Socialoscope

Sensing User Loneliness and Its Interactions with Personality Traits

By

Gauri Pulekar

A Thesis

Submitted to the Faculty

Of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

In

Computer Science

April 2016

APPROVED:

Dr. Emmanuel O. Agu, Thesis Advisor

Dr. Charles Rich, Thesis Reader

Dr. Craig E. Wills, Head of Department

Abstract

Loneliness and social isolation can have a serious impact on one's mental health, leading to increased stress, lower self-esteem, panic attacks, and drug or alcohol addictions. Older adults and international students are disproportionately affected by loneliness. This thesis investigates Socialoscope, a smartphone app that passively detects loneliness in smartphone users based on the user's day-to-day social interactions, communication and smartphone activity sensed by the smartphone's built-in sensors. Statistical analysis is used to determine smartphone features most correlated with loneliness. A previously established relationship between loneliness and personality type is explored. The most correlated features are used to synthesize machine learning classifiers that infer loneliness levels from smartphone sensor features with an accuracy of 90%. These classifiers can be used to make the Socialoscope an intelligent loneliness sensing Android app. The results show that, of the five Big-Five Personality Traits, emotional stability and extraversion personality traits are strongly correlated with the sensor features such as number of messages, number of outgoing calls, number of late night browser searches, number of long incoming or outgoing calls and number of auto-joined trusted Wi-Fi SSIDs. Moreover, the classifier accuracy while classifying loneliness levels is significantly improved to 98% by taking these personality traits into consideration. Socialoscope can be integrated into the healthcare system as an early warning indicator of patients requiring intervention or utilized for personal self-reflection.

Acknowledgements

I would first like to express my gratitude to my advisor, Prof. Emmanuel O Agu, for his continuous guidance and mentoring that helped me shape and steer my research, learning and understanding of mobile sensing, user studies, machine learning, loneliness and personality detection in the right direction. In this entire process, he introduced to me and patiently taught me the process of conducting research, to which I was completely new. For this, I will be forever grateful. I would like to thank Prof. Charles Rich for being the reader for this thesis. He provided me with his invaluable feedback, suggestions and a much-needed alternate perspective.

I would like to express my appreciation to Prof. Kymberlee O’Brien, Prof. Kent J. Rissmiller, and Ruth McKeogh for their support and guidance with the Institutional Review Board approval and Participant Pool user study. I would like to express my gratitude to Worcester Polytechnic Institute and the Computer Science department for providing me with this golden opportunity of pursuing my thesis. I would like to thank all the members of the Performance Evaluation of Distrusted Systems (PEDS) research group at WPI for their feedback and suggestions.

I would like to thank my friends for their constant encouragement and support, for helping me with my user study and for letting me test my code on their phones. Last but not the least, I would like to thank my family members for their perpetual support and motivation throughout my years of study. For this, I will be forever grateful.

I sincerely thank you all.

Table of Contents

List of Tables

List of Figures

1. Introduction

“*The most terrible poverty is loneliness, and the feeling of being unloved*”, is an apt saying by Mother Teresa [36]. Man, is a social animal, who needs rewarding social contact and relationships to make him feel comfortable. When it comes to our wellbeing, people around us, people we feel connected to and people we communicate with, matter.

Evidence shows that healthy relationships with family members, extended family, friends, colleagues and community members are important for our social wellbeing [47]. Building stronger, healthier and rewarding social connections helps us build a sense of belonging, gives us a greater sense of purpose and self-worth. These relationships allow us to share our feelings and experiences, and make us feel that we are understood. They give us emotional strength and support, and a chance to support others, thereby keeping us happier and promoting mental wellbeing. Social wellbeing thus means feeling good about ourselves and functioning well within our social world [47]. When this need is not met, we feel isolated, leading to thoughts of not fitting in, not being understood, feeling empty and isolated [33]. This poem by the 17th Century English poet, John Donne, best summarizes this social need:

“*No man is an island,*
Entire of itself,
Every man is a piece of the continent,
A part of the main.”

Loneliness, is not the same as being alone. Aloneness is finding freedom in the isolation. It's a state of serenity you feel in your own company. It is an awareness that you are complete and very connected to life, needing nothing and no one [49]. It is a pleasant and blissful feeling. You can be alone and very happy at the same time. Being alone, thus can be experienced as a positive emotion. On the other hand, you can be surrounded by people and still feel lonely. Loneliness is an assertion that you are somehow disconnected or incomplete, that you need someone or perhaps something to feel complete [49]. Loneliness leaves you craving for someone or something to pass your time and to fill the void in your heart. You hope for company and feel depressed when you are by yourself. You can feel lonely even when you are surrounded by people, because loneliness is more about the quality of relationships rather than their quantity [45].

1.1. Effects of Loneliness

Studies show that loneliness and social isolation can have a serious impact on one's mental health. Loneliness can lead to increasing levels of stress, anxiety, panic attacks, and drug or alcohol addiction [35]. It can lead to lack of confidence and lowering self-esteem [40]. A survey done by Mental Health Foundation shows that more than a third of people surveyed had felt depressed as a result of chronic loneliness [45].

The effects of chronic loneliness and social isolation aren't limited to one's mental health. Studies suggest loneliness is associated with a higher rate of death in older people [40]. It can weaken the immune system, increase sleep problems, increase blood pressure, lead to irregular heartbeat and can increase the chances of stroke and cardiovascular disease [35]. Research shows that Perceived Social Isolation (PSI), the scientific term for loneliness, increases the exposure to chronic diseases [38], mortality risk by 26% [45] and the risk of premature death by around 30% [46]. Research states that a lack of face-to-face networking could alter the way our genes work, upset immune responses, hormone levels and the function of arteries [41]. It could increase the risk of serious health problems like cancer, strokes, heart disease, and dementia [41]. Chronic loneliness impacts health in a greater way than smoking 15 cigarettes a day or being obese [45, 46].

Research shows that loneliness increases signaling in the sympathetic nervous system, which is responsible for controlling our fight-or-flight responses, and affects the production of white blood cells. White blood cells are critical to the immune system and defending the body against bacteria and viruses [38]. Hormone oxytocin is believed to be the chemical process underpinning the relationship between social contact and healthy hearts. Studies suggest that physical presence is needed for the release of hormone oxytocin, thereby associating physical social contact with healthy hearts [44]. Social pain is as real a sensation as physical pain. Research show that loneliness and social isolation activates the same parts of the brain as physical pain [45].

1.2. Increasing Levels of Loneliness

Despite of the hyper-connected world we live in, chronic loneliness seems to have slowly become a persistent problem [37]. The number of people saying there is no-one with whom they discuss important matters have nearly tripled in less than two decades [41]. The number of older adults aged 65 to 74 living alone and those aged 75 or over living alone has risen by 22 percent and 5 percent over the past decade [39]. WRVS warned more than 360,000 older people feel lonely because their children were too far away and too busy to visit them [40].

The Mental Health Foundation shows that one in 10 Britons is lonely and 48 per cent of people think the world is getting lonelier in general[40, 44]. It's not a problem faced by the elderly alone. Report by the Mental Health Foundation shows that loneliness is rising among the young as well [40]. Nearly 60% of those aged between 18 to 34 questioned spoke of feeling lonely often or sometimes, compared to 35% of those aged over 55 [44]. Only 11 per cent of single men and 15 per cent for married men in their early 20s to late-middle age have a friend to turn to in a time of crisis [37]. A survey carried out by The Movember Foundation in Britain shows alarming results of the increasing levels of loneliness and single-person households [37]. A Finnish study suggest that working people who live alone increase their risk of depression by up to 80% in comparison with people living in families [40]. The proportion of single-person households in Western countries has increased during the past three decades, with one in every three people in the US and the UK living alone [42]. The number of people aged between 45 and 64 living on their own has increased by 23 per cent over the past decade [39]. 58 per cent of the 2.47 million people living alone between the ages of 45 and 64 are men [37]. The reasons cited for this include waiting until they are older to get married, failed marriages usually leading to children living with their mother [37, 39]. Overall, women are still more likely to be alone. Of the 7.7m single-occupant households across all ages in Britain, 54 percent are comprised by women [37].

1.3. Reasons for Increasing Levels of Loneliness

Social changes such as the rise of the solo dweller, nuclear families and the surge in social networks, decline of community belonging, growing focus on work combined with an ageing population, are changing the way people interact with each other [40]. People's social networks are becoming smaller and families are not providing the same level of social context they may have done 50 years ago [40]. Increasing geographical distance between family members, marriage breakdown, multiple caring responsibilities and longer working hours are affecting the quality of family relationships [40]. Bereavement is also a major cause - losing your husband or wife can have a detrimental impact on one’s wellbeing [40].

Social networking websites like Facebook, Instagram, and Twitter were created to enrich our social life by letting us stay in touch despite of the geographical distance. They enable people to make connections they could not otherwise have been made, and virtual friendships can evolve into real-life relationships [44]. But such technologies are also replacing genuine human in-person interaction. Nearly a third of young adults questioned said they spent too much time communicating with friends and families online when they should see them in person [44]. They end up reducing people's real social networking involving quality personal interaction thereby leaving them with superficial, unrewarding relationships, feeling more isolated [40,41].

With the rising use of electronic media, the number of hours people spend interacting face-to-face has fallen dramatically since 1987 [41]. In-person interactions has an effect on the body that is not seen with exchange of emails or text messages [41].

Increasing rates of divorce, early deaths that leave the significant other alone, longer working hours and commute duration are some factors that have recently led to a tremendous increase in loneliness [32].

1.4. Groups Disproportionately Affected by Loneliness

A teenager transitioning from school to college struggling to make friends, divorced single parents, a bachelor relocating to a new city, bereaved older adults can all be inflicted by loneliness. Loneliness transcends all ages and gender. But two groups are particularly susceptible to loneliness: older adults and international students [33].

Loneliness is a widespread problem among older people and the absence of regular visits from children is an important factor in this sense of isolation [43]. The changing nature of the family, with fewer children who themselves often move away, has increased the prospect of elderly isolation. Working children move away leading to older parents seeing less of their family than they would like [43]. Fragmented family life can leave older people stranded with little contact from their own children. A report by Women's Royal Voluntary Service shows that thousands of elderly people are left isolated because their grown-up children live too far away [40]. One in 10 older people with children do not have family within an hour's drive distance of where they live [43]. 50% have visits from their children only once every two to six months [43]. There are 15% of such parents who only see their children only once a year. These elderly parents depend on phone calls from their children for social contact [43]. Study shows that two-thirds of old people who feel lonely prefer not admitting their loneliness to their own children in order to save their children from the bothering [43]. In a rather bleak view of family life during Christmas, many older people depend on television as their main form of company [43]. Almost a quarter of a million older people spend Christmas Day on their own [43]. Thus, loss of a spouse, death of friends, children moving away, feeling cut off from family, friends and community, financial issues, deteriorating health and decreased mobility are the reasons which increase the likelihood of loneliness in older adults.

On the other hand, distant family members, culture shock, climate change, academic stress and anxiety are some issues that cause loneliness in international students.

1.5. Coping With Loneliness

Research suggests that this experience of loneliness can be useful to us if we take it as motivation to reconnect with others and to seek out new friendships in order to reduce the social pain that we feel. But for some, when reconnection is not easy or not possible, if a person is socially isolated, people can remain in this uncomfortable loneliness state for a number of years [45]. Reports vary but typical numbers of people experiencing loneliness in this prolonged way range from three to 30 per cent [45].

Strategies to reduce loneliness that target negative thought processes were the most successful. For some lonely people, reducing social isolation and helping them link up with others reduces loneliness. But those who have been lonely for a number of years will have anxiety about making new friends, they may be distrustful of others and feel low about their own social skills. They need support to change their view of themselves, and how they feel others will react to them [45].

Effective ways to cope with loneliness include medication, meditation and yoga, exercise classes, walking groups, behavioral activation, cognitive behavioral therapy, interpersonal psychotherapy and group therapy [34]. Although they are many such known effective ways to cope with loneliness, these treatments reach less than half of those afflicted with loneliness worldwide due to social stigma associated with mental disorders, lack of resources, lack of skilled therapists and misdiagnosis [34].

1.6. Technologies to Detect Loneliness

There is a lot of scope to use technology such as wearable devices, Bluetooth, Infrared and other sensors to detect, monitor and track loneliness and similar psychological factors. Section 2 covers a few of such related studies and products in a more detailed way. There are a few studies take make use of wearable body gadgets that one can keep in close proximity. Though such technologies give higher accuracy, they are not suited for an everyday use because most users find the task of wearing and carrying an additional device cumbersome. A loneliness detecting technology should be an unobtrusive, so that it can be easily integrated in day-to-day life. Other approaches use short distance sensors like Bluetooth and Infrared, which require the surrounding devices to have Bluetooth or Infrared enabled. Any technology that monitors psychological factors should ideally not put restriction on the people the user communicates with or is surrounded by. Lastly, an ideal technology should be in the reach of the millions affected by loneliness and should be cost-effective. Socialoscope, which is a mobile sensing app that can passively detect loneliness in a cost-effective way, meets these requirement. Using such an app, therapists can track the interactions and loneliness levels of their patients. This app can be used by children of old adults to track the social isolation of their parents.

1.7. The Link between Loneliness and Personality

Loneliness is a psychological phenomenon. Prior research has reported individual links between loneliness and personality variables like self-esteem, depression, anxiety, neurosis, psychoticism and mistrust. A study by Hojat [14] establishes a relationship between loneliness and personality variables in a multivariate statistical model. This study shows that personality variables such as depression, anxiety, neuroticism, psychoticism, misanthropy positively predict loneliness score, while, personality variables such as self-esteem and extraversion negatively contribute to loneliness score.

The study uses UCLA Loneliness Scale, along with Rosenberg's Self-Esteem Scale, Rosenberg's Misanthropy Scale, Taylor Anxiety Scale, and Beck Depression Inventory.

1.8. Objective

This thesis investigates a smartphone app that passively detects loneliness in smartphone users based on the user's social interactions sensed by the smartphone's built-in sensors. Due to the ubiquitous ownership of smartphones, such a solution will likely be cost effective and have a global reach. Social interaction and activities detected by a smartphone do not have a one-to-one mapping with social wellness and loneliness levels, since human personality type plays a major role in determining how lonely each person feels for a given level of social interaction [35]. The average level of social interactions of a socially healthy extroverted person is not the same as that of a socially healthy introvert. In this work, like prior work, our goal was to detect loneliness levels of smartphone users passively. However, we also aimed to explore interactions of loneliness with personality, which will allow us to factor in personality into our loneliness inferences. Our loneliness and personality inferences will be deduced from users' smartphone interactions such as phone calls, text messages, Wi-Fi, Bluetooth, web browsing and app usage. This work makes the following specific research contributions:

- Investigate what sensed smartphone features (call logs, contacts, SMS logs, location, Bluetooth, Wi-Fi devices, app usage, browser usage, emails and social media) are statistically correlated with loneliness questions on the clinically validated UCLA loneliness scale.

- Extend the list of features explored by prior work on smartphone loneliness and personality sensing by including new Internet (emails, browser usage) and social media features (e.g. Facebook).

- Explore whether smartphone sensed loneliness is correlated with the big five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience). For instance, prior work by Hojat [14] has found some personality traits (e.g. extroverts) to be more vulnerable to loneliness.

- Synthesize machine learning classifiers that detect lonely smartphone users, while factoring in personality traits.

2. Related Work

Prior research on detecting social wellness and other psychological findings using sensors and smartphones is reviewed here.

2.1. Sociometer

Choudhury and Pentland [16, 17] describe the Sociometer, a wearable sensor device that acquires knowledge of how groups of people interact. This knowledge can be put to use in disciplines like organizational behavior, social network analysis and knowledge management applications like expert finding. The authors develop a computational framework to model face-to-face interactions within a community and identity leaders and connections with a group.

The Sociometer is an adaptation of the hoarder board, a wearable data acquisition board. It has an IR transceiver, a microphone, two accelerometers, on-board storage, and power supply. It captures and stores (i) the identity of people using IR transceiver that exchanges unique IDs with other people in its proximity, (ii) speech information using microphone detected conversation and (iii) motion information detected by the accelerometer. Thus using the built-in infrared transmitters and receivers, microphone and accelerometer, the Sociometer can communicate with another individual wearing the same device, thereby monitoring the people around, as speech information and motion information.

In comparison to Socialoscope, the Sociometer is a specialized wearable device, and comes with an additional manufacturing cost, and the discomfort of carrying a shoulder mounted device around, while our goal is to use off-the-shelf smartphones. Sensors used in Sociometer such as the IR and accelerometer are active sensors requiring more power, while Socialoscope focuses on minimizing power consumption. A set of four AAA batteries is required to power a Sociometer device for 24 hours. Also, Sociometer, focuses only on in- person communication, while Socialoscope focusses on calls, messages, and other means of social interaction. Moreover, Socialoscope captures a broader range of features and sensor data in comparison to Sociometer. Additionally, Sociometer can communicate only with people wearing the Sociometer, while Socialoscope does not require the people who the user is communicating with have the app installed on their smartphones. Furthermore, infrared works over a limited distance and its success depends on the line-of-sight between the transmitter-receiver pair.

