BACHELOR OF SCIENCE IN
COMPUTER SCIENCE AND ENGINEERING

Inspiring Excellence

An experiential investigation on Internet addiction and specific disorders like Depression, Self-esteem and Introversion among University Students

AUTHORS
Nadia Rubaiyat
Anika Islam
Lamiah Israt
Lamiya Kabir

SUPERVISOR
Hossain Arif
Assistant Professor
Department of CSE

A thesis submitted to the Department of CSE
in partial fulfillment of the requirements for the degree of
B.Sc. Engineering in CSE

Department of Computer Science and Engineering
BRAC University, Dhaka - 1212, Bangladesh

December 2018

Declaration

We, hereby declare that the content of this thesis is the result of work which has been carried out by the authors alone.All the guidelines have been followed properly while preparing this and materials of work from researches conducted by others are mentioned in references.

Authors:

Nadia Rubaiyat
Student ID: 15101079

Anika Islam
Student ID: 15101082

Lamiah Israt
Student ID: 15101088

Lamiya Kabir
Student ID: 15101124

Supervisor:

Hossain Arif
Assistant Professor, Department of Computer Science and Engineering
BRAC University

December 2018

The thesis titled An experiential investigation on Internet addiction and specific conditions like Depression, Self-esteem and Introversion among University Students
Submitted by:
Nadia Rubaiyat Student ID: 15101079
Anika Islam Student ID: 15101082
Lamiah Israt Student ID: 15101088
Lamiya Kabir Student ID: 15101124
of Academic Year 2018 has been found as satisfactory and accepted as partial fulfillment of the requirement for the Degree of B.Sc. Engineering in CSE

1. Hossain Arif
Assistant Professor
BRAC University

2. Dipankar Chaki
Lecturer
BRAC University

3. Md. Abdul Mottalib
Professor and Chairperson
BRAC University

Acknowledgements

We would like to express our sincere gratitude to our supervisor Hossain Arif and co-supervisor Dipankar Chaki for their utmost attention and valuable time. We would also like to thank them for giving us the opportunity to work on this topic and guiding us throughout the process.

Abstract

The tormenting hurdle that is slowly isolating the youth of this generation from being leaders of tomorrow, is depression. In recent times, students have been observed to indulge in compulsive use of Internet which has a relation to the massive upsurge of depressed individuals among the youth. Hence, the main and foremost objective of this research is to find correlation among the leading disorders which are Internet addiction, depression, self-esteem and introversion which can diminish these predicaments. In order to serve this purpose, 461 undergraduate students have been selected arbitrarily from several institutions of Bangladesh and were solicited to complete a standard questionnaire which was prepared based on the self-reported measures concerning the disorders mentioned before. The correlational survey design which was employed through social media contains Internet Addiction Test by Dr. Kimberly Young, Rosenberg Self-Esteem Scale by M. Rosenberg, PROMIS Emotional Distress-Depression short scale and Introversion scale by McCroskey. In pursuit of extracting desired results from the survey as well as to check accuracy, numerous methods have been utilized. Cronbach alpha provided the proof that the data retrieved from the survey is consistent. Subsequently, Chi-square test and ANOVA procured that positive correlations exists among internet addiction, depression, self-esteem and introversion. To corroborate this newly established correlation, multifarious machine learning techniques have been adopted and experimented. These experiments revealed that Internet addiction can predict self-esteem, depression can predict self-esteem and depression can predict introversion. After applying these methods, we came to the conclusion that there exists correlation among the psychological syndromes mentioned above and that this correlation can be exploited and manoeuvred to a positive outcome whose aim will be to reduce the severity of these disorders.

Table of contents

List of figures

List of tables

Chapter 1

Overview

1.1 Introduction

The increasing impact of communication media especially the internet has a significant effect on people and their day to day life. Nowadays, it is observed that every aspect of people's lives is being affected profoundly by the internet. According to Lukeoff (2004), certain factors grow people's tendency to the Internet. They want to connect comfortably, freely to establish an identity, and to have significant relationships with others. This study observes the prevalence of factors such as self-esteem, personality traits and environmental determinants like parental behaviour on internet addiction. People having psychological instability and issues with impulsiveness are more likely to be addicted to the internet as per their findings [1]. Although our life is made easy and straightforward through the internet, it also brings a considerable risk of causing depression, introversion and self-esteem among students [2], [3]. An individual's beliefs and perceptions about themselves reflect in their behaviour and actions while using the Internet. It is shown in studies that Internet addiction ratings are positively interrelated with stress and depression scores, and those who are more addicted to the internet are likely to be depressed, have low self-esteem and act like introverts. Depression is considered as one of the leading mental issues in the current social environment [3]. Internet addiction has been experienced by roughly 1.4-20.8% of teenagers and 8-13% of college students in their lifetime [2].

Depression or major depressive disorder is a common drawback in human psychology that leads to cynical effects in our behaviour, mood, thought process and finally our health. Depression is not as simple as feeling low for a day or two. The persistent feeling of being nonchalant about any kind of social, professional or even personal activity for a long period of time, leads to serve depression. It corrupts a person's mind and damages them from the inside. To such extent that, in 2016 World Health Organisation (WHO) in 2016 performed a

detailed study that shows that around 350 million people from all around the world are being affected by depression [4]. Apart from basic behavioural change and short term emotional reactions, depression becomes a very threatening and major health problem especially when intense depressive feelings are shown for long period of time. Depression can degrade a person's ability to work efficiently, his response to others, relationships and financial losses. This psychological disorder influences individuals' capability to think, sleep, feel, eat, work and study [4]. If categorised according to the scale of age, depression occurs more among the youth, while according to gender categorisation, it is more common among women than man [5]. Emotional vulnerability, isolation and lack of proper counselling have a pivotal role in causing depression.

Many factors affect depression among which self-esteem is a major one. It reveals an individuals' confidence towards themselves [4]. Self-esteem reflects one's appreciation and approval of own characteristics through individuals' self-assessment. People who tend to have a low appreciation towards them and feel they have lack of qualities to do anything, also view themselves insufficient against their surroundings are considered to have low self-esteem [4]. Reports display that self-esteem explains 38% of internet addiction with contentment and remoteness, and correlate with time allocation troubles, interpersonal communications and health issues in internet addiction. Also different stages of self-esteem leads to self-disbelief, addicted personality, consciousness of loss of self control and sense of failure[3]. Vulnerability model is built to explain the connection between depression and self-esteem. As depicted in this model, low self-esteem is a crucial factor for depression in the upcoming days [4]. Besides, Internet addiction is found strongly related with introversion, depression and self-esteem and these factors must be considered in the prevention and intervention programs for people with attention deficit and hyperactivity issues [3].Another study has found that Internet addiction and self-esteem are negatively associated and they have used "Internet Addiction Scale (IAS)" and "Coopersmith Self- Esteem Scale" for finding correlation among them [6]. However, some researches have also shown that Internet addiction has a positive relationship with extroversion. . In these studies, they approached with the fact that extroverts feel more need of communicating to express their feelings whereas introverts mainly use the Internet because by using this they can reduce their fear of being rejected [7].

According to the studies mentioned before, depression, self-confidence, introversion and parental behaviour can be predictive of extreme internet usage illustrating some hypothesis like introversion will be positively related to internet addiction, high self-esteem will be negatively connected with internet addiction, depression will be positively correlated with internet addiction and depression will be positively interconnected with self-esteem.

The primary objective of this research is to provide a fundamental review of any inter-

connection between Internet addiction, self-esteem, depression and introversion by using different kinds of measures. Additionally, the study will provide insights on whether Internet addiction can be predicted using depression, self-esteem and introversion features and also try to gain the most accurate result with the help of different machine learning algorithms. Although there remains much confusion about whether these variables have positive or negative relationship between them, previous studies have proven that these factors have a significant impact on social and personal life and the correlations can be a way of predicting the possible threats from these factors in an individual's life.

