Prediction on Digital Addiction among Undergraduate Students during Pandemic using Machine Learning

by

Nur Diana Binti Jamaludin

17001471

Dissertation submitted in partial fulfilment of

the requirements for the

Bachelor of Information Technology (Hons)

SEPTEMBER 2021

Universiti Teknologi PETRONAS

32610 Seri Iskandar

Perak Darul Ridzuan

Malaysia

CERTIFICATION OF APPROVAL

Prediction on Digital Addiction among Undergraduate Students during Pandemic using Machine Learning

By

Nur Diana Binti Jamaludin

17001471

A project dissertation submitted to the
Information Technology Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF INFORMATION TECHNOLOGY (Hons)

Approved by,

Ts Dr Norshakirah Ab. Aziz
Senior Lecturer
Computer Information & Sciences Department
Universiti Teknologi PETRONAS

(Dr. Norshakirah Bt. A. Aziz)

UNIVERSITI TEKNOLOGI PETRONAS
BANDAR SERI ISKANDAR, PERAK
September 2021

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own expect as specifies in the reference and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

NUR DIANA BINTI JAMALUDIN

ABSTRACT

The purpose of the final year project is to do detailed research on digital addictions on how it affects undergraduate students during the pandemic, and to come out with a analytical analysis using technical skills. The research on the project had shown that digital addiction is not only focus on playing online video games, but it can also be in any types of online activities. The study also shown that digital addiction has grown throughout the pandemic and had affect the student's personal health. In most recent cases, it had shown that university students went through multiple types of mental health issues due to limitations of movement and had spent hours on online activities. By having this project, may help the respected bodies and university association to take further action in order to take care the students physical and mental health. Other than that, it might also help the government bodies to take early actions to handle student's welfare.

ACKNOWLEDGEMENT

Firstly, I would like to express my gratitude to my university, Universiti Teknologi Petronas, for giving me opportunity to complete my Final Year Project and provide enough resources for the future research on the project.

Next, I want to manifest my gratitude to my supervisor, Dr. Norshakirah Bt. A. Aziz, who has given me full support throughout my project by having a great patience in guiding me for any tasks that was been given. Furthermore, he also has shown a great leadership skill by spread positiveness and gives full support. I also able to hone my skills that are not only on technical skills.

Furthermore, I would like to express my gratitude towards the FYP coordinator, Ts. Dr Said Jadid, internal and external examiners for guiding me and gives positive feedback to improve myself in the future.

Lastly, I would like to thank my parents for giving me full support throughout my completion of the project and provide financial support at the same time.

Table of Contents

List of Figures.

List of Tables.

CHAPTER 1
INTRODUCTION

1.1 Introduction

1.1 Background

1.1.1 Digital Addiction During Pandemic

Digital addiction can be in any types of addictions that related with listening, watching, or playing games by using electronic devices such as computer, laptop, or mobile phone for their entertainment purposes (Aziz et al., 2021). There are three different types of digital addiction that people can be diagnosed with, one of the addictions is computer game and Internet addiction, the second type is social media addiction, and the third type is smartphone addiction (Aziz et al., 2021). Addiction is where the person cannot live without the particular activities they did. Therefore, it is really dangerous as the addiction can affect people either mentally or physically. Furthermore, and addiction is something uncontrollable once it started for an example, an addiction towards nicotine and drugs, the addicts cannot simply stop taking nicotine and drugs, but they need to stop slowly or taking any medicines.

The pandemic has hit the whole world since March 2020 until today. It has affected a lot of activities and industries at the same time due to limitations for social and social distancing practicing. Due to pandemic, most people had to spend their time at home youngster and elderly, students to working people. Especially during quarantine and lockdown, everyone will have limitations for movements and not allowed to go out because of government orders.

Consequently, since the pandemic started most people has become addicted towards digital because they cannot spend their time socializing or doing activities and works virtually. Since of the limitations, most activities need to be online such as online class for students and online meeting employee. In conjunction to that, they spend most of their time on computer screen or mobile phone screen since there is no other choices to comply their studies or working without doing it so. It has affected the student the most because schools and university are not allowed to open, and student are not allowed to come so they had no choice but to face online learning at home.

1.1.2 Machine Learning Methods and Modelling.

Machine learning is a data analysis method that automates the creation of analytical models. It has been widely used for data analysis since it is the best method to show the analysis by using the correct models that suit with the dataset and the type of studies. There are a few types of machine learning methods, which are supervised learning and unsupervised learning. Supervise learning is mapping functions predict dependent variables from independent variables (Mak et al., 2019). Technically it is for train a model for prediction. Supervised learning has two types of modelling which are regression and classification. In contrast to supervised learning, when labels are provided along with the data, unsupervised learning is a form of machine learning technique used to generate inferences from datasets without human interaction. The modelling in unsupervised learning is clustering.

1.1.3 Undergraduate Students in Malaysia.

An undergraduate is either college or university students who are not yet graduated that just finish high school. The majority age of undergraduate students in Malaysia in within 17 to 25 years old. The addiction studies are mainly focus on undergraduate students. This is because this group of people are most likely to have high possibilities to become addicted towards digital

and other addiction studies as well. It is because of their dilemma between being a teenager and adult which means not too young or not too old to think and make their own decisions.

Other than that, undergraduate students are highly focus on their studies in order to graduate. This group are willingly to put all of their effort towards their studies by hook or by crook. In conjunction to that, due to current situations undergraduate students are facing online classes for their studies. By having online classes, they are mainly spending their time on computer, laptop or tablet screen. They spend hours to complete their task, assignment, test and quizzes. This will lead to other consequences where students might spend too much time on computer screen for studies or they are distracted with other online activities in order to release stress. The other online activities can be online shopping, playing computer games or scrolling on social media since there is no one watching what they do or stop them.

1.2 Problem Statement

First and foremost, according to (Rahayu et al., 2020) digital addiction can have negative impacts on specific person life. The negative impacts vane be a huge scope such as family conflict, work performance issues, invasion of others' privacy, dietary-related issues, violence (self-harm, harm to others, and harm from others), emotional issues, personal issues, and social issues. These negative impacts will affect the person life emotionally, the worst part of these impacts are deaths. As shown by the latest statistics, the number of people who commit suicide has increased from time to time and it has not only happened among adult but has a high number among students.

Other than that, physical health problem has been increase like back discomfort, eye strain, and carpal tunnel syndrome (Rahayu et al., 2020). The person physical health is not only lead them to minor pain, but it can also turn into death. There is a recent case where university students were dead at his parent's house due to cerebral haemorrhage which was lead from studies pressure and had mainly spend their time on digital until does not get enough rest. This has clearly shown, students physical health is also important to be taken care of especially during this pandemic situation.

Besides that, students have not choices but to socialise using digital due to current situations. The only way to communicate with their friends and respective lecturer either at college or university are only through digital. Therefore, they had no choice but will spend hours on gadgets to keep their interactions. Moreover, there are also no other activities that can be done at home due to loneliness, introversion, and boredom inclination but to spend their time on online activities. Certain people who used to communicate with their online friends through digital has become more obsessive where they will spend hours on online activities with their online friends. These actions will lead to negative impacts where students might not be able to communicate or socialise where because they might have anxiety or low emotional intelligence (Griffiths, 2014).

1.3 Objectives

The aim of the projects is to create analytical analysis of the digital addiction among undergraduate student in Malaysia either private or public university during pandemic using machine learning. This project will be able to come out a dashboard of the analysis how digital addiction has become whichever worser or better during pandemic. The analytical analysis can be use by any parties that relate with healthcare in order to come out with the approaches that can be helping to resolve the digital addiction.

Therefore, the objectives that applied for the project are:

- To predict the number of undergraduate students that has digital addiction which are within age range 17 to 25 years old.
- To train data using Python language.
- To develop data visualization using Power Bi with analytical analysis dashboard.

1.4 Scope of Study

The boundaries and limitations of this digital addiction for undergraduate students with machine learning will include the user level functions, project completion period, focus and the hardware and software requirements for this system to run.

1.4.1 Focus

The focus of this project is to study how digital addiction has been growing among undergraduate student which will lead to mental and physical health. The boundaries and limitations of this project is to create an analytical analysis of the digital addiction by

a) Conduct survey based on digital addiction model provided by experts.
b) Collect the survey by distribute to undergraduate's student for datasets collections.
c) Predict data using k-means modelling by using Python languages with the datasets.
d) Prepare a dashboard for digital addiction using Power Bi.

1.4.2 Target user

The target user for the projects is undergraduate students who are within the age range of 17- 25 years old that has been exposed with digital addiction or not.

1.4.3 Time Limitations

The time limitation giving to complete this project are 8 months, which is 4 months for FYP1 and 4 months to FYP 2 in total 24 weeks

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

A proper method and modelling for the project is the most important part for machine learning. Therefore, research is important in order to support the proposed project and determine the specific method that can be used for the project development. Other than that, the importance on the research will personally help me to get a better idea on how to implement digital addiction studies into machine learning. It is because there are multiple types of method that can be used for machine learning. Furthermore, the research purpose also to make sure the validation of the symptoms, constraints and effect of digital addiction among undergraduate students, without proper research the project can be made up and irrelevant of the proposed project. Generally, the success toward the success towards the project can be derived from the literature review research result.

