ABSTRACT

Mental disorders may have various causes, often these diseases have a complex mix of causes. However, certain aspects of mental health conditions affect particular individuals. Childhood abandonment, depression, extreme or long-term stress are all causes that can contribute to poor mental health. This project proposed a classification based on mental health disorders. The task of classification of mental disorders is vital to neuroscientist and clinicians because it helps to identify some of the key indicators for mental health disorders. K- Nearest Neighbours (KNN) algorithm is used to solve the classification problem. A dataset of mental health disorders is obtained by conducting a mental health survey is used for training and testing medium. The objective, problem statement, scope, and significance of this project are explained in detail. The dataset is split into a few categories in order to determine the highest accuracy in the modelling phase. Experiments are tested rigorously with a few tweaks on the value of the parameter, which produce a few results. The highest accuracy for training data is 97.75% in 80% of train data. The accuracy is 75% in 20% of the test data. Besides, conclusion and recommendation are explained further in this report.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE **PAGE**

LIST OF TABLES

TABLE **PAGE**

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
DM	Data Mining
DT	Decision Trees
EDA	Exploratory Data Analysis
KNN	K Nearest Neighbour
ML	Machine Learning
NB	Naive Bayes
SVM	Support Vector Machines

1.0 Introduction

Mental health disorders, also known as mental illness, refer to a wide range of mental health conditions that negatively impact the mood, thought and behaviour of a person. Most people have mental health concerns from time to time. Nevertheless, mental health concerns became a mental illness when persistent signs and symptoms cause recurring stress and affect someone's ability to function. Several disorders have been identified with signs and symptoms that vary widely across different disorders, depending on the condition, circumstances and other factors involved.

Some of the main groups of mental health disorders are mood disorders that include depression or bipolar disorder in which the general emotional state is distorted and inconsistent with the situations. As a consequence, daily activities may be interrupted as the sufferers become extremely sad, depressed, or having periods of depression alternating with being excessively happy. The next group of mental disorders is anxiety disorders, an emotional disorder that can affect anyone at any age and is characterised by intense feelings of anxiety and fear. Anxiety can be described as constant and chronic worrying about future events, and persistent fear reaction in the current events that may cause physical symptoms, such as a fast heart rate and shakiness.

In the context of personality disorders, the sufferers might have a rigid and unhealthy pattern of thinking, functioning, and behaving, as well as difficulties in perceiving and relating to situations and people. Conversely, psychotic disorders such as schizophrenia are severe mental disorders that cause abnormal thinking and perceptions. People with psychoses lose touch with reality, and it can be portrayed by two main symptoms, delusions, and hallucinations. Delusions are false beliefs, such as the illusion that someone is conspiring against others, or that the Television is sending subliminal messages. On the contrary, hallucinations are false perceptions such as hearing, seeing, or sensing something that does not exist. Eating disorders are also a category of mental disorders characterised by irregular eating habits and severe anxiety or concern over body weight or shape, along with inadequate or excessive intake of food, which may ultimately harm the well-being of the individual. The last group of mental disorders were substance abuse disorders that may affect the brain and behaviour of a person. This type of disorder may lead to an inability to control the use of a legal or illegal drug or medication. Substances such as alcohol, marijuana and nicotine are considered to be drugs.

Issues related to mental health can affect different people in different ways. Factors such as age, lifestyle, and genetics may contribute to a person's likelihood of developing a mental disorder. Physical and mental health are equally important components of overall health at every stage of life, from childhood and adolescence through adulthood. In order to assess the status and frequency of mental health disorders, a survey is conducted by recording age group, gender, states, education level, employment status, marital status, family history, dietary behaviours, physical activity, early symptoms and past medical history of the respondents. Data relating to the survey are then evaluated using a machine-learning algorithm to gain insights into groups that are more prone to develop a mental disorder. Moreover, by identifying causes, signs and symptoms of mental disorder, this project is capable of discovering preventative measures to combat mental illnesses.

2.0 Problem Statements

Mental Health illness is expected to be the second-largest health problem affecting Malaysians after heart disease. According to the recent National Health and Morbidity Survey, every three of 10 adults aged 16 years and above in Malaysia suffered from some form of mental health issues. Mental health problems can have a wide range of causes. These disorders sometimes have a complicated combination of factors, although different people may be more likely to be affected by certain factors of mental health disorders.

Factors that can lead to poor mental health are childhood abuse, loneliness, severe or long-term stress, and others. In most cases, people who suffered from mental health disorders do not realise that they have symptoms. This can lead to a bigger problem as what they suffered might be worsening. However, many people are unaware of their level of mental health and did not respond quickly to their problems.