1.2 Literature Review

Several research efforts have been directed towards the everyday use of the internet and variables that are directly or indirectly related with this. In one of the study, a significant relationship is found between self-esteem and Internet addiction. This research shows that in the way of dealing with Internet addiction, self-respect and psychopathology can be considered as important factors. They have used Internet Addiction Scale (IAS), Symptom checklist (SCL) and Rosenberg self-esteem scale and they discovered that self-esteem decreases as Internet addiction increases [2]. Additionally, unique relationships have been seen among depression, self-esteem, and regular internet use. The results show that depression is positively but weakly related to internet addiction and it is moderately but negatively associated with self-esteem. The Tested model that is used in this paper, can predict 28 percent of the depression among adolescents. It also shows that daily internet use and self-confidence can affect depression significantly. Moreover, depression failed in predicting the low self-esteem levels of adolescents, contrarily low self-esteem lead the individuals to depression [3]. In another case-study, Kruskal-Wallis test, Spearman's correlation and Mann-Whitney U test analysis are used to correlate Internet addiction with depression of the university students and found the mean score of 8.28 of the students on the Internet Addiction Scale. Their finding is a significant positive relation between internet addiction and depression with the Spearman's correlation coefficient (rs) of 0.804. The reseachers of one study stated the relationship between anxiety disorders and Internet Addiction through a correlation coefficient which shows the fact that behaviours related to anxiety have an effect on Internet usage [8]. Also, another study revealed that the young generation is more prone to use the internet. Their survey among 217 students of a regional university in Australia reveals that maturity and self-ability to produce intended result do not affect internet dependency [9]. Moreover, individuals can have virtual social assistance from the internet along with the real social support that they get from families and friends. This study shows that both male and female has a risk of greater

depressive symptoms if they get more virtual social support than the actual social assistance. Also, the research indicates that adults with Internet Addiction Disorder appear to be more disposed to aggression than those who are non addicted [10]. Also self-esteem is considered as an predecessor to Internet addiction which leads to determining individuals behaviours and activities [4].

One of the investigations claims that Internet Addiction affects stress, anxiety, and depression directly. Students with more Internet addiction are more likely to emotional or mental breakdown attack to stress, anxiety, and depression [11]. Another study shows that having depression co-morbid with another chronic physical disease lowers health state substantially, despite of a respondent's age, sex, and other demographic variables. Adjusting the socioeconomic factors and health conditions it shows that depression had the largest effect on decreasing mean health scores compared with the other chronic conditions.[12].

Self-esteem is also an essential factor, and several studies have shown a significant impact of self-esteem in Internet addiction. As an example, low self-esteem may exaggerate the exposure of Internet addiction. One of the studies has worked with psychological inflexibility/experiential avoidance and stress coping measures for excessive internet use. The results confirmed that less effective stress coping strategy predicted a high risk of significant depression and suicidality. Also, The result of the mentioned study indicates that PI/EA at the beginning enhanced the risk of Internet addiction (OR = 1.087, 95% CI: 1.042–1.135), suicidality (OR = 1.099, 95% CI: 1.053–1.147) and depression (OR = 1.125, 95% CI: 1.081–1.170), where OR means the odds ratio and CI indicates the confidence interval. Students who have high psychological inflexibility/experiential avoidance should be the target of prevention programs for Internet addiction, depression and suicidality [13]. One study tells that those who uses internet more than one times per day were less likely to be addicted to the Internet than other people. In addition, duration is an important factor in Internet addiction and spending longer periods of time on the Internet allows individuals to explore. As a result, this may have more effects on psychological engagement and motivation. An important factor in the adolescents' internet addiction is the behaviour of parents towards their offspring. Negligence and over-protectiveness of parents can cause a child to be more lenient and involved with the internet and thus become addicted to it. Researches show that we can explain the variance of Internet addiction with the analysis of self-esteem and depression [1]. Moreover, studies reveal that personal attributes can explain around 8.6% prospect of internet addiction. This case study discovers that internet addiction is negatively associated with introversion and has a strong relationship with outward attitude and openness. Those who like to communicate with others more are further addicted to the internet [7]. Results of a study indicate that 40.7% of the students that they examined have an

Internet addiction.Among which 2.2% had severe addiction and 38.5% had moderate Internet addiction. A significant connection observed between depression, internet addiction and self-esteem. Additionally, they have used regression analysis which reveals that depression and self-esteem are capable of predicting the variance of daily internet use [14]. Only openness to experience is found to be statistically significant in the analysis (probability value, $p < .050$). Also, the results reveal that male university students experience more internet addiction and depression than the female university students and internet addiction and depression are strongly and positively correlated with each other [15]. On the contrary, researches conducted by Niemz and colleagues and Sanders and colleagues resulted in no significant relationship between depression and internet addiction. Their findings show that there is no dependency of internet addiction on depression, stress and anxiety [16].

Chapter 2

Methods & System Structure

2.1 Dataset and preprocessing

2.1.1 Dataset

For research purpose, the dataset was created by collecting the answers of 461 undergraduate students who were chosen arbitrarily using various social medias and networking sites. They were asked to partake in a 61-item survey measuring the levels of depression, internet addiction, self-esteem and introversion. The responses from Individuals between the age of 20 and 25 were used to select the participants. The participants were chosen based on their will to provide rejoinders to the questionnaire who are residing within Dhaka city. The responses of those contributors were gauged using demographic data, Internet Addiction Scale (IAT) [17] by Dr. Kimberly Young, PROMIS Emotional Distress-Depression short form [18] by PROMIS Health Organization, Introversion Scale [19] by James McCroskey and Rosenberg Self-esteem Scale [20]. The survey contains 61 questions and data collection was utterly anonymous. Cross-sectional survey methodology is utilized to assess depression, self-esteem, internet addiction and introversion level of those who partook in the survey. Table 2.1 depicts the demographic data of the respondents:

2.1.2 Preprocessing

The preprocessing of the dataset was done according to the scales explained below:

1. Internet Addiction Scale (IAT):
   The Internet Addiction Test measures the stage of web usage of an individual [21]. It contains 20 questions, which is further divided into subcategories (salience, excessive use, ignorance towards labor, anticipation, lack of restrain and negligence in social

Table 2.1 DEMOGRAPHIC PROFILE OF THE POPULATION

		number	%
Gender	Female	212	46.98
	Male	249	53.02
Year	1st year	115	24.95
	2nd year	64	13.88
	3rd year	92	19.96
	4th year	190	41.21
Age	20	1	0.22
	21	19	4.12
	22	56	12.15
	23	68	14.75
	24	33	7.16
	25	13	2.82
Current Employment	none	347	75.27
	part-time	105	22.78
	full-time	9	1.95
Sleeping Hours per day	less than 8 hours	285	61.92
	8 to 10 hours	167	36.22
	more than 10 hours	9	1.95

life). Based on self-assesed score, the scale provides a result ranging from 0 to 100. The higher score an individual achieves, the more addicted the person is. Those who acquire a score between 0-30 are normal internet user; scores in between 31-49 represents a mild level of Internet usage; 50-79 is considered a moderate level; and scores between 80-100 represents a severe level of dependency upon the internet. The question selected for this test is given below.

(a) How often do you find that you stay on-line longer than you intended?

(b) How often do you neglect household chores to spend more time on-line?

(c) How often do you prefer the excitement of the Internet to intimacy with your partner?

(d) How often do you form new relationships with fellow on-line users?

(e) How often do others in your life complain to you about the amount of time you spend on-line?

(f) How often do your grades or school work suffers because of the amount of time you spend on-line?

(g) How often do you check your email before something else that you need to do?

(h) How often does your job performance or productivity suffer because of the Internet?

(i) How often do you become defensive or secretive when anyone asks you what you do on-line?

(j) How often do you block out disturbing thoughts about your life with soothing thoughts of the Internet?

(k) How often do you find yourself anticipating when you will go on-line again?

(l) How often do you fear that life without the Internet would be boring, empty, and joyless?

(m) How often do you snap, yell, or act annoyed if someone bothers you while you are on-line?

(n) How often do you lose sleep due to late-night log-ins?

(o) How often do you feel preoccupied with the Internet when off-line, or fantasize about being on-line?

(p) How often do you find yourself saying “just a few more minutes” when online?

(q) How often do you try to cut down the amount of time you spend online and fail?

(r) How often do you try to hide how long you've been on-line?

(s) How often do you choose to spend more time on-line over going out with others?

(t) How often do you feel depressed, moody or nervous when you are off-line, which goes away once you are back on-line?