2.2 Digital addiction

According to (Aziz et al., 2021) Addiction is characterised as an overwhelming craving, typically accompanied by a lack of control, and continued use despite the fact that the conduct is causing a problem. This is shown that an addiction can be anything people cannot control or live with, or else they tend to act negatively and gave them the negative impact. Meanwhile, due to arising in digital technology there is a new consequences called digital addiction where it can be in any type of technological addiction (Rahayu et al., 2020). There are a few different types of digital addiction which are internet addiction, online or video game addiction, scrolling social media and smartphone addiction. There is a gap between using smartphone or laptop for daily use and being addicted towards it. The addiction is where, a person willingly to spend hours on mobile or PC screen for hours for the entertainment use without having any limitations and if they stop

doing it, they might to react differently or unable to find their happiness. Meanwhile, using the technological advancement for working purpose is where they spend their times to complete their work and task, then take a break after spending hours on it.

2.3 Digital addiction parameters.

There are a few symptoms and measures in order to conclude that a person are digital addicts. These are the important parameters and components that are used to create a survey for dataset collection. The addiction components are from the medical experts such as tolerance, salience, mood modification, harm, relapse, conflict, withdrawal, physical health, and loss of control (Aziz et al., 2021).

Based on the components stated, we will conclude based on how bad the undergraduate students facing each of the components while doing their online activities. The digital addicts tend to at least be facing 5 of the listed components as they are spending hours on the screen doing their online activities. According to (Aziz et al., 2021) the experience of streaming, satisfaction, and involvement in computer games influence the player's attitudes and real use of computer games but when it comes to evaluating actual gameplay, enjoyment and social influence become persuasive evidence.

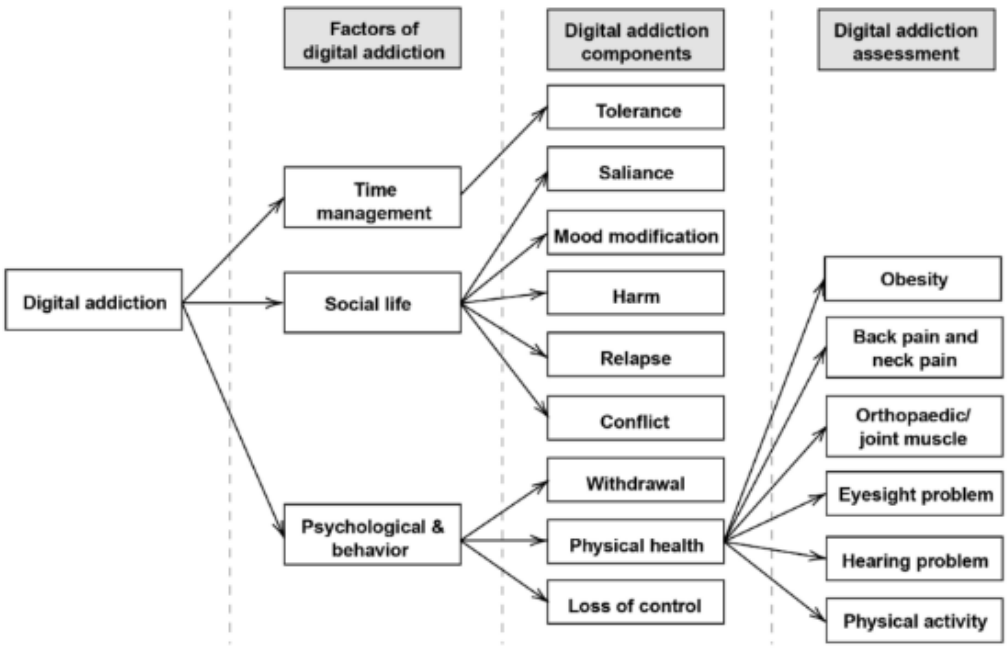


Figure 1 Digital addiction relationship diagram

Other than that, the figure below shows how a particular activity will give impact towards a particular physical health. This will give a better idea on how an online activity can gives impact towards them and also are considered in creating the survey for the distribution purpose. Furthermore, these impacts are easier to be identified since it happens on the undergraduate students' physicals where it can be clearly seen without needs to diagnose with doctors and experts.

Factor of DA	Description of Activities	Consequences on Physical Health
Psychological behavior	Playing computer games is a sedentary activity. Gamers tend to spend time playing games indoors instead of performing outdoor activities. Hence, they are prone to the risk of obesity, especially when they eat while playing computer games.	Obesity
	Prolonged physical immobility will lead to muscle pain such as back and neck pain.	Back pain and neck pain
	Using a mouse and keyboard for a long time causes muscle problems in fingers and hands.	Orthopaedic/ joint muscle
	Having a long on-screen time can cause dry eyes and eyesight problems.	Eyesight problem
	Continuous exposure to loud noise from headphones can reduce hearing ability.	Hearing problem
	Computer gamers tend to have much less physical activity than other people as they spend more time playing computer games in a room.	Physical inactivity

Figure 2 The factors of digital addiction, description and how it effects on physical health.

2.4 Digital addiction effects on undergraduate students.

Spending hours on online activities can gives negatives impact towards the particular person either physically or mentally. Mainly, I will affect the person's health problem such as mental health, psychology, physical health and anxiety (Aziz et al., 2021). Mental health can be in a wide scope such as depression, personality disorder, psychotic disorder and more, different online activities a person do can derived different types of negative impact towards them.

These negative effect can impact them in three different areas which are time management, social life and emotions (Latif et al., 2017). An individual who are addict with online activities tend to spend hours and does not count when they are going to do it, it can be either during their study time, eating time or break time since their main purpose is to entertain themselves. Usually, the worst effect is on students and single person because they do not have big responsibility to be taken care of. It is different with working people, where they only have after working hour time or during weekend only to spend on their online activities. Meanwhile, for a person who already got married they have big responsibility to take care of their partner and family on the same time. These group of people had no choice but to limit their online activities or a worst case can be happen for an example, "an internet addicts left their babies starving and suffering to death" (Latif et al., 2017). Meanwhile, for social life impacts, the digital addicts definitely have their own online friend whom they never met before, but they knew each other through social medias or online video games. These actions can arise to have problems on handling serious life relationship with their partners or family members. It is because they do not have a great communications and social skills in real life and tend to avoid a serious conversation. In terms of impact on emotions, a digital addict unable to control their emotions where they can become very intense by being irritable, anxious, or depressed. The worst-case scenarios are where the digital addict unable to control themselves and will act violently either through speaking or actions.

2.5 Machine learning applications in digital addictions.

According to (Akhter, 2017) Machine learning applications can help with non-deterministic situations that are too complex or large to be broken down into step-by-step instructions. It is mentioned that machine learning can be used for any sort of problems and can be narrow down to make it more complex. In the project, the objectives are to prepare the data for making an analytical analysis afterwards. The process will require a few steps in order to make sure that the data are using a correct model and method. Furthermore, there are various types of methods and modelling that can be used and implement in machine learning that suitable with data related project.

The reason why I use machine learning in the project is because, it is mentioned by (Mak et al., 2019) to use machine learning in an addiction studies. There are a few methods that can be used for unsupervised learning. Unsupervised learning method is suitable for a study that has low number of datasets since it will be hard to split the training and test data due to the small number. The suitable method for the case study is clustering method because the number of students that involve in the survey are most likely to be independent variable. According to (Klochko, n.d.) the fundamental concept is to use a variety of clustering approaches to conduct an empirical comparison study and discover which methods provide the best data grouping while solving a given problem.

2.6 K-Mean modelling on predicting addicted students.

The K-means algorithm is an iterative technique that attempts to split a dataset into K separate non-overlapping subgroups (clusters), each of which contains only one data point. K-means will assign the intra-cluster data points as comparable as possible while keeping the cluster different at the same time. K-means will assign data points to each cluster such that the sum of the squared distances between the data points and the cluster's centroid equals one.

