CLASSIFICATION OF PSYCHOLOGICAL ILLNESSES USING NAÏVE BAYES

KAY KHAING SOE

M.C.Sc. **SEPTEMBER 2022**

CLASSIFICATION OF PSYCHOLOGICAL ILLNESSES USING NAÏVE BAYES

By

KAY KHAING SOE

B.C.Sc(Hons:)

A dissertation submitted in partial fulfillment of the requirement for the degree of

Master of Computer Science

M.C.Sc.

University of Computer Studies, Yangon
September 2022

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

………………………… ……………………………….

Date Kay Khaing Soe

ACKNOWLEDGEMENTS

To complete this thesis, many things are needed like my hard work and the supporting of many people who gave a lot of idea to me to complete this thesis.

Firstly, I am to express my appreciation and thanks to **Dr. Mie Mie Khin**, Rector of the University of Computer Studies, Yangon, for her kind permission to submit this dissertation.

I am thankful to my thesis supervisors, **Dr. Win Lai Hnin**, for their close supervision, proper guidance, valuable suggestions, guidance, advice and encouragement during the course of this work.

I would like to thank course coordinators, **Dr. Si Si Mar Win** , **Dr. Tin Zar Thaw**, Professor and Head of Faculty of Software Department of the University of Computer Studies, Yangon, **Daw Myo Myo Khin,** Associate Professor and Head of Information Technology Supporting and Maintenance of Computer Science Department of the University of Computer Studies, Hinthada and **Dr. Tun Myat Aung**, Pro- Rector of the University of Computer Studies, Hinthada for their superior suggestions and administrative supports during my academic study.

I also would like to express my gratitude and appreciation to **U Aung Myint Than**, Assistant Lecturer and Head of Department Language Development (English) of University of Computer studies, Yangon, for her advice, editing and suggestion from the language point of view.

I also like to acknowledge my thanks to **all my teachers** who taught me throughout the master's degree course and **my friends** for their corporation and wish to express my gratitude to **my beloved family** for their invaluable support and encouragement to fulfill my wish.

Finally, I wish to express my thankful. If not for the above mentioned people, my seminar would never have been completed in such a successfully manner. I once again extend my sincere thanks to all of them.

ABSTRACT

Physical and psychological illness are important for people in their daily life. If people face risks, conflicts and difficulties, they suffer from psychological illness problems such as depression, bipolar disorder, schizophrenia, and anxiety. Psychological illness professionals ask the patient concerned with the present conditions (risk and symptoms) and mood or feelings. So, the patients caused with the risks or symptoms are classified (psychological illness Yes or No) according to their cases. There are many techniques of data mining applied for performance of many application areas. This system contributes to classify psychological illness by using one of Data mining techniques, Naïve Bayes and to test performance result with Random Split Test, K-Folds, Test and Unlabeled Test. The system's performance is evaluated in terms of average accuracy is over 80%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

In Data mining, there is a variety of techniques to identify suggest of information or decision-making knowledge in the database and extracting these in a way such as decision support, and predictions. In the late of 1960s, model-oriented decision support system became practical. Decision support system is to help complete decision makers with managerial decisions based on personal intuition and experience. Naive Bayes classifier is based on Bayes theorem. This classifier algorithm used conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes.

Psychological illness, called mental health disorder, is the feelings, thoughts and behavior in mind. Psychological illness concern when ongoing conditions and symptoms cause frequent stress and affect the ability of people to function. A mental illness can make the user miserable and can cause problems in your daily life, such as at school or work or in relationships. People join and take the help of Mental Health professionals and mental health professionals ask their conditions. In this system, the patients are classified that he/she suffers or not mental health based on the patients' answers with the mental conditions using Naïve Bayes.