The data gathering study for Sociometer was conducted at the MIT Media Lab with 8 user subjects who wore the wearable device both indoors and outdoors for six hours a day for 10 days, amounting to 60 hours of data per subject. The study for Socialoscope, on the other hand scoped for 9 user subjects for a duration of 14 days amounting to 336 hours of data per subject. To tackle privacy issues, Sociometer extracted only speech features such as spectral peaks and energy, and did not process the speech content. Instead of encrypted audio, garbled audio which makes the audio content unintelligible but maintains the identity and pitch of the speaker was used for this purpose. Socialoscope tackles similar privacy issue with regards to content of messages, contact names being called or messaged, etc. One-way encryption was utilized to protect the privacy of the user subjects.

2.2. Sociable Sense

Rachuri et al [19] presented a social sensing smartphone tool, SociableSense, which detects the sociability levels of smartphone users and the strength of their relations with their colleagues in their workplace. It captures user behavior in work environments, and provides the users with a quantitative real-time feedback of their sociability and that of their colleagues. This feedback includes sociability, strength of their relations, activity levels, and alerts about the users in sociable locations. The aim of this feedback is to help users in fostering their interactions and improving their relations with colleagues.

SociableSense computes the sociability values based on interaction and colocation patterns extracted from the sensed data at run-time. It uses accelerometer, Bluetooth, microphone, learning methods and decision theory. A sensor sampling component is used which adaptively controls the sampling rate of accelerometer, Bluetooth, and microphone sensors while balancing energy-accuracy-latency trade-offs based on reinforcement learning mechanisms. Additionally, a computation distribution component based on multi-criteria decision theory dynamically decides where to perform computation of tasks such as data analysis and classification, by considering the importance given to each of the dimensions: energy, latency, and data sent over the network.

The psychological study for SociableSense was conducted in an office environment that include work locations as users work spaces and sociable locations that include a common room and a cafeteria. It surveyed 10 user subjects who carried a Nokia 6210 Navigator mobile for two working weeks.

SociableSense brings mobile sensing and social network theory together in order to provide real-time feedback to users for enhancing their sociability with colleagues. In comparison, Socioloscope uses data from more variety of sensors. It is not restricted to a workplace environment. Additionally, Socialoscope gives quantity scores of loneliness, while the quality feedback mechanism for Socialoscope requires psychological expertise like constructive feedback to normal users and non-destructive feedback to lonely users. Moreover, Socialoscope also investigates the correlation of personality traits. In contrast, SociableSense takes computation optimization and sampling rate in focus, which is not handled in Socialoscope.

2.3. StudentLife

Wang et al [20] presented a smartphone sensing system, StudentLife that correlates sensor data from smartphones with mental wellbeing and academic performance. It is a continuous sensing app that assesses the day-to-day and week-by-week impact of academic workload on stress, sleep, activity, mood, sociability, mental well-being and academic performance. Results from this study show significant correlations between the automatic sensor data collected from smartphones and mental health and academic performance of the students.

The user study for StudentLife consists of 48 students voluntarily recruited from a computer science programing class at Dartmouth College across a 10 week term using Android phones. As a part of the study, data is automatically sensed and collected without any user interaction and uploaded to the cloud when the phone is being recharged and under Wi-Fi. The app performs automatic activity detection, conversation detection and sleep detection.

Simultaneously, the user subjects were asked to respond to various EMA questions at multiple times during the day. These questions administer the patient health questionnaire, UCLA loneliness scale, perceived stress scale, flourishing scale to users via pop-up questionnaires thereby providing additional state information such as stress, mood, happiness and current events. A team of medical doctors and psychologists assisted in providing the EMA reports.

To protect the privacy of the user subjects, the identity of each participant is anonymize with a random user id which is kept separate from all other project data so that the data cannot be traced back to individuals. Call logs and SMS logs are one-way hashed in order to protect the phone numbers or messages from the data. Participants' data is uploaded using encrypted SSL connections to ensure that their data cannot be intercepted by third parties. Data is then stored on secured servers. Similar techniques of random unique IDs, one-way hashed, secure servers are used in Socialoscope. In addition, Socialoscope used an encryption for the entire dataset before anonymization and after hashing.

In comparison to StudentLife, Socialoscope is not restricted to an academic environment, uses additional privacy and security mechanisms, is focused on loneliness levels and personality traits and uses a few different and a few common features for sensing.

2.4. Who’s Who with Big-Five

Chittaranjan et al [25] explored the relationship between behavioral characteristics derived from smartphone usage and self-perceived personality type. Like Socioloscope, it is based on categorization into Big-Five Personality Traits of Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness to Experience.

The user study included a data set collected from 83 individuals collected over a continuous period of 8 months in Switzerland. For this a continuous non-intrusive, passive data collection software running on Nokia N95 phones was used. This software collected anonymized logs of calls, SMS, Bluetooth scans, and application usage. Self-perceived personality was measured using the TIPI questionnaire [56]. Factors such as number and length of SMS messages, call duration and count, contact associated, missed calls, physical proximity via Bluetooth Scans and application usage – frequency and duration of using Office Apps, Internet Apps, Mail, Maps, Calendar, Camera are considered.

This work thus shows that aggregated features obtained from smartphone usage data can be indicators of the Big-Five personality traits and an automatic method is developed to infer the personality type of a user based on smartphone usage using supervised learning. In comparison, Socialoscope determines user loneliness levels by taking into account user personality traits derived from smartphone usage and daily social interactions. It takes into consideration privacy issues. Our work also uses the TIPI questionnaire based on Big-Five personality traits to determine the user's’ personality.

2.5. Vive: Inferring Loneliness in Older Adults

Sanchez et all [60] determine loneliness levels in old age population by monitoring various attributes through a smartphone app called Vive. Vive focuses on four main modules that originate from four factors of loneliness - family, spouse, social and existential crisis, and proposes four predictive models to determine level of loneliness of each factor focusing on the activities that can be monitored using a Smartphone. For data gathering of level of loneliness ESTE-R scale is used. This scale is targeted to Spanish speakers and targets different dimensions of loneliness and groups them into four factors: family loneliness, spousal loneliness, social loneliness and existential crisis. Encouragement messages are used to boost morale. For data gathering of activities performed by older adults, questionnaires referring to four activity categories: cellular phone use, number of outings from home, number of activities performed at home and time spent inside/outside of home are used. The final Android app, Vive, uses a configuration module that requests demographic information such as name, age, sex, civil status, health condition and working situation. Phone numbers of frequently contacted friends and family members, call log, geographic coordinates of home are GPS location used.

While Vive focuses on old adults, Socioloscope targets all lonely people including older adults and international students. The data gathering tool used in Socialoscope is an app, unlike Vive which is completely questionnaire based.

2.6. Detecting Loneliness in Smart Homes

Petersen, Johanna et al [59] developed a methodology to detect loneliness amongst elderly. Passive and unobtrusive in-home sensing technologies was used for this purpose. Loneliness is correlated to with decreased physical activity, decreased motor functioning, and a decline in activities of daily living. As all of these factors results in decrease in the amount of time spent outside home, time spent out-of-home was measured based on logistic regression. Motion sensors, contact sensors, computer sensors, and phone sensors were used to monitor the covariates of loneliness like sleep quality, frequency of visitor contact, total activity and time out-of-home. Just like Socialoscope, UCLA Loneliness Index was used for the loneliness detection.

In comparison to this work, Socialoscope, being an mobile sensing app, is not restrictive to a particular location, is cost-effective and accessible and feasible to all age groups

2.7. Moodscope

Moodscope [53] is an online mood-tracking system which helps you measure, record and track your daily mood scores. For measuring mood, previously well-validated daily mood measure, PANAS is used. Graphs are used for tracking your mood history which can also be shared via system generated emails with another person - a friend, colleague, partner, therapist - you nominate to act as a buddy for your tracking. On a daily basis, users are shown twenty double-sided playing cards daily, each of which represents an emotion like ‘alert’ or ‘nervous’. To answer the degree to which one might be feeling that emotion, on a scale from 'Very slightly or not at all' to 'Extremely', the cards can be flipped back to front or spun head to toe. Using these answers, a Moodscope score is calculated which ranges from 0 and 100, indicating how happy or sad you are. 100% is very happy, and 0% suggests that your spirits are extremely low.

Figure 1. Setting number of beep and beep hours in Mappiness [54]

2.8. Mappiness

Mappiness [54] is a part of a current research project at the London School of Economics that measures and records happiness. The Mappiness iPhone app prompts the user one to five times a day during agreed hours, as per their chosen setting as shown in Figure 1. It asks them how they are feeling, who they are with, where they are, what they are doing as shown in Figure 2. Photos can be submitted as well when outdoors. User can voluntarily report on their feeling too. These answers, photos, along with the location and ambient noise levels captured using microphone are sent to a data store to process the results of the users' happiness scores which are then rendered to them via graphs.

Figure 2. Giving your feedback on your feelings in Mappiness [54]

2.9. AutoEmotive

AutoEmotive [55] is a part of a current research project at MIT aiming to add emotion sensing technologies, specially stress sensing, inside cars, thereby improving driving experience, safety of drivers, and social awareness. Wireless biosensors, camera to capture heart rate, respiration rate, and heart rate variability, voice features from interactions of the person with the phone and GPS, amount of contact and forceful grasping of the door handle, the steering wheel and touch interactions with the navigation system, along with contextual data like the amount of acceleration, average speed, and amount of gas in the tank are used together to measure the stress levels of the driver. Figure 3 shows an overview of the relevant sensors used is AutoEmotive.

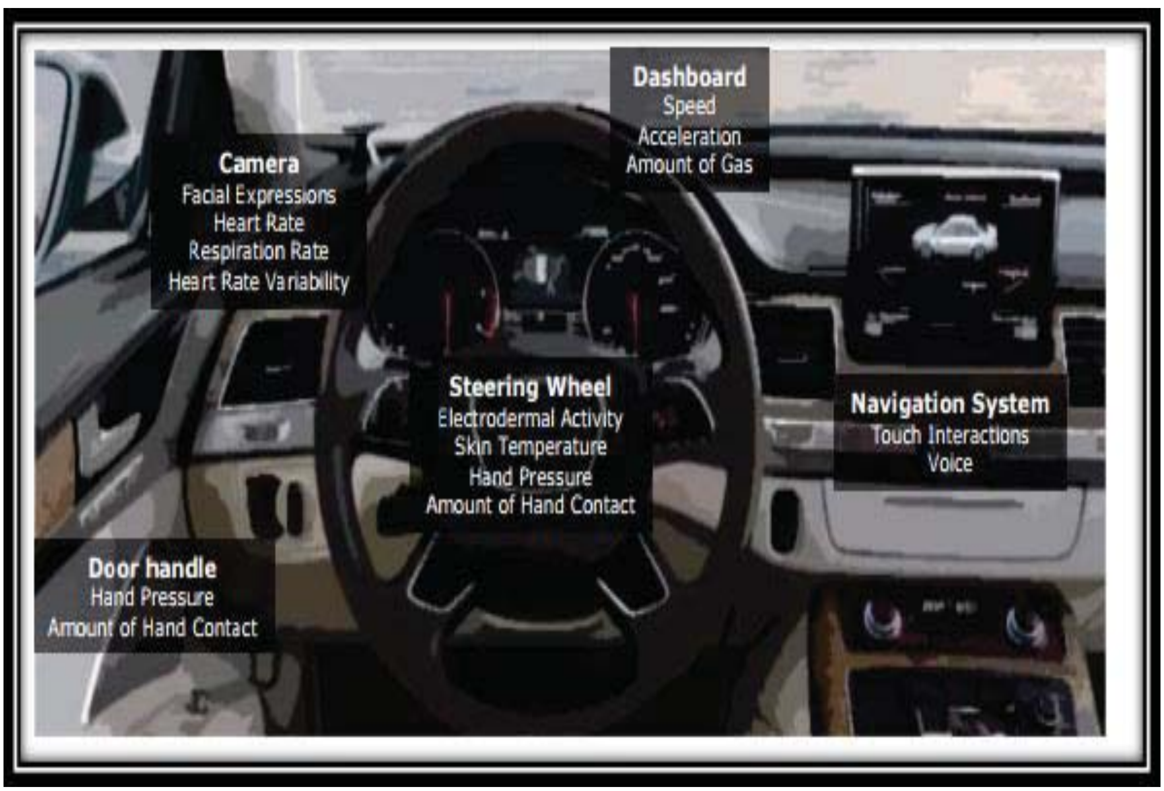

Figure 3. Sensors in AutoEmotive [55]

3. Approach

3.1. Overview of our Approach

This section describes the approach we followed to prove our hypothesis. In order to create an intelligent smartphone app that could passively detect user loneliness levels, we needed classifiers that could take as input the daily activities and social interactions the user has on his smartphone, factor in his personality traits using this and thereby predict his loneliness levels. To build the classifiers, we required the training data of the smartphone activities of the user subjects, their predetermined loneliness levels and personality trait scores.

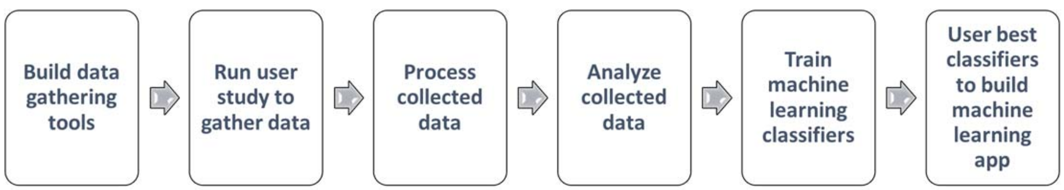

Figure 4. Workflow of the approach followed for Socialoscope

To gather such data we conducted a pilot study which would collect the required information. For gathering this data, we required appropriate tools that could gather such sensitive data correctly. The details of these tools and the study are described in section 3.2 and Section 3.3 respectively. The next step was to extract useful knowledge from this information. This step is described in detail in Section 3.4. After extracting the required information from the data, we performed statistical analysis on it to determine the features the correlated the best. This analysis is described in detail in Section 3.5. The most correlated features were then used to train machine learning classifiers and the most optimal one could then be used for building the final app. Figure 4 describes the workflow of the approach followed for Socialoscope at a higher level.

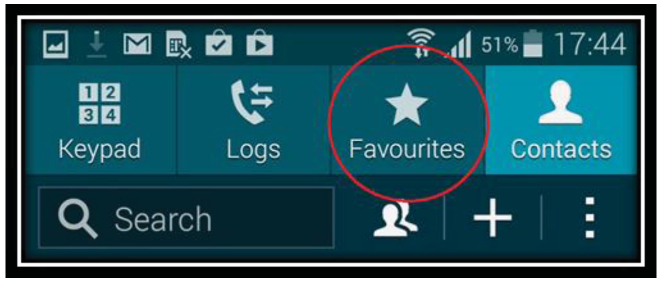

Figure 5. Marking favorite or starred contacts

3.2. Data Gathering

To gather the data required for training the machine learning classifiers, we required to run a user study by means of which we could collect the user subjects' smartphone activity and daily social interactions along with his loneliness and personality trait scores. To run this user study, we required to build the appropriate data gathering tools.

3.2.1. Data Gathering Tools

The aim was to create Android apps that could passively monitor these features and activities, log them on a daily basis and upload them to a server. We started with building such tools from scratch. We used Android Studio IDE for developing these Android apps. To keep the app running in the background, Android Services were used and Broadcast Receivers were used. The app informed the user of the permission required by it during the installation of the app. These included READ_CONTACTS,

ACCESS_FINE_LOCATION, READ_CALL_LOG, READ_SMS, ACCESS_WIFI_STATE, etc. Figure 6 shows the list of contacts marked as favorite or starred contacts using the favorite contacts option on any Android phone as shown in Figure 5. Figure 7 show the app usage on the phone along with their time of launch. Figure 8 show the bookmarks extracted from the browser while Figure 9 shows the browser history.