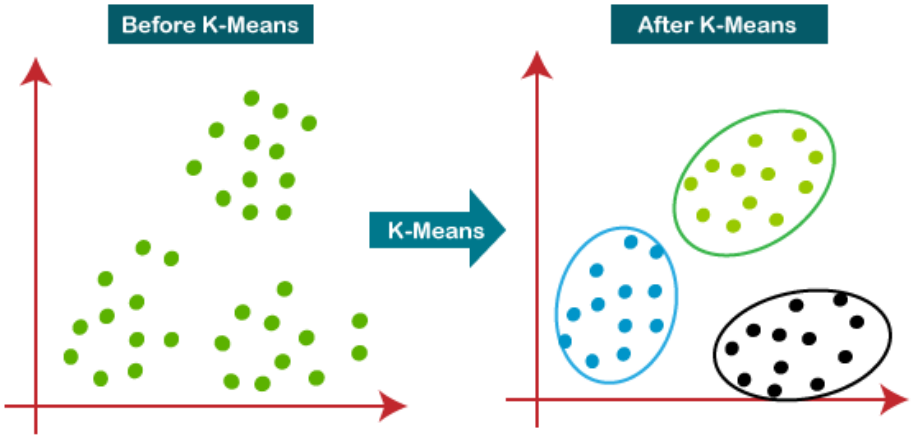


Figure 3 K-Mean clustering

According to (Shi et al., 2010), clustering is a technique for logically classifying raw data and searching for hidden patterns in datasets .Therefore, K-Mean modelling is suitable for the case study since it will cluster the students according to the level of addicted students based on the parameter given. Furthermore, the objective of the project is to predict how many undergraduate students that are addicted with digital activities. Since clustering is going to group the students based on the cluster size and score, it is easier to find the group of students that has high, medium and low addiction towards digital activities.

CHAPTER 3
METHODOLOGY & PROJECT ACTIVITIES

3.1 Methodology

Methodology is a logical and cohesive structure based on views, attitudes, and values that directs the choices researchers [or other users] make. This allows us to keep all initiatives on track and repeat successful parts while learning from mistakes, resulting in a continual improvement process. The purpose of using methodology for a project is to organize project time which will help the project to reach the objectives and able to minimize risk. Other than that, it will also help the project to be on track and keep on the timeline since the durations for the projects are only for 2 consecutive semesters and one semester for the development part.

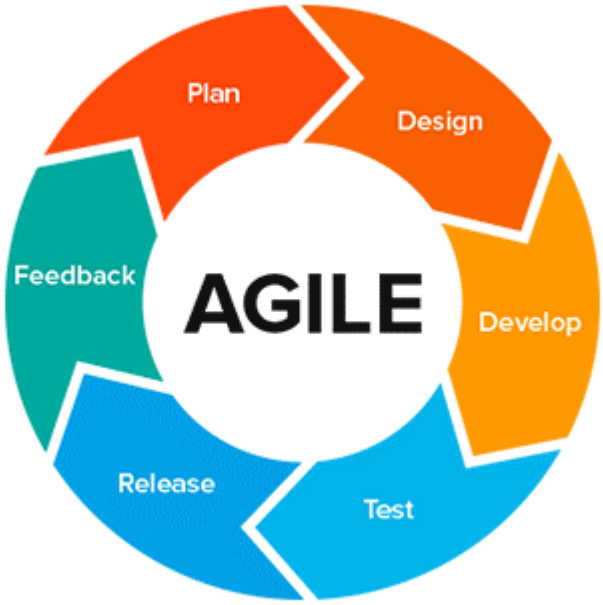


Figure 4 Agile methodology

The perfect methodology that matches with the project is agile methodology. Agile methodologies seek to create the optimal product by delivering small bits of functionality incrementally and often. The five phases in agile methodologies are plan, design, develop, test, release, and feedback:

i. **Planning:** Identify the problem statement and the objectives of the project by narrow down the scope. In this phase, the project planning and what will be done need to be clear and has been identify the purpose of the project. Furthermore, need to identify the software will be use and the methods for machine learning.

ii. **Design:** Plan how to build the project that will reach the objectives and able to solve the problem of the project into a product. The design phase is including the steps of create the survey and distribute to the target user for dataset collection purposes. Other than that, has identify how the dashboard will look like with the desired filters.

iii. **Develop:** In this phase, need to start train the dataset using Python language and using the correct method for modelling such as regression.

iv. **Test:** Testing will require to verify if the product function as per designed. The project will require to use multiple types of scoring methods to make sure the model is accurate.

v. **Release:** Deployment the stage of initial development, where the project is put into development and runs actual state-run.

3.2 Project activities

The project activities that will be implement in the project is CRISP-DM. This is because the project will require machine learning methods and modelling, and CRISP-DM is the suitable methodology for the dataset training.

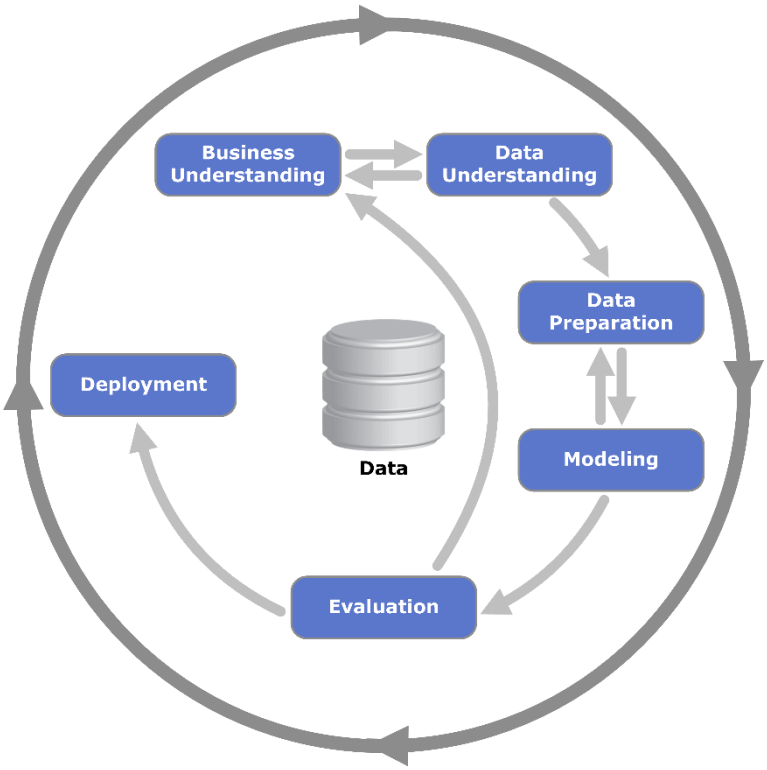


Figure 5 CRISP-DM methodology

CRISP-DM defines a framework for denoting data mining projects and sets out activities to be performed to complete a product or service. The activities consist of six phases:

i. **Business understanding:** The first stage in CRISP-DM process is the need to understand the whole project objectives and the goals to make sure the project is successful. In this process, need to set the project objectives which refers to the primary goal of the objective which is the objective is to create a Power Bi dashboard that relate with digital addiction during pandemic. Other than that, the activity on creating and distribute survey for collecting dataset are also a part of the business understanding process.

ii. **Data understanding:** In this stage, the process needs to acquire the data listed in the project resources. The activity in this phase acquires to have

data description report, explore data, verify data quality and data quality report. The activities for this phase in the project are to gather all the data through the responses from the survey and make sure it follows the data formats. In order to standardize the forms and avoid different answers from respondent, I use rating for the questions that correlate with the respondents' online activities. Other than that, need to find the quality of the data from the survey, either it can be used or remove. For example, the unnecessary questions from the survey are removed and replaced with questions that has more correlation with digital addiction. This step will be able to understand the datasets ad build the modelling.

```
plt.hist(df["Level SL"], bins=10)
plt.style.use('ggplot')
plt.title("Number of students affected by Social Life ")
plt.xlabel("Level of addiction")
plt.ylabel("Frequency")
plt.show()
```

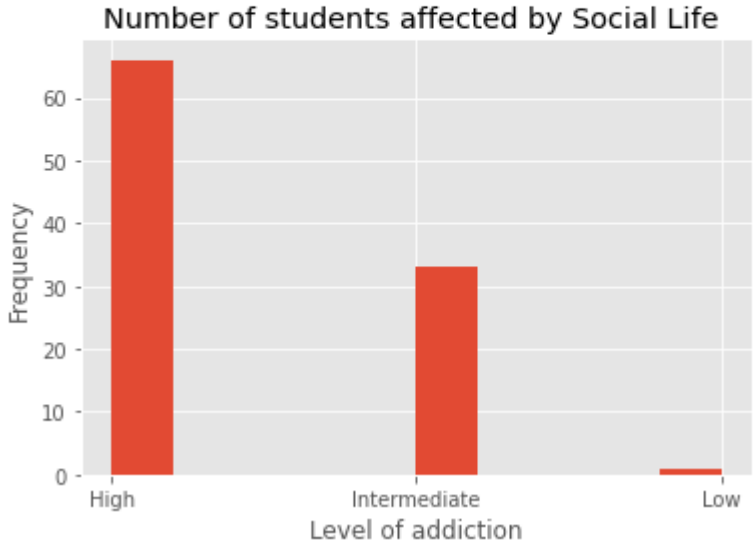


Figure 6 Create visualization for Addiction parameter.

iii. **Data preparation:** This is the stage where the need to decide the data that are needed for the analysis. The criteria that need to be consider are data mining goals, the quality of the data, and the technical constraints of the data. The activities in this phase are data cleaning, construct required data, and integrate data. As shown in the Figure 6 below, remove the null values columns is a part of the data preparation.

```
df.drop('Name', axis=1, inplace=True)
df.head()
```

Figure 7 Remove null values.

iv. **Modelling:** There are multiple types of data modelling that can be used for the machine learning, but the suitable model for the dataset and the constraints acquires a few steps. The activities in this phase are select modelling technique, generate test design, build a model, and assess model.

v. **Evaluation:** The evaluation purpose is to find the accuracy and generality of the model that has constructed from the previous activities where it will require to evaluate the result through assessment of data mining result. Next, it is also needed to review the process and determine the next steps. If the evaluation fails or does not fulfilling the criteria, the activities need to be repeated until it reaches the requirements.

vi. **Deployment:** In this deployment stage, after the completion of evaluation process need to determine the strategy for deployment. Other than that, it will require to plan monitoring and maintenance of the important issues of the data mining. The deployment stage will involve data visualization in the Power bi by creating a dashboard.