2. PROMIS Emotional Distress-Depression Short scale:
The PROMIS depression instrument includes an eight-item short questionnaire. It is shortened from the initial descriptive scale formed by PROMIS organization. At first, the raw scale is calculated by adding all the questions and then finding the corresponding t-score from the t-score table assess the depression level of an individual. Equation 2.1 shows the necessary formula in order to calculate t-score.

$$t = \frac{x - \mu}{s/\sqrt{n}} \tag{2.1}$$

Here,

- x means sample mean
- μ means total mean of population
- s represents the standard deviation
- n represents total sample size

The t-score is a type of a standardized test statistics. It works by taking an individual value and then translates it in a standardised format. The score range of 0-54.9 is marked as having none to slight depression, 55-59.9 is considered to have mild depression, 60-69.9 represents moderate level, and above 70 interprets as a high level of depression. The questionnaire to evaluate depression is given below.

(a) I felt worthless.

(b) I felt that I had nothing to look forward to.

(c) I felt helpless.

(d) I felt sad.

(e) I felt like a failure.

(f) I felt depressed.

(g) I felt unhappy.

(h) I felt hopeless.

3. Introversion Scale:
   James McCroskey developed introversion scale that can measure the social phobia of communication in an individual [20]. It contains 18-items with 5-Likert scoring. 6 of the questions are used for reducing redundancy and are not included in the calculation of introversion level. The score ranges from 12-60 with three categorizations: low, moderate and high. Low introversion ranges from 5-23, moderate introversion between 24-48 and high introversion considered for 49-60 [20]. The eighteen statement to assess introversion in given below. Participants are asked to imply how strongly they agree or disagree with the statements

   (a) Are you inclined to keep in the background on social occasions?

   (b) Do you like to mix socially with people?

   (c) Do you sometimes feel happy, sometimes depressed, without any apparent reason?

   (d) Are you inclined to limit your acquaintances to a select few?

   (e) Do you like to have many social engagements?

   (f) Do you have frequent ups and downs in mood, either with or without apparent cause?

   (g) Would you rate yourself as a happy-go-lucky individual?

   (h) Can you usually let yourself go and have a good time at a party?

   (i) Are you inclined to be moody?

   (j) Would you be very unhappy if you were prevented from making numerous social contacts?

   (k) Do you usually take the initiative in making new friends?

   (l) Does your mind often wander while you are trying to concentrate?

   (m) Do you like to play pranks upon others?

   (n) Are you usually a "good mixer?"

   (o) Are you sometimes bubbling over with energy and sometimes very sluggish?

   (p) Do you often "have the time of your life" at social affairs?

   (q) Are you frequently "lost in thought" even when you should be taking part in a conversation?

   (r) Do you derive more satisfaction from social activities than from anything else?

4. Rosenberg Self-esteem Scale:
   The Rosenberg Self-esteem Scale (RSE) is a well-renowned measurement for assessing self-esteem [22]. The RSE, compared to any other scales to measure self-esteem , has encountered increased psychometric interpretations and empirical authentication. It contains 10-items with a 4-point Likert scale with the score that stretches from 4-40 [19]. Greater score acquired by an individual represents higher self-esteem. The score ranging from 0-15 indicates that one has low self-esteem. The questions used to create the self-esteem scale is mentioned below. The participants answer the questions by indicating how strongly they agree or disagree to the statements.

   (a) On the whole, I am satisfied with myself.

   (b) At times I think I am no good at all.

   (c) I feel that I have a number of good qualities.

   (d) I am able to do things as well as most other people.

   (e) I feel I do not have much to be proud of.

   (f) I certainly feel useless at times.

   (g) I feel that I'm a person of worth, at least on an equal plane with others.

   (h) I wish I could have more respect for myself.

   (i) All in all, I am inclined to feel that I am a failure.

   (j) I take a positive attitude toward myself.

2.2 Applied Methods

The noted results were analyzed using SPSS (Statistical Product and service Solutions) version 25.0 to meet the purpose of the study and perform the statistical analysis more accurately. For precise statistical analysis, the Pearson correlation coefficient, Cronbach alpha, Chi-square test, Median and Standard deviations are measured. The Pearson correlation coefficient determines how strongly two variables are associated with each other. Chi-square test is used to determine the interrelation between the variables used in this research. Cronbach alpha is a measure of core reliability, which is how closely interconnected a collection of components are as a group. It simply represents how stable the scale of measurement is. The significant statistical amount was determined at alpha level or p-value less than 0.05 which is considered as significant. After that, the findings are analysed carefully and the correlations among the variables are interpreted.

2.2.1 Finding internal consistency via Cronbach Alpha

Cronbach alpha which is introduced by Lee Cronbach measures the reliability of the dataset that is, how correctly tests identifies what it should [23]. It measures the reliability of Likert scale-based survey. Likert scales are a conventional assessments format for survey and dataset. A Dataset is considered reasonable if the alpha value is between 0.8 and 0.9. It is considered excellent if the value is 0.9 and above. In simple words, it substantiates how dependable the dataset is by assessing the value generated by said method of measurement. The spreadsheet software, Microsoft Excel has been used to calculate the Cronbach alpha which improved the efficiency. The Cronbach alpha is computed by evaluating the value for each item with the total value of each remark and next it is weighed against the variance of every individual item's value. Equation 2.2 shows the formula to calculate Cronbach alpha.

$$\alpha = (\frac{K}{K-1})(\frac{\sum_{i=1}^{N} \sigma_{y_i}^2}{\sigma_x^2}) \tag{2.2}$$

Here,

- α represents Cronbach alpha
- K denotes the number of scale items
- σy refers to the variance associated with item i
- σx refers to the variance associated with total score

2.2.2 Median & Standard Deviation

Median defines the value that is situated on the midpoint of a dataset. Median denotes the mid value that separates the higher half and the lower half of values in a dataset. The advantage of using a median value above average is that it enhances the accuracy of the mid-value because it is not skewed very much by large or small values. Standard deviation is a method of measurement that evaluates the variance of a data set which means it describes how measurement of the variables spread from the average or median value. A low standard deviation implies that the value is closer to the mean value and a high standard deviation refers to how far the data has spread from the mean

or median value. Equation 2.3 illustrates how to calculate standard deviation.

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{2.3}$$

Here,

- (x_1,x_2,...,x_N) are the observed values
- $\bar{x}$ is the mean value
- N is the number of observations

2.2.3 Pearson Correlation Coefficient

Pearson correlation coefficient is also known as bi-variate correlation or the Pearson product-moment correlation coefficient. It allows one to measure the linear correlation between any two variable. This coefficient is a number between -1 to 1, where 1 represents a complete linear relationship, 0 represents no relationship between the variables and -1 denotes a total negative relationship.Pearson Correlation coefficient is undeniably a very useful method to obtain the linear connection between two variable. However, to reduce complications, A stable version of SPSS has been used in order to achieve results in a concise amount of time. It offers a much flexible and easy environment to work with huge amount of data. The coefficient is calculated by dividing the co-variance of the two variable by the product of each of the variable's standard deviation. Equation 2.4 depicts the formula to calculate Pearson correlation coefficient. The Pearson Correlation coefficient is commonly referred to as ρ.

$$\rho(XY) = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} \tag{2.4}$$

Here,

- ρ is the Pearson correlation coefficient
- cov represents the co-variance
- σX denotes the standard deviation of first variable
- σY denotes the standard deviation of second variable

2.2.4 Chi Square Test

A chi-square independence test compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of categorical variables differ from each another.A very small chi square test statistic means that the observed data fits the expected data extremely well. In other words, there is a relationship.A very large chi square test statistic means that the data does not fit very well. In other words, there is no relationship among the variables.The chi-square test is known as a goodness of fit test and is very useful for finding relationships between categorical variables.A chi square test will give a p-value which is known as the probability,under the null hypothesis.The null-hypothesis refers to the condition that the variables in the dataset do not have any relationship and that they are completely independent. The p-value will tell if the test results are significant or not.The smaller the p-value the higher will be the significance.In order to perform a chi square test and get the p-value,two pieces of information is needed.The first one is known as the Degrees of freedom which is the number of categories minus 1.The second one is the alpha level which is chosen by the researcher. The usual alpha level is 0.05 which is the significant p-value.The p-value is mostly used in statistical hypothesis testing,specifically in null hypothesis significance testing.Microsoft Excel, a software to handle data in a spreadsheet has been utilised to calculate the p-value for the efficient result. Equation 2.5 shows the formula to evaluate the chi-square statistic used in the chi square test.

$$X_c^2 = \sum (\frac{(O_i - E_i)^2}{E_i}) \tag{2.5}$$

Here,

- c are the degrees of freedom
- O refers to the observed value
- E refers to the expected value
- i refers the ith position in the contingency table

2.2.5 ANOVA Test

Normality Test

Checking of the dataset's normality is the most important step in any research as many statistical tests have predefined assumptions and they do not deliver effective result if the assumptions are not met fully. Some statistical methods like ANOVA demands the dependent variable to be normally distributed. In other words, the data should be linear or else the result can be biased sometimes for the violation of the assumption. Some statistical methods like Kruskalwallis, Chi square test etc require non-parametric dataset. To check whether the dataset is normal or not, a plethora of existing statistical tests can be used. In this research,Kolmogorov-Smirnov Test and Shapiro-Wilk Test is used to check the normality of the variables. Both of the tests is based on the p value denoted by sig. value and the significance level used for both of them is 0.05.The null hypothesis for both of the test is that the variables are normally distributed. If p value is less than 0.05 then the assumption of normality of the dataset is rejected and if it is greater than 0.05, it stands as the substantiation of the fact that the dataset is normally distributed.