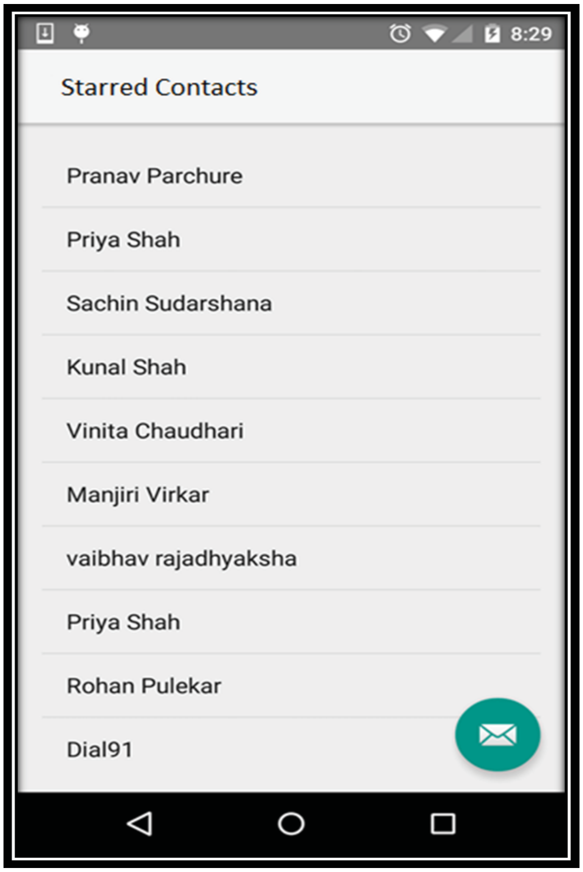

Figure 6. List of contacts marked as favorite or starred contacts

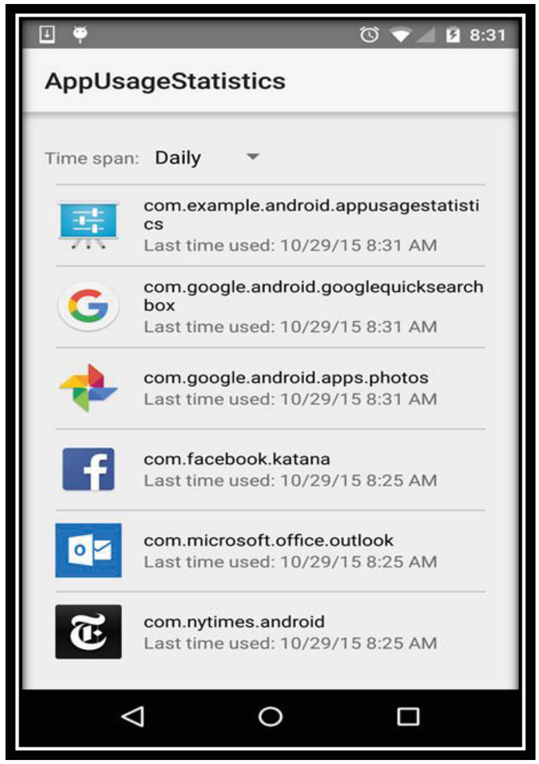

Figure 7. App usage log

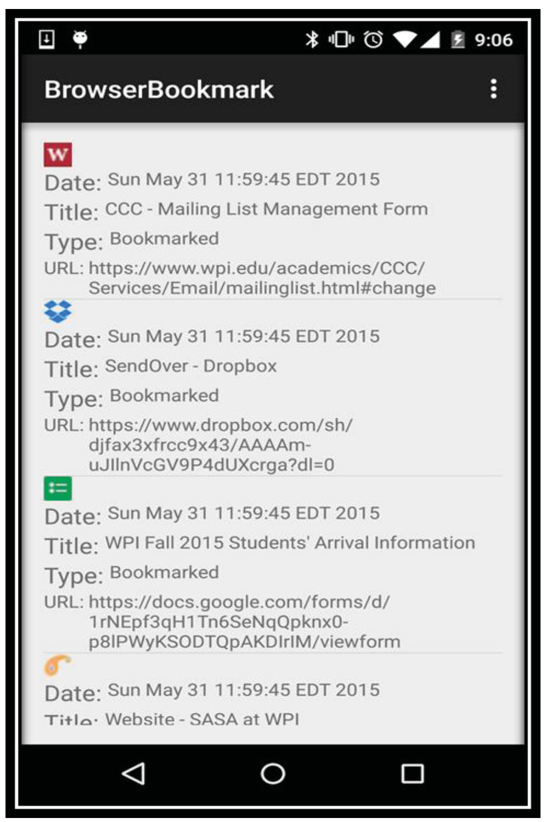

Figure 8. Browser Bookmark Log

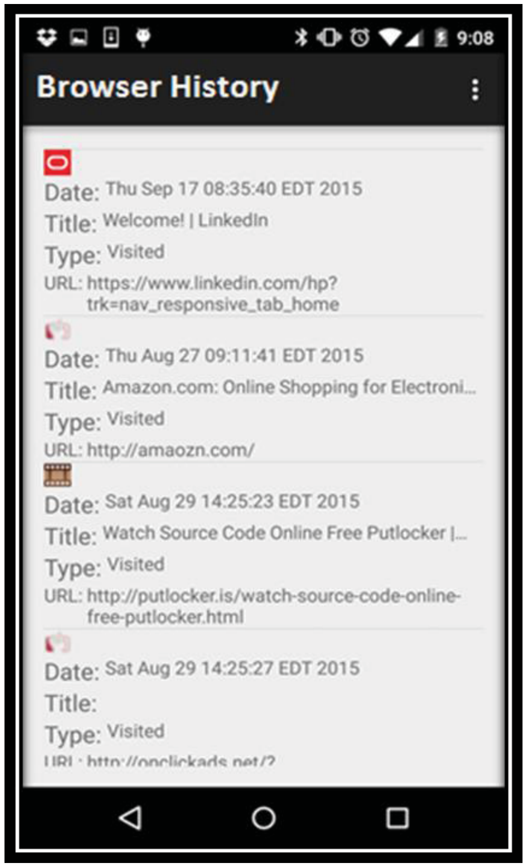

Figure 9. Browser History Log

At a later point, we utilized Funf in a Box [50]. This MIT created tool helps to create customized Android apps for gathering sensor data. There is a provision to select the features and sensors which one would like to collect in your app and the rate of probing the sensors. Figure 10 shows a part of the interface of Funf-in-a-Box showing the customization options. The app monitors the selected features for the specified frequency and uploads it a Dropbox account. You need to integrate your app to a Dropbox account for the same. This eliminates the need to host your own server to collect the data. You also need to provide the frequency of uploading the data and whether to upload automatically over Wi-Fi and/or mobile network. There is also an option to sync up manually. Since the app monitored and uploaded user data for a couple of sensors on a daily basis, the data scaled up to a few gigabytes at the end. We required a premium Dropbox account for the same. Given the vast variety of features targeted and sorted time span of the thesis, we landed up using this third-party tool to fasten data collection. The software is still in beta version, so there were a couple of issues that required to be handled. These are described in the Discussion section in further detail.

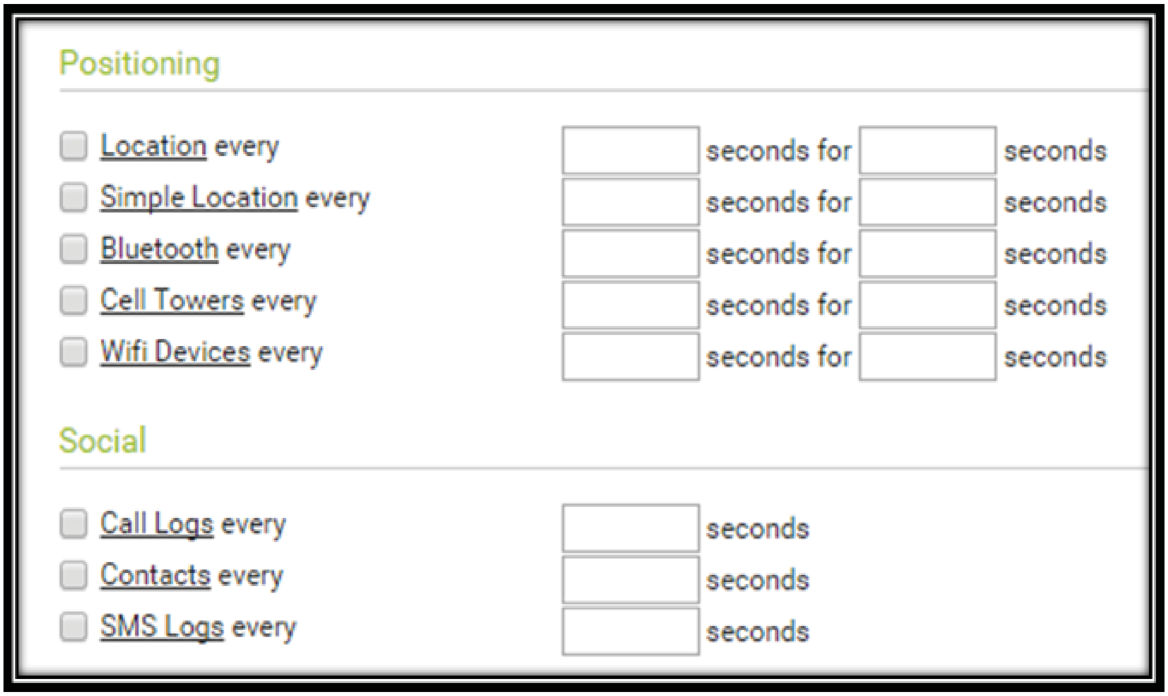

Figure 10. Interface of Funf-in-a-Box

3.3. User Study to Gather Training Data

A two weeks user study was conducted to gather data required for training machine learning classifiers. The user study consisted of three parts:

(a) Capturing the social interactions and activity on the user subjects' smartphone,

(b) Detecting the personality traits of the user subjects and

(c) Detecting the loneliness scores of the user subjects

3.3.1. Detecting Personality Traits

For capturing the user subjects' personality traits, we used a one-time personality detection survey based on Big-Five Personality Traits [25]. Many psychologists believe that there are five basic dimensions of personality namely Extraversion, Agreeableness, Conscientiousness, Emotional Stability or Neuroticism, and Intellect or Openness to Experience. These traits are commonly referred to as the Big-Five personality traits. It is a commonly used theory to describe human personality.

3.3.1.1. Other Theories on Personality Traits

Earlier theories have suggested a various number of possible traits. For example, Gordon Allport suggested a list of 4,000 personality traits that can be categorized into three levels namely, cardinal traits, central traits and secondary traits [61]. Raymond Cattell reported 16 personality factors namely Abstractedness, Apprehension, Dominance, Emotional Stability, Liveliness, Openness to Change, Perfectionism, Privateness, Reasoning, Rule Consciousness, Self-Reliance, Sensitivity, Social Boldness, Tension, Vigilance and Warmth [61]. Hans Eysenck's proposed a model of personality using three universal traits namely, Introversion/Extraversion, Neuroticism/Emotional Stability and Psychoticism [61].

Many psychologists and researchers had an opinion that Cattell's theory was complex and Eysenck's theory was limited in scope. The research for Big-Five dimensions of personality began in 1949 by D. W. Fiske. It was later expanded upon by other researchers such as Norman (1967), Smith (1967), Goldberg (1981), and McCrae & Costa (1987). Big-Five personality traits is proven universal by McCrae and his colleagues. Psychologist David Buss has proposed an evolutionary explanation for the big-five personality traits, there by stating that they have biological origins.

3.3.1.2. Big-Five Personality Traits

Let's look into the details of the five dimensions of personality covered by Big-Five Personality Traits [25].

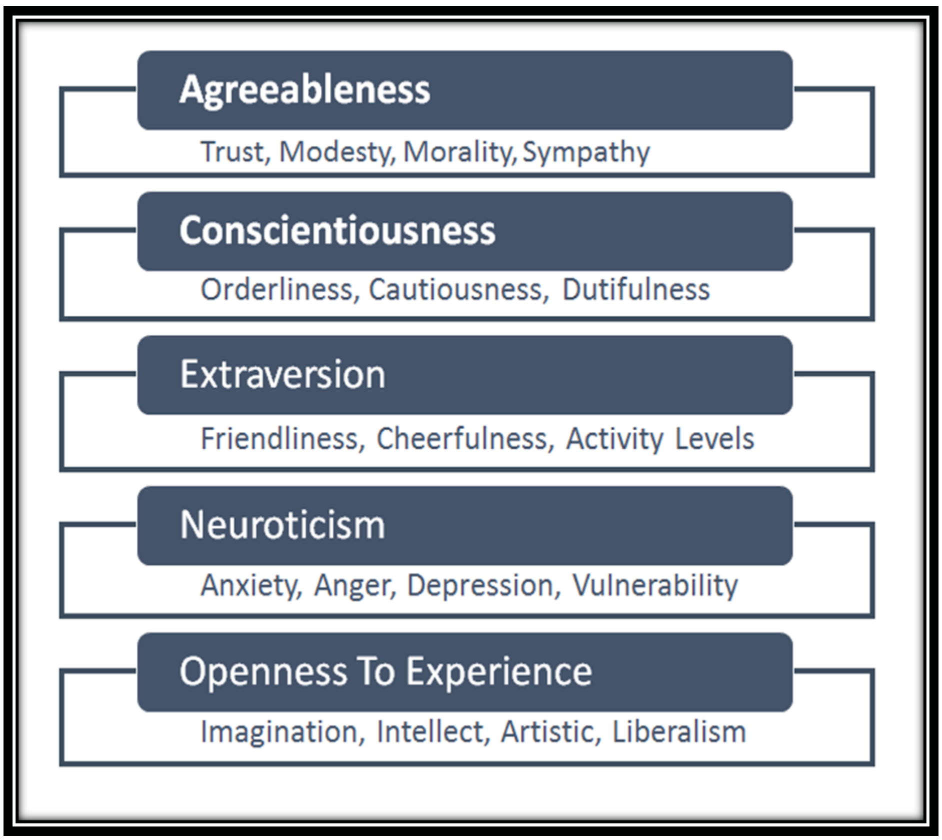

Figure 11. Big-Five Personality Traits

Extraversion describes energy assertiveness, sociability, gregariousness, excitement, and cheerfulness and the tendency to seek stimulation in the company of others. Agreeableness is the tendency to be compassionate, modest and cooperative towards others rather than suspicious and skeptical. Conscientiousness is a tendency to show self-discipline, plan, organize and act orderly, strive hard and aim for achievement.

Neuroticism describes an individual's level of emotional stability and impulse control, hence it is something referred to as emotional stability. Neuroticism describes vulnerability to negative emotions like anger, anxiety and depression. Openness to experience refers to one’s degree of creativity and imagination, intellectual curiosity, and preference for novelty and variety. Figure 11 summarizes these five traits.

Agreeableness

+ keyed	Am interested in people.
	Sympathize with others' feelings.
	Have a soft heart.
	Take time out for others.
	Feel others' emotions.
	Make people feel at ease.
– keyed	Am not really interested in others.
	Insult people.
	Am not interested in other people's problems.
	Feel little concern for others.

Figure 12. Questions for Agreeableness Big-Five Personality Trait [58]

Conscientiousness

+ keyed	Am always prepared.
	Pay attention to details.
	Get chores done right away.
	Like order.
	Follow a schedule.
	Am exacting in my work.
– keyed	Leave my belongings around.
	Make a mess of things.
	Often forget to put things back in their proper place.
	Shirk my duties.

Figure 13. Questions for Conscientiousness Big-Five Personality Trait [58]

Figures 12-16 list the questions from the Big-Five Personality Traits grouped together by the traits. Additionally, the reverse scored questions are marked with '–keyed' tags. Each trait has 10 questions amounting to 50 total questions for the entire questionnaire.

Emotional Stability

+ keyed	Am relaxed most of the time.
	Seldom feel blue.
– keyed	Get stressed out easily.
	Worry about things.
	Am easily disturbed.
	Get upset easily.
	Change my mood a lot.
	Have frequent mood swings.
	Get irritated easily.
	Often feel blue.

Figure 14. Questions for Emotional Stability Big-Five Personality Trait [58]

Extraversion

+ keyed	Am the life of the party.
	Feel comfortable around people.
	Start conversations.
	Talk to a lot of different people at parties.
	Don't mind being the center of attention.
– keyed	Don't talk a lot.
	Keep in the background.
	Have little to say.
	Don't like to draw attention to myself.
	Am quiet around strangers.

Figure 15. Questions for Extraversion Big-Five Personality Trait [58]

Intellect

+ keyed	Have a rich vocabulary.
	Have a vivid imagination.
	Have excellent ideas.
	Am quick to understand things.
	Use difficult words.
	Spend time reflecting on things.
	Am full of ideas.
– keyed	Have difficulty understanding abstract ideas.
	Am not interested in abstract ideas.
	Do not have a good imagination.

Figure 16. Questions for Intellect Big-Five Personality Trait [58]

A survey based on Big-Five Personality Traits was administered to subjects using WPI Qualtrics [51], an online survey software solution. This questionnaire consisted of 50 questions. Each personality trait had 10 questions corresponding to it, some of which are reverse scored. For each question, the user subjects had to mark their response from a choice of "Very Inaccurate", "Moderately Inaccurate", "Neither Inaccurate nor Accurate", "Moderately Accurate", and "Very Accurate". For the positively scored questions, the response "Very Inaccurate" is assigned a value of 1, "Moderately Inaccurate" a value of 2, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 4, and "Very Accurate" a value of 5. For the negatively keyed items, the response "Very Inaccurate" is assigned a value of 5, "Moderately Inaccurate" a value of 4, "Neither Inaccurate nor Accurate" a 3, "Moderately Accurate" a 2, and "Very Accurate" a value of 1. All the values of the responses of all the questions are summed to get the total score as well as score for each trait. The total score range from 50 to 250, while the score for each trait ranges from 10 to 50. The questions for each of the five personality traits is listed in the Appendix A.