3.3 GANTT CHART

TASK	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23	W 24
Project proposal																								
Background & introduction																								
Objectives and literature review																								
Data collection and analysis																								
Methodology																								
Proposal defence																								
Interim report submission																								
Data understanding																								
Data Preparation																								
Modelling																								
Evaluation																								
Viva																								
Soft Bound																								
Hard Bound																								

The figure shows the complete planning for FYP 1 of the project from week 1 until week 24.

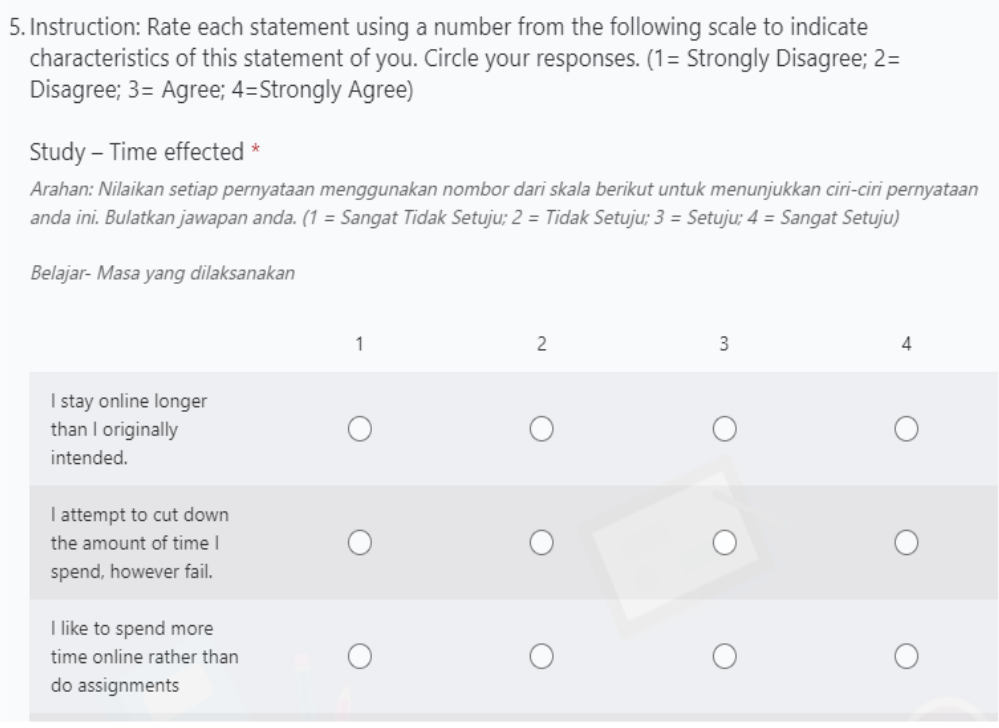
3.4 Tools and Software

The tools and software that are used in the project development are Anaconda, Jupyter notebook, and Power Bi. Meanwhile, the language used for the project development in Jupyter notebook is Python language.

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Dataset

The dataset is collected through survey that are distributed among undergraduate students in Malaysia. The questionnaire for the survey is build using the five parameters of digital addiction that has been approved by the experts which are Study-Time effected, Social Life, Psychological-Emotion, Addiction and Physical.



5. Instruction: Rate each statement using a number from the following scale to indicate characteristics of this statement of you. Circle your responses. (1= Strongly Disagree; 2= Disagree; 3= Agree; 4=Strongly Agree)

Study – Time effected *

Arahan: Nilaikan setiap pernyataan menggunakan nombor dari skala berikut untuk menunjukkan ciri-ciri pernyataan anda ini. Bulatkan jawapan anda. (1 = Sangat Tidak Setuju; 2 = Tidak Setuju; 3 = Setuju; 4 = Sangat Setuju)

Belajar- Masa yang dilaksanakan

	1	2	3	4
I stay online longer than I originally intended.	○	○	○	○
I attempt to cut down the amount of time I spend, however fail.	○	○	○	○
I like to spend more time online rather than do assignments	○	○	○	○

Figure 8 Example of Survey Questions

The survey distributed manage to collect 100 students' data in total. The survey was distributed for any undergraduate students that are currently studying.

4.2 K-Means Modelling

In order to build the model, the dataset need to build accordingly to make sure all the methods use will be accurate. Firstly, need to transform all the data into numerical data from strings since the scoring method will require to calculate the columns. Next, select the respective columns that will be used for building the modelling.

```
        'Level P_Intermediate',
data= df [cols_of_interest]
data.head()
```

Figure 9 Select specific column.

The next step is to scale the values so that it will give equal weight. Scaling is also crucial in clustering since the distance between points has an impact on how clusters form. By using the StandardScaler to convert our data frame into the numpy arrays below.

```
X = StandardScaler().fit_transform(data)
X
```

```
array([[-1.71481604,  0.12803688, -0.1498018 , ...,  0.57735027,
        -0.531085  , -0.17586311],
       [-1.68017329,  0.12803688, -0.1498018 , ...,  0.57735027,
        -0.531085  , -0.17586311],
       [-1.64553055,  1.40840568, -0.1498018 , ...,  0.57735027,
        -0.531085  , -0.17586311],
       ...,
       [ 1.64553055,  1.40840568, -0.1498018 , ..., -1.73205081,
         1.88293774, -0.17586311],
       [ 1.68017329, -1.15233192, -0.1498018 , ...,  0.57735027,
        -0.531085  , -0.17586311],
       [ 1.71481604,  0.12803688, -1.30212333, ...,  0.57735027,
        -0.531085  , -0.17586311]])
```

Figure 10 Standard column.

Next step is implementing the k-mean algorithm using sci-kit learn with the cluster size initiation of 12.

```
#Set number of clusters at initialisation time
k_means = KMeans(n_clusters=12)
#Run the clustering algorithm
model = k_means.fit(X)
model
#Generate cluster predictions and store in y_hat
y_hat = k_means.predict(X)
```

Figure 11 Implementation of K-Mean modelling.

4.3 Clustering Method

4.3.1 Silhouette Score

The separation distance between the generated clusters can be studied using silhouette analysis. The silhouette plot shows how close each point in one cluster is to points in neighbouring clusters, and so allows you to visually examine factors like cluster count. The range of this metric is [-1, 1]. When dealing with higher dimensions, the silhouette score comes in handy for validating the clustering algorithm's operation, as no other sort of visualisation can be used to validate clustering when the dimensions exceed three.

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

The Silhouette Coefficient is a two-score system that is defined for each sample:

a: The mean distance between a sample and the rest of the cluster's points.

b: The mean distance between a sample and the nearest cluster's other points.

```
1 from sklearn import metrics
2 labels = k_means.labels_
3 metrics.silhouette_score(X, labels, metric = 'euclidean')
```

```
0.03306434895117573
```

Figure 12 Silhouette score for 12 cluster sizes.

```
1 k_means_8 = KMeans(n_clusters=8)
2 model = k_means_8.fit(X)
3 y_hat_8 = k_means_8.predict(X)
```

```
1 labels_8 = k_means_8.labels_
2 metrics.silhouette_score(X, labels_8, metric = 'euclidean')
```

```
0.030366554245848727
```

Figure 13 Silhouette score for 8 cluster sizes

```
k_means_3 = KMeans(n_clusters=3)
model = k_means_3.fit(X)
y_hat_3 = k_means_3.predict(X)
```

```
labels_3 = k_means_3.labels_
metrics.silhouette_score(X, labels_3, metric = 'euclidean')
```

```
0.09653980347547625
```

Figure 11 Silhouette score for 3 cluster sizes

From the silhouette score the following table is the comparison of the scores.

Table 1 Silhouette score

Cluster size	**Silhouette score**
12	0.03306
8	0.0303667
3	0.0965398

It can be concluded that the cluster size 3 silhouette score is closer to 1. When the value is closer to one the clusters are well apart from each other. But if the score is -1 it is shown that means cluster are assigned in wrong way.