1.1 Objectives of the Thesis

The objectives of the thesis are as follows:

- To study the concept of decision support system
- To aware the psychological illness is very important for the health of people
- To know one of the machine learning techniques (supervised machine learning)
- To learn Naïve Bayesian Classification algorithm
- To evaluate the performance of the system with accuracy, precision, and recall

1.2 Contributions of the Thesis

The system is developed to classify mental or psychological illness acting upon thoughts and mind of people. Two mental health datasets are used for the thesis and dataset1: 159 instances are used for this system from Kaggle.com. 159 instances of

Mental Health are described as dataset, 27 attributes: questions, and Prediction_status: Class (Yes) or (No) and dataset2: 1147 instances, 21 attributes and Prediction_status is same. The system displays the accuracy result of whether the disease causes or not accuracy result is calculated.

1.3 Motivation of the Thesis

Everyone has difficulties. People have much stress between families, friends, other relationships and in workplaces, universities, environments, in their daily lives. These stress makes them illness and diseases in their mind. They join to doctors or professionals for psychological illness. When doctors and professionals are absent, or travel, the patients have difficulties. So, the system is proposed to help and to solve their difficulties. This system is developed by using Naïve Bayes and, system performance is measured with accuracy (%) of Random Split Test, and K-Folds Test. Finally, accuracy result is displayed with table and bar chart.

1.4 Organization of the Thesis

This thesis is organized into five chapters, abstract, acknowledgment and references.

The classification of psychological illness system is introduced in chapter 1. Objectives, contribution, motivation and organization of thesis are described in this chapter.

The psychological illness disease, symptoms of disease, KDD process in Data Mining, data mining concepts, data, information, knowledge, data mining techniques, and Naïve Bayes Classifier in chapter 2.

The model of classification is explained in detailed in chapter 3. First of all, classification and prediction, comparing classification methods, classification methods, data collection, and data preprocessing are explained.

Design and implementation of classification of psychological illness are presented in chapter 4. The system overview design, two datasets with attributes are described with tables in this chapter. And, the implementation for the proposed system is explained with Graphical User Interfaces details. Also predicting a class label by using Naïve Bayes classifier.

As the conclusion, advantages of the system and further extensions that propose some improvements which could be made are presented in chapter 5. Finally, the experimental results are shown by charts and tables in this chapter.

CHAPTER 2
BACKGROUND THEORY

This chapter presents what is psychological illness, and its common symptoms. As background theory, Data Mining's concepts, the importance of data, information, and knowledge, Data Mining's techniques, and Naïve Bayes classifier are explained.

2.1 Psychological Illness

Psychological illness or Mental illness, also called mental health disorders, refers to a wide range of mental health conditions — disorders that affect your mood, thinking and behavior. Examples of mental illness include depression, anxiety disorders, schizophrenia, eating disorders and addictive behaviors. Many people have mental health concerns from time to time. But a mental health concern becomes a mental illness when ongoing signs and symptoms cause frequent stress and affect your ability to function. A mental illness can make you miserable and can cause problems in your daily life, such as at school or work or in relationships. In most cases, symptoms can be managed with a combination of medications and talk therapy (psychotherapy).

2.1.1 Common Symptoms of Psychological Illness

Signs and symptoms of mental illness can vary, depending on the disorder, circumstances and other factors. Mental illness symptoms can affect emotions, thoughts and behaviors. Examples of signs and symptoms include:

- Feeling sad or down
- Confused thinking or reduced ability to concentrate
- Excessive fears or worries, or extreme feelings of guilt
- Extreme mood changes of highs and lows
- Withdrawal from friends and activities
- Significant tiredness, low energy or problems sleeping
- Detachment from reality (delusions), paranoia or hallucinations
- Inability to cope with daily problems or stress
- Trouble understanding and relating to situations and to people
- Problems with alcohol or drug use
- Major changes in eating habits

