simple to calculate. This makes it feasible for large-scale dataset analysis and iterative assessments of clustering approaches.
However, It's crucial to remember that the Calinski-Harabasz index has its restrictions. It makes the supposition that clusters are small and clearly divided, which

may not always be the case. It would be inappropriate to compare Calinski-Harabasz Index to other types of clustering algorithms because it is often higher for density-based clustering techniques.

Davies-Bouldin Index: When assessing the caliber of clustering findings, the Davies-Bouldin Index (DBI) is a measure of cluster validity. It evaluates both the compactness inside clusters and the distance between clusters. This index represents the average ”similarity”—a metric that contrasts the size of the clusters with the distance between them—between clusters. The lowest possible score is zero. Values that are nearer to 0 denote better partitions. A lower score indicates that the cluster is well-defined and relatively close to another cluster in terms of size.

The difference measure, denoted as R_{ij}, is calculated as:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{4.7}$$

Finally, the Davies-Bouldin Index (DB) is calculated as:

$$\text{DB} = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} R_{ij} \tag{4.8}$$

Where,
k: Number of clusters
s_i: Average distance between each point of cluster i to cluster center c_i
d_{ij}: Distance between cluster centers c_i and c_j

The Davies-Bouldin Index is used as a measure of the clustering quality. Lower values indicate better-defined clusters with greater separation.
Compared to Silhouette scores, Davies-Bouldin calculation is less complicated. As its computation exclusively makes use of point-wise distances, the index is purely reliant on elements and characteristics that are present in the dataset.It’s crucial to remember that the DBI has some restrictions. It is predicted that clusters are spherically shaped, comparable in size, and that the distances between their centroids may be used to calculate the distances between clusters. These presumptions might not always be true, thus it’s best to take the DBI into account along with other cluster validity metrics to provide a more thorough assessment of clustering quality.
For our Clusters the Intrinsic Measures are (Table 4.2),

	With Activity Data		**Without Activity Data**	
	K- Means	**K - Medoids**	**K- Means**	**K - Medoids**
Silhouette Coefficient	0.56	0.55	0.51	0.51
Calinski-Harabasz Index	245.44	237.40	217.23	217.23
Davies-Bouldin Index	0.71	0.72	0.78	0.78

Table 4.2: Intrinsic Measures

As the Intrinsic Measures classify how well the data points have clustered we can see that here K- Means Algorithm performs slightly better than K - Medoids. Silhouette Coefficient, which measures the compactness and separation of clusters, is higher for both K-Means (0.56) and K-Medoids (0.55) when activity data is included compared to when it is not included (K-Means: 0.51, K-Medoids: 0.51). This suggests that incorporating activity data improves the clustering performance for both algorithms. The Calinski-Harabasz Index, which evaluates the ratio of between-cluster dispersion to within-cluster dispersion, also indicates better clustering results when activity data is used. The index is higher for both K-Means (245.44) and K-Medoids (237.40) with activity data compared to without activity data (K-Means: 217.23, K-Medoids: 217.23).
On the other hand, the Davies-Bouldin Index, which measures the average similarity between clusters, shows slightly higher values for both K-Means (0.71) and K-Medoids (0.72) when activity data is included compared to when it is not included (K-Means: 0.78, K-Medoids: 0.78). Although the values are slightly higher, they are still relatively low, indicating reasonable cluster quality.

4.3 Discussion

Based on the discussion and analysis of both the intrinsic and extrinsic measures as shown in Tables 4.2 and 4.1, it can be concluded that the inclusion of activity data significantly improves the clustering performance of the K-Means and K-Medoids algorithms. The intrinsic measures, which evaluate the separation and dispersion of clusters, indicate that incorporating activity data leads to better cluster separation and reduced overlap between clusters. This suggests that activity data provides valuable information for distinguishing different patterns of smartphone usage and identifying distinct groups of users with specific behavioral characteristics related to smartphone addiction.