3.3.2. UCLA Loneliness Scale

Statement	Never	Rarely	Sometimes	Often
*1. How often do you feel that you are "in tune" with the people around you?	1	2	3	4
2. How often do you feel that you lack companionship?	1	2	3	4
3. How often do you feel that there is no one you can turn to?	1	2	3	4
4 How often do you feel alone?	1	2	3	4
*5. How often do you feel part of a group of friends?	1	2	3	4
*6. How often do you feel that you have a lot in common with the people around you?	1	2	3	4
7. How often do you feel that you are no longer close to anyone?	1	2	3	4
8. How often do you feel that your interests and ideas are not shared by those around you?	1	2	3	4
*9. How often do you feel outgoing and friendly?	1	2	3	4
*10. How often do you feel close to people?	1	2	3	4
11. How often do you feel left out?	1	2	3	4
12. How often do you feel that your relationships with others are not meaningful?	1	2	3	4
13. How often do you feel that no one really knows you we11?	1	2	3	4
14. How often do you feel isolated from others?	1	2	3	4
*15. How often do you fee1 you can find companionship when you want it?	1	2	3	4
*16. How often do you feel that there are people who really understand you?	1	2	3	4
17, How often do you feel shy?	1	2	3	4
18. How often do you feel that people are around you but not with you?	1	2	3	4
*19. How often do you feel that there are people you can talk to?	1	2	3	4
*20. How often do you feel that there are people you can turn to?	1	2	3	4

Figure 17. Questions of UCLA Loneliness Scale - Version 3

For capturing the user subjects' loneliness values, we used a loneliness detection survey based on UCLA Loneliness Scale [21]. UCLA Loneliness Scale is a commonly used scale for quantitatively measuring one's subjective feelings of loneliness as well as feelings of social isolation. It's a 20 item scale, where participants rate each item on a scale as Never (1), Rarely (2), Sometimes (3), or Often (4). Some of the questions are reverse scored similar to Big-Five Personality Questionnaire. This scale was revised twice, once to make the items reverse scored and then to simplify the verbiage on the scale so that it is easy to comprehend. We used the latest revision, UCLA Loneliness Scale Version 3. Figure 17 lists the questions in this. The items marked with an asterisk are reverse scored. The score ranges from 20 to 80.

3.3.3. Android App

Figure 18. Summary of features tracked in Socialoscope

For collecting the user's smartphone activities and sensor data, we used the Android app created using Funf in a Box [50]. Since Funf in a Box is in beta version, we did not publish the app in the android market as recommended by the developers. Instead we distributed the app using Dropbox. On downloading the app, the user subjects gave permission to the app to access the required permissions before installing it on their

Android smartphones. Thus, they required to own and carry the Android smartphone during the period of the study. The app increased the battery consumption on these phones for the duration of the study. But it did not affect the data consumption as all uploads would take place only over Wi-Fi. The app monitored call logs, message log, app usage, email usage, Bluetooth and Wi-Fi connections and browser usage. Figure 18 summarizes the features tracked while Table 1 gives more details on the measures for each feature type.

Data Type	**Measured By**	**What it measures**	**Comments**
Phone Calls	Call count	If user has any phone communication channel	Used
	Call duration	If user has any prolonged phone communications, or keeps them to the minimum	Not used because of encryption
	Call from/to: Is Starred	If user has any phone communication with favorite contacts or others	Not used because of encryption
	Call type	If the user is the one calling or receiving calls, or is trying to avoid calls	Used
SMS	SMS count	If user has any message communication channel	Used
	SMS character count	If user has few short worded message interaction, or long chats	Not used because of encryption
	SMS from/to: Is Starred	If user has any message communication with favorite contacts or others	Not used because of encryption
	SMS type	If the user is the one sending or receiving messages	Used
App	No of launches	How frequently the user interacts with apps	Used
	App duration	How long does the app has the user’s focus	Used
	App category	What category of apps the user is visiting: Internet, Music, Audio, Video, Maps, Calendar, Camera, Social Networking, Games, Chat	Used
Bluetooth	No of unique BT IDs	No of people the user is around	Used

	No of times saved BT IDs are seen	No of frequently seen people the user is around	Not used because of encryption
	Duration of availability	For how long the user is around other people	Not used because of encryption
Wi-Fi	No of SSIDs	No of people networks the user is around	Used
	Duration of SSIS connectivity	For how long the user is around people networks	Not used because of encryption
	Type: Public/Home/Work	Whether the user is in a home or public of work network	Used
Browser	Browser history	What webpages were visited	Used
	Browser favorites	Whether a visited webpage was bookmarked	Used
	Browsing time of day	Whether the user browsed it during office hours, later in the evening, etc.	Used
	Browsing duration	For how long the user browsed the webpage	Used
	Category of website	What was the category of the webpage, social media, technical, news, etc.	Not used because of encryption
Social Media	Number of times social media app is launched	How frequently the user attempts to access the social world.	App launch along with category of app was used for this.

Table 1. Features analyzed in Socialoscope

3.3.3.1. Privacy and Security Concerns

Given the sensitivity of the data we collected, there were privacy concerns. To tackle these, two levels of encryption inbuilt in Funf were used. First level of encryption encrypted all the uploaded data, and was decrypted by the investigators. The second level of encryption was a one-way hash that could not be decrypted. This included certain private data like message text, website URL, message text, calling number, etc. Thus we had access to details of how many calls were made, but now which contact or number was called. Thus all the personal level information is hidden. Moreover, the data collected was uploaded to a Dropbox account which provides for a secure cloud storage.

Additionally, the data was mapped to a random user id. Thereby, any data that could be used for personal identification was not recorded. Also, at any given point, the user subjects had an option to quit the study in case of any risks or discomfort. Despite of this, we faced difficulty in reaching the targeted count for user subject.

3.4. Processing

The data collected from the user study had to be decrypted, processed and features had to be extracted before it could be analyzed. Funf in a Box creates a folder for your app at this location Dropbox/Apps/Funf in the associated Dropbox account. The name of this folder is the same as the name of app we specify will creating the app. Figure 19 shows the structure of the app directory.

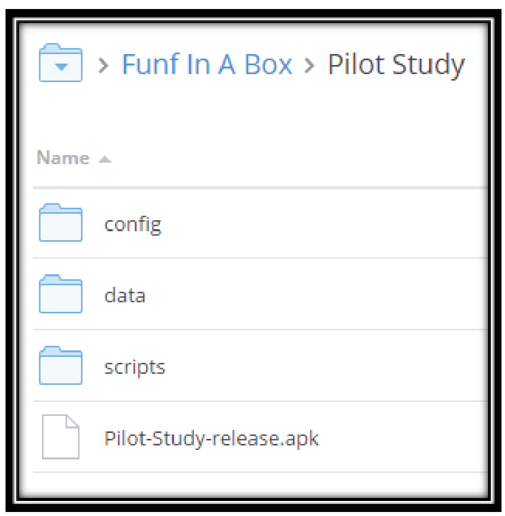

Figure 19. Dropbox folder structure for Funf Directory

This folder contains the APK for the app along with three sub-folders namely config, data and scripts. The data folder contains a subfolder name raw which contains the raw encrypted .db files uploaded from the app in a decrypted format. The config file contains the two JSON config files: ‘app_config.json’ specifying the app name, id, contact email address and description and ‘funf_config.json’ specifying the configuration parameters we chose while creating the app. It also contains a text file specifying the auto-generated encryption key used for the encryption. The scripts subfolder contains two files executables each for Windows OS and Mac OS for decrypting and extracting the features into csv files. For Linux, raw python scripts are given, which are located in the raw_scripts.zip file. On running the ‘/scripts/process_data’ script, all the files in the /data/raw folder are decrypted and merged into one SQLite database file. Processing time depends on the size of the data collected, and may take a few minutes as the size of your data increases. After completion, an ‘all_data.db’ file is created inside the /data directory. Every time you run this script this file will be overwritten, but will always include all of the data available in the raw directory. All the features are then extracted from this file into individual csv files using the ’/scripts/convert_to_csv’ script. A CSV file will be generated for each probe, which is then processed by java code to perform aggregation to get the counts. For example, get the outgoing calls made by the user per day using the CSV file of call log probe.

Similarly UCLA loneliness scores and scores for big-five personality traits had to be computed from the results of the loneliness and personality questionnaires. For both the questionnaires hosted on WPI Qualtrics, the results were extracted into CSV files, and the scores were computed. Java code that would read the CSV files, determine the reverse scored questions and scoring pattern, compute the corresponding scores and write it back to a result CSV file was used. Along with this, moving averages were computed for the loneliness scores as loneliness is a slow changing value.

3.5. Analysis

On processing the gathered data, the next step was to analyze the data to find out which features correlate the best with loneliness and personality scores. The most correlated features were then used to train machine learning classifiers. Correlation based feature selection (CFS) was used for finding out the most correlated features. CFS evaluates subsets of features on the basis of the hypothesis that "Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other" [52]. Loneliness scores, personality scores, and the feature values were used to do this feature selection as shown in Figure 20.

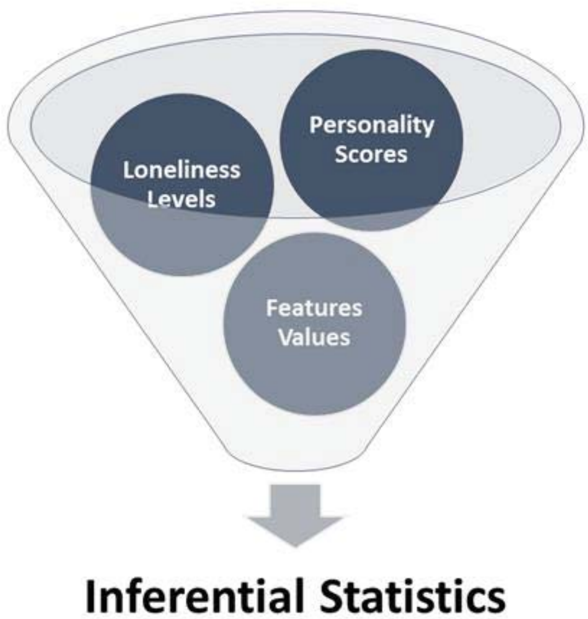

Figure 20. Inferential Statistics for Socialoscope

The correlation coefficient between each sensed feature and loneliness or personality were first computed. Correlation quantifies the extent to which two quantitative variables, say x and y, relate. If x increases proportionally as y does, a positive correlation exists between the two variables. On the other hand, if y decreases as x increments, a negative correlation exists between x and y. Correlation coefficient ranges from -1 to +1, where in a stronger positive coefficient is closer to +1 while a stronger negative coefficient is closer to -1. Table 2 summarizes this result. The formula used to calculate correlation coefficient is:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)(n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)}}$$

Equation 1. Calculating correlation coefficient

Here, n is the total number of samples, x_i is the i^{th} x value and y_i is the i^{th} y value.

Value of Correlation Coefficient (r)	Interpretation
-1 <= r < -0.7	Strong negative correlation
-0.7 < r < -0.3	Moderate negative correlation
-0.3 < r < 0	Weak negative correlation
0 < r < 0.3	Weak positive correlation
0.3 < r < 0.7	Moderate positive correlation
0.7 < r < 1	Strong positive correlation

Table 2. Correlation value and its interpretation [57]

The same was performed on normalized features values. After this, standard error of correlation coefficient, T-score, and degree of freedom were calculated. The formula used to calculate standard error of correlation coefficient and t-score or t-statistic are given below:

$$t_{stat} = \frac{r}{se_r}$$

Equation 2. Calculating t-score

$$se_r = \sqrt{\frac{1-r^2}{n-2}}$$

Equation 3. Calculating standard error of correlation coefficient

This was followed by hypothesis test. For this, p-value were computed to determine the significance of the results. Two-tailed hypothesis at a significance level of 0.05 was used for this. Table 3 shows the interpretation of p-values at significance level 0.05.

p-value	Interpretation
Small p-value (<= 0.05)	Strong evidence against the null hypothesis
P-value close to 0.05	Could go either way
Large p-value (> 0.05)	Weak evidence against the null hypothesis

Table 3. p-value and its interpretation

After calculating the p-value, all the feature values were then filtered by the p-value threshold and sorted as per the correlation coefficient. The result was the most correlated features. Figure 21 summarizes this statistical analysis workflow.

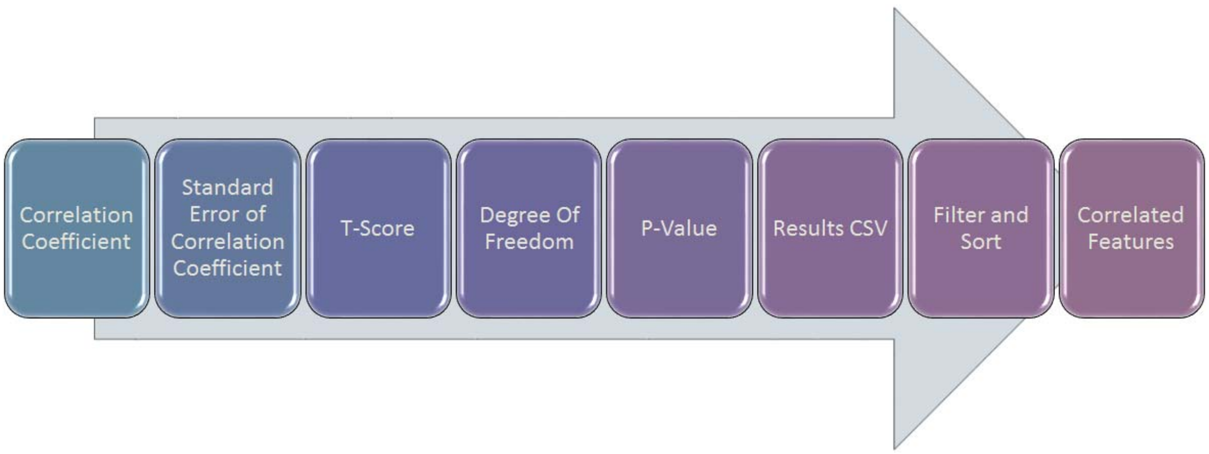

Figure 21. Workflow of statistical analysis for Socialoscope

4. Results

4.1. Overview of Results

The user study was conducted for a duration of two weeks for nine user subjects. We recruited user subjects using word of mouth, IRB and SONA Participant Pool. Probably because of the sensitivity of data collected, it was difficult to reach the targeted count of user subjects. Of the 9 subjects, 7 were students and 2 recent graduates. Of the 7 user subjects who were students, 6 were WPI students. Of the 7 user subjects who were students, 6 were master's students and 1 was a PhD candidate. Of the 9 users, 8 were international students. 6 were males and 3 females. The age range of the user subjects was 23 to 28 years.

This section summarizes the results drawn in this thesis by following the approach described in section 3. First, the results of personality scores are shown and are followed by the results of correlation and inferential statistics of the features with personality scores. Secondly, the results of UCLA loneliness scores are shown which are followed by the results of correlation and inferential statistics of the features with UCLA loneliness levels. Lastly, the results of the classifiers are shown for classifying loneliness levels using only the significant features. This is followed by the results of the classifiers for classifying loneliness levels taking into consideration significant personality traits along with the features.

4.2. Results of Personality Scores

Now, we analyze the Big-Five Personality Scores of the user subjects. Figure 22 shows the total Big-Five Personality Score for each of the nine user subjects. As a total score, we can see this value is not very useful for our classification. Figure 23, Figure 24, Figure 25, Figure 26, Figure 27 show the score of Agreeableness personality trait, Conscientiousness personality trait, Emotional Stability personality trait, Extraversion personality trait and Intellect personality trait for each of the nine user subjects respectively. Extraversion and Emotional Stability have a significant difference between User 1 and the rest of the users. As we will see in the later sections, this user corresponds to a user with high loneliness levels. No such conclusions can be made from the other three traits namely – Agreeableness, Conscientiousness and Intellect. This results are statistically proved in the next section.