4.3.2 Calinski-Harabasz Score.

The Calinski-Harabasz index also known as the Variance Ratio Criterion, is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters, the higher the score, the better the performances.

For a set of data E of size n_E which has been clustered into k clusters, the Calinski-Harabasz score s is defined as the ratio of the between-clusters dispersion mean and the within-cluster dispersion:

$$s = \frac{\mathrm{tr}(B_k)}{\mathrm{tr}(W_k)} \times \frac{n_E - k}{k - 1}$$

where $\mathrm{tr}(B_k)$ is trace of the between group dispersion matrix and $\mathrm{tr}(W_k)$ is the trace of the within-cluster dispersion matrix defined by:

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T$$

$$B_k = \sum_{q=1}^{k} n_q (c_q - c_E)(c_q - c_E)^T$$

with C_q the set of points in cluster q, c_q the center of cluster q, c_E the center of E, and n_q the number of points in cluster q.

Figure 14 Calinski-Harabasz formula.

The advantages of Calinski-Harabasz score are because the score is higher when cluster are well dense and separated. Furthermore, it is faster to compute.

```
1 metrics.calinski_harabasz_score(X, labels)
```

```
4.168367351428027
```

Figure 12 Calinski Harabasz score for 12 cluster sizes.

```
1 metrics.calinski_harabasz_score(X, labels_8)
```

```
4.724464202216688
```

Figure 13 Calinski Harabasz score for 8 cluster sizes

```
1 metrics.calinski_harabasz_score(X, labels_3)
```

```
8.121480613228686
```

Figure 15 Calinski Harabasz score for 3 cluster sizes

Table 2 Calinski-Harabasz score

Cluster	Calinski-Harabasz score
12	4.16837
8	4.724464
3	8.1214806

From the table above, it can be concluded that cluster size 3 has the highest score with 8.1214806. Therefore, it is suitable to use cluster size 3 for the clustering. Compare with cluster size 12 and 8 which has huge gap in the scoring.

4.3.3 Elbow plot graph.

The distance between the mean of a cluster and the other data points in the cluster is at its shortest at what value of k, according to an elbow plot. There are important values which are inertia and distortion. The average of the Euclidean squared distance from the centroid of the respective clusters is used to calculate distortion. The sum of squared distances between samples and their closest cluster centre is inertia.

```
#finding optimal number of clusters using the elbow method
from sklearn.cluster import KMeans
wcss_list= []  #Initializing the list for the values of WCSS

#Using for loop for iterations from 1 to 10.
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state= 42)
    kmeans.fit(x)
    wcss_list.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss_list)
plt.title('The Elobw Method Graph')
plt.xlabel('Number of clusters(k)')
plt.ylabel('wcss_list')
plt.show()
```

Figure 16 Elbow Plot

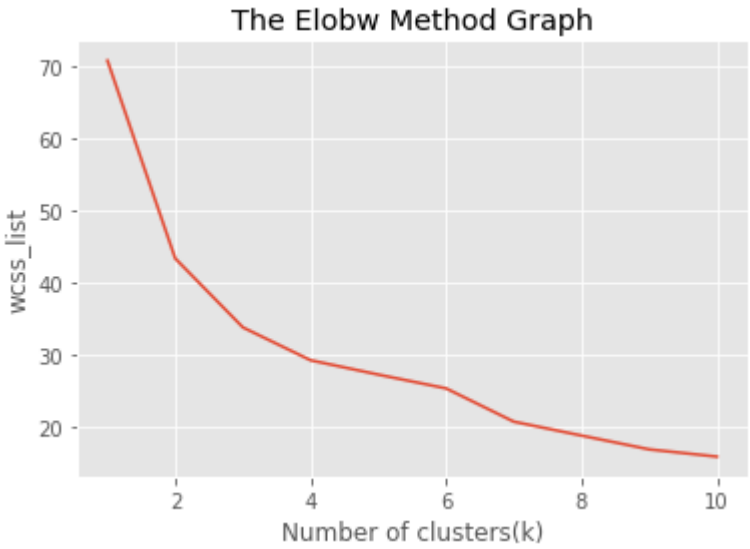


Figure 17 Elbow Plot Graph.

From the figure above, it shows that the drop of the square distance starts to slow down at 3. Therefore, the optimal number of k is 3 by using elbow plot.

4.3.4 Cluster each of the students based on the cluster score.

From the Silhouete score and Harabasz-Calinski scoring method, it is suitable to use cluster size of 3 as it fits with the dataset. Therefore, the cluster size of 3 is use along with the K-mean modelling in order to cluster the group of students.

```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df [cols_of_interest])
y_predicted
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

```
df['cluster']=y_predicted
df.head()
```

Figure 18 Predict each student cluster.

The average of all points (elements) in the cluster makes up the cluster's centre. If the points are pixels in an image, then the centre of the cluster will be a pixel from that image.

```
km.cluster_centers_
```

```
array([[ 1.75000000e+01,  2.05882353e+00,  2.23529412e+00,
         1.70588235e+00,  3.35294118e+00,  3.02941176e+00,
         3.05882353e+00,  1.70588235e+00,  2.73529412e+00,
         3.11764706e+00,  2.32352941e+00,  2.23529412e+00,
         2.55882353e+00,  2.32352941e+00,  2.50000000e+00,
```

Figure 19 Get each cluster center.

The next step is to get the total number of students of each cluster size. As shown in the figure below, cluster 0 has 34 students, cluster 1 has 31 students and cluster 3 has 35 students. Each of the cluster size will be label based on the level of addiction. For an example, cluster 0= Low level of addiction, cluster 1= Intermediate level of addiction, and cluster 2 = High level of addiction.

```
1 cluster = df['cluster']
2 counts = cluster.value_counts()
3 counts
```

```
2    35
0    34
1    31
Name: cluster, dtype: int64
```

Figure 20 Total number of students for each cluster size.

4.4 Data Visualization

Data visualization for the prediction is using PowerBi tools by creating an interactive dashboard. The tools able to filter based on preferences to see the total number of students that has low, intermediate, and high level of addiction. The process in doing so is by exporting the data from jupyter notebook into a csv file. The csv file will be used to create the dashboard.