- Sex drive changes
- Excessive anger, hostility or violence
- Suicidal thinking

2.2 KDD Process in Data Mining

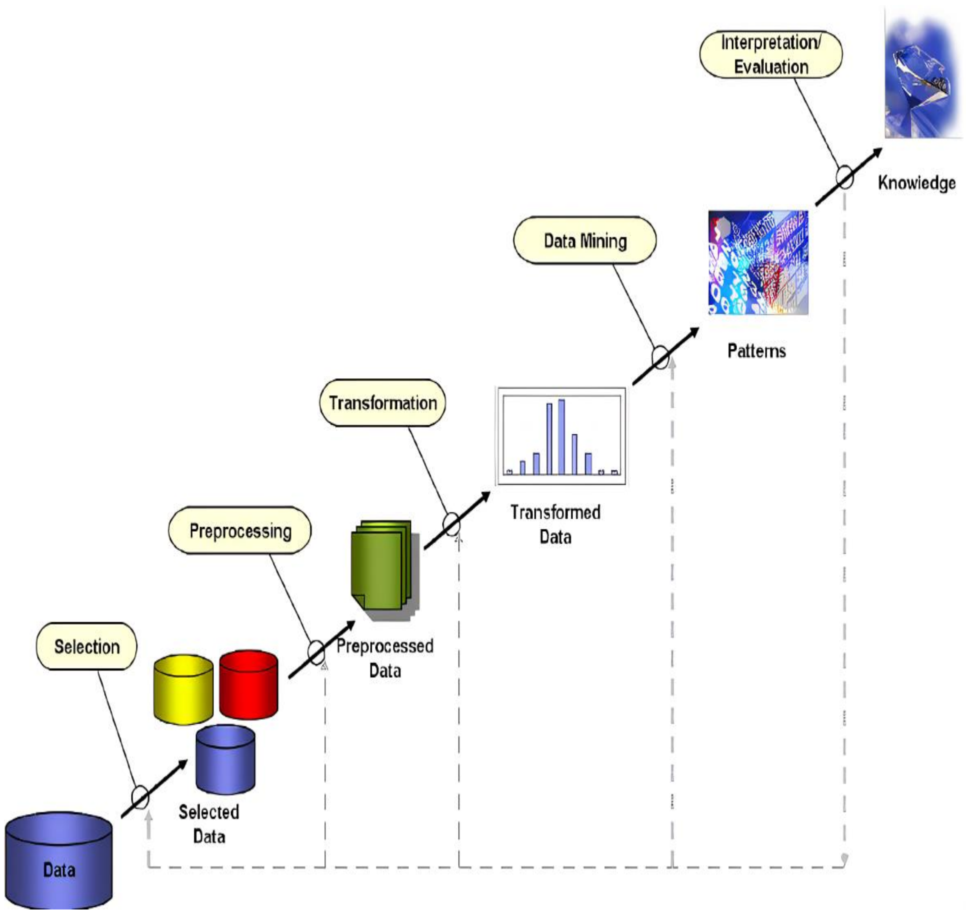


Figure 2.1 KDD Process in Data Mining

In Figure 2.1, KDD is the knowledge Discovery in Databases in data mining. In this process, the following steps are essential.

- Data Cleaning: Data cleaning is the process of removing noisy data, and irrelevant data from collection with discrepancy detection and data translation tools
- Data Integration: Data integration is the process of combining us data from multiple sources in a common source (Data Warehouse).

- Data Selection: Data selection is the process of selecting relevant data and retrieved data from dada collection by using Neural Network, Decision Trees, Naïve Bayes, Clustering, and Regression.
- Data Transformation: Data transformation is the process of transforming data into appropriate form required by mining procedure. Data transformation are two step process: Data Mapping, and Code Generation.
- Data Mining: Data mining is the technique that are applied to extract task relevant data into patterns and decides purpose of model using classification or characterization.
- Pattern Evaluation: Pattern Evaluation is the process of identifying strictly increasing patterns representing knowledge based on given measures by finding interestingness score of each pattern and using summarization and visualization to make data understandable by user.
- Knowledge Representation: Knowledge representation is the technique by utilizing visualization tools to represent data mining results with reports, tables, discriminant rules, classification rules, characterization rules.