When it comes to the conventional way of dealing with mental health disorders, doctors or medical practitioners are having difficulty such as they have to consume much time when they are trying to analyse with a patient's data regarding their mental health.

Apart from that, there will usually be multiple sources of information when it comes to handling the patient. As a consequence, there will be a possibility of discrepancy when deducting new data. Thus, the problem could arrive when the solution is much needed to treat the patient. In the long run, it will be harder to conclude to a centralised system where all the patients are prated considerably.

Lastly, to the advancement of the data technology, it is only a matter of time before the medical and psychological time are implemented in Artificial Intelligent (AI) and Big data. Therefore, the conventional system of managing Mental Health Disorders will be a hurdle because everything will be assisted with machine learning. The age of data science is the future, and hence, this practice should be a part of technological advancement alike.

3.0 Project Objective

Identifying objectives can help what needs to be done and as a general guideline for the model development. The objectives are:-

i. To design the model that trains the datasets of mental health disorders collected from citizens in Malaysia using a machine-learning algorithm. ii. To develop a prediction model that can analyse the data related to surveys in which groups are more likely to suffer from mental health disorders.

iii. To test the accuracy of the prediction model of mental health disorder in ensuring that the machine recognises correct patterns in the data.

4.0 Project Scope

The scope of users of the system will be neuroscientists and clinicians that mainly practised in mental health disorder domain. The machine learning system will use KNN algorithm to model data for classification and clustering. Furthermore, the system will also use python computer language and will be developed using Jupyter application. One of the advantages is that machine learning helps clinicians identify who could be at risk for a specific condition. The main scope of this project is, therefore, to identify some of the key indicators for mental health disorders before they may set in.

5.0 Significant of Project

Through the advancement of machine learning technology in psychological science, the humankind could benefit prominently in assisting the study of mental health. The usage of Artificial Intelligence (AI), especially in machine learning prediction, could support mental health advocates such as doctors or practitioner to understand better how to cure and cooperate patients in a more efficient way. Computerised technology has gone a long way in biomedical science. It is only a matter of time before AI is implemented in psychological science. This application would result in better decision making and searching for a suitable cure alike. This study could, in a way assist in understanding better of implementing machine learning algorithms in health and medicinal sector in our country, especially when it comes to mental health.

6.0 Literature Review

The research on applying machine learning techniques in mental health diagnoses have started in the nineteen eighties (Sumathi, & Poorna, 2016). The advances in technology, such as social media, smartphones, wearables, and neuroimaging, have allowed mental health researchers and clinicians to collect a vast range of data at a rapidly growing rate as mentioned by Mohr, Zhang, and Schueller (2017). A robust technique that has emerged to analyse the mental health data is machine learning (ML), which aims to construct systems that can automatically improve through experience using advanced statistical and probabilistic techniques as stated by Jordan & Mitchell (2015).

The study stated that most of the articles implemented one technique only on mental health disorders data for prediction and investigation which is using machine learning techniques (Shatte, Hutchinson & Teague, 2019). However, some authors combined the use of classification, unsupervised learning, and other novel techniques. ML techniques are including supervised learning and classification approaches such as support vector machines (SVM), naive Bayes (NB), decision trees (DT).

While unsupervised and clustering approaches such as k-nearest neighbors (kNN), k-means clustering, text analysis, sentiment analysis, and novel techniques, including techniques based on deep learning and a range of custom ML methods devised for specific domains (Shatte et al., 2019). It was observed that ML applications were also evident across a range of mental health conditions, including depression, Alzheimer's disease, and another cognitive decline, schizophrenia, stress, and suicide (Bone, Chaspari, Gibson & Narayanan, 2017). The data types used to develop ML models included imaging data, survey data, mobile and wearable sensor data, and social media data as referred to Luo, Wu, Gopukumar & Zhao (2016).

Overall, machine learning (ML) demonstrates the potential to improve the efficiency of mental health clinical and research processes and to assist in generating new insights into health and well-being. Other than that, there also some challenges for consideration when using the ML techniques in mental health applications (Srividya, Mohanavalli & Bhalaji, 2018). ML models are inevitably limited by the quality of the data used to develop a model. Moreover, Shatte et al., (2019) also stated that many ML techniques also require access to training the data sets, which may require a greater collaboration of data.

7.0 Methodology

In this section, all the design and development in the methodology including Exploratory Data Analysis (EDA), Data Pre-Processing or Data Mining, Features or Parameters and Target or label will be described.