```
1 df.to_csv('Cluster_Output', sep='\t')
```

Figure 21 Change to CSV file.

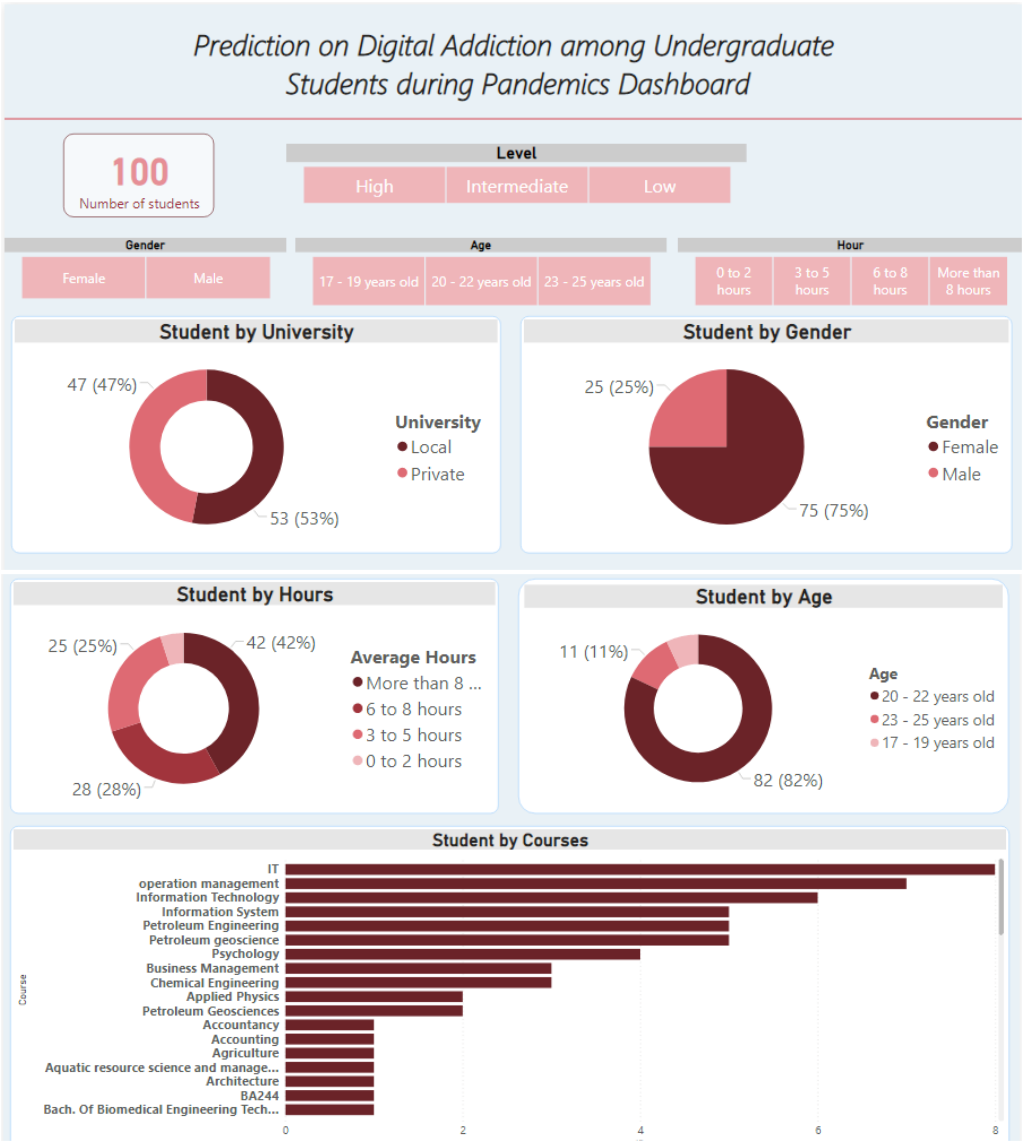


Figure 22 Dashboard for the prediction.

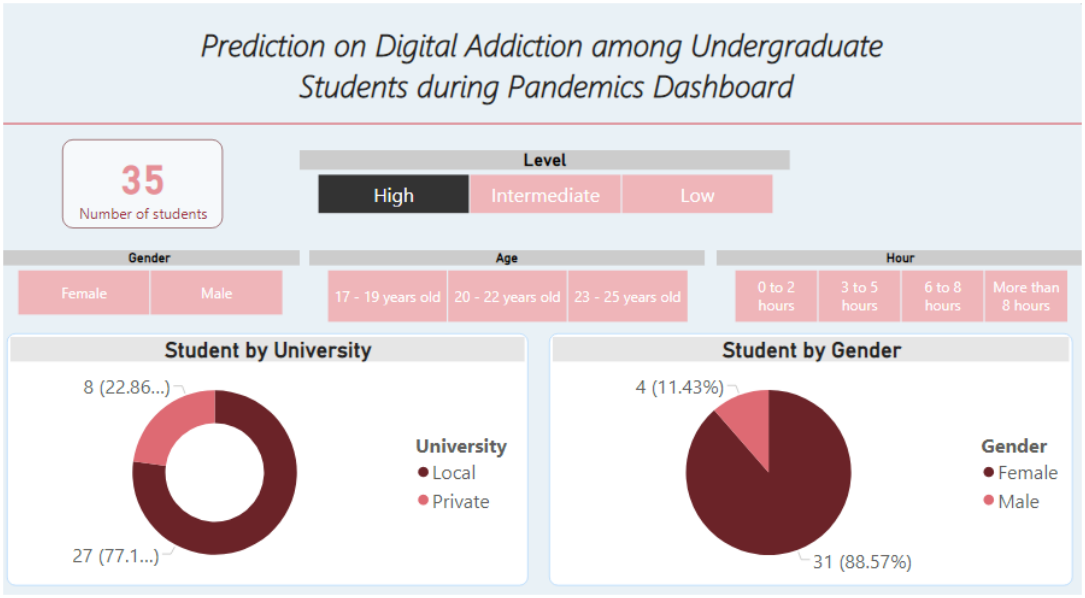


Figure 23 Students that has High level of addiction.

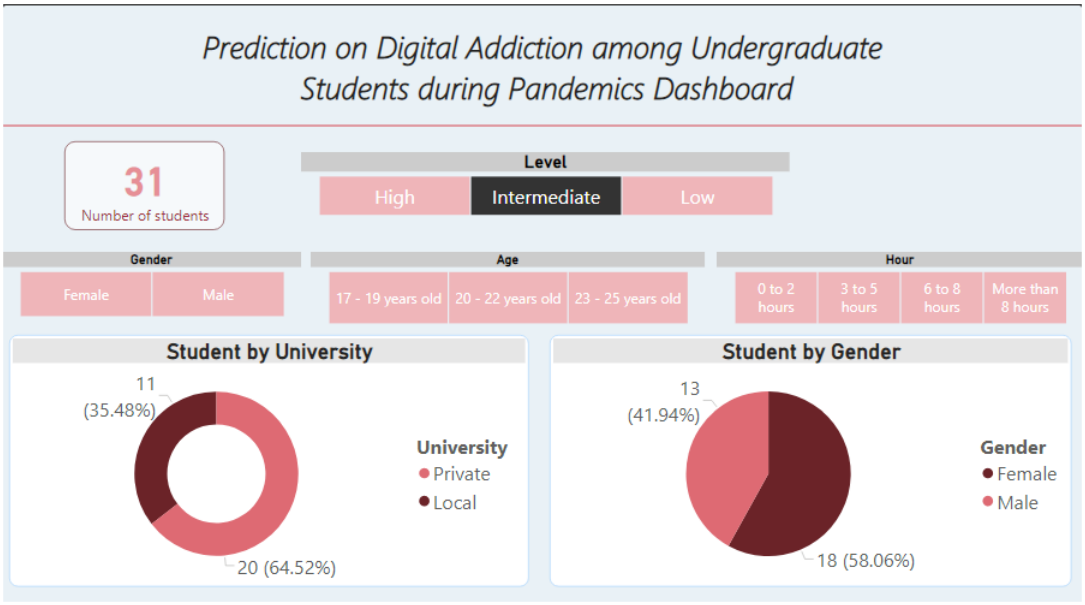


Figure 24 Students that has Intermediate level of addiction

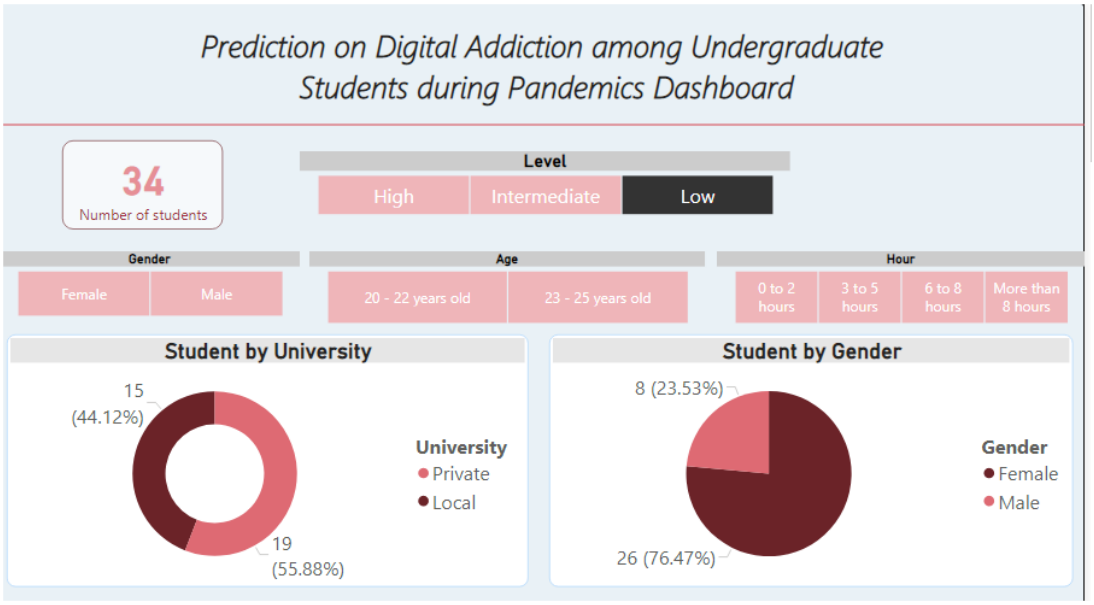


Figure 25 Students that has Low level of addiction

4.5 Prediction on Digital Addiction

Table 3 Total addicted students based on the level

Level of Addiction	Number of students
Low	34
Intermediate	31
High	35
Total	100 students

4.5.1 Low cluster of addiction (Cluster = 0)

Table 4 Students that addicted based on gender for Low addiction

Gender	Number of students
Female	26
Male	8

Table 5 Students that addicted based on University for Low addiction

Type of university	Number of students
Private	19
Local	15

Table 6 Students that addicted based on Hour Spend for Low addiction

Hours Spend	Number of students
0 to 2 hours	1
3 to 5 hours	10
6 to 8 hours	11
More than 8 hours	12

Table 7 Students that addicted based on Age for Low addiction

Age	Number of students
17 - 19 years old	0
20 – 22 years old	32
23 – 25 years old	2

Table 8 Students that addicted based on Parameter for Low addiction

Rating\ Parameter	Psychological- Emotion	Addiction	Study-Time effected	Social Life	Psychological - Emotion
1	8	12	0	8	1
2	6	10	3	13	3
3	16	11	14	7	15
4	4	1	16	6	15

4.5.2 Intermediate cluster of addiction (Cluster = 1)

Table 9 Students that addicted based on Gender for Intermediate addiction

Gender	Number of students
Female	18
Male	13

Table 10 Students that addicted based on University for Intermediate addiction

Type of university	Number of students
Private	20
Local	11

Table 11 Students that addicted based on Hours Spend for Intermediate addiction

Hours Spend	Number of students
0 to 2 hours	3
3 to 5 hours	6
6 to 8 hours	10
More than 8 hours	12

Table 12 Students that addicted based on Age for Intermediate addiction

Age	Number of students
17 - 19 years old	2
20 – 22 years old	23
23 – 25 years old	6

Table 13 Students that addicted based on Parameter for Intermediate addiction

Rating\ Parameter	Psychological - Emotion	Addiction	Study-Time effected	Social Life	Psychological - Emotion
1	5	6	2	5	2
2	11	11	8	7	11
3	13	12	14	13	14
4	2	2	7	6	4

4.5.3 High cluster of addiction (Cluster = 3)

Table 14 Students that addicted based on Gender for High addiction

Gender	Number of students
Female	31
Male	4

Table 15 Students that addicted based on University for High addiction

Type of university	Number of students
Private	8
Local	27

Table 16 Students that addicted based on Hours Spend for High addiction

Hours Spend	Number of students
0 to 2 hours	1
3 to 5 hours	5
6 to 8 hours	11
More than 8 hours	18

Table 17 Students that addicted based on Age for High addiction

Age	Number of students
17 - 19 years old	5
20 – 22 years old	27
23 – 25 years old	3

Table 18 Students that addicted based on Parameters for High addiction

Rating\ Parameter	Psychological- Emotion	Addiction	Study-Time effected	Social Life	Psychological - Emotion
1	9	9	4	10	4
2	12	15	4	12	9
3	10	10	14	12	9
4	4	1	13	1	13

CHAPTER 5
CONCLUSION AND RECOMM ENDATION

CONCLUSION

As a conclusion, the project will acquire to use machine learning algorithm to predict the number of undergraduate students who has a digital addiction. Machine learning modelling are the best implementation because it will help to identify the trend and pattern on how many students has digital addiction and how bad their addiction towards online activities based on the parameters. Besides, machine learning algorithm can be easily use in Power Bi to create data visualization with the dataset available. This research also has concluded how digital addiction effect undergraduate students negatively if there is no limitations and controls of their online activities. From the prediction above, it shows that the number of students that most of the student has high addiction with the digital activities with total number of 35 students out of 100 students. It is clearly shown that student's life is at risk if it is not taken care of in early stages. The impact of digital addiction has the same negative impact with other addictions such as drugs, gambling and more towards a particular person.

Furthermore, from the project we can conclude that k-means algorithm is suitable for the prediction studies as it will cluster them into a group using the suitable scoring method. In additional, using Power Bi for data visualization able to capture the details of the students that has digital addiction and what are the most affected areas based on the five parameters.

REFERENCES

Akhter, S. A. (2017). *Using machine learning to predict potential online gambling addicts* .

Aziz, N., Nordin, M. J., Abdulkadir, S. J., & Salih, M. M. M. (2021). Digital addiction: Systematic review of computer game addiction impact on adolescent physical health. *Electronics (Switzerland)*, *10*(9), 1–18. https://doi.org/10.3390/electronics10090996

Bhardwaj, A. (2020, May 27). *Silhouette coefficient : Validating clustering techniques*. Medium. Retrieved November 28, 2021, from https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c.

Piech, C. (n.d.). CS221. Retrieved November 28, 2021, from https://stanford.edu/~cpiech/cs221/handouts/kmeans.html.