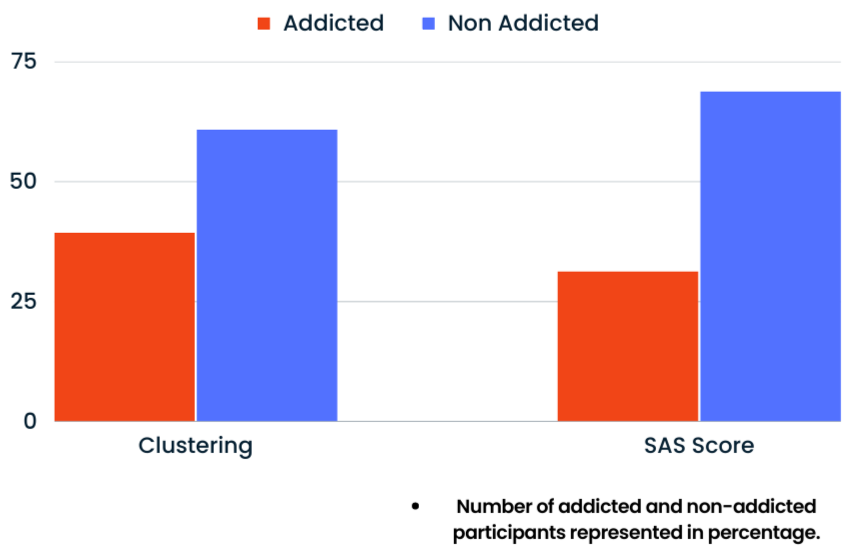


Figure 4.8: Percentage of Addicted And Non-Addicted Participants In Proposed Method And SAS-Score

In the figure - 4.8, it was found 39.28 percent of the participants were addicted and 60.72% participants were non addicted in the proposed model, in contrast there were 31.25% participants were addicted and 68.75% participants were found non addicted. Moreover, the extrinsic measures, which assess the agreement between the clusters and known class labels, reveal that including activity data enhances the agreement and similarity between the derived clusters and the actual addiction status of the participants. This implies that the activity data contributes to more accurate clustering results that align with the participants' addiction levels. Considering these findings, it can be concluded that incorporating activity data into the clustering process is beneficial for achieving higher accuracy and more meaningful results. In this study, the final accuracy achieved was **87%**, demonstrating the effectiveness of the approach in detecting smartphone addiction.

Chapter 5

Conclusion

In conclusion, this research presents an alternative approach for detecting smartphone addiction. Smartphones can be incredibly useful tools, compulsive usage of them can harm one's relationships, career, or academic performance. There are countless negative effects of excessive smartphone use. The effects of smartphones on everything from physical health to mental health are significant. The use of this device compulsively has an impact on people. In addition to utilizing survey questionnaires as a means of assessment, it emphasizes the importance of gathering sufficient empirical data as in App usage and Smartphone Sensor data to comprehend the behavior's development, manifestations, and associated dysfunctions. By using this model of detecting smartphone addiction we are able to correctly classify a person with **87%** accuracy. We ran into restrictions like the barrier brought up by the variety in smartphone devices. As many iOS users as there, we had to exclude them to do this study due to the limitations of gathering their sensor data. In our research we had a particular age group that is mostly teenagers and data collection time was 7 days. It is recommended that further studies be conducted, preferably with larger demographics and with a more robust method to collect sensor data and app uses data, to better conceptualize problematic smartphone use or addiction. Currently, there is still much to be learned about the point at which smartphone use becomes problematic and for whom it becomes an issue.

Bibliography

[1] F. M. Sirois, M. L. Melia-Gordon, and T. A. Pychyl, ""i'll look after my health, later": An investigation of procrastination and health," en, *Pers. Individ. Dif.*, vol. 35, no. 5, pp. 1167–1184, Oct. 2003.

[2] F. M. Sirois, M. L. Melia-Gordon, and T. A. Pychyl, ""i'll look after my health, later": An investigation of procrastination and health," en, *Pers. Individ. Dif.*, vol. 35, no. 5, pp. 1167–1184, Oct. 2003.

[3] D. J. Stein, E. Hollander, and B. O. Rothbaum, *Textbook of anxiety disorders*, 2nd ed. American Psychiatric Publishing, Dec. 2010.

[4] M. Kwon, D.-J. Kim, H. Cho, and S. Yang, "The smartphone addiction scale: Development and validation of a short version for adolescents," en, *PLoS One*, vol. 8, no. 12, e83558, Dec. 2013.

[5] M. Kwon, J.-Y. Lee, W.-Y. Won, *et al.*, "Development and validation of a smartphone addiction scale (SAS)," *PLoS ONE*, vol. 8, no. 2, J. Laks, Ed., e56936, Feb. 2013. DOI: 10.1371/journal.pone.0056936. [Online]. Available: https://doi.org/10.1371/journal.pone.0056936.

[6] L. Lapointe, C. Boudreau-Pinsonneault, and I. Vaghefi, "Is smartphone usage truly smart? a qualitative investigation of IT addictive behaviors," in *2013 46th Hawaii International Conference on System Sciences*, Wailea, HI, USA: IEEE, Jan. 2013.

[7] BBC News, "Smartphone overuse may 'damage' eyes, say opticians," en, *BBC*, Mar. 2014.