One-way Non-parametric ANOVA

One way non parametric Analysis of variance (ANOVA) is conducted on our dataset and it is used to differentiate between means of three or more category of independent variables. It is also called Kruskalwallis Test. Before conducting this test, every assumptions that are necessary for the dataset to become qualified for this test should be checked properly. The independent variable should be categorical and dependent variable should be numeric in order to get appropriate result in one way non-parametric ANOVA test.Also, the data should be non-parametric which means non-normal. As the dataset of this research indicates the behaviors of non-normality, Kruskal Wallis Test has been conducted on this. If the dependent variable is not normally distributed then the best alternative for ANOVA is Kruskal Wallis Test.In this research, SPSS is used to convert the independent variables categorical by automatically recoding it so that the dataset should work properly for ANOVA test. One of the important assumption of Kruskalwallis test is independence of groups and in this research independent groups are compared which means this assumption is met accordingly. After checking all of the assumptions, our dataset is used to conduct Kruskal Wallis test to find correlations between different dependent and independent variables and the whole test is done

by using SPSS version 25. First of all, a null hypothesis is assumed as there is no significant difference between the median of the variables that is the median of the different groups are same. After conducting the test the null hypothesis is either rejected or accepted based on the Test statistics and P value(probability value) of the analysis. If the null hypothesis is rejected then the alternate or research hypothesis is selected which says that there is significant difference between the median of the variables.If p-value is less than .05, then a statistically significant difference in the continuous outcome variable between the two independent groups remains which means they have correlation between them.If the p-value is more than .05, then there is no statistical significant difference in the continuous outcome variable between the two independent groups which means no interrelation remains between the variables.

2.2.6 Algorithms and Machine Learning

Machine learning is a widely used method of data analysis that allows the computer systems to automatically learn and act without doing any explicit programming. Machine learning algorithms can be classified into supervised and unsupervised. Supervised machine learning algorithms are applied to the dataset where the outcome or target is known or labelled and machine is trained on the given dataset and tested out to predict correctly in the future. On the other hand, unsupervised machine learning algorithms aims to find hidden relations of a dataset which is unlabelled and does not contain the outputs. It makes decisions on the basis of reasoning from the dataset to describe the hidden layers of the unlabelled dataset.

Pre-processing

The dataset used in this research required a supervised machine learning approach since it contains labelled training data. It is a machine learning method where the input is mapped to output labels. After collecting all the data, at first the dataset was divided into 4 parts: depression, internet addiction, introversion and self-esteem. Next, some features were excluded from the introversion dataset which were mainly used to remove redundancy from the responses and did not contribute to the overall result. In order to run machine learning algorithms on our dataset, all the comma, hyphen, single and double quotation marks were removed from the feature labels and it was converted to csv files. Since our dataset did not contain any missing or null values, it did not require removal or fixing data. Also, no outliers or extreme values that did not

fit or exceed the scoring scale were found in the dataset. The whole dataset was used for sampling and was explored to provide the best accuracy.

Data Partitioning

The 10-fold Cross-validation method has been used for data partitioning. Cross-validation method is used for the evaluation of different predictive machine learning models. For a limited amount of data, it is best to use the K-Fold Cross Validation estimator since it has a lower variance. If we split the data 90 percent for training and 10 percent for testing, the test set becomes very small and a lot of variation is generated for different samples of data. However, cross validation partitions the data in k-folds and then generates average over these partitions and therefore it reduces this variance. Generally 10-folds(k=10) results in a model evaluation with a low modest variance.

Data evaluation

Many parametric and non-parametric algorithms were run on the dataset to predict the outcome of one attribute with the features of another attribute; for example, depression features have been used to predict the outcome of self-esteem. Parametric algorithms are those algorithms which have presumptions before running tests and non- parametric means it does not make any prediction regarding the underlying data distribution.. Multiple algorithms have been tested out to find the best suited prediction model for the dataset. Features were mixed and matched to improve accuracy and Receiver Operating Characteristics value (ROC value) used to define how many targets are being predicted correctly and provide a better model. The algorithms which provided the best results are Logistic regression, Random forest, C4.5 decision tree, K-Nearest Neighbour and Naive Bayes classifier.

Logistic Regression

Logistic regression is the most renowned machine learning algorithm. Logistic regression is mainly used for estimating the parameters of a logistic model.It is also a type of binomial regression and it works as a predictive analysis. The primary goal of this method is to find the most suitable model to explain the correlation between the categorical dependent variable and one or more independent variables. Logistic

regression is also known for classification tasks.A popular choice of equation that has been used in weka to estimate the parameters is given in equation 2.6 below.

$$y = \frac{1}{1 + exp(-w_0 - w_1 a_1 - ... - w_k a_k)} \tag{2.6}$$

here

(a) a represents input features

(b) w is the model coefficient or the feature weight

(c) y represents the output

The algorithm transfigures the input to an output value which can only be between 0 to 1. Logistic regression algorithm uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity.

Random Forest Tree

Random forests tree are a learning technique specifically used for classification, regression that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classification or regression of the individual trees. When it is rather important to work with predictions, the random forest creates an average of all the individual decision tree assessments. Random Forest tree is very flexible, multipurpose and easy to use machine learning algorithm that produces a great result most of the time. It is an ensemble of Decision Trees and on more than average occasions, trained with the "bagging" method. The bagging method shows that a combination of learning models increases the overall result. This methods allows the user to work free of other cross- validation or individual test to get unbiased estimations of any error. Random forest is widely used because of its simple nature.

Decision Tree

A decision tree is a kind of tree where each node represents an attribute, each link represents a decision and each leaf represents an output. The outcome of the tree can be categorical or continues value. The main reason is to create a tree like this for the entire data and process a single output at every leaf to minimise the error in every leaf.

A decision tree can be used to visually and explicitly represent decisions and decision making for decision analysis. Moreover,a decision tree describes data in case of data mining. The goal is to create a model that predicts the value of a target variable based on several input variables.

Naive Bayes

Naive Bayes classifiers are a type of simple "probabilistic classifiers". It is based on applying Bayes theorem with strong independence assumptions between the features. In a learning problem, Naive Bayes classifiers are favourably scalable, requiring a number of constraints linear in the number of characteristics. For constructing classifiers,models that assign class labels to problem instances that are represented as vectors of feature values Naive Bayes can be used as a simple technique. The equation that has been used to in order to utilise this classifier is given in equation 2.7.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{2.7}$$

Here,

- p(c|x) is the posterior probability
- P(c) represents the existing likelihood of the class c
- P(x) is the previous probability of the predictor
- P(x|c) depicts the probability of predictor or the likelihood

Naive Bayes theorem has plethora of purposes and is particularly effective for very large dataset. In order to use this algorithm, the data set needs to be converted into a frequency table, creating likelihood or probability and then by employing the equation. Since this algorithm performs well in categorical inputs rather than numerical variables, it was preferable for the dataset collected for this research.

C4.5

The C4.5 algorithm is used in Data Mining as a Decision Tree Classifier which can generate a decision based on a certain sample of data. C4.5 is also known as a statistical classifier.This is not considered to be one of best methods, however it has proven itself to be effective in a number of situations. C4.5 constructs decision trees from a set of

training data using the concept of information entropy.At each node of the tree C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets. The splitting criterion is the normalised information gain. The attribute with the highest normalised information gain is what makes the decision. C4.5 can work with both Discrete and Continuous Data.It can also handle the issue of incomplete data.