2.3 Data Mining Concepts

Data mining is the technique to extract the hidden predictive information from large databases. It is a powerful and popular technology and has a great potential to provide enterprises focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors and allow knowledge driven decisions. Data Mining consists of five major elements. They are:

- Extract, transform, and output transaction data onto data warehouse system.
- Store and manage data in a multidimensional database system.
- Provide data access to analysts and information technology professionals of business, education, medicine, and others.
- Analyze data by using software.
- Present data in useful formats, such as graphs, tables, charts or others.

2.4 Data, Information, and Knowledge

The essential and main components in every research of enterprises, organizations, and others are data, information and knowledge.

2.4.1 Data

Data are raw facts: text or numbers, that are processed by a computer with program. Nowadays, organizations, business enterprises, and others are accumulating many data with different formats in different databases. These data are:

- Operational or transactional data
- Non-operational data
- Meta data
- Big data

2.4.2 Information

Information are meaningful data that related to the specific domains the users want. They can be provided by the patterns, associations, and relationships between data. For example, World Health Organization (WHO) catalogued the people affected by COVID-19, and death all over the world via the medical offices.

2.4.3 Knowledge

Knowledge is the final confirmation on cases and it can be converted from information about historical trends and future cases. For example, WHO is announcing final confirmation based on the certain information day by day on media for present and can be analyzed to provide others knowledge of the users' wanted.

2.5 Data Mining Techniques

The primary data mining tasks are:

- Classification
- Regression
- Clustering
- Summarization
- Dependency Modelling and
- Change and Deviation Detection

There are several major Data Mining techniques have been developed and used in Data Mining Field's projects including association, classification, clustering, prediction and sequential patterns. These are briefly explained as follows:

2.5.1 Classification

Classification is the mining technique based on machine learning. It is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

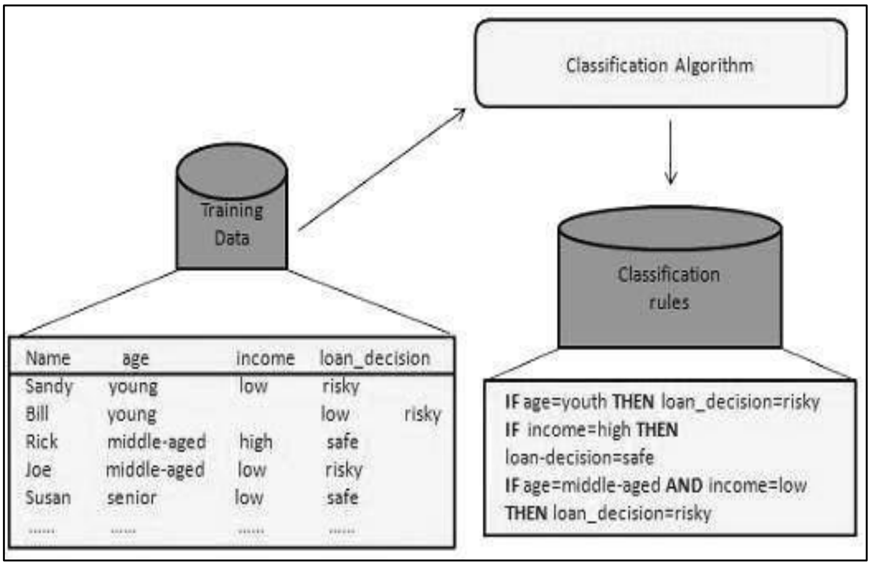


Figure 2.2 Sample Classification of Finance System

2.5.2 Association

Association is one of the well-known data mining techniques. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. Association rule mining, at a basic level, involves the use of machine learning models to analyze data for patterns, or co-occurrences, in a database. It identifies frequent if-then associations, which themselves are the association rules. An association rule has two parts: an antecedent (if) and a consequent (then).