[8] F. M. Sirois, "Procrastination and stress: Exploring the role of self-compassion," *Self Identity*, vol. 13, no. 2, pp. 128–145, Mar. 2014.

[9] S. Arora and D. Okunbor, "Social media applications: Are the youth addicted?" In *2015 International Conference on Cyberspace (CYBER-Abuja)*, Abuja, Nigeria: IEEE, Nov. 2015.

[10] J. Billieux, P. Maurage, O. Lopez-Fernandez, D. J. Kuss, and M. D. Griffiths, "Can disordered mobile phone use be considered a behavioral addiction? an update on current evidence and a comprehensive model for future research," en, *Curr. Addict. Rep.*, vol. 2, no. 2, pp. 156–162, Jun. 2015.

[11] W. Jun, "An analysis study on correlation of internet addiction and smartphone addiction of teenagers," in *2015 2nd International Conference on Information Science and Security (ICISS)*, Seoul, Korea (South): IEEE, Dec. 2015.

[12] S.-E. Kim, J.-W. Kim, and Y.-S. Jee, "Relationship between smartphone addiction and physical activity in chinese international students in korea," en, *J. Behav. Addict.*, vol. 4, no. 3, pp. 200–205, Sep. 2015.

[13] University of Waterloo, "Reliance on smartphones linked to lazy thinking," en, *Science Daily*, Mar. 2015.

[14] C. Chen, K. Z. K. Zhang, S. J. Zhao, M. K. O. Lee, and T. Cong, "The impact of mere exposure effect on smartphone addiction," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, USA: IEEE, Jan. 2016.

[15] C. Wang, M. K. O. Lee, C. Yang, and X. Li, "Understanding problematic smartphone use and its characteristics: A perspective on behavioral addiction," in *Lecture Notes in Information Systems and Organisation*, ser. Lecture notes in information systems and organisation, Cham: Springer International Publishing, 2016, pp. 215–225.

[16] R. Zhou and D. Shi, "Smartphone usage pattern based on geographical location and addiction issue," in *2016 5th International Conference on Computer Science and Network Technology (ICCSNT)*, Changchun, China: IEEE, Dec. 2016.

[17] Y. Sanal and Ö. Ozer, "Smartphone addiction and the use of social media among university students," en, *Mediterr. J. Humanit.*, vol. 7, no. 2, pp. 367–377, Dec. 2017.

[18] A. Shenhav, S. Musslick, F. Lieder, *et al.*, "Toward a rational and mechanistic account of mental effort," en, *Annu. Rev. Neurosci.*, vol. 40, no. 1, pp. 99–124, Jul. 2017.

[19] H. H. Wilmer, L. E. Sherman, and J. M. Chein, "Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning," en, *Front. Psychol.*, vol. 8, p. 605, Apr. 2017.

[20] A. Turner, *How many people have smartphones worldwide (sept 2022)*, en, https://www.bankmycell.com/blog/how-many-phones-are-in-the-world, Accessed: 2022-9-17, Jul. 2018.

[21] A. Alruban, H. Alobaidi, N. Clarke, and F. Li, "Physical activity recognition by utilising smartphone sensor signals," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS - Science and Technology Publications, 2019. DOI: 10.5220/0007271903420351. [Online]. Available: https://doi.org/10.5220%5C%2F0007271903420351.

[22] R. Latifa, E. F. Mumtaz, and I. Subchi, "Psychological explanation of phubbing behavior: Smartphone addiction, emphaty and self control," in *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, vol. 7, Jakarta, Indonesia: IEEE, Nov. 2019, pp. 1–5.

[23] S. W. Pienaar and R. Malekian, *Human activity recognition using lstm-rnn deep neural network architecture*, 2019. arXiv: 1905.00599 [cs.LG].

[24] L. Podo, B. Taccardi, A. Colonna, and L. Patrono, "A smart approach able to face distraction issue due to smartphone usage running social networks," in *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, Split, Croatia: IEEE, Jun. 2019.

[25] M. Tateno, A. R. Teo, W. Ukai, *et al.*, "Internet addiction, smartphone addiction, and hikikomori trait in japanese young adult: Social isolation and social network," *Front. Psychiatry*, vol. 10, Jul. 2019.

[26] B. Harris, T. Regan, J. Schueler, and S. Fields, "Problematic mobile phone and smartphone use scales: A systematic review," *Frontiers in Psychology*, vol. 11, May 2020. DOI: 10.3389/fpsyg.2020.00672.