K-nearest neighbours

K-nearest neighbours algorithm is a non-parametric method. It is mainly used for classification and regression. The output is a class membership in case of k-nearest neighbour classification. Furthermore, the output is the property value for the object for k-nearest neighbour regression. It is considered to be very useful as there is no assumption of data. Also it is flexible and has high accuracy and can be used for handwriting detection, image recognition and video recognition. There are a number of approaches to calculate the data based on this method. However, a method that is more relevant to this dataset has been chosen to apply which is depicted in equation 2.8 and explained below.

$$d(x,x') = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + ... + (x_n - x_n')^2} \tag{2.8}$$

Here,

- d is the distance metric between two points
- x_n represents each observation in the training dataset

This method computes the whole dataset through computing the distance metric between each of the observations. Next, it estimates the probability for each case of the given dataset. It is the most simple and versatile algorithm and should be the first choice if there is no prior knowledge about the dataset.

All the algorithms has been applied through Anaconda application with python coding and rechecked with the data analysis tool Weka. Logistic regression, Naive Bayes and Random Forest Tree are supervised algorithms so they ware applied to the dataset where the outcomes were labelled. Decision Tree algorithm can be used for both supervised and unsupervised learning. K-nearest neighbours algorithm is a unsupervised algorithms so it was used to find hidden relations of a dataset.

Chapter 3

Evaluation & Result

The task of evaluating something as a person's cognitive behaviour is quite difficult as well as nearly capricious. However, with the help of the aforementioned methods, it was less hectic to anticipate the results of the research. The Pearson correlation coefficients, Cronbach alpha, Median value, Standard Deviation performed as the structure in order to gain the desired result. First of all, Cronbach's alpha is a measure of internal consistency, which is how closely related a set of items are as a group. It is a measure of scale reliability. The significant statistical value was determined at a probability value, $p<0.01$ level. Figure 1 displays the sequence of operation of this research. Secondly, The Pearson correlation coefficient is a measure of the strength of the linear relationship between two variables. Also, Chi-square test is used to determine the interrelationship between the variables used in this research. Another important analysis is to determine whether the data set is non-parametric or not. Non-parametric data refers to data that is not obliged to fit normal distribution. The dataset used in this research is non-parametric which is proved in the tests for checking normality of the dataset by Kolmogorov-Simrov and by Shapiro-wilk. Both of the tests denotes uniform result of the data to be non-parametric that means it is not normally distributed. A significant value below 0.05 results into the statement that the dataset is non-parametric.

Table 3.1 KOLMOGOROV-SMIRNOV TEST FOR CHECKING NORMALITY OF THE DATASET

Variables	Statistics	sig
Internet Addiction	0.45	0.027
Depression	0.077	0.0
Self-esteem	0.074	0.0
Introversion	0.069	0.0

In table 3.1 from the Kolmogorov-Smirnov test, it is quite visible that each of the sig value of Internet addiction, depression , self-esteem and introversion is significantly below 0.05. They are 0.027, 0.000, 0.000 and 0.000 respectively.

Table 3.2 SHAPIRO-WILK TEST FOR CHECKING NORMALITY OF THE DATASET

Variables	Statistics	Sig
Internet Addiction	0.991	0.006
Depression	0.973	0.000
Self-esteem	0.990	0.003
Introversion	0.983	0.000

Similarly, the table 3.2 denotes the P-value acquired from the Shapiro-wilk test which also evaluates each of the p value to be below 0.05. The results of p value are 0.006, 0.000, 0.003 and 0.000 for internet addiction, depression, self-esteem and introversion respectively.

3.1 Cronbach Alpha

In this survey, the internal consistency for each of the variables is good with a Cronbach alpha value greater than 0.7 (Rumsey, 2013). Table 3.3 represents the internal consistency for each of the variables where it is evident that the data collected to conduct this research is undoubtedly consistent and is reliable enough for further processing.

Table 3.3 INTERNAL CONSISTENCY FOR EACH OF THE VARIABLES

Variables	Cronbach Alpha
Internet Addiction	0.89
Depression	0.94
Self-esteem	0.83
Introversion	0.72

3.2 Median & Standard Deviation

Standard deviation is a method which allows the user to understand how deviated their result is from the average. Median is a value that provides the central tendency of a dataset. According

to Table 3.2, the maximum number of respondents are found to be moderately addicted to internet. Also, majority of students are observed to be moderately depressed and moderately introverted. Again, most of the participants have a low self-esteem. The participants deviated 18.11 percent from the mean score for Internet addiction; 8.9 percent from depression; 4.93 percent from self-esteem and 6.7 percent people deviated from introversion .

Table 3.4 MEDIAN & STANDARD DEVIATION OF THE VARIABLES

Variables	Median	Standard Deviation
Internet Addiction	40	18.11
Depression	60.7	8.9
Self-esteem	24	4.93
Introversion	37	6.7

3.3 Pearson Correlation Coefficient

Pearson correlation coefficient was used to find the linear correlation between two variables, It provides us with astounding results regarding self-esteem and depression. Figure 3.1 depicts a positive relation between these two variables and also that the questionnaire used for these two survey has a significant similarity between them.

It is evident that four statement of the PROMIS depression short form scale have a moderate and positive correlation with self-esteem with a p-value greater than 0.65. The statement “I felt worthless” can significantly influence the result of self-esteem influence the result of self-esteem with a p-value of 0.71 which represents the strong association between the variables. Also, if a person feels that he is hopeless, a failure or depressed, then he is moderately prone to have low self-esteem as the correlation coefficients are greater than 0.6 for those statements. Again, three questions from Rosenberg self-esteem scale can influence the ultimate result of depression of an individual with the p-value greater than 0.55. The statement “I certainly feel useless at times” can moderately affect the results of the depression questionnaire with a p-value of 0.59. Also, the other two statements mentioned in the figure are also reasonably connected with depression. Furthermore, the similarity in the question pattern between the self-esteem scale and depression scale can be seen.