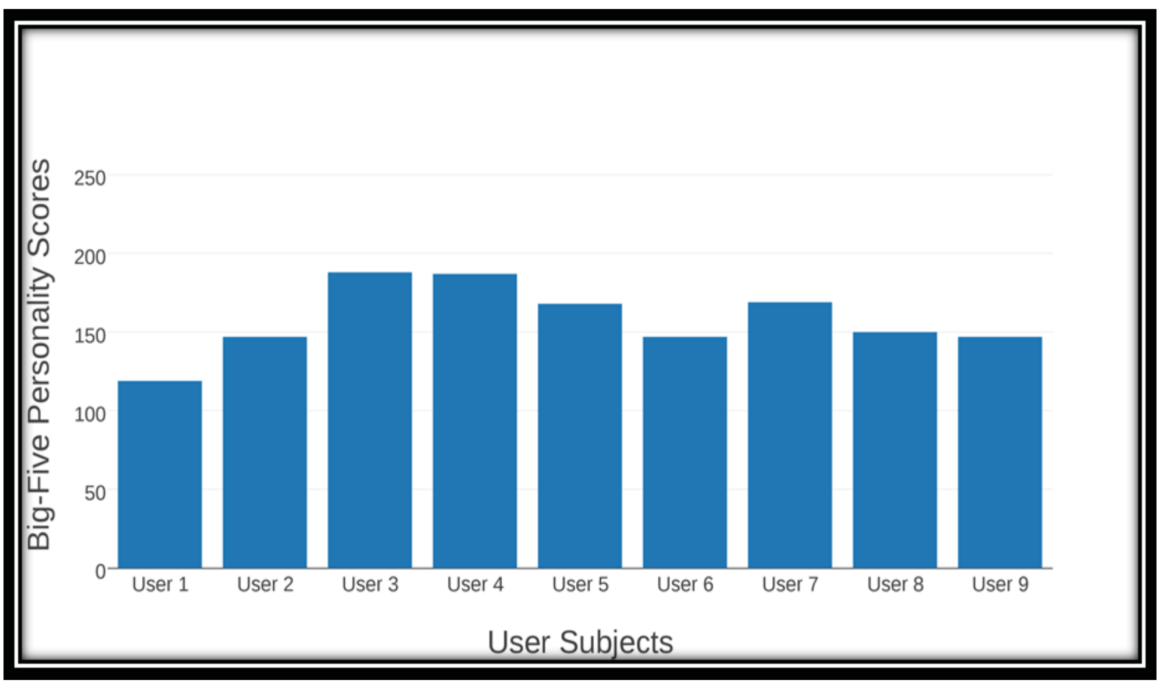

Figure 22. Big-five personality scores of user subjects

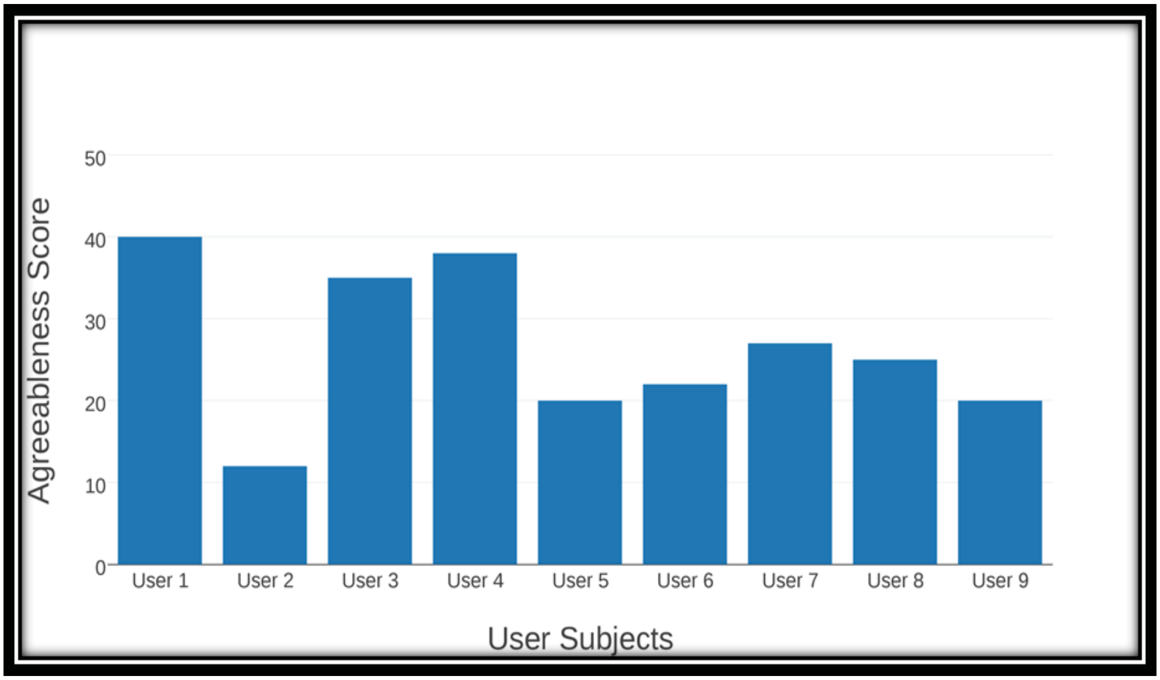

Figure 23. Agreeableness scores of user subjects

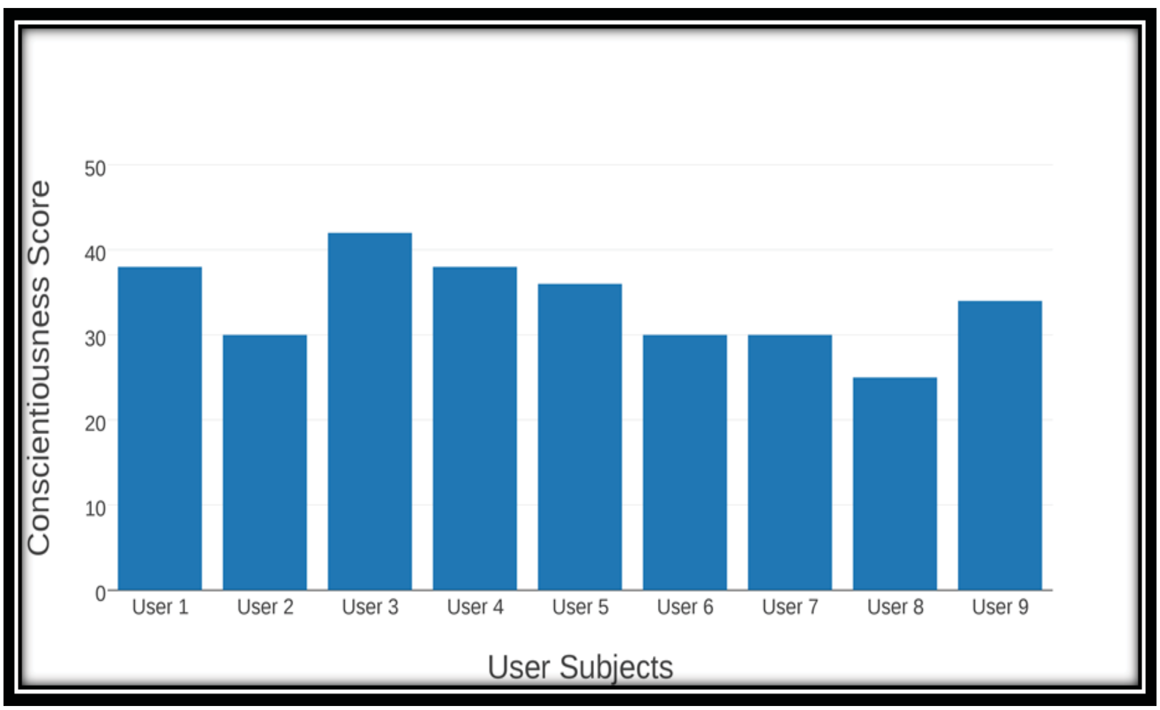

Figure 24. Conscientiousness scores of user subjects

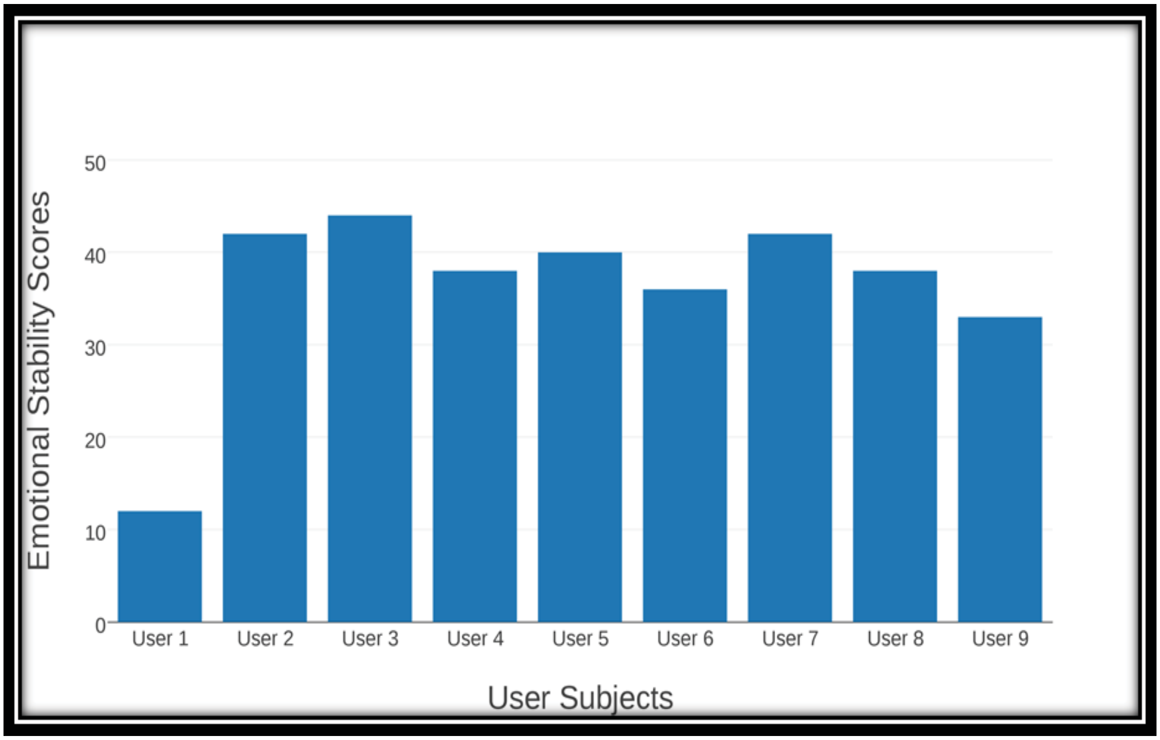

Figure 25. Emotional Stability scores of user subjects

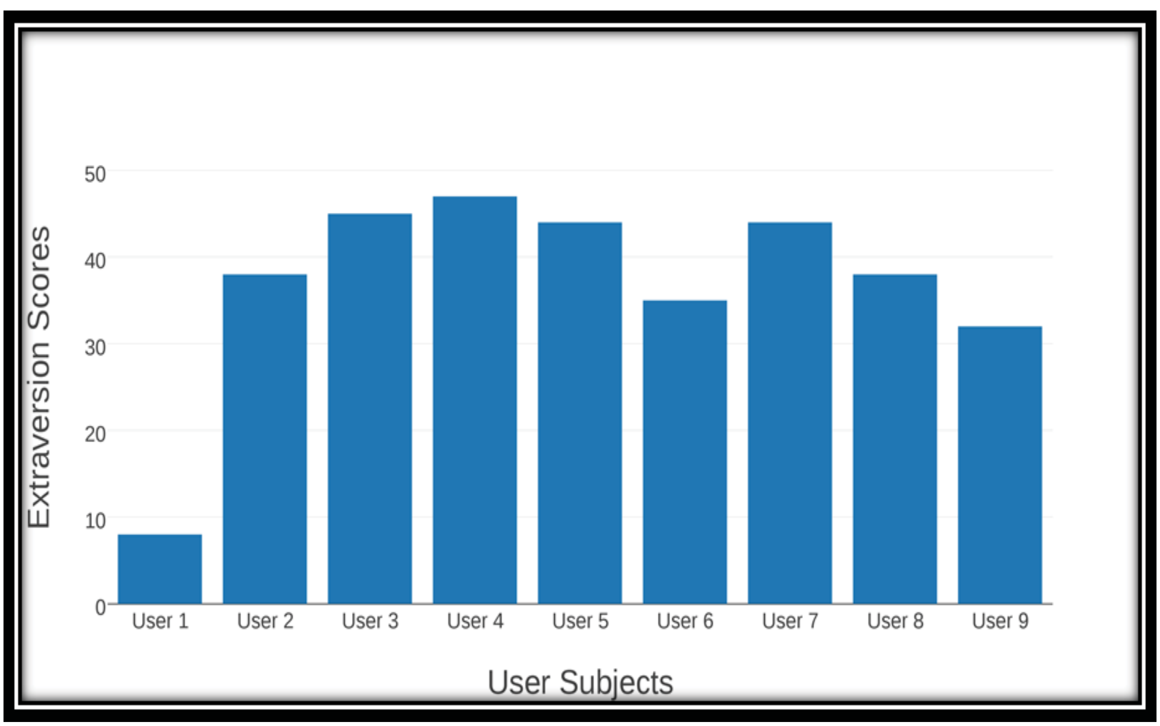

Figure 26. Extraversion scores of user subjects

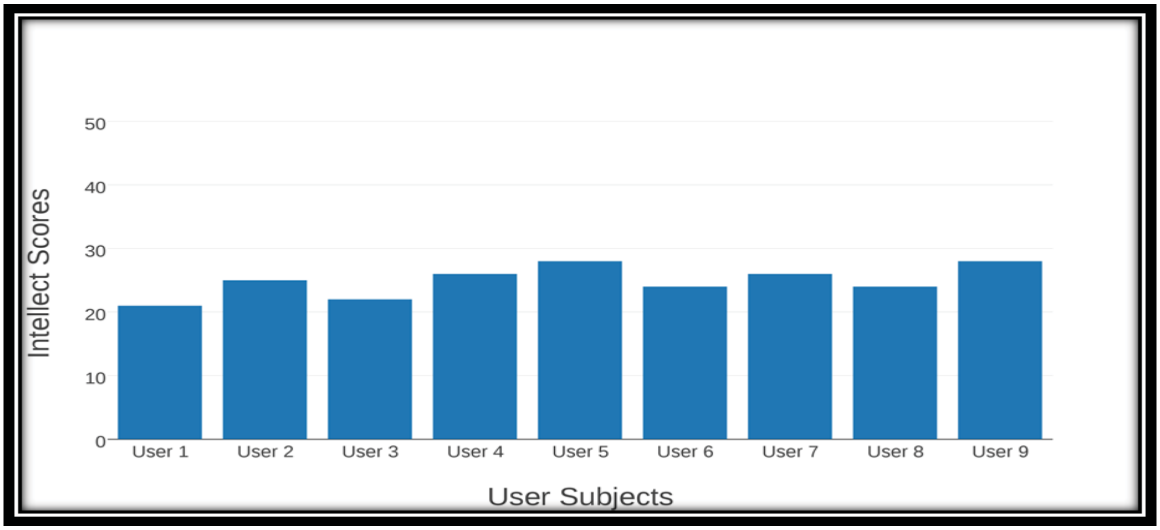

Figure 27. Intellect scores of user subjects

To understand the significance of the above conclusions, we performed statistical analysis of the feature values with each of these personality traits. We calculated the correlation coefficient, standard error of correlation coefficient, t-score followed by p-value for each of the five traits for all the features values as described in the analysis section. The results of this analysis is summarized in Tables 4-8 for Agreeableness personality trait, Conscientiousness personality trait, Emotional Stability personality trait, Extraversion personality trait and Intellect personality trait respectively. The significant factors are listed first, followed by the non-significant ones, in the decreasing order of their absolute value of correlation coefficient.

On analysis, we conclude that Emotional Stability personality trait and Extraversion personality trait are strongly correlated with most of the features, while Agreeableness personality trait, Conscientiousness personality trait and Intellect personality trait are weakly correlated with most of the features. This result proves the finding by Hojat that the felling of loneliness is accompanied by depression, anxiety, neuroticism and that loneliness is linked inversely to self-esteem and extraversion [14].