[27] J. Ramos-Diaz, Q. Jiang, R. Ramos-Sandoval, and M. D. Griffiths, "Problematic mobile phone use and risky behaviors among college students in peru," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia: IEEE, Nov. 2020.

[28] J. Ramos-Diaz, Q. Jiang, R. Ramos-Sandoval, and M. D. Griffiths, "Problematic mobile phone use and risky behaviors among college students in peru," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia: IEEE, Nov. 2020.

[29] S. Subramaniam, J. S. Dhillon, A. C. Kah Hoe, M. Shanmugam, and S. S. Gunasekaran, "Evaluating smartphone addiction disorder among university students," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, Selangor, Malaysia: IEEE, Aug. 2020.

[30] S. Subramaniam, J. S. Dhillon, A. C. Kah Hoe, M. Shanmugam, and S. S. Gunasekaran, "Evaluating smartphone addiction disorder among university students," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, Selangor, Malaysia: IEEE, Aug. 2020.

[31] X. Yang, P. Wang, and P. Hu, "Trait procrastination and mobile phone addiction among chinese college students: A moderated mediation model of stress and gender," *Front. Psychol.*, vol. 11, Dec. 2020.

[32] M. X. Zhang and A. M. S. Wu, "Effects of smartphone addiction on sleep quality among chinese university students: The mediating role of self-regulation and bedtime procrastination," en, *Addict. Behav.*, vol. 111, no. 106552, p. 106 552, Dec. 2020.

[33] R. Behzad, "The relationship between smartphone addiction and procrastination," en, *Annals of RSCB*, pp. 8316–8323, Apr. 2021.

[34] Y. Y. Al-Kandari and M. M. Al-Sejari, "Social isolation, social support and their relationship with smartphone addiction," en, *Inf. Commun. Soc.*, vol. 24, no. 13, pp. 1925–1943, Oct. 2021.

[35] B. Khan, U. I. Janjua, and T. M. Madni, "The identification of influential factors to evaluate the kids smartphone addiction: A literature review," in *2021 4th International Conference on Computing & Information Sciences (ICCIS)*, Karachi, Pakistan: IEEE, Nov. 2021.

[36] J. H. Park and M. Park, "Smartphone use patterns and problematic smartphone use among preschool children," en, *PLoS One*, vol. 16, no. 3, e0244276, Mar. 2021.

[37] J. Tian, J.-Y. Zhao, J.-M. Xu, *et al.*, "Mobile phone addiction and academic procrastination negatively impact academic achievement among chinese medical students," *Front. Psychol.*, vol. 12, Nov. 2021.

[38] H. Yang, B. Liu, and J. Fang, “Stress and problematic smartphone use severity: Smartphone use frequency and fear of missing out as mediators,” *Front. Psychiatry*, vol. 12, Jun. 2021.

[39] K. Yasudomi, T. Hamamura, M. Honjo, A. Yoneyama, and M. Uchida, “Usage prediction and effectiveness verification of app restriction function for smartphone addiction,” in *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, Shenzhen, China: IEEE, Mar. 2021.

[40] J. Aru and D. Rozgonjuk, “The effect of smartphone use on mental effort, learning, and creativity,” en, *Trends Cogn. Sci.*, vol. 26, no. 10, pp. 821–823, Oct. 2022.

[41] D. Chaffey, *Global social media statistics research summary 2022*, en, https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/, Accessed: 2022-9-17, Aug. 2022.

[42] T. Li, Y. Li, M. A. Hoque, T. Xia, S. Tarkoma, and P. Hui, “To what extent we repeat ourselves? discovering daily activity patterns across mobile app usage,” *IEEE Trans. Mob. Comput.*, vol. 21, no. 4, pp. 1492–1507, Apr. 2022.

[43] C. Odhiambo, S. Saha, C. Martin, and H. Valafar, *Human activity recognition on time series accelerometer sensor data using lstm recurrent neural networks*, Jun. 2022.

[44] *Can Disordered Mobile Phone Use Be Considered a Behavioral Addiction? An Update on Current Evidence and a Comprehensive Model for Future Research - Current Addiction Reports — link.springer.com*, https://link.springer.com/article/10.1007/s40429-015-0054-y, [Accessed 19-May-2023].

[45] *Identifying Indicators of Smartphone Addiction Through User-App Interaction — sciencedirect.com*, https://www.sciencedirect.com/science/article/pii/S0747563219301712, [Accessed 19-May-2023].

[46] *Smartphone ownership in the U.S. 2015-2021, by age group*, en, https://www.statista.com/statistics/489255/percentage-of-us-smartphone-owners-by-age-group/, Accessed: 2022-9-17.