Elbow method¶. Elbow Method - Yellowbrick v1.3.post1 documentation. (n.d.). Retrieved November 28, 2021, from https://www.scikit-yb.org/en/latest/api/cluster/elbow.html.

Evaluation metrics for machine learning for Data scientists. Analytics Vidhya. (2020, November 23). Retrieved November 28, 2021, from https://www.analyticsvidhya.com/blog/2020/10/quick-guide-to-evaluation-metrics-for-supervised-and-unsupervised-machine-learning/.

Kaushiva, M. (2020, January 3). *Clustering with K-means*. Medium. Retrieved November 28, 2021, from https://towardsdatascience.com/clustering-with-k-means-1e07a8bfb7ca.

Klochko, O. (n.d.). *An empirical comparison of machine learning clustering methods in the study of Internet addiction among*. 58–75.

Latif, R. A., Aziz, N. A., Taufik, M., & Jalil, A.-D. (2017). Impact of Online Games Among Undergraduate Students. *ICOCI Kuala Lumpur. Universiti Utara Malaysia*, *028*, 25–27. http://icoci.cms.net.my/PROCEEDINGS/2017/Pdf_Version_Chap11e/PID28-523-532e.pdf

Mak, K. K., Lee, K., & Park, C. (2019). Applications of machine learning in addiction studies: A systematic review. *Psychiatry Research*, *275*(January), 53–60. https://doi.org/10.1016/j.psychres.2019.03.001

ML clustering: When to use cluster analysis, when to avoid it. Explorium. (2021, May 18). Retrieved November 28, 2021, from https://www.explorium.ai/blog/clustering-when-you-should-use-it-and-avoid-it/.

Rahayu, F. S., Nugroho, L. E., Ferdiana, R., & Setyohadi, D. B. (2020). Research trend on the use of it in digital addiction: An investigation using a systematic literature review. *Future Internet, 12*(10), 1–23. https://doi.org/10.3390/fi12100174

Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010*, 63–67. https://doi.org/10.1109/IITSI.2010.74

APPENDICES

Appendix 1. Survey Questionnaire for dataset collection (1)

Dear Respondent,

The questionnaire is a part of study being conducted under Computer and Information Sciences Department of Universiti Teknologi PETRONAS. This study related to Digital Addiction Assessment. Digital addiction (DA) can be described as an addiction towards listening, watching, or playing for entertainment purposes using an electronic device. It would be highly appreciated if you kindly fill in the questionnaire. All information will be kept confidential and neither any name nor details will be identified in reports of this research. Thank you in advance for your co-operation.

Respondent Details.

University	
Course	
Steam ID (online game ID)	
Age	☐17-19 ☐ 20- 22 ☐ 23-25
Gender	

1. I spend an average of _____ on the Internet per day.

☐	0 to 2 hours	☐	6 to 8 hours
☐	3 to 5 hours	☐	More than 8 hours

2. I usually connect to the Internet at

☐	Home	☐	Hostel
☐	Cyber cafe	☐	College
☐	Mall	☐	Others _____

Appendix 2. Survey Questionnaire for dataset collection (2)

3. Instruction: From range 1 to 3, choose your favorite online activities. No. 1 indicates the most favorite.

☐	Online Shopping
☐	Online computer games e.g., PUBG , Call of Duty, Fortnite Battle Royale. Mobile Legend, Counter Strike: Global Offensive
☐	Social networking sites e.g., Instagram, Twitter, Tiktok

4. Did you have experience perform any online game addiction test?

☐	Yes
☐	No

Instruction: Rate each statement using a number of from the following scale to indicate characteristic this statement of you. Circle your responses. (1= Strongly Disagree; 2= Disagree; 3= Agree; 4=Strongly Agree)

	Items	1	2	3	4
Study – Time effected					
1.	I stay online longer than I originally intended				
2.	I attempt to cut down the amount of time I spend, however fail.				
3.	I like to spend more time online rather than do assignments.				
4.	I choose to skip classes and misses assignment deadline in order to spend more time online				
5.	I have insufficient time to study for exam due to online activities.				
6.	I lose sleep due to late-night online activities				

Appendix 3. Survey Questionnaire for dataset collection (3)

Social Life					
7.	I feel more connected to my online friends than real world friends.				
8.	I spend less time with my family due to spend more time playing online games online activities				
9.	I spend my saving on buying unnecessary stuff on online platform.				
10.	I share my online activities with my friend more than assignments and tasks.				
11.	I do other online activities during online class and meeting				
Psychological-Emotion					
12.	I become annoyed, angry or yell if someone bothers me while on online activities				
13.	I feel depressed, moody, or nervous when I am off-line, but these feelings go away once I am back online.				
14.	I choose to spend times on online activities when I get stress or to avoid thinking of life problems.				
15.	Focus on online activities lets me vent and relieve stress from the day.				
Addiction					
16.	Others in my life complain to me about the amount of time I spend on online activities.				
17.	I find myself saying 'just a few minutes' for online activities.				
18.	Online activities let me forget some of the real-life problems I have.				
19.	I fear that life without online activities would be boring, empty, and joyless.				
20.	The way I am in online platform is the way I am in real life.				
Physical					
21.	I feel back pain or neck pain if I spend hours on online activities.				
22.	I ate more at midnight than normal eating schedule.				
23.	After spending long hours on-screen time, my eyes become dry and eyesight problems (e.g. fatigue, blurry vision, and headaches)				

Appendix 4. Import Dataset to Jupyter.