	Correlation Coefficient	**Standard Error of Correlation Coefficient**	**T score**	**p-value**	**Is Significant**
No of outgoing calls	-0.403	0.082	-4.914	< 0.00001	Yes
No of messages	-0.36	0.084	-4.286	0.00003	Yes
No of late night browser searches	0.344	0.084	4.096	0.000075	Yes
No of long incoming or outgoing calls	-0.33	0.085	-3.882	0.000167	Yes
High Activity Level	-0.319	0.085	-3.753	0.000267	Yes
No of short incoming or outgoing calls	-0.303	0.086	-3.523	0.0006	Yes
Percentage of calls that are missed	0.256	0.087	2.943	0.0038	Yes
No of incoming calls	-0.253	0.087	-2.908	0.0043	Yes
No of auto-joined trusted Wi-Fi SSIDS	-0.236	0.087	-2.713	0.0076	Yes
Difference between no of outgoing and incoming messages	-0.222	0.088	-2.523	0.0129	Yes
No of Calls	-0.381	0.83	-0.459	0.647	No
Difference between no of outgoing and incoming calls	-0.162	0.089	-1.821	0.71	No
No of Contacts	-0.159	0.089	-1.787	0.764	No
No of missed calls	0.139	0.089	1.562	0.1208	No
Low Activity Level	0.14	0.089	1.573	0.118	No
No of browser searches	0.114	0.089	1.281	0.203	No
Moving Travel State	-0.111	0.089	-1.247	0.215	No
No of Bluetooth devices around	-0.097	0.089	-1.089	0.278	No

Table 4. Correlation coefficient and p-values of factors with Agreeableness Score

	Correlation Coefficient	**Standard Error of Correlation Coefficient**	**T score**	**p-value**	**Is Significant**
High Activity Level	-0.435	0.08	-5.438	< 0.00001	Yes
No of long incoming or outgoing calls	-0.298	0.086	-3.465	0.000729	Yes
Low Activity Level	0.233	0.087	2.678	0.0084	Yes
No of auto-joined trusted Wi-Fi SSIDS	-0.184	0.089	-2.068	0.0407	Yes
No of late night browser searches	0.161	0.089	1.809	0.728	No
No of Contacts	-0.144	0.089	-1.618	0.108	No
Percentage of calls that are missed	0.111	0.089	1.248	0.214	No
No of outgoing calls	-0.106	0.0893	-1.187	0.2374	No
No of messages	-0.099	0.089	-1.112	0.268	No
No of browser searches	0.096	0.089	1.079	0.283	No
No of Calls	-0.092	0.089	-1.034	0.303	No
Difference between no of outgoing and incoming calls	-0.063	0.0897	-0.708	0.48	No
Difference between no of outgoing and incoming messages	-0.058	0.089	-0.652	0.516	No
No of incoming calls	-0.042	0.089	-0.472	0.6377	No
No of short incoming or outgoing calls	0.042	0.089	0.472	0.6377	No
No of Bluetooth devices around	-0.017	0.089	-0.191	0.849	No
No of missed calls	0.0135	0.089	0.151	0.8802	No
Moving Travel State	0.011	0.089	0.124	0.902	No

Table 5. Correlation coefficient and p-values of factors with Conscientiousness Score

	Correlation Coefficient	**Standard Error of Correlation Coefficient**	**T score**	**p-value**	**Is Significant**
No of messages	0.845	0.048	17.604	< 0.00001	Yes
No of outgoing calls	0.673	0.067	10.045	< 0.00001	Yes
No of Calls	0.656	0.068	9.648	< 0.00001	Yes
No of late night browser searches	-0.627	0.069	-9.087	< 0.00001	Yes
No of short incoming or outgoing calls	0.555	0.075	7.4	< 0.00001	Yes
No of long incoming or outgoing calls	0.504	0.077	6.546	< 0.00001	Yes
No of auto-joined trusted Wi-Fi SSIDS	0.474	0.079	6	< 0.00001	Yes
Moving Travel State	0.469	0.079	5.937	< 0.00001	Yes
No of incoming calls	0.425	0.081	5.247	< 0.00001	Yes
Percentage of calls that are missed	-0.399	0.082	-4.866	< 0.00001	Yes
Difference between no of outgoing and incoming messages	0.393	0.083	4.735	< 0.00001	Yes
Difference between no of outgoing and incoming calls	0.267	0.087	3.069	0.0026	Yes
No of missed calls	-0.185	0.089	-2.079	0.3967	Yes
No of browser searches	-0.182	0.088	-2.069	0.406	Yes
High Activity Level	0.1315	0.089	1.478	0.142	No
No of Bluetooth devices around	-0.0648	0.089	-0.731	0.466	No
Low Activity Level	-0.013	0.089	-0.146	0.884	No
No of Contacts	0.003	0.089	0.034	0.973	No

Table 6. Correlation coefficient and p-values of factors with Emotional Stability Score

	Correlation Coefficient	Standard Error of Correlation Coefficient	T score	p-value	Is Significant
No of messages	0.795	0.055	14.455	< 0.00001	Yes
No of late night browser searches	-0.601	0.072	-8.348	< 0.00001	Yes
No of outgoing calls	0.589	0.072	8.181	< 0.00001	Yes
No of Calls	0.584	0.073	8	< 0.00001	Yes
No of short incoming or outgoing calls	0.505	0.078	6.474	< 0.00001	Yes
No of auto-joined trusted Wi-Fi SSIDS	0.484	0.079	6.126	< 0.00001	Yes
Moving Travel State	0.458	0.079	5.798	< 0.00001	Yes
No of long incoming or outgoing calls	0.442	0.081	5.457	< 0.00001	Yes
No of incoming calls	0.393	0.083	4.735	< 0.00001	Yes
Percentage of calls that are missed	-0.379	0.083	-4.566	0.000012	Yes
Difference between no of outgoing and incoming messages	0.368	0.084	4.381	0.000025	Yes
No of browser searches	-0.231	0.087	-2.656	0.0089	Yes
Difference between no of outgoing and incoming calls	0.216	0.088	2.455	0.1547	Yes
No of Contacts	-0.211	0.088	-2.398	0.0179	Yes
No of missed calls	-0.176	0.089	-1.978	0.5015	No
No of Bluetooth devices around	-0.108	0.089	-1.213	0.227	No
High Activity Level	0.099	0.089	1.112	0.268	No
Low Activity Level	0.009	0.089	0.101	0.919	No

Table 7. Correlation coefficient and p-values of factors with Extraversion Score

	Correlation Coefficient	Standard Error of Correlation Coefficient	T score	p-value	Is Significant
No of Contacts	-0.438	0.081	-5.408	< 0.00001	Yes
No of late night browser searches	-0.373	0.083	-4.494	0.000016	Yes
High Activity Level	0.351	0.084	4.179	0.000055	Yes
No of messages	0.344	0.084	4.095	0.000076	Yes
No of browser searches	-0.339	0.084	-4.036	0.000094	Yes
No of auto-joined trusted Wi-Fi SSIDS	0.318	0.085	3.741	0.00027	Yes
Percentage of calls that are missed	-0.258	0.087	-2.966	0.00362	Yes
No of short incoming or outgoing calls	0.222	0.088	2.523	0.0129	Yes
No of Calls	0.221	0.088	2.511	0.0133	Yes
No of outgoing calls	0.219	0.087	2.517	0.0131	Yes
No of incoming calls	0.196	0.089	2.202	0.295	Yes
Difference between no of outgoing and incoming messages	0.193	0.089	2.16	0.327	Yes
No of long incoming or outgoing calls	0.169	0.088	1.921	0.057	No
Moving Travel State	0.172	0.088	1.955	0.528	No
No of missed calls	-0.154	0.089	-1.731	0.0869	No
Low Activity Level	-0.151	0.089	-1.697	0.922	No
Difference between no of outgoing and incoming calls	0.039	0.089	0.438	0.662	No
No of Bluetooth devices around	-0.026	0.089	-0.292	0.771	No

Table 8. Correlation coefficient and p-values of factors with Intellect Score

4.3. Results of Loneliness Levels

Next, we analyze the UCLA Loneliness Levels of the user subjects. Figure 28 to Figure 31 are box plots showing the trends in the feature values for some of the significant features for all the nine user subjects. Figure 28 shows the number of messages sent by user per day for the duration of two weeks of the study. The box plot show the trends of messages sent for each user – the quartiles, the median, the minimum and the maximum values and the quartiles. User 1 who corresponds with high loneliness levels can be seen here with relatively low messages counts. Figure 29, 30 and 31, on similar lines, show the number of outgoing calls made by each user per day for the duration of the study, the number of calls made by each user per day for the duration of the study, and the number of short calls made by each user per day for the duration of the study. Similar trends can be seen in these figures for user with high loneliness levels.

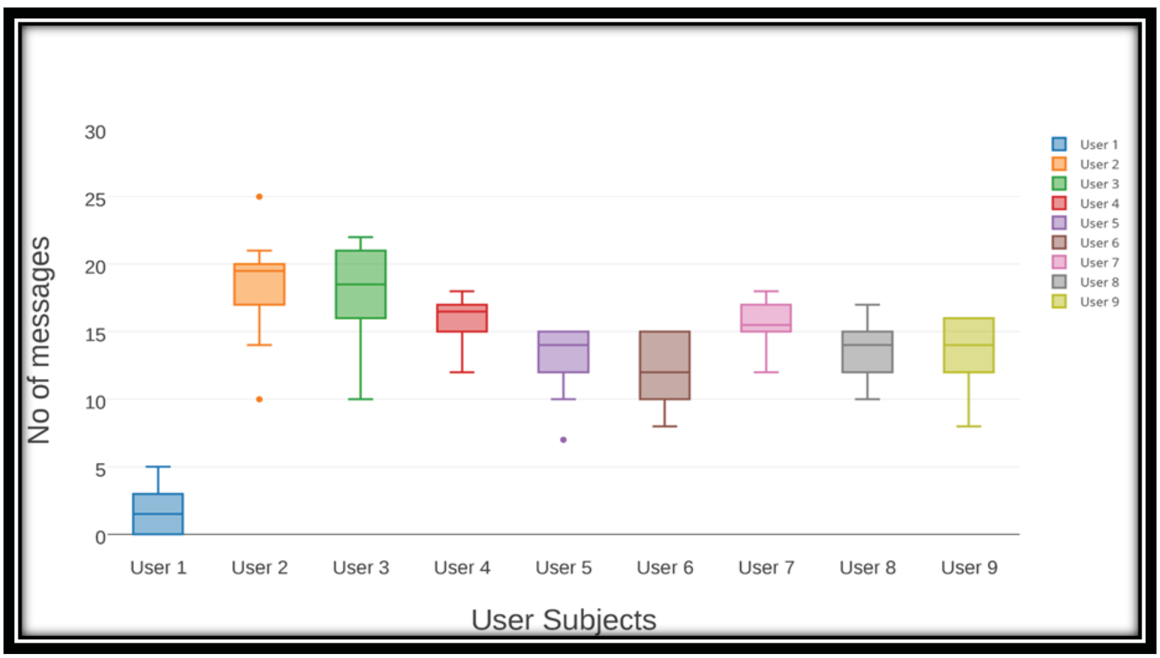

Figure 28. Number of messages per user subject

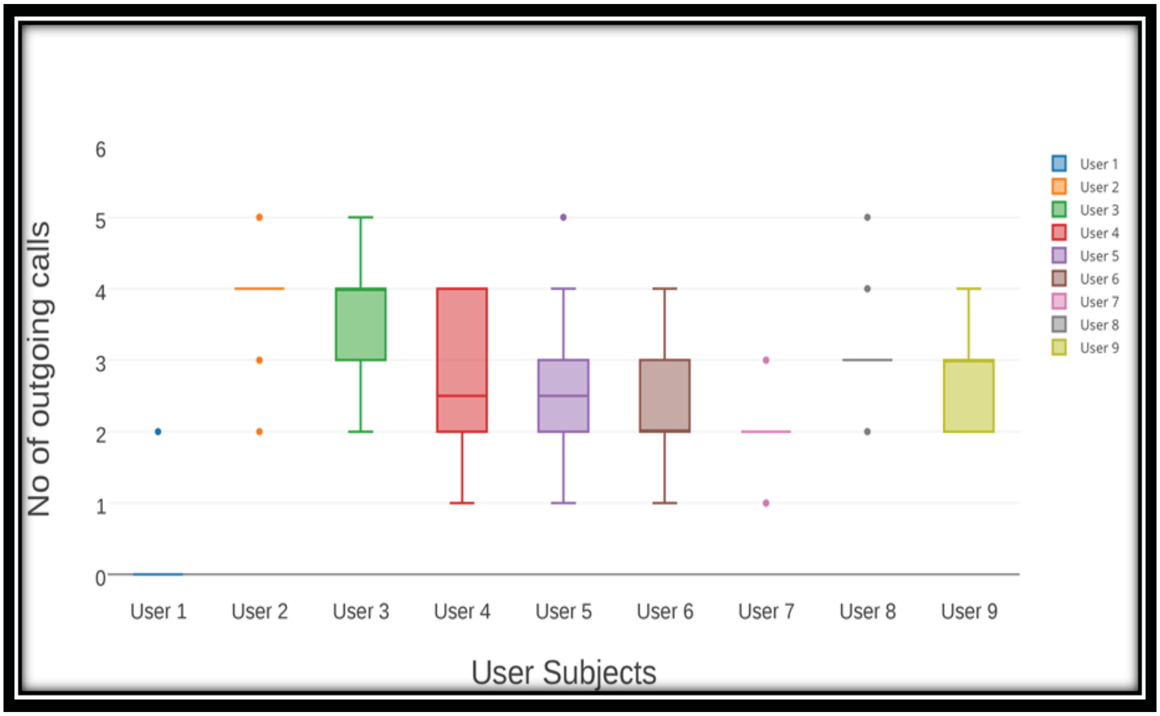

Figure 29. Number of outgoing calls per user subject

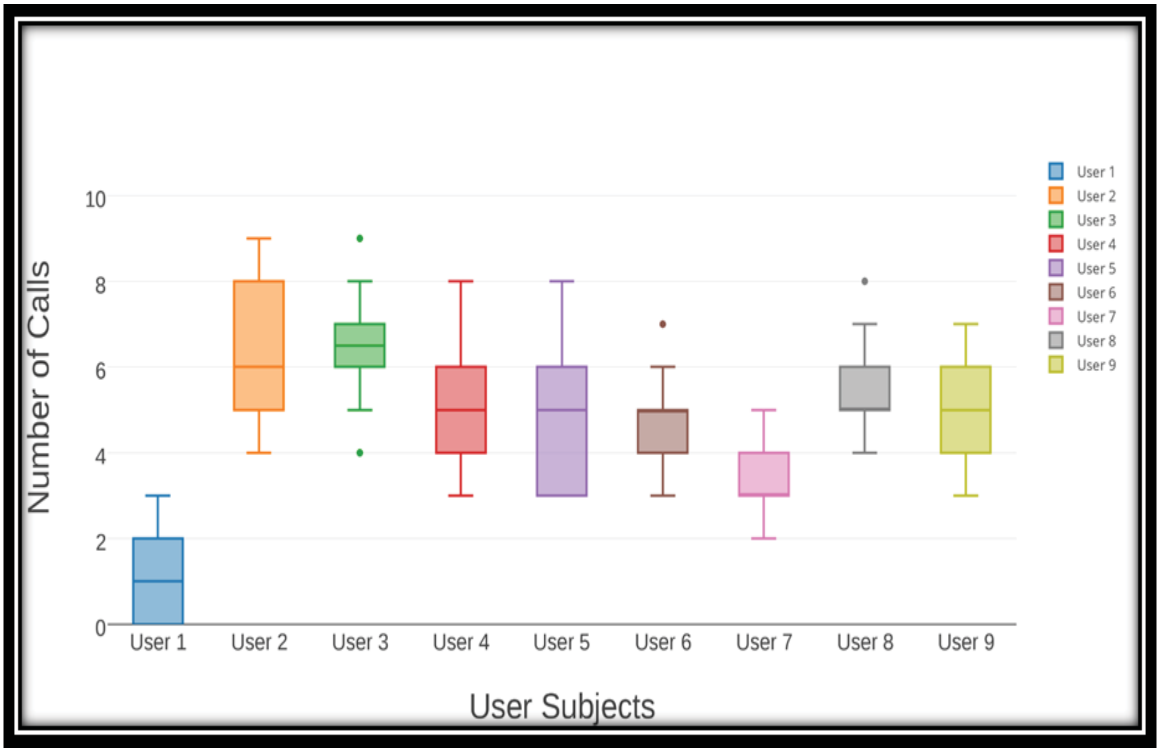

Figure 30. Number of calls per user subject

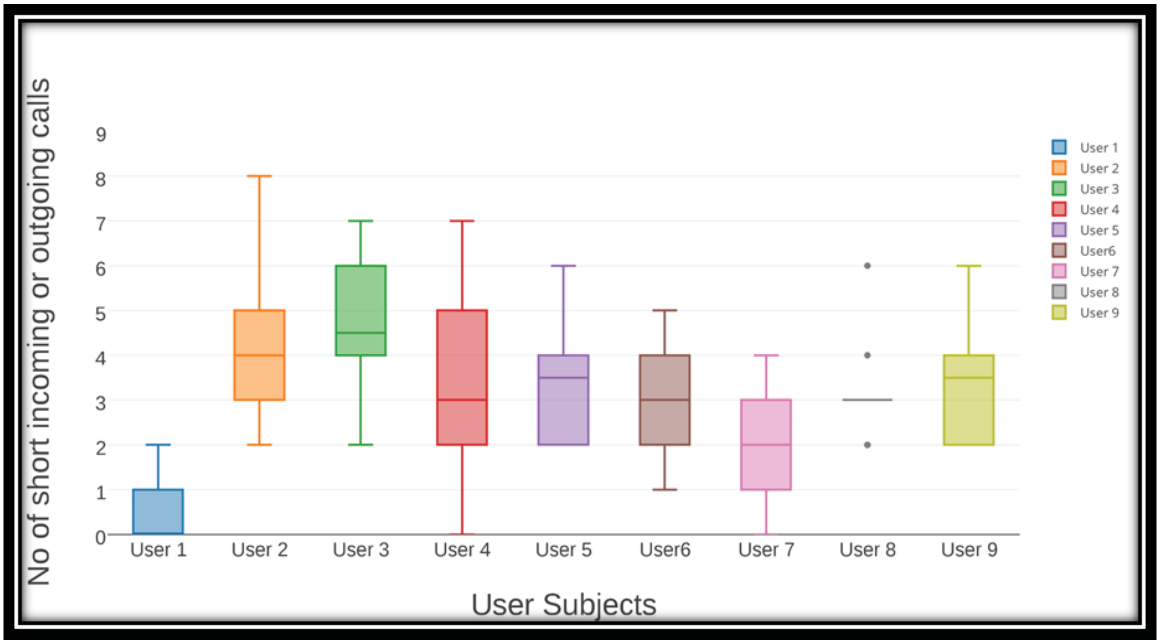

Figure 31. Number of short incoming or outgoing calls per user subject

To test the significance of the features with UCLA Loneliness Levels, we performed statistical analysis of the feature values with each of these personality traits. Figure 32 to Figure 35 are scatter plots showing the correlation of some of the significant features with UCLA Loneliness Levels. Figure 32 shows the correlation between number of messages and UCLA Loneliness Levels. A positive slope tending towards +1 shows a strong positive correlation of the features with UCLA Loneliness Levels. We see user with high loneliness levels and low number of messages on the extreme left of the fit line. Figure 33 and Figure 34 shows similar results for the number of outgoing calls and number of calls. Users with high loneliness levels with lower number of messages, calls, and outgoing calls. Figure 35 on the other hand shows the correlation of number of late night searches with UCLA Loneliness Levels. A negative slope tending towards -1 shows a strong negative correlation of the features with UCLA Loneliness Levels. Users with high loneliness levels tend to browse more at night.