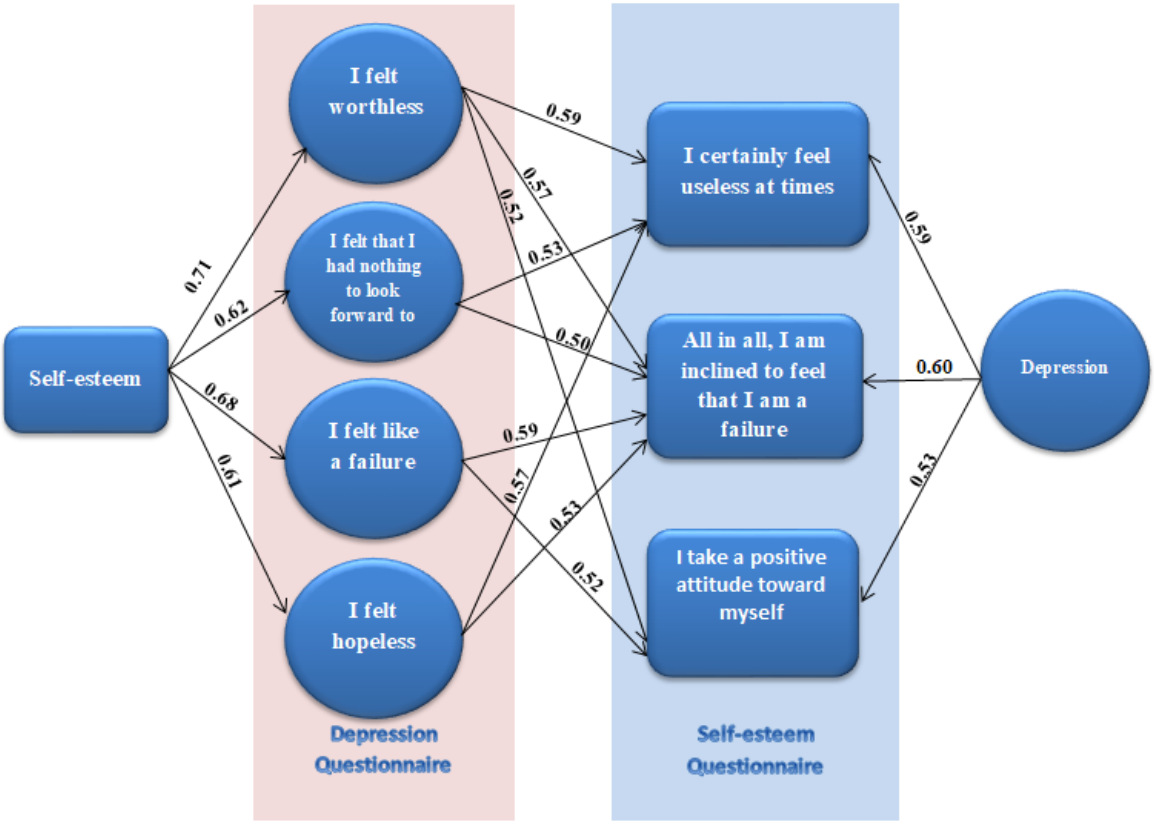

Fig. 3.1 Pearson Correlation Coefficient between depression & self-esteem

3.4 Chi Square Test

Chi square test represented how closely related the variables are. Table 3.5 shows that gender and introversion have a p-value of 0.04771 which is less than 0.05 that means there is a significant relationship between them. The analysis revealed that gender and introversion have a relationship with each other. Also, Depression and self-esteem are found to be correlated with each other as the p-value is 1.49E-36 which is distinctly less than 0.05. Likewise, we can conclude that there is a statistically significant relationship between depression-introversion, depression-internet addiction and self-esteem-internet addiction with a p-value less than 0.05. Table 3.5 shows the Chi-square test analysis with a significance value of 0.05 .

Table 3.5 CHI-SQUARE TEST ANALYSIS

Variables	Chi-square test analysis	Critical Point	Degree of freedom	P-value
Gender -Introversion	2.798	5.99	2	0.04771
Depression -Self esteem	181.677	12.59	6	1.49E-36
Depression Introversion	18.193	12.59	6	0.006
Depression -Internet Addiction	61.567	16.92	9	6.68E-10
Self esteem -Internet Addiction	36.269	12.59	6	2.44E-06

3.5 Results from ANOVA test

ANOVA assists its user to figure out if the survey results are noteworthy or not. It contemplates if the user can reject the null hypothesis, i.e. there is no significant relation among the groups or variables. The decision can be made by perceiving if the P value here is greater than 0.05 or not. If the value recorded is more than 0.05, it indicates that there is no remarkable correlation between the varibales. However, if it is the other way around, that is,the value is below 0.05, then correlation exists between the two variables.

The One-way parametric ANOVA test divulges quite intriguing outputs regarding the correlations. Table 3.6 shows that Internet-addiction and depression has a correlation with a p value of 0.000 and that rejects the null-hypothesis. Furthermore, after calculating the

Table 3.6 ONE-WAY NON PARAMETRIC ANOVA (KRUSKAL WALLIS TEST)

Variables	Degree of Freedom	H-test Statistics	Sig	Decision
Internet Addiction -Depression	3	66.627	0.000	Reject the null-hypothesis
Internet Addiction -Self-esteem	3	40.994	0.000	Reject null-hypothesis
Internet Addiction -Introversion	2	0.089	0.956	Retain the null-hypothesis
Depression -Self-esteem	3	215.845	0.000	Reject the null-hypothesis
Depression -Introversion	3	26.094	0.000	Reject the null-hypothesis
Self-esteem -Introversion	2	229.098	0.000	Reject the null-hypothesis

correlation of depression with the other two variable self-esteem and introversion, they also denote that the null-hypothesis can be rejected with a p value of 0.000. However, the only exception is the correlation between internet addiction and introversion obtained a p value of 0.956 which depicts that the null-hypothesis has to be retained. Self-esteem and introversion, like other variables, showed a positive correlation and that the null-hypothesis can be replaced with the alternate hypothesis which is that there is a notable relation between these variables.

3.6 Outcomes Of Algorithms

In order to apply machine learning algorithms, the dataset of this research requires supervised methods of approach since it has labelled training data. By using these techniques, the accuracy of the newly established correlation can be assessed. It is observable that using cross-validation 10-fold helped achieve better results for these variables with a lower variance. Each of the correlation found through the study is discussed below along with corresponding figures with a view to better understand the outputs.

3.6.1 Depression & Self-esteem

Figure 3.2 depicts the accuracy rate found by the five algorithms that has been deliberated in methods & system structure. Here, eight features of depression and two features of

self-esteem has been engaged as independent variable and the final score of self-esteem is taken as dependent variable. the accuracy generated by each of the algorithms is 0.83, 0.82, 0.86, 0.82 and 0.84 for C4.5 Decision tree, K-nearest neighbour, Logistic regression, Naive Bayes and random forest respectively.

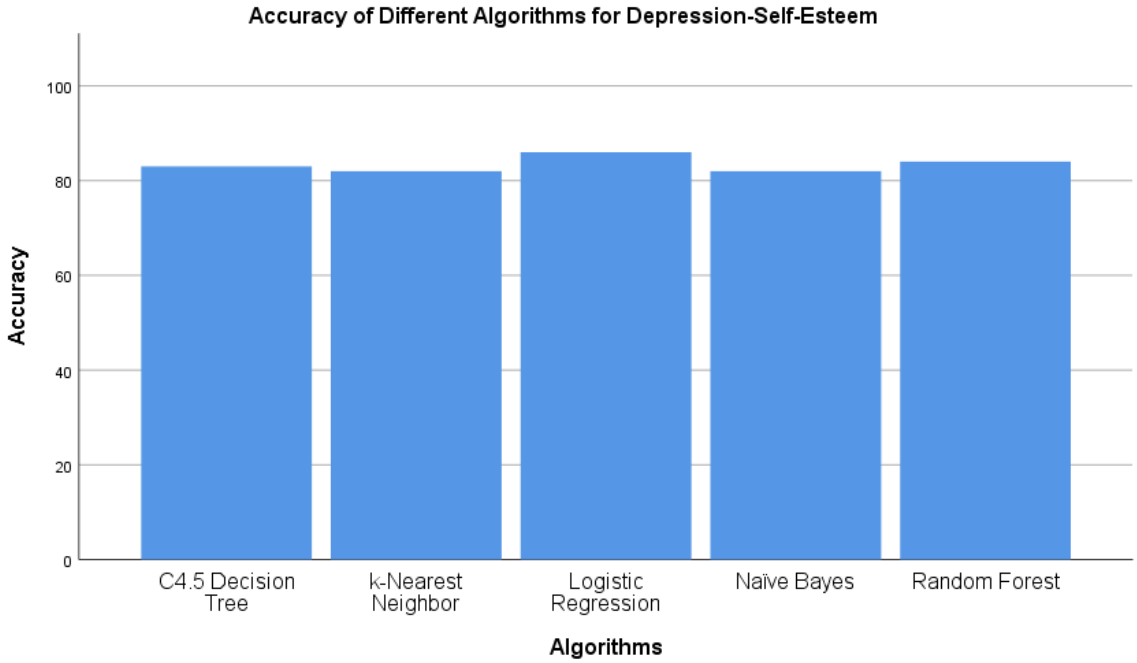

Fig. 3.2 Accuracy of different algorithms for Depression & self-esteem

3.6.2 Internet Addiction & Self-esteem

The results produced by the prediction of Internet Addiction & Self-esteem has an accuracy rate that is slightly below depression and self-esteem. Here, twenty features of internet addiction and three features of self-esteem are taken as the independent value whereas the final value achieved for self-esteem is taken as the dependent value, again. In figure 3.3 it is visible that logistic regression and random forest delivers the best results which are 0.82 and 0.83 respectively. The other techniques, Naive-Bayes produces an accuracy rate of 0.76 and K-nearest neighbour produces 0.69 whereas C4.5 decision tree has a value of exactly 0.80 which are nevertheless, very intriguing values.