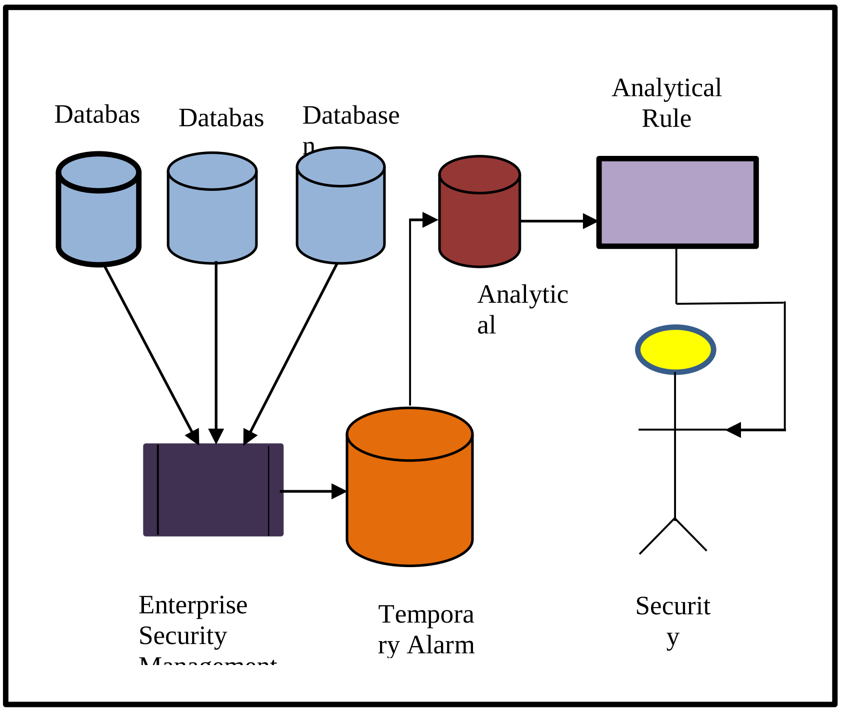


Figure 2.3 Association of Security Management System

2.5.3 Clustering

Clustering is a division of data into groups of similar objects. Representing data by fewer clusters necessarily loses certain fine details, but achieves simplification. It models by its clusters. From machine learning, perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning. A practical perspective clustering plays an outstanding role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, CRM, and so on.

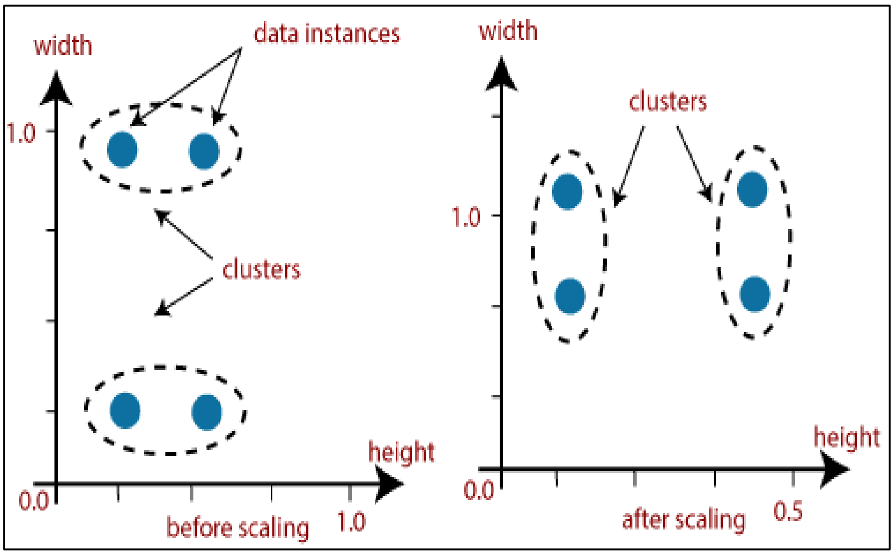


Figure 2.4 Clusters of Map's Scale System

2.5.4 Prediction

Prediction is a technique to discover relationship between independent variables and that between dependent and independent variables. Predictive data mining is data mining that is done for the purpose of using business intelligence or other data to forecast or predict trends. This type of data mining can help business leaders make better decisions and can add value to the efforts of the analytics team.