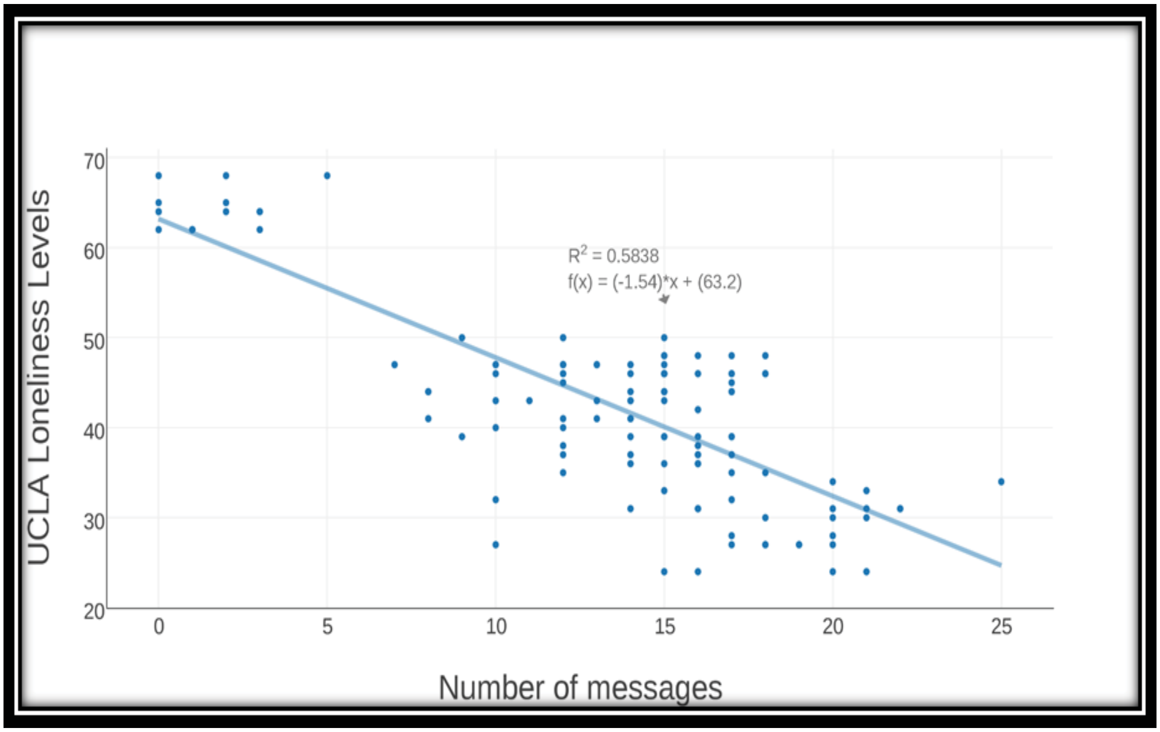

Figure 32. Number of messages vs. UCLA Loneliness Levels

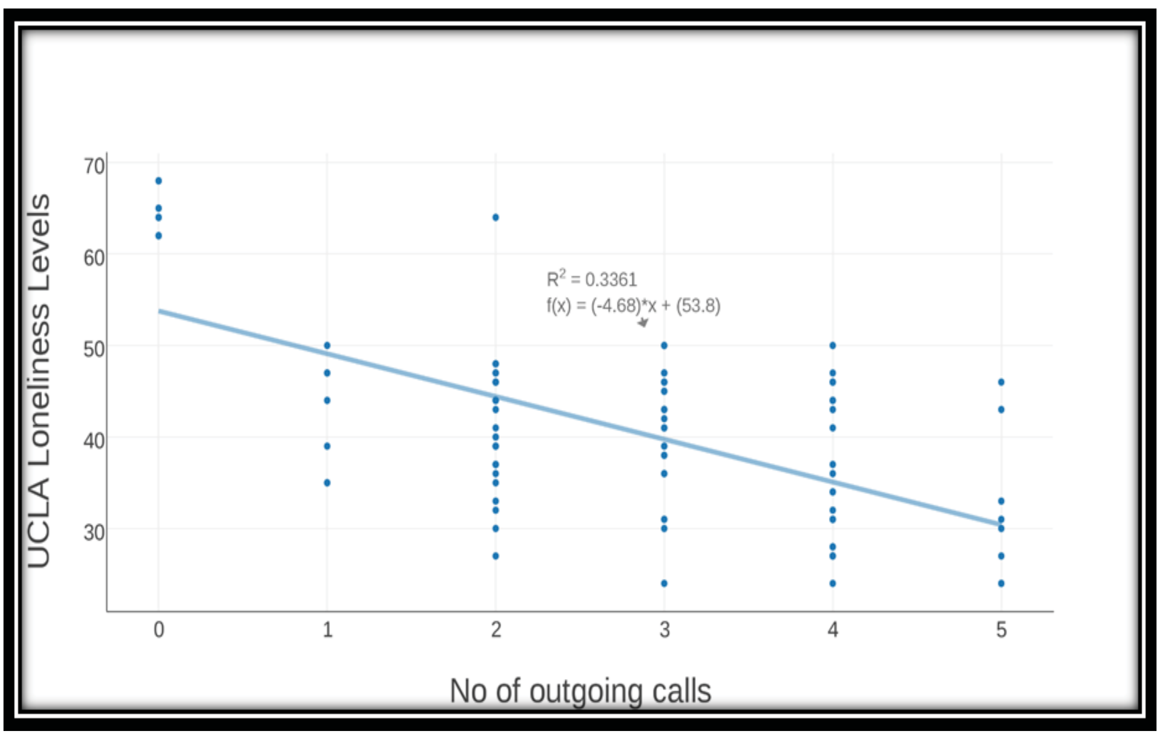

Figure 33. Number of outgoing calls vs. UCLA Loneliness Levels

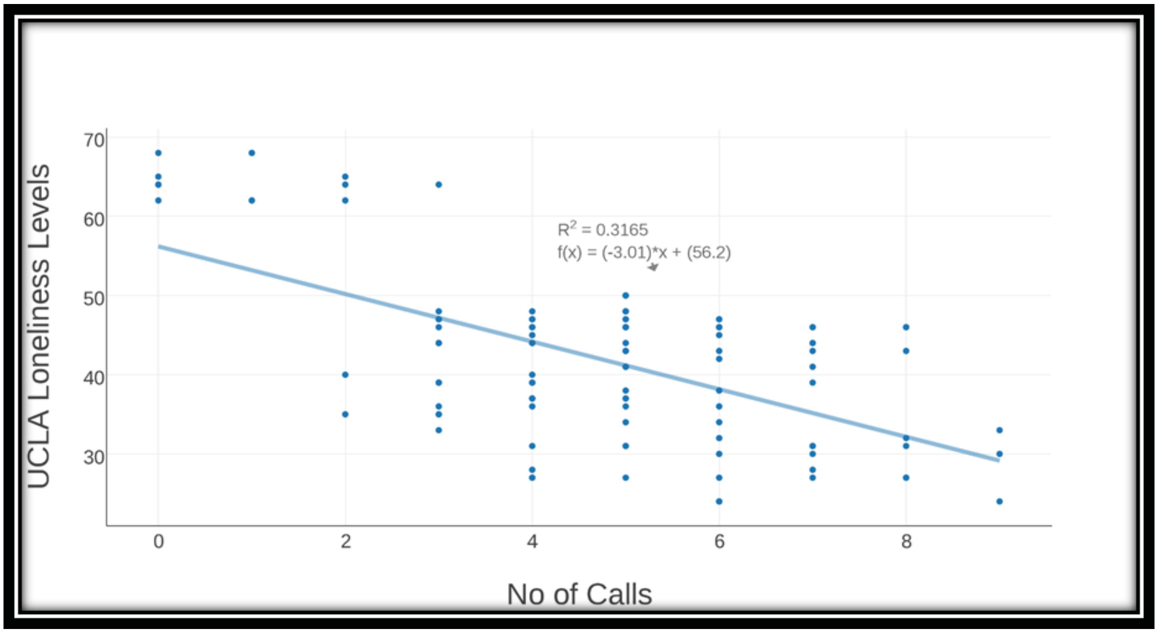

Figure 34. Number of calls vs. UCLA Loneliness Levels

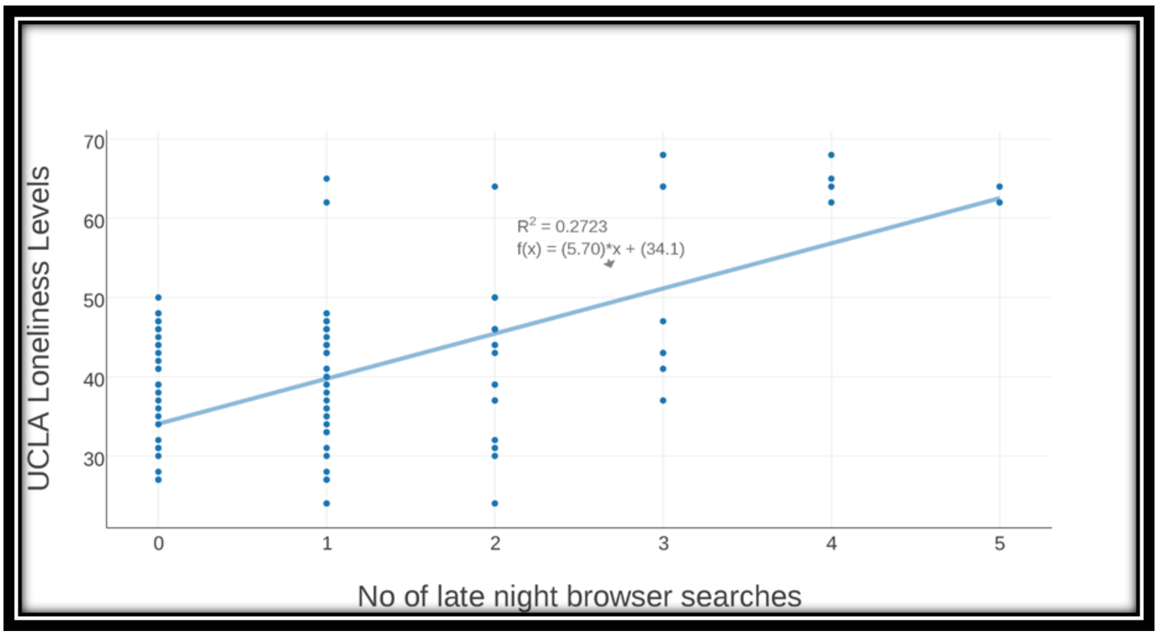

Figure 35. Number of late night browser searches vs. UCLA Loneliness Levels

We calculated the correlation coefficient, standard error of correlation coefficient, t-score followed by p-value for each of these features with UCLA Loneliness Levels as described in the analysis section. The results of this analysis is summarized in Table 9.

The significant factors are listed first, followed by the non-significant ones, in the decreasing order of their absolute value of correlation coefficient.

On analysis, we conclude that features like number of messages, number of outgoing calls, number of calls, number of short incoming or outgoing calls, number of late night searches, number of long incoming or outgoing calls, moving travel state, number of contacts, number of incoming calls, difference between number of outgoing and incoming messages, difference between no of outgoing and incoming calls, percentage of calls that are missed and number of auto-joined trusted Wi-Fi SSIDS are the features which are strongly correlated to UCLA Loneliness Levels.

	Correlation Coefficient	Standard Error of Correlation Coefficient	T score	p-value	Is Significant	Is novel feature
No of messages	-0.793	0.05468	-14.5025	< 0.00001	Yes	
No of outgoing calls	-0.681	0.0657	-10.3653	< 0.00001	Yes	
No of Calls	-0.626	0.07003	-8.939	< 0.00001	Yes	
No of short incoming or outgoing calls	-0.548	0.07511	-7.296	< 0.00001	Yes	Yes
No of late night browser searches	0.51	0.0772	6.6062	0.00001	Yes	Yes
No of long incoming or outgoing calls	-0.448	0.08029	-5.5798	< 0.00001	Yes	Yes
Moving Travel State	-0.412	0.08183	-5.0343	< .00001	Yes	
No of Contacts	-0.386	0.08284	-4.6596	< .00001	Yes	
No of incoming calls	-0.363	0.083677	-4.3381	0.000029	Yes	
Difference between no of outgoing and incoming messages	-0.338	0.084504	-4.004544	0.000107	Yes	
Difference between no of outgoing and incoming calls	-0.327	0.08486	-3.8534	0.000186	Yes	
Percentage of calls that are missed	0.326	0.08489	3.84262	0.000193	Yes	Yes
No of auto-joined trusted Wi-Fi SSIDS	-0.297	0.08575	-3.4635	0.000734	Yes	Yes
No of missed calls	0.162	0.0886	1.82844	0.069893	No	
No of browser searches	0.078	0.08953	0.8712	0.3853	No	
Low Activity Level	0.022	0.08978	0.245	0.80686	No	
High Activity Level	0.001	0.0898	0.0111	0.99116	No	

Table 9. Correlation coefficient and p-values of factors with UCLA Loneliness Levels

4.4. Results of Classification

Next using the significant features and personality traits, we performed classification to classify UCLA Loneliness Levels. First, we ran classifiers for classifying loneliness levels using only the significant features. Then using the significant features and the most significant personality traits, we performed two layer classification. We first classified the personality traits using the feature values, and then classified the loneliness level using the feature values and the results of the personality trait classification combined together.

As mentioned in sections 3.3.2 UCLA Loneliness levels range from 20 to 80. Users with have loneliness levels have values above 60. Average values fall in the range 40-60 and the most content, at peace users have loneliness levels below 40. Our aim is to correctly distinguish in which of these ranges the user falls in, and not the precise value. Therefore, we used three buckets namely, class 1: 20-40, class 2: 40-60 and class 3: 60-80 for classifying the UCLA Loneliness Levels.

The accuracy of the classifiers without taking personality traits into account and by taking personality traits into account is summarized in Table 10 and Table 13 in decreasing order of accuracy. J48 classifier has the most optimal accuracy at 90%, followed by Random Forest at 86% when personality traits are not take into consideration. This accuracy is improved to 98% and 94% for J48 and Random Forest respectively when personality traits are take into consideration. Table 11 and Table 14 gives detailed accuracy of the classifiers by class in decreasing order of accuracy without taking personality traits into account and by taking personality traits into account respectively. Table 12 and Table 15 gives the confusion matrix for J48, the best performing classifier in both the cases. Thus we can see that the classifier accuracy is significantly improved by taking personality traits into consideration.