3.6.3 Depression & Introversion

The highest outcomes were recorded in terms of depression and introversion. Almost all of the algorithms produced a value greater than 90% except Naive-Bayes. After using a number

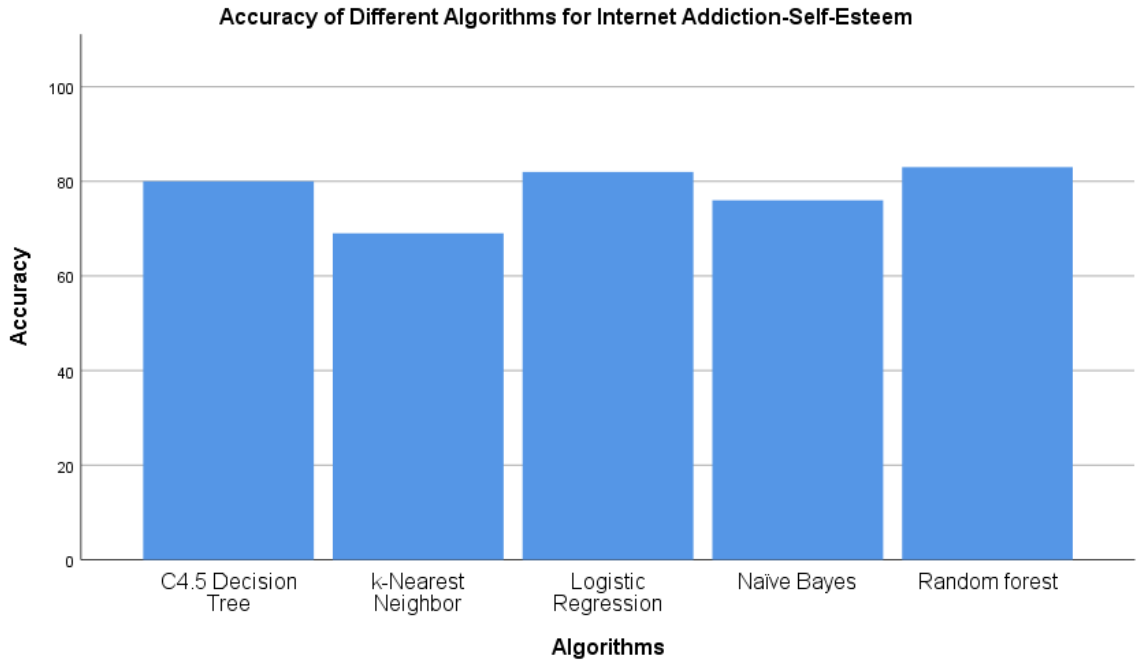

Fig. 3.3 Accuracy of Algorithms for Internet Addiction & Self-esteem

of techniques, the study found that if 8 features of depression are taken as independent variable, the final value of introversion can be generated as the dependent variable. This correlation gave an astounding result. Logistic regression, K-nearest neighbour, Random forest and C4.5 tree, each of these methods delivered an accuracy rate of 94.0% which is notably higher than any of the values produced before. However, Naive-Bayes theorem assessed the prediction of depression and introversion with an accuracy of 0.84.

Precision

The precision is ability of the classifier not to declare a sample as positive when it is actually negative. The precision is the ratio that can be written as:

$$Precision = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \tag{3.1}$$

The more the precision value is close to 1, the more it is correct and the model has less chance of having false positives.

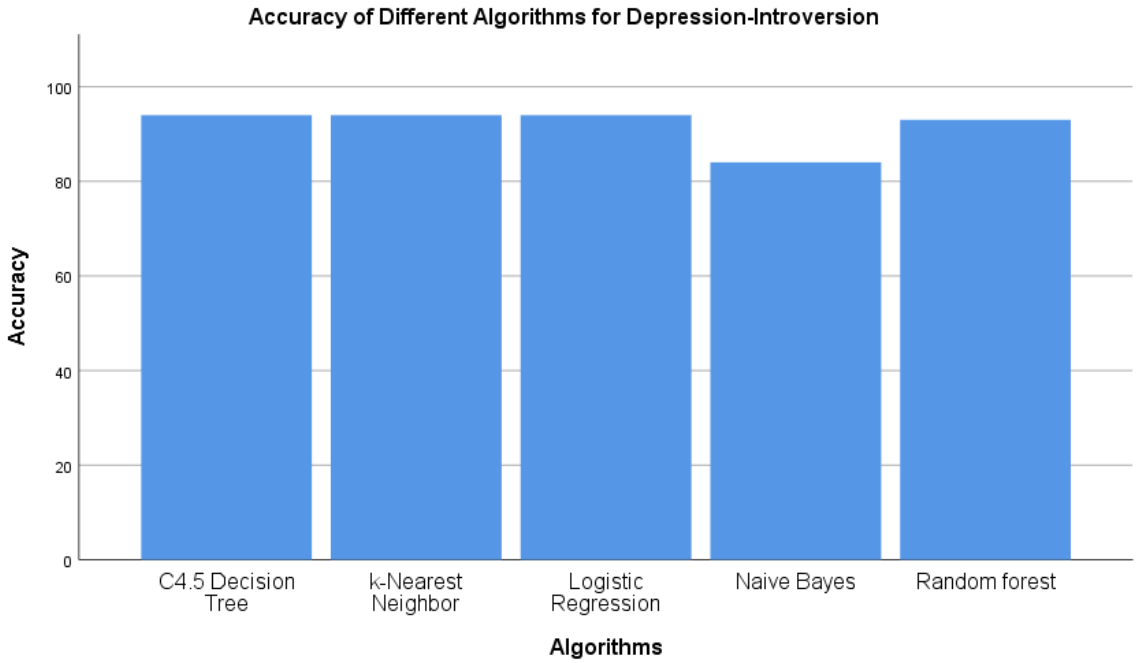

Fig. 3.4 Accuracy of Algorithms for Depression & Introversion

Recall

The recall is ability of the classifier to find all the positive samples.The recall is the ratio that can be described as:

$$Recall = \frac{\text{True} \quad \text{positive}}{\text{True} \quad \text{positive} + \quad \text{False} \quad \text{negative}} \tag{3.2}$$

Recall describes the amount of actual positives that were identified correctly and the more the value of the recall, the model performs better and gives less false negatives.

F-measure

The F-measure can be calculated as a weighted mean of the precision and recall, where 1 is its best value and 0 is its worst value.It considers both false positives and false negative. It is more useful than accuracy in some cases where uneven class distribution occurs.

$$F - measure = \frac{2 * (Recall * Precision)}{Recall + Precision} \tag{3.3}$$

Roc Area (Receiver Operator Characteristics Curve)

It is a curve used to measure the performance of binary classifiers. Accuracy is measured by the area under the roc curve. It's best value is 1 and the performance of the classifier is considered as good when roc area value is greater than 0.5.

Accuracy

To measure the performance of any model, accuracy score plays a significant role as an evaluation factor. Also, confusion matrix, mean absolute errors are some noteworthy evaluation metric. Classification accuracy is the ration of number of correct samples to the total number of samples.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions made}} \tag{3.4}$$

But accuracy only gives the best result when there are almost equal number of samples in each class.

In the Table 3.7, precision, recall and ROC area values are greater than 0.5, therefore the prediction results generated in the three cases (Depression - Self-esteem, Internet addiction - Self-esteem and Depression - Introversion) are not random and they provide the actual accuracy.

After comparison of accuracy and the ROC value of different algorithms in each of the prediction case, the best suited algorithm for determining the dependable variable is found. The algorithm that gives the best accuracy for predicting self-esteem from depression and additional 2 features of self-esteem is Logistic Regression with the highest accuracy (0.86) and ROC area (0.933). Even though Random forest algorithm generates a higher accuracy of 0.83 that Logistic Regression with an accuracy of 0.82 for predicting self-esteem from internet addiction with an additional 3 features from self-esteem, since ROC area of Logistic Regression (0.94) is distinctively higher than ROC area of Random forest (0.823), therefore it can be concluded that Logistic Regression is the superior algorithm for this data. For the prediction of introversion from the features of depression, K-nearest neighbour (KNN=1) is the best suited algorithm with an accuracy of 0.94 and ROC area of 0.901.