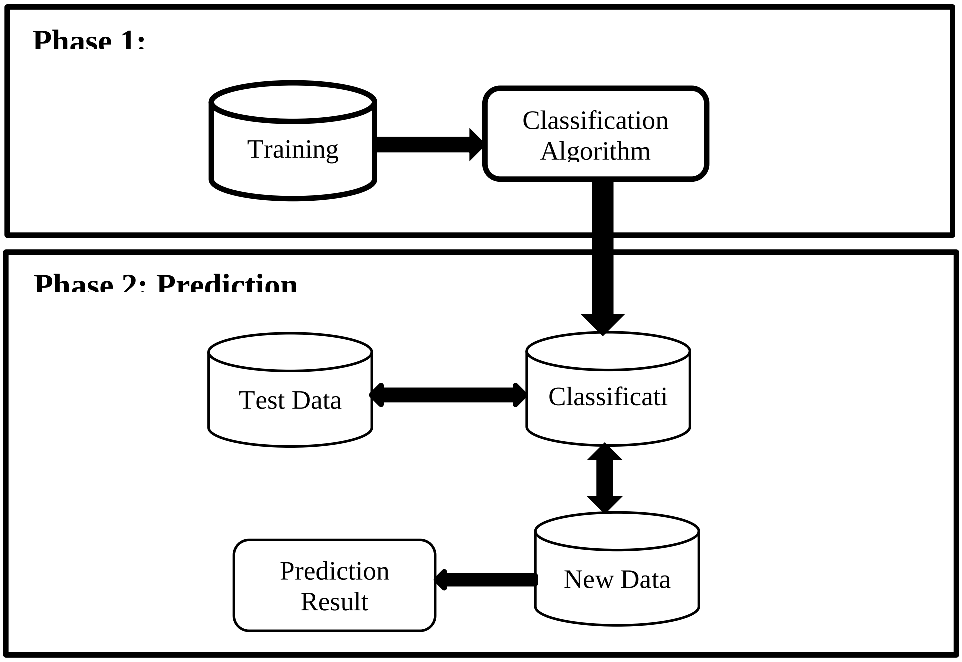
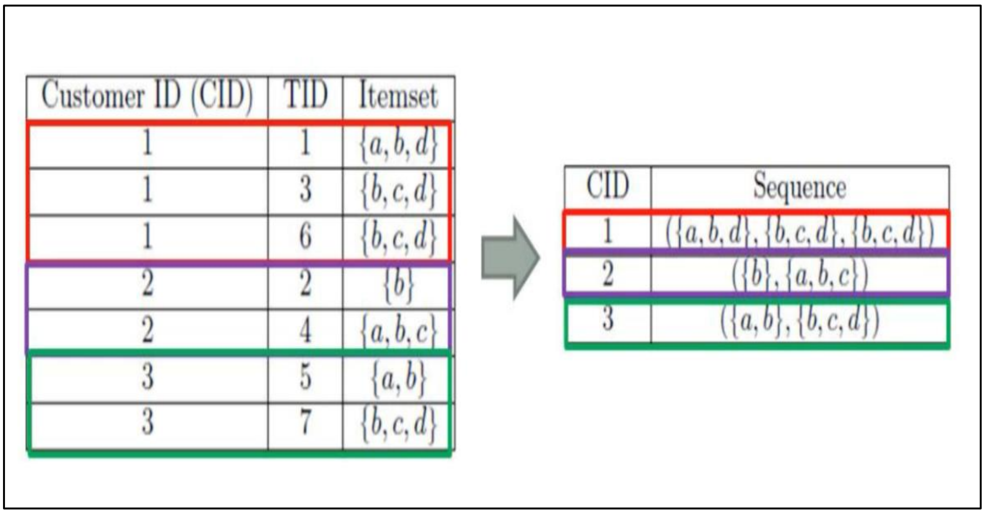


Figure 2.5 Architecture of Prediction Technique

2.5.5 Sequential Patterns

Sequential patterns analysis seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data.