	Correctly Classified Instances	**Incorrectly Classified Instances**	**Kappa statistic**
J48	90%	10%	0.8258
Random Forest	86%	14%	0.7611
Bayes Net	88%	12%	0.7973
AdaBoostM1	88%	12%	0.7973
Naive Bayes	80%	20%	0.6479

Table 10. Summary of Classifier Accuracy without Considering Personality Traits

		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
J48		0.8	0.033	0.941	0.8	0.865	0.883	1
		0.96	0.16	0.857	0.96	0.906	0.9	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.86	0.107	0.868	0.86	0.86	0.97	
Bayes Net		1	0.2	0.769	1	0.87	0.955	1
		0.76	0	1	0.76	0.864	0.957	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.88	0.08	0.908	0.88	0.88	0.96	
Random Forest		0.9	0.167	0.783	0.9	0.837	0.965	1
		0.8	0.08	0.909	0.8	0.851	0.968	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.86	0.107	0.868	0.86	0.86	0.97	
AdaBoostM1		1	0.2	0.769	1	0.87	0.9	1
		0.76	0	1	0.76	0.864	0.904	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.88	0.08	0.908	0.88	0.88	0.912	
Naive Bayes		0.6	0.067	0.857	0.6	0.706	0.89	1
		0.92	0.32	0.742	0.92	0.821	0.894	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.8	0.187	0.814	0.8	0.793	0.903	

Table 11. Detailed Accuracy of Classifiers by Class without Considering Personality Traits

A	B	C	<-- CLASSIFIED AS
16	4	0	a=1
1	24	0	b=2
0	0	5	c=3

Table 12. Confusion Matrix for J48 without Considering Personality Traits

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic
J48	98%	2%	0.8258
Random Forest	94%	6%	0.7611
Bayes Net	94%	6%	0.7973
AdaBoostM1	92%	8%	0.7973
Naive Bayes	82%	18%	0.6479

Table 13. Summary of Classifier Accuracy by Considering Personality Traits

		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
J48		0.95	0	1	0.95	0.974	0.975	1
		1	0.04	0.962	1	0.98	0.98	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.98	0.02	0.981	0.98	0.98	0.98	
Bayes Net		0.95	0.0067	0.905	0.95	0.927	0.99	1
		0.92	0.04	0.958	0.92	0.939	0.99	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.94	0.047	0.941	0.94	0.94	0.991	
Random Forest		0.95	0.067	0.905	0.95	0.927	0.992	1
		0.92	0.04	0.958	0.92	0.939	0.992	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.94	0.047	-.941	0.94	0.94	0.993	
AdaBoostM1		0.8	0	1	0.8	0.889	0.98	1
		1	0.16	0.862	1	0.926	0.981	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.92	0.08	0.931	0.92	0.919	0.982	
Naive Bayes		0.65	0.067	0.867	0.65	0.743	0.913	1
		0.92	0.28	0.767	0.92	0.836	0.917	2
		1	0	1	1	1	1	3
	Weighted Avg.	0.82	0.167	0.83	0.82	0.815	0.924	

Table 14. Detailed Accuracy of Classifiers by Class by Considering Personality Traits

A	B	C	<-- CLASSIFIED AS
19	1	0	a=1
0	25	0	b=2
0	0	5	c=3

Table 15. Confusion Matrix for J48 by Considering Personality Traits

5. Discussion

We now discuss a few issues faced in the duration of this research.

Difficulty recruiting subjects: As mentioned in the previous sections, one of the key issue faced in this thesis was recruiting user subjects. Given the sensitivity of the data collected, we had a hard time convincing people to let us monitor their smartphone activity for a duration of two weeks. Despite of the two layer encryption, one-way hashed private data like message content, and anonymous surveys, it was difficult to recruit subjects without any other incentive.

Frequency of administering Loneliness Scale: Another key factor to consider for future researchers is whether loneliness questionnaire should be administered daily or on a less frequent basis since loneliness is a slow-changing factor.

Reinstallation issue: Two technical issues were encountered while administering the pilot study app through Funf in a Box [50]. Funf uses the device ID for mapping and uniquely identifying the records of each device. This device ID is derived from Utils.getInstallationId(context). The installation ID changes every time one uninstalls and reinstalls an app. Therefore the Device ID changes on reinstallation. During the user data, such an issue was encountered when the user had to uninstall the app for a couple of days and reinstall it back after a few days to resume the study, which lead to records with different Device IDs mapping to the same user. We had to manually identify the corresponding Device IDs using the SIM Serial ID to tackle the issue.

Timestamp issue: Another issue encountered was in timestamp field, where in a future timestamp was shown. The source of this issue lies in the way the particular

Android device measures time either using System.currentTimeMillis(), uptimeMillis() or using elapsedRealtime() and elapsedRealtimeNanos().

Scope for more features: Additionally, we would like to highlight the features we could not take into consideration given the time constraint, but should be considered for future work. Many of the target international students communicated with their family using web messaging and VOIP services like WhatsApp, Hike and Skype. The usage of these would help improve the classifier accuracy.

Weighing false positives and true positives: Moreover, the final app needs to be more specific in its target audience to choose the accuracy of false positives or false negatives. An app customized for a therapist could render more false positives at the cost of reduced false negatives. Similarly an app customized for a therapist’s patients or normal users trying to monitor his social isolation should not render more false positives at the cost of reduced false negatives. Cost sensitive learning could be used to tune the ratio of false positives to true positives.

6. Future Work

In this thesis, we completed the user study, research and investigation, a preliminary app, and a model to develop the final Socialoscope app. A future researcher, could close the development of the Socialoscope app by putting the classifier into the app. Moreover, the personality and loneliness surveys can be added to the Socialoscope app for the user study. This ensures the user subjects have all the required tools for the study at one place. Furthermore, the loneliness survey can administered to the users as prompts at the end of the day.

Additionally, more features can be added to the list of features tracked by Socialoscope. Features tracking the audio and video call counts of VOIP services such as Skype, FaceTime and Viber can be investigated. On similar lines, message interaction on web messaging services such as WhatsApp, Hike, WeChat, and Google Hangouts can be explored. Due to the increasing quality and cost-effectiveness of such VoIP and web messaging services, many people have shifted their regular usage to such services. Moreover, many international students prefer such services for connecting with their friends and family staying in different countries and continents.

Furthermore, if the entire data gathering can be done by relaxing the encryption layer a bit, thereby having access to the contact favorite, message character count, social media post counts, browser URLs. Additionally, text analysis can also be performed on social media posts and messages. This will make the analysis of the study more robust and accurate. At the same time, this increases the privacy and security issue, thereby increasing the difficulty recruiting subjects for the user study and users for the final app.

Therefore, a balance between the privacy and security issue and accuracy of the app must be maintained as per the use case.

Due to the privacy constraint of the data monitored during the user study, we faced a hard time recruiting more user subjects. The accuracy of the classifiers can be further improved with data gathered from more user subjects, and from a diverse background including older adults.

Another key factor that can be considered for future researchers is the frequency of administering loneliness questionnaire. Since loneliness is a slow-changing factor, should the loneliness questionnaire be administered daily or on a less frequent basis? Furthermore, investigation on whether administering questionnaire loneliness on a daily basis distorts the results of the survey should be explored.

This app aims at detecting loneliness levels amongst smartphones users based on the social interactions and activities they have on their smartphones. This app thus bridges the technological gap by detecting loneliness levels using mobile sensing. The results of this app, at present is a loneliness level. A finished app, should ideally, convert this loneliness score into psychological feedback, like encouragement messages, alerts to psychotherapists and family members, if required. This feedback must be carefully designed with the help of a team with the necessary medical and psychological background. Telling a socially healthy user that he is social healthy is not a difficult task, but determining a constructive way to give a beneficial feedback to a lonely user and alerting his psychotherapists, friend and/or family members, if necessary should be further explored.

7. Conclusion

As loneliness increases at an alarming rate in the modern times, it is critical to devise easy, user-friendly ways to detect, monitor, and tackle the social isolation. A smartphone that can passively monitor users' social interactions and day-to-day activity and then detect his loneliness levels can support a wide range of interventions. This thesis analyzes steps to research and develop Socialoscope, an Android app that can passively monitors the day-to-day social interactions, communication and smartphone activity sensed by the smartphone's built-in sensors in terms of calls, messages, social media usage, Wi-Fi devices, emails and browsing. Previously discovered relationships between personality and loneliness were explored. The app will be especially useful to disproportionality affected groups like international students who keep their smartphones at a close proximity and have busy schedules that cannot permit them to go out of their way to monitor or ponder on their social wellness or loneliness levels. The app will also be useful to old people, who face challenges in monitoring themselves and using smartphones (since the app monitors autonomously). The Socialoscope app can be integrated into the healthcare system as an early warning indicator of patients requiring intervention or utilized for personal self-reflection. It will directly impact smartphone social health monitoring.

8. References

[1]. N. D. Lane, M. Mohammod, M Lin, X Yang, H Lu, S Ali, A Doryab, E Berke, T Choudhury, A T. Campbell, “BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing", in Proc. Pervasive Health Conference 2011, May 2011, pp. 23-26.

[2]. H Lu, W Pan, N D. Lane, T Choudhury and A T. Campbell, “SoundSense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones”. In Proc MobiSys 2009 (2009), 165-178.

[3]. N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure Using Mobile Phone Data”, in Proc National Academy of Sciences (PNAS) Vol. 106(36), pp. 15274-15278.

[4]. M C Gonzalez, C Hidalgo and A Barabasi, “Understanding individual human mobility patterns”, Nature 453 (7196), 779-782.

[5]. D. Olguin, P. Gloor, and A. Pentland, “Capturing Individual and Group Behavior Using Wearable Sensors”, in Proc AAAI Spring Symposium on Human Behavior Modeling (2009), Palo Alto.

[6]. T. Abdelzaher, Y. Anokwa, P. Boda, J. Burke, D. Estrin, L. Guibas, A. Kansal, S. Madden, and J. Reich (2007), “Mobiscopes for human spaces”, in IEEE Pervasive Computing, pp 20-29.

[7]. S. Avancha, A. Baxi, and D. Kotz, “Privacy in mobile technology for personal healthcare”, ACM Computing Surveys (2012).

[8]. N. Christakis and J. Fowler, “The spread of obesity in a large social network over 32 years”, New England Journal of Medicine (2007), 357(4):370

[9]. N. Christakis and J. Fowler, “The collective dynamics of smoking in a large social network”, New England Journal of Medicine (2008), 358(21):2249.’

[10]. J. Fowler and N. Christakis, “Dynamic Spread of Happiness in a Large Social Network: longitudinal analysis over 20 years” in Framingham Heart Study British Medical Journal.

[11]. K. George, D.G. Blazer, D.C. Hughes, and N. Fowler, “Social Support and the Outcome of Major Depression”, in British Journal of Psychiatry (1989), vol. 154. No. 4, pp 478.

[12]. George Forman, “An extensive empirical study of feature selection metrics for text classification”, Journal of Machine Learning Research 3 (2003), 1289-1305.

[13]. "AudioRecord Android Media Recording", http:// developer.android.com/reference/android/media/AudioRecord.html

[14]. M Hojat, "Loneliness as a function of selected personality variables", Journal of Clinical Psychology, Volume 38, Issue 1, pages 137–141, January 1982.

[15]. "Android Service Component", http://developer. android.com/guide/ components/services.html

[16]. T Choudhury, A Pentland, "Characterizing Social Networks using the Sociometer", in Proc NAACOS 2004.

[17]. T Choudhury, A Pentland, "The Sociometer: A Wearable Device for Understanding Human Networks", In Proc CSCW 2012.

[18]. A Ghose, C Bhaumik and T Chakravarty, "BlueEye - A system for Proximity Detection Using Bluetooth on Mobile Phones", in Proc UbiComp 2013.

[19]. K Rachuri, C Mascolo, M Musolesi, P Rentfrow, "SocialableSense: Exploring the Trade-Off of Adaptive Sampling and Computation Offloading for Social Sensing", in Proc MobiCom 2011.

[20]. R Wang, F Chen, Z Chen, T Li, G Harari, S Tignor, X Zhou, D Ben-Zeev, and A T. Campbell, "StudentLife: Assessing Mental Health, Academic Performance and Behavioural Trends of College Students using Smartphones", in Proc Ubicomp 2014.

[21]. D Russell, "UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure", in Journal of Personality Assessment (1996), 66(1):20-40.

[22]. A Ong, B Uchino and E Wethington, "Loneliness and the health of older people" in Gerontology.

[23]. "Determining if one is a social butterfly", Unpublished manuscript.

[24]. W Lane, C Manner, "The Impact of Personality Traits on Smartphone Ownership and Use", Int'l Journal Business and Social Science, Vol. 2 No. 17.

[25]. G Chittaranjan, J BlomDaniel, and Gatica-Perez "Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones", in Proc ISWC 2011, Washington, DC, USA.

[26]. "Funf Sensing Framework", https://code.google.com/p/funf-open-sensing-framework/source/checkout

[27]. Ohmage open mobile data collection platform, http://ohmage.org

[28]. “5 Medical Technologies Revolutionizing Healthcare” by Forbes (2013), http://www.forbes.com/sites/stevenkotler/2013/12/19/5-medical-technologies-revolutionizing-healthcare/2

[29]. “Top 10 Medical Gadgets” by Technology Personalized” (2012), http://techpp.com/2012/03/26/ top-medical-gadgets

[30]. D Howard, “Effect of Temperature on the Intracellular Growth of Histoplasma Capsulatum, J Bacteriol”, 1967, Jan; 93 (1): 438-444.

[31]. “Depression Toolkit”, by University of Michigan Depression Center http://www.depressiontoolkit.org/aboutyourdiagnosis/depression.asp

[32]. “The Loneliness of American Society” by The American Spectator, http://spectator.org/articles/ 59230/loneliness-american-society

[33]. “Campaign to End Loneliness”, http://www.campaign toendloneliness.org

[34]. “Mind for Better Mental Health”, http://www.mind. org.uk/information-support/tips-for-everyday-living/loneliness/about-loneliness

[35]. “Psychologist Anywhere Anytime”, "http://www. psychologistanywhereanytime.com/relationships_psychologist/psychologist_loneliness.htm

[36]. “Thought Catalog”, http://thoughtcatalog.com/lorenzo-jensen-iii/2015/03/36-absolutely-heartbreaking-quotes-about-loneliness/

[37]. “Why have men become so lonely – and how does it affect their health?”, by Newstatesman, http://www.newstatesman.com/2015/11/why-have-men-become-so-lonely-and-how-does-it-affect-their-health

[38]. Steven W. Colea, John P. Capitanio Katie Chun, Jesusa M. G. Arevalo, Jeffrey Ma and John T. Cacioppo (2015), “Myeloid differentiation architecture of leukocyte transcriptome dynamics in perceived social isolation”, in PNAS.

[39]. “Loneliness: Number of baby boomers living alone soars by nearly a quarter since 2005” by Independent, http://www.independent.co.uk/news/uk/home-news/loneliness-number-of-baby-boomers-living-alone-soars-by-nearly-a-quarter-since-2005-a6722981.html

[40]. “Is modern life making us lonely?” by BBC News, http://www.independent.co.uk/life-style/health-and-families/features/the-loneliness-epidemic-more-connected-than-ever-but-feeling-more-alone-10143206.html

[41]. “Online networking 'harms health” by BBC News ,http://news.bbc.co.uk/2/hi/7898510.stm

[42]. “People living alone 'are more depressed”, by BBC News, http://www.bbc.com/news/health-17475240

[43]. “Many elderly isolated from far-away family, charity says” by BBC News, http://www.bbc.com/news/education-20664760

[44]. “Young more lonely than the old, UK survey suggests” by BBC News ,http://news.bbc.co.uk/2/hi/health/8701763.stm

[45]. “The loneliness epidemic: We're more connected than ever - but are we feeling more alone?” by Independent, http://www.independent.co.uk/life-style/health-and-families/features/the-loneliness-epidemic-more-connected-than-ever-but-feeling-more-alone-10143206.html

[46]. “Loneliness 'increases risk of premature death” by NHS, http://www.nhs.uk/news/2015/03March/Pages/Loneliness-increases-risk-of-premature-death.aspx

[47]. “Connect for mental wellbeing” by NHS, http://www.nhs.uk/Conditions/stress-anxiety-depression/Pages/Connect-for-mental-wellbeing.aspx

[48]. “Why Loneliness May Be the Next Big Public-Health Issue” by Time, http://time.com/3747784/loneliness-mortality

[49]. “Aloneness versus Loneliness” by Dr. Lawrence Wilson in The Center For Development, http://drlwilson.com/ARTICLES/LONELINESS.htm

[50]. “Funf-In-A-Box”, http://inabox.funf.org

[51]. “WPI Qualtrics”, https://wpi.qualtrics.com

[52]. “Feature Selection” by Wikipedia, https://en.wikipedia.org/wiki/Feature_selection

[53]. “Mood Scope”, https://www.moodscope.com/

[54]. “Mappiness”, http://www.mappiness.org.uk/more/

[55]. “AutoEmotive”, http://autoemotive.media.mit.edu/

[56]. S. D. Gosling, P. J. Rentfrow, and W. Swann. A very brief measure of the big-five personality domains. Journal. of Research in Personality, 37:504 – 528, 2003

[57]. “Correlation” by B. Burt Gerstman, http://www.sjsu.edu/faculty/gerstman/StatPrimer/correlation.pdf

[58]. “Big-Five Personality Trait Questions” by IPIP, http://ipip.ori.org/newBigFive5broadKey.htm

[59]. Petersen J, Austin D, Kaye JA, Pavel M, Hayes TL. “Unobtrusive in-home detection of time spent out-of-home with applications to loneliness and physical activity” in IEEE journal of biomedical and health informatics. 2014;18(5):1590-1596. doi:10.1109/JBHI.2013.2294276.

[60]. Sanchez, Wendy, Martinez, Alicia, Campos, Wilfrido, Estrada, Hugo, Pelechano, Vicente “Inferring loneliness levels in older adults from smartphones” in Journal of Ambient Intelligence and Smart Environments, vol. 7, no. 1, pp. 85-98, 2015.

[61]. “Trait Theory of Personality - The Trait Approach to Personality”, www.verywell.com/trait-theory-of-personality-2795955