Table 3.7 ACHIEVED VALUES AFTER APPLYING DIFFERENT ALGORITHMS

Independent	Dependent variable	Algorithm variable	Precision	Recall	F-measure	ROC area	Accuracy
8 features of depression and 2 features of self-esteem	Self-esteem	Logistic Regression	0.86	0.859	0.859	0.933	0.86
		K-nearest Neighbour	0.825	0.824	0.825	0.853	0.82
		Random Forest	0.838	0.837	0.838	0.918	0.84
		C4.5	0.834	0.831	0.832	0.849	0.83
		Naive-Bayes	0.822	0.816	0.818	0.889	0.82
20 features of Internet addiction and 3 features of self-esteem	self-esteem	Logistic Regression	0.827	0.82	0.8823	0.94	0.82
		K-nearest Neighbour	0.681	0.688	0.683	0.678	0.69
		Random Forest	0.823	0.829	0.829	0.826	0.83
		C4.5	0.805	0.805	0.804	0.758	0.80
		Naive-Bayes	0.762	0.761	0.762	0.841	0.76
8 features of depression	Introversion	Logistic Regression	0.573	0.946	0.565	0.607	0.94
		K-nearest Neighbour	0.916	0.937	0.924	0.901	0.94
		Random Forest	0.908	0.905	0.933	0.669	0.93
		C4.5	0.572	0.946	0.565	0.466	0.94
		Naive-Bayes	0.905	0.837	0.867	0.901	0.84

Chapter 4

Conclusion & Future Work

4.1 Conclusion

The leading disorders in this era of digital revolution are very conspicuous as well as an important area of research. The primary goal of this research is to find if these malignant syndromes like internet addiction, depression, self-esteem and introversion are interconnected and influential on each other. This type of experiment requires immense computational effort which is why a number of intriguing results have surfaced in the study conducted in contemplation of acquiring correlation. First of all, it is evident that the dataset is reliable and consistent which led us to work with a number of methods and machine learning techniques to find astounding results. Using logistic regression, decision tree, random forest tree and naive bayes provides the information that each of the correlation has an accuracy rate of more than 80%. These results stand as a proof that the studies conducted in this research accommodates us with precise correlations and appropriate outcomes. The research demonstrates that gender and introversion have been found to be dependent on each other. Again, depression has a relation to self-esteem, internet addiction and introversion. Self-esteem is also correlated with internet dependency among the university students. The most sensational and astounding finding of this study is that depression and self-esteem questionnaire are interrelated to each other on a moderate level. A similar type of questionnaire can derive any of these two disorders. Subsequently, the factors of self-esteem can influence the overall result of depression as well as depression can persuade the results of self-esteem. Particular questions of both self-esteem scale of Rosenberg and specific questions of PROMIS depression scale-short form are moderately interrelated with each other. Therefore it is safe to denote that this research provides substantial results that can be used to provide proper treatment to reduce these disorders.

4.2 Future Work

The addition of the new feature would be to conduct more in-depth analysis of particular mechanisms and new proposals to try different methods. The scope of age, extent of area and the number of respondents can be broadened to get a more accurate result. The scope of this study has predominantly emphasised on statistical analysis, therefore in future experiments,other regression methods or machine learning techniques can be applied to broaden the horizon of the outcomes. New technologies can be discovered in the future to establish the true extent of correlation and predict the indispensable procedures that can be immensely beneficial to minimize the effects of these leading disorders. Again, the finding replicates previous studies and results, therefore contribution and uncovering new factors can be generated in order to find more preferable solutions to these malicious illness.

References

[1] M. Yao, J. He, D. Ko, and K. Pang. The Influence of Personality, Parental Behaviors, and Self-Esteem on Internet Addiction: A Study of Chinese College Students. In *Cyberpsychology, Behavior, and Social Networking*, volume 17, pages 104–110. National Center for Biotechnology Information,U.S, 2014.

[2] E. Budak, I. Taymur, R. Askin, B. Gungor, H. Demirci, A. Akgul, and Z. Sahin. Relationship between internet addiction, psychopathology and self-esteem among university students. In *The European Research Journal*, volume 1, page 128. Elsevier Science Publishers B. V. Amsterdam, The Netherlands, 2015.

[3] K. Kırcaburun. Self-Esteem, Daily Internet Use and Social Media Addiction as Predictors of Depression among Turkish Adolescents. In *Journal of Education and Practice*, volume 7. IISTE, 2016.

[4] B. Aydm and S. San. Internet addiction among adolescents: The role of self-esteem. In *Procedia - Social and Behavioral Sciences*, volume 15, pages 3500–3505. ELSEVIER, 2011.

[5] Paul R. Albert. Why is depression more prevalent in women? In *J Psychiatry Neurosci*, volume 40, page 219–221. National Center for Biotechnology Information,U.S, July 2015.

[6] T. Pedersen. Introverts spend more Time on the Internet IMing. Psych Central. URL https://psychcentral.com/news/2013/12/29/introverts-spend-more-time-on-the-internet-iming/63886.html.

[7] C. Öztürk, M. Bektas, D. Ayar, B. ÖzgüvenÖztornacı, and D. Yağcı. Association of personality traits and risk of internet addiction in adolescents. In *Asian Nursing Research*, volume 9, pages 120–124. ELSEVIER, 2015.

[8] O. Orsal, O. Orsal, A. Unsal, and S. Ozalp. Evaluation of Internet Addiction and Depression among University Students. In *Procedia - Social and Behavioral Sciences*, volume 82, pages 445–454. National Center for Biotechnology Information,U.S, 2013.

[9] W. WANG. Internet dependency and psychosocial maturity among college students. In *International Journal of Human-Computer Studies*, volume 55, pages 919–938. ELSEVIER, 2001.

[10] Y. Yeh, H. Ko, J. Wu, and C. Cheng. Gender Differences in Relationships of Actual and Virtual Social Support to Internet Addiction Mediated through Depressive Symptoms

among College Students in Taiwan. In *CyberPsychology& Behavior*, volume 11, pages 485–487. SAGE Journals, 2008.

[11] L. Rabadi, M. Ajlouni, S. Masannat, S. Bataineh, G. Batarseh, A. Yessin, K. Haddad, M. Nazer, S. Hmoud, and G. Rabadi. The Relationship between Depression and Internet Addiction among University Students in Jordan. In *Journal of Addiction Research & Therapy*, volume 8. 2017.

[12] S. Moussavi, S. Chatterji, E. Verdes, A. Tandon, V. Patel, and B. Ustun. Depression, chronic diseases, and decrements in health: results from the world health surveys. In *The Lancet*, volume 370, pages 851–858. National Center for Biotechnology Information, U.S., 2007.

[13] C. Yen W. Chou and T. Liu. Predicting Effects of Psychological Inflexibility/Experiential Avoidance and Stress Coping Strategies for Internet Addiction, Significant Depression, and Suicidality in College Students: A Prospective Study. In *International Journal of Environmental Research and Public Health*, volume 15, page 788. National Center for Biotechnology Information, U.S., 2018.

[14] SA Baharainian., K. Haji Alizadeh, MR. Raeisoon, O. Hashemi Gorji, and A. Khazaee. Relationship of Internet addiction with self-esteem and depression in university students. In *Journal of Preventive Medicine and Hygiene*, volume 55, pages 86–89. National Center for Biotechnology Information, U.S., 2014.

[15] S. Ahmad Bhat and M. Hussain Kawa. A Study of Internet Addiction and Depression among University Students. In *International Journal of Behavioral Research & Psychology*, pages 105–108. SciDoc Publishers,Lewes, Delaware, 2015.

[16] K. Niemz, M. Griffiths, and P. Banyard. Prevalence of Pathological Internet Use among University Students and Correlations with Self-Esteem, the General Health Questionnaire (GHQ), and Disinhibition. In *CyberPsychology& Behavior*, volume 8, pages 562–570. National Center for Biotechnology Information, U.S., 2005.

[17] K. YOUNG. Internet Addiction: The Emergence of a New Clinical Disorder. In *CyberPsychology Behavior*, volume 1, pages 237–244. Mary Ann Liebert, Inc., publishers,U.S.A, 1998.

[18] Healthmeasures.net, 2018. [online]. Available:. URL http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Depression_Scoring_Manual_02222017.pdf. [Accessed: 17- Aug- 2018].

[19] W. Petersen. Society and the Adolescent Self-Image. Morris Rosenberg. Princeton University Press, Princeton, N.J., 1965. xii 326 pp. 6.50. In *Science*, volume 148, pages 804–804. American Association for the Advancement of Science, Jul. 1965.

[20] V. P. Richmond, J. C. McCroskey, and J. S. Wrench. Communication apprehension, avoidance, and effectiveness. Boston: Pearson, ResearchGate, 2013.

[21] D. Goel, A. Subramanyam, and R. Kamath. A study on the prevalence of internet addiction and its association with psychopathology in indian adolescents, May 2017.

[22] W. J. Hagborg. The Rosenberg Self-Esteem scale and Harters Self-Perception profile for adolescents: a concurrent validity study. In *Psychology in the Schools*, volume 30, pages 132–136. John Wiley Sons, Inc, 1993.

[23] D. Rumsey. Statistics I & II For Dummies 2 ebook Bundle. Hoboken: Wiley, Goodreads Inci, 2013.