Customer ID (CID)	TID	Itemset
1	1	{a, b, d}
1	3	{b, c, d}
1	6	{b, c, d}
2	2	{b}
2	4	{a, b, c}
3	5	{a, b}
3	7	{b, c, d}

CID	Sequence
1	({a, b, d}, {b, c, d}, {b, c, d})
2	({b}, {a, b, c})
3	({a, b}, {b, c, d})

Figure 2.6 Extraction of Sequential Patterns in Sale Transaction System

2.6 Naïve Bayes

Naïve Bayesian classifiers is based on Data Mining Techniques. This algorithm works quickly and can save a lot of time. Naive Bayes is suitable for solving multi-class prediction problems. Naive Bayes is better suited for categorical input variables than numerical variables.

Naïve Bayesian classifiers is one of Data Mining Techniques. This algorithm works quickly and can save a lot of time. Naive Bayes is the most suitable classifier for solving multi-class prediction problems and for categorical input variables than numerical variables. Naïve Bayesian classifiers (Formula):

$$P(c|x) = \frac{P(x \mid c)P(c)}{P(x)} \qquad (1.1)$$

where,

P(c) =The prior probability of class

P(x) =The prior probability of predictor

P(c|x) =The posterior probability of class (target)given predictor (attribute)

P(x | c) =The likelihood which is the probability of predictor given class

X can be written as follow:

$X = (x1, x2, x3, . . . , xn)$

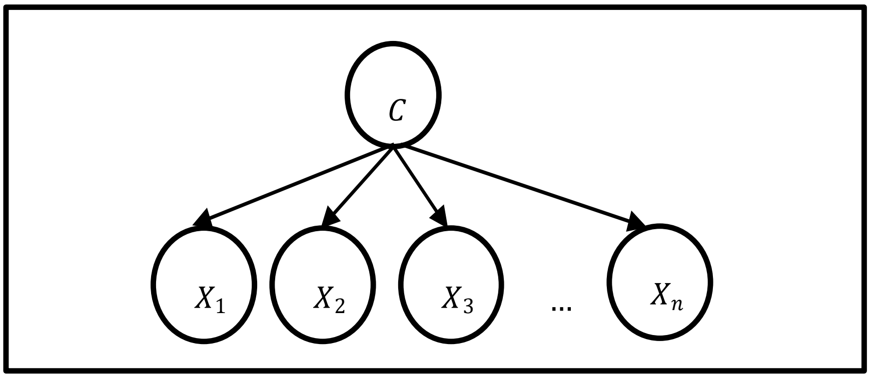


Figure 2.7 Naïve Bayes

In Figure 2.1, the relationship between C(class) and X (predictor) are described.

$$P(C|x_1, x_2, \ldots, x_n) = \frac{P(C)P(x_1, x_2, \ldots, x_n|C)}{P(x_1, x_2, \ldots, x_n)} \quad (2.2)$$

In this formula, with the substitution of X, the Bayes formula can be printed as follows complex factors of probability values which is nearly impossible to analyze one by one. As a consequence, the calculation develops hard to do.

$$P(C|x_1, x_2, \ldots, x_n) = P(C) \prod_{i=1}^{n} P(x_i|C) \quad (2.3)$$

In Naïve Bayes Classifiers essential to maximize the probability value of individually class, which is expressed as the Hypothesis Maximum a Posteriori (HMAP):

$$