7.1 EDA

Exploratory Data Analysis (EDA) refers to the vital process of conducting preliminary investigations of data to find patterns, to test hypotheses, and to examine assumptions with the help of summary statistics and graph representations as stated by Patil (2018). Exploratory data analysis with visualisation was performed to understand the number of reported cases.

There are several useful tools for EDA. Most EDA techniques are graphical with some quantitative methods. In our project Prediction of Mental Health Disorders, the typical graphic techniques used in EDA are histogram, scatter plot, pie chart, and confusion matrix. To build on all this graphical data, we use Python as software that is an open-source programming language widely used in machine learning.

Figures 7.1 through Figure 7.5 show the EDA graphical techniques used in our project. Figure 7.1 and Figure 7.2 show the histogram of the applicant by gender and the probability histogram for the mental health disorders by gender. According to the histogram above, our survey respondents were more female than male, and the probability of mental disorders was similar for both genders.

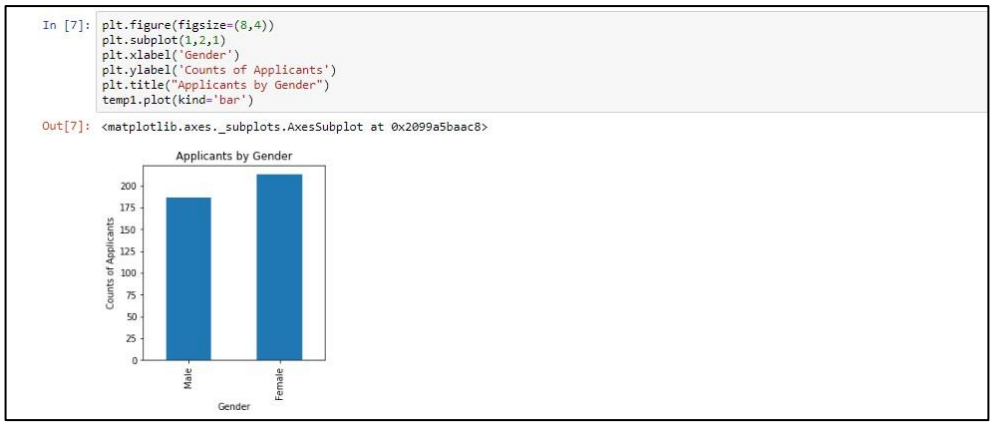

Figure 7.1 Histogram of Applicants by Gender

Figure 7.2 Histogram of Probability of Mental Disorder by Gender

Besides, Figure 7.3 shows the respondents' age pie charts containing Children, Teens, Adults, and Elderly responding to our survey. The pie chart above shows that the age group with a higher response rate is Adults.

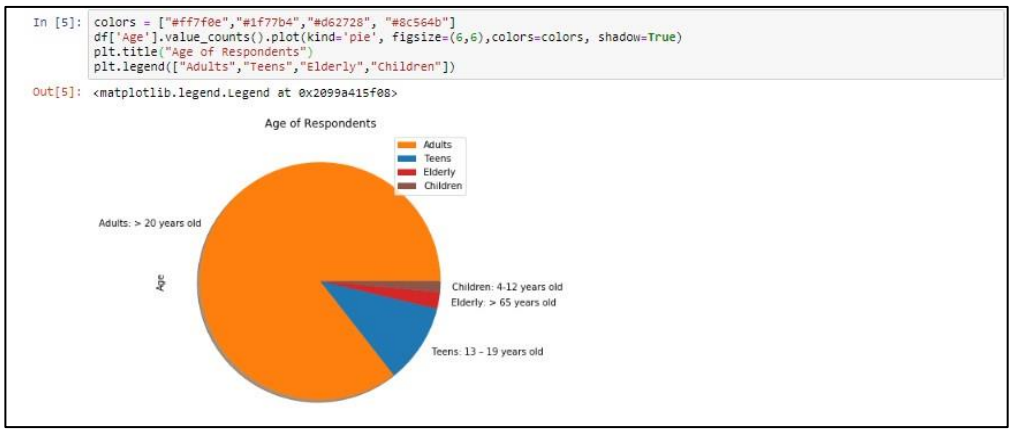

Figure 7.3 Pie Chart of Respondent's Age

Meanwhile, Figure 7.4 shows the plot of the scatter plot of respondents' marital status, and Figure 7.5 shows the confusion matrix of the testing data. All types of graphical techniques in Exploratory Data Analysis (EDA) were used in this project study.

```
In [15]: x = df.iloc[:, 1:12]
         y = df.loc[:, ['Class_label']]
         plt.figure(figsize=(14,7))
         colormap = np.array(['red', 'lime'])
         plt.subplot(2, 4, 2)
         plt.scatter(x.Employment_status, x.Marital_status, c=colormap[y.Class_label], s=10)
         plt.title('Status')

Out[15]: Text(0.5, 1.0, 'Status')
```

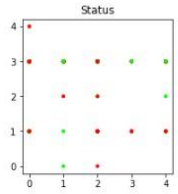

Figure 7.4 Scatter Plot of Marital Status

```
In [22]: import seaborn as sns
         plt.figure(figsize=(9,9))
         sns.heatmap(sm.confusion_matrix(y, yPred), annot=True, fmt=".3f", linewidths=.5, square = True, cmap = 'PuRd_r');
         plt.ylabel('Actual label');
         plt.xlabel('Predicted label');
         all_sample_title = 'Accuracy Score: {0}'.format(sm.accuracy_score(y, yPred))
         plt.title(all_sample_title, size = 15);
```

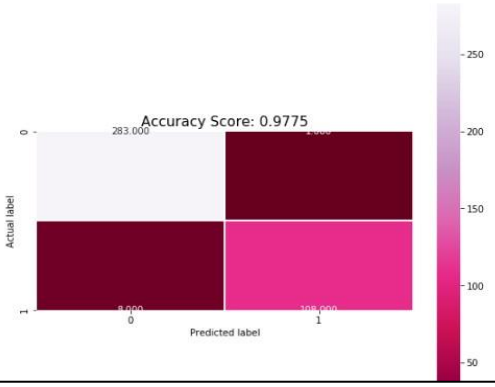

Figure 7.5 Confusion Matrix of Testing Data

7.2 Data Pre-Processing / Data Mining

Data pre-processing in machine learning is a crucial step that helps enhance the quality of data collected in order to facilitate the extraction of meaningful insights from the data. It refers to the technique of preparing the raw data that makes it ideal for building and training machine learning models. The first step is importing the libraries that will be needed in the program, as in Figure 7.6. A library is essentially a collection of modules that can be called and used.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn import datasets
from sklearn.neighbors import KNeighborsClassifier
from sklearn.preprocessing import LabelEncoder
import sklearn.metrics as sm
from sklearn import metrics
from collections import Counter
%matplotlib inline
```

Figure 7.6 The snippet of Code to Import Libraries

Data Collection and Selection

A mental health survey was conducted by creating a questionnaire comprising 11 questions through the Google form. The total of 503 responses was recorded. Next, the dataset which has been collected is

imported by locating the directory of the CSV file, as shown in the snippet code of Figure 7.7. The dataset was accessed by using a method called read_csv which can be found in the pandas library.

```
df = pd.read_csv("TrainData.csv") #Reading the dataset in a dataframe using Pandas
df.head()
```

Figure 7.7 The snippet of Code to Read the Dataset

All of the attributes in Figure 7.8 are relevant for the analysis and can help with the classification task was imported into pandas.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 401 entries, 0 to 400
Data columns (total 14 columns):
Age                 400 non-null object
Gender              400 non-null object
State               399 non-null object
Education_level     398 non-null object
Employment_status   400 non-null object
Marital_status      400 non-null object
Intake_sweet        396 non-null object
Intake_salty        396 non-null object
Intake_fruit        398 non-null object
Intake_vegetables   393 non-null object
Exercise            399 non-null object
Early_symptoms      398 non-null object
History_family      398 non-null object
Class_label         395 non-null object
```

Figure 7.8 The Attributes of the Dataset

Data Cleansing

Missing data have been identified as in Figure 7.9. As null values are found in some columns of the data, data cleansing should be performed to prevent inaccurate and faulty conclusions and inferences from the data.

```
# To report missing data
df.apply(lambda x: sum(x.isnull()),axis=0)
```

```
Age                 1
Gender              1
State               2
Education_level     3
Employment_status   1
Marital_status      1
Intake_sweet        5
Intake_salty        5
Intake_fruit        3
Intake_vegetables   8
Exercise            2
Early_symptoms      3
History_family      3
Class_label         6
dtype: int64
```

Figure 7.9 The snippet of Code to Check Missing Values

The missing data were handled by taking the mode, which is the set of values that most frequently occurs in all values of the same column and have it to replace the missing data. Figure 7.10 shows the code snippet to replace the missing data in the dataset.

```
df['Age'].fillna(df['Age'].mode()[0], inplace=True)
df['Gender'].fillna(df['Gender'].mode()[0], inplace=True)
df['State'].fillna(df['State'].mode()[0], inplace=True)
df['Education_level'].fillna(df['Education_level'].mode()[0], inplace=True)
df['Employment_status'].fillna(df['Employment_status'].mode()[0], inplace=True)
df['Marital_status'].fillna(df['Marital_status'].mode()[0], inplace=True)
df['Intake_sweet'].fillna(df['Intake_sweet'].mode()[0], inplace=True)
df['Intake_salty'].fillna(df['Intake_salty'].mode()[0], inplace=True)
df['Intake_fruit'].fillna(df['Intake_fruit'].mode()[0], inplace=True)
df['Intake_vegetables'].fillna(df['Intake_vegetables'].mode()[0], inplace=True)
df['Exercise'].fillna(df['Exercise'].mode()[0], inplace=True)
df['Early_symptoms'].fillna(df['Early_symptoms'].mode()[0], inplace=True)
df['History_family'].fillna('No',inplace=True)
df['Class_label'].fillna('No',inplace=True)
```

Figure 7.10 The snippet of Code to Fill in Missing Values

Data Transformation

Before the dataset can be used in model prediction, label encoding is used to represent categorical values as the last step of data pre-processing. A label encoding transforms each value in a column to a number. The code snippet to encode the data is shown in Figure 7.11.

```
var_mod = ['Age','Gender','State','Education_level','Employment_status','Marital_status','Intake_sweet','Intake_
le = LabelEncoder()
for i in var_mod:
    df[i] = le.fit_transform(df[i])
df.dtypes
df.head()
```

Figure 7.11 Encoding Categorical Values

Following the previous step, the columns of the dataset are read by using iloc of pandas, which is used to fix the indexes for selection. As in Figure 7.12, the selection of rows takes : as an argument that indicates the parameter selects all the rows, while the column takes the first 1 rows until the 11th column of the data frame.

```
df.iloc[:, 1:12]
df.loc[:, ['Class_label']]
```

Figure 7.12 The snippet of Code to Read the Columns and Rows of the Dataset

Data Split

The dataset consists of 2 classes, 11 attributes, and 503 samples of data were divided into train data by 80 percent or 401 data and test data by 10 percent or 102 data.

7.3 Model Design and Development

After thorough planning and reading, the next phase is model design and development. This phase is especially tricky as compared to other phases to navigate and manage (Wynn & Clarkson, 2018). There are a few activities or processes involved in this phase of methodology that include classification model, architecture, target, data preparation features data and machine learning model.

7.3.1 Classification Model

Machine Learning is a natural progenitor of computer science and statistics (Mohammed, Khan & Bashier, 2017). There are several applications of machine learning, the most important of which is data mining. People are often prone to errors in analysis or possibly in trying to establish relationships between several characteristics (Akinsola, 2017). The classification model plays a significant role in overcoming this problem in machine learning. Classification can be defined as the process of categorising characteristics into exactly the class to which a particular case belongs (Luis & Medina, 2013). In this

study of the mental health prediction model, KNN (K Nearest Neighbours) is used to classify monitored data into the class to which the majority of KNN belongs.

7.3.2 Algorithm / Computational

K nearest neighbors or KNN algorithm is a simple algorithm that uses the entire data set in its training phase. Whenever a prediction is required for an unseen data instance, it searches the entire training data set for k-like instances, and the data with the most similar instance is finally returned as a prediction. The algorithm for the k-nearest neighbors uses a straightforward approach to perform the classification. When tested with a new example, it searches the training data and finds the k training examples that are most similar to the latest example. It then assigns the most common class seal (among these k-training examples) to the test example. The k in the KNN algorithm represents the number of nearest neighboring points that match the class of the new test data. Based on this project, the results shown in its subchapter are the calculated values induced by the KNN calculation process.

7.3.3 Architecture

KNN is simplistic in its nature. It only has to assume the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. Therefore, its architecture only revolves around comparative testing through some calculations with k value. The Euclidean distance formula is used ($dist((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2}$) and the value will determine where the data point belong in which category. The figure below shows the basic architecture of KNN data processing algorithm.

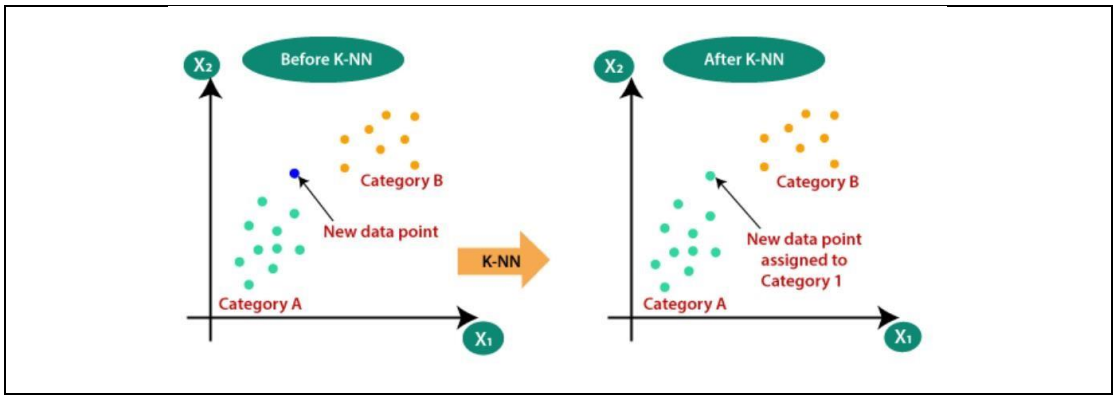

Figure 7.13 The KNN architecture
(*Source*: JavaTpoint, 2018)

7.3.4 Features/Parameters

Parameter, k is the number of nearest neighbours. The parameter is important and is the core when deciding factor. 'k', generally an odd number if the number of classes is 2. When k=1, the algorithm is known as the nearest neighbours algorithm. Suppose when P1 is the point, for which label needs to predict. First steps are we need to find the closest point to P1 and then the label of the nearest point assigned to P1. Suppose P1 is the point for which label needs to predict. Find the k closest point to P1 and then classify points by majority vote of its k neighbours. Each object votes for their class and the class with the most votes is taken as the prediction. For finding closest similar points, the distance between points was found using distance measures such as Euclidean distance.

Data is collected through the survey among Malaysian citizens. From the data collected such as age, gender, states that the respondents currently live, level of education, employment status, marital status, weekly food intake frequency, exercise frequency, symptoms experienced, and history of mental disorder, the prediction about mental health disorders based on their life background and lifestyle are made.

Table 7.1 The Features, Description and Value

Features	Description	Value
Age	Respondent's age	(Children: 4-12 years old, Teens: 13 – 19 years old, Adults: > 20 years old, Elderly: > 65 years old)
Gender	Respondent's gender	(Female, Male)
States	Respondent's states	(Perlis, Pulau Pinang, Kedah, Perak, Kelantan, Terengganu, Pahang, Selangor, W.P Kuala Lumpur, Negeri Sembilan, Melaka, Johor, Sabah, Sarawak)
Level of Education	Respondent's highest level of education	(Primary or less, Secondary, Post-Secondary or tertiary (Diploma, Graduate, Undergraduate, Doctorate), and others)
Employment Status	Respondent's employment status	(Seeking employment, Temporary or Parttime employment, Self-employed, Full-time employment, Others (Student or Pursuing further studies)
Marital Status	Respondent's marital status	(Single, Married, Separated, Divorced, and Widowed)
Food Intake Frequency	Respondent's weekly food intake frequency against sweet foods, salty foods, fresh fruit, and fresh vegetables	(Several times a day, Once a day, Several times a week, Less often, Never)
Exercise Frequency	Respondent's exercise frequency	(Every day, At least 5 times a week, At least 3 times a week, Once a week, Never)
Symptoms Experienced	Respondent's experienced in any related mental disorders symptoms.	(Yes, I have never experienced any of the symptoms stated above)
History of Mental Disorder in Family	Whether the respondents have a history of mental disorder in family or not.	(Yes, No)
Experiencing in Diagnosed Mental Disorder	Whether the respondents have been diagnosed with or experiencing mental disorder such as shown in Figure 7.14 below or not.	(Yes, No)

7.3.5 Target/Label

Based on the survey conducted, a data collection of Mental Health Survey was made. The total of 503 respondents answered in this survey, including UiTM students and public community. Based on the survey done, a mental illness prediction based on their food categories intake frequencies can also be made.

Based on the survey conducted, it was seen that 228 from 503 respondent took sweet foods several times a day. 98 respondents took sweet foods once a day, 109 respondents took sweet foods several times a week, 59 respondents less often took sweet foods in a week, and 3 respondents never take any sweet food in a week. Next, 187 from 503 respondents took salty foods several times a day. There are 111 respondents took salty foods once a day, 113 respondents took salty foods several times a week, 81 respondents less often took salty foods in a week, and 5 respondents never took any salty food during the weeks. Moreover, there are 54 out of 503 respondents took fresh fruits several times a day. 104 respondents took fresh fruits once a day, 141 respondents took fresh fruits several times a week, 195 respondents less often took fresh fruits in a week, and 5 respondents never took any fresh fruits during the weeks.

Among the 503 respondents that answered the survey, 70 respondents took fresh vegetables several times a day. 115 respondents took fresh vegetables once a day, 117 respondents took fresh vegetables several times a week, 175 respondents less often took fresh vegetables in a week, and 17 respondents never take any fresh vegetables during the weeks. Finally, 44 respondents exercised every day, 22 respondents exercised at least 5 times a week, 61 respondent exercised at least 3 times a week. 171 respondents exercised once a week. 9 respondents occasionally exercised, and 191 respondents were never exercised.

8.0 Result

After dividing the dataset into train and test data, the KNN classifier was carried out with different sets of parameters which include *n_neighbors, weights, and p*. This was done to obtain the optimal value of the parameter with the highest accuracy. Train data was first tested to determine the effect of the variable *n_neighbors* value on the performance of the KNN classifier on the data. In order to obtain the accuracy of the model in all experiments, the range of *n_neighbors* was progressively increased. The *n_neighbors* values used are 2,3,4, and 5. Experiment 1 sets the default parameters with *p = 2* and *weights = 'uniform'*. Next, repeat the testing process to the other experiments by using the same default value of *p* and *weights*. Table 7.2 displays the average accuracy value of all the experiments conducted with different *n_neighbors* values.

Table 7.2 Train Data Accuracy Values using the Default Value of *p* and *weights*

List of experiment	n_neighbors	p	weights	Accuracy(%)
Experiment 1	2	2	uniform	81.05%
Experiment 2	3	2	uniform	82.04%
Experiment 3	4	2	uniform	75.31%
Experiment 4	5	2	uniform	76.56%

By using the same train data with *n_neighbor* followed by the previous value, the *weights* of the parameter are changed to *'distance'* to observe if the accuracy of the data can be improved. The *p* variable is set as the default value for all experiments. The results of the accuracy values are shown in Table 7.3.

Table 7.3 Train Data Accuracy Values After Changing *weights* to *'distance'*

List of experiment	n_neighbors	p	weights	Accuracy(%)
Experiment 1	2	2	distance	97.75%
Experiment 2	3	2	distance	97.75%
Experiment 3	4	2	distance	97.75%
Experiment 4	5	2	distance	97.75%

Finally, the improved k nearest neighbor classifier model was used to classify the test samples. By applying weight = '*distance'* to test data, the accuracy values of the test data are significantly improved as can be seen in Table 7.4.

Table 7.4 Test Data Accuracy Values using the Previous Parameters

List of experiment	n_neighbors	p	weights	Accuracy(%)
Experiment 1	2	2	distance	75.55%
Experiment 2	3	2	distance	70.59%
Experiment 3	4	2	distance	69.61%
Experiment 4	5	2	distance	66.67%

9.0 Discussion

Based on the results from the above subsection, it can be concluded that it is expected from a KNN data processing algorithm to compute the outcome in such manner. By further increasing the data quality using certain measures, the results will yield a higher accuracy which is on par on the capabilities of the KNN algorithm technique. Nevertheless, not all the datasets acquired are up for the task of testing since the existence of outliers, which is most probably due to inappropriate survey results. The discussion answers the problem statement by analysing the results produced from the given solution, which means that in a way, implementation of machine learning in health science has proven to be very valuable. It

can be proven that major research has already been done in this subject, which can be found on countless research papers.

The implementation of machine learning does provide some invaluable advantage based on the results given if done with a higher quality dataset. In a way, all the results computed will give a better understanding of machine learning improvement and can be used as a future reference for upcoming machine learning system creation.

10.0 Conclusion and Recommendation

Based on the obtained results, it can be concluded that machine learning is useful for the implementation of health-based problems, especially when it comes to data processing and analysis. The objective of using machine learning in this field is to avoid the human effort to waste their time manually analysing data and therefore, could rely on technology to aid their workflow. The datasets acquired were used extensively and data processing and analysing indicated that machine learning is mature enough for important data processing only if the datasets are accurate and have good quality.

In conclusion, with the modelling process and computation, we are not just able to predict mental illness, but we also can help other experts in their subject matter in assisting their work with the use of machine learning. Machine Learning has proven to be developed enough to be used for more important tasks, especially when it comes to a more tech-induced world where everything is automated and uses an algorithm to solve problems digitally. Hopefully, more data can be retrieved to assist with better machine learning soon.

REFERENCES

Bone, D., Lee, C. C., Chaspari, T., Gibson, J., & Narayanan, S. (2017). Signal Processing and Machine Learning for Mental Health Research and Clinical Applications [Perspectives]. *IEEE Signal Processing Magazine*, *34*(5).

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomedical Informatics Insights*, *8*, BII.S31559.

Sumathi, M.R., & Poorna, B. (2016). Prediction of Mental Health Problems Among Children Using Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications*, *7*(1), 552–557.

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, *13*(1), 23–47.

Patil, P. (2018). *What is Exploratory Data Analysis? - Towards Data Science*. Towards Data Science. https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448.

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Journal of Medical Systems*, *42*(5).

Akinsola, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138.

Mohammed, M., Khan, M. B., & Bashie, E. B. M. (2016). Machine learning: Algorithms and applications. In *Machine Learning: Algorithms and Applications* (Issue July).

Wynn, D. C., & Clarkson, P. J. (2018). Process models in design and development. *Research in Engineering Design*, *29*(2), 161–202.

APPENDICES

Mental Health Survey

Assalamualaikum and Hye guys!

We're students of University Teknologi Mara (UiTM) Jasin, Melaka conducting a survey on CSC649 (Special Topics in Computer Science) Project about Prediction of Mental Health Disorders using Machine Learning Algorithm. All the information gathered will be used for academic purposes only. We hope you can spend your time answering this survey with honor and respect in order to help us gather the information and data needed for our project.

Thank you for your cooperation.

1. Which category below includes your age?

- ◯ Children: 4-12 years old
- ◯ Teens: 13 – 19 years old
- ◯ Adults: > 20 years old
- ◯ Elderly: > 65 years old

2. Select your gender

- ◯ Female
- ◯ Male

3. Which states are you currently living in?

Choose ▾

4. What is the highest level of education you have attained?

- ◯ Primary or less
- ◯ Secondary
- ◯ Post-secondary or tertiary (diploma, graduate, undergraduate, doctorate)
- ◯ Other: .....

5. Select your current employment status

- ◯ Seeking employment
- ◯ Temporary/ Part-time employment
- ◯ Self-employed
- ◯ Full-time employment
- ◯ Others (Student / Pursuing further studies)

6. Choose one the following option that describes your marital status

- ◯ Single
- ◯ Married
- ◯ Separated
- ◯ Divorced
- ◯ Widowed

7. What is your weekly food intake frequency of the following food categories?

	Several times a day	Once a day	Several times a week	Less often	Never
Sweet foods	◯	◯	◯	◯	◯
Salty foods	◯	◯	◯	◯	◯
Fresh fruit	◯	◯	◯	◯	◯
Fresh vegetables	◯	◯	◯	◯	◯

8. How often do you exercise?

- ◯ Every day
- ◯ At least 5 times a week
- ◯ At least 3 times a week
- ◯ Once a week
- ◯ Never

9. Have you experienced any symptom(s) stated below?

Long-lasting sadness or irritability

Extremely high and low moods

Confused thinking or reduced ability to concentrate

Excessive fears or worries, or extreme feelings of guilt

Extreme mood changes of highs and lows

Significant tiredness, low energy or problems sleeping

Detachment from reality, paranoia or hallucinations

Inability to cope with daily problems or stress

Excessive anger, hostility or violence

Suicidal thoughts

- ◯ Yes
- ◯ I have never experienced any of the symptoms stated above

10. Is there a history of mental disorder in your family?

- ◯ Yes
- ◯ No

11. Have you been diagnosed with/experiencing mental disorder(s) as described below?

Mood disorder – Depression or bipolar disorder in which the general emotional state is distorted and inconsistent with the situations

Personality disorder - Experiencing a rigid and unhealthy pattern of thinking, functioning and behaving, as well as difficulties in perceiving and relating to situations and people

Psychotic disorder - Severe mental disorders that cause abnormal thinking and perceptions such as schizophrenia

Eating disorder - Irregular eating habits and severe anxiety or concern over body weight or shape, along with inadequate or excessive intake of food

Substance abuse disorder – Drugs and alcohol habituation or addiction

- ◯ Yes
- ◯ No