```
1 import pandas as pd
2 df=pd.read_excel ('Digital Addiction.xlsx')
3 df
```

	ID	Start time	Completion time	Email	Name	I spend an average of _____ on the Internet per day.	I usually connect to the Internet at	Online Shopping e.g: Shopee, Lazada, Zalora, etc	Online computer games e.g: PUBG , Call of Duty, Fortnite Battle Royale. Mobile Legend, Counter Strike: Global Offensive	Social networking sites e.g: Instagram, Twitter, Tiktok	...	Online activities let me forget some of the real-life problems I have.\t\t\t\t	I fear that life without online activities would be boring, empty, and joyless.\t\t\t\t	The way I am in online platform is the way I am in real life.
0	1	2021-07-02 04:59:37	2021-07-02 05:02:13	anonymous	NaN	3 to 5 hours	Cyber cafe ;Mall;	2	2	2	...	3	3	2
1	2	2021-07-09 12:45:28	2021-07-09 12:46:55	anonymous	NaN	3 to 5 hours	Hostel ;College/University;Home ;	2	2	2	...	3	2	3

Appendix 5. Remove null columns.

```
In [2]:
1 df.drop('Name', axis=1, inplace=True)
2 df.head()
```

Out[2]:

	ID	Start time	Completion time	Email	I spend an average of _____ on the Internet per day.	I usually connect to the Internet at	Online Shopping e.g: Shopee, Lazada, Zalora, etc	Online computer games e.g: PUBG , Call of Duty, Fortnite Battle Royale. Mobile Legend, Counter Strike: Global Offensive	Social networking sites e.g: Instagram, Twitter, Tiktok
0	1	2021-07-02 04:59:37	2021-07-02 05:02:13	anonymous	3 to 5 hours	Cyber cafe ;Mall;	2	2	2
1	2	2021-07-09 12:45:28	2021-07-09 12:46:55	anonymous	3 to 5 hours	Hostel ;College/University;Home	2	2	2

Appendix 6. Rename columns columns.

```
In [5]:
1 df.rename(columns= {'I spend an average of _____ on the Internet per day.' : 'HourSpend'}, inplace = True)
2 df.head()
```

Out[5]:

	ID	Start time	Completion time	HourSpend	Connect	Online Shopping e.g: Shopee, Lazada, Zalora, etc	Online computer games e.g: PUBG , Call of Duty, Fortnite Battle Royale. Mobile Legend, Counter Strike: Global Offensive	Social networking sites e.g: Instagram, Twitter, Tiktok	Did you have experience perform any online game addiction test?	I stay online longer than I originally intended.	...	Online activities let me forget some of the real-life problems I have.\t\t\t\t	I fear that life without online activities would be boring, empty, and joyless.\t\t\t\t
0	1	2021-07-02 04:59:37	2021-07-02 05:02:13	3 to 5 hours	Cyber cafe ;Mall;	2	2	2	No	3	...	3	3

Appendix 7. Datasets describe.

```
In [8]:  1 df.describe()
```

Out[8]:

	ID	OnlineShopping	OnlineGame	SocialNetwork	I stay online longer than I originally intended.	I attempt to cut down the amount of time I spend, however fail.	I like to spend more time online rather than do assignments \t\t\t\t	I choose to skip classes and misses assignment deadline, to spend more time online	I have insufficient time to study for exam due to online activities	I lose sleep due to late-night online activities\t\t\t\t	...	I choose spe times onl activit when I stress to av thinking probler
count	100.000000	100.00000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	...	100.0000
mean	50.500000	1.90000	2.130000	1.750000	3.080000	2.830000	2.880000	1.890000	2.570000	2.930000	...	2.7700
std	29.011492	0.78496	0.872185	0.845368	0.872475	0.910711	0.879394	1.072098	0.934793	1.007597	...	0.8859
min	1.000000	1.00000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.0000
25%	25.750000	1.00000	1.000000	1.000000	3.000000	2.000000	2.000000	1.000000	2.000000	2.000000	...	2.0000
50%	50.500000	2.00000	2.000000	1.000000	3.000000	3.000000	3.000000	1.500000	3.000000	3.000000	...	3.0000
75%	75.250000	3.00000	3.000000	3.000000	4.000000	4.000000	4.000000	3.000000	3.000000	4.000000	...	3.0000
max	100.000000	3.00000	3.000000	3.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	...	4.0000

8 rows × 27 columns

Appendix 8. Dataset's info.

```
In [9]:  1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 36 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   ID
100 non-null    int64
 1   Start time
100 non-null    datetime64[ns]
 2   Completion time
100 non-null    datetime64[ns]
 3   HourSpend
100 non-null    object
 4   Connect
100 non-null    object
 5   OnlineShopping
100 non-null    int64
```

Appendix 9. Calculate parameters.

```
dflist = ['I stay online longer than I originally intended.', 'I attempt to cut down the amount of time I spend, however
df['Study-Time effected'] = df[dflist].sum(axis=1,skipna=False)
df['Level Study']= ["Low" if i<9 else "Intermediate" if i==9 or i<17 else "High" for i in df['Study-Time effected']]
df
```

	ID	Start time	Completion time	HourSpend	Connect	OnlineShopping	OnlineGame	SocialNetwork	Did you have experience perform any online game addiction test?	I stay online longer than I originally intended.	...	The way I am in online platform is the way I am in real life.	I pai pain activi
0	1	2021-07-02 04:59:37	2021-07-02 05:02:13	3 to 5 hours	Cyber cafe ;Mall;	2	2	2	No	3	...	2	
1	2	2021-07-09 12:45:28	2021-07-09 12:46:55	3 to 5 hours	Hostel ;College/University;Home ;	2	2	2	No	3	...	3	
2	3	2021-07-09 12:45:30	2021-07-09 12:47:48	3 to 5 hours	Cyber cafe ;Mall;Hostel ;	3	2	2	No	4	...	4	

Appendix 10. Group based on level.

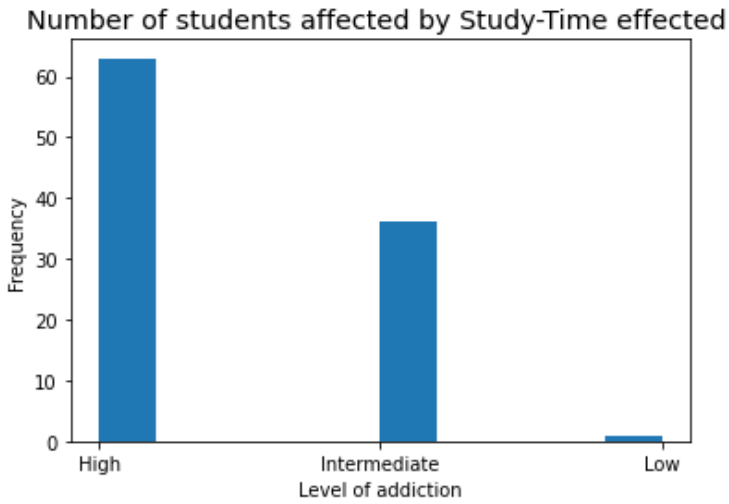
```
o = df['Level Study']
counts = o.value_counts()
counts
```

```
High            63
Intermediate    36
Low              1
Name: Level Study, dtype: int64
```

Appendix 11. Plot graph for parameters.

In [13]:

```
import matplotlib.pyplot as plt
plt.hist(df["Level Study"], bins=10)
plt.style.use('ggplot')
plt.title("Number of students affected by Study-Time effected ")
plt.xlabel("Level of addiction")
plt.ylabel("Frequency")
plt.show()
```



Appendix 12. One hot encoded for string columns.

```
g= pd.get_dummies(df["Age"],prefix= 'Age' )
df= df.join(g)
df
```

stay nline nger nan l nally ded.	...	Course_Psychology	Course_Science geomatics	Course_business	Course_civil eng	Course_engineering	Course_law of accounting	Course_operation management	Age_17 - 19 years old	Age_20 - 22 years old	Age_23 - 25 years old
3	...	0	0	0	0	0	0	0	0	1	0
3	...	0	0	0	0	0	0	0	0	1	0
4	...	0	0	0	0	0	0	0	0	1	0
3	...	0	0	0	0	0	0	0	0	1	0

Appendix 13. Import libraries.

```
import seaborn as sns
import sklearn
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn import metrics
from sklearn.cluster import AgglomerativeClustering
```

```
df.columns
```

```
Index(['ID', 'Start time', 'Completion time', 'HourSpend', 'Connect',
       'OnlineShopping', 'OnlineGame', 'SocialNetwork',
       'Did you have experience perform any online game addiction test?',
       'I stay online longer than I originally intended.',
       ...
       'Course_business ', 'Course_civil eng', 'Course_engineering',
       'Course_law of accounting', 'Course_operation management',
       'Age_17 - 19 years old', 'Age_20 - 22 years old',
       'Age_23 - 25 years old', 'Gender_Female', 'Gender_Male'],
      dtype='object', length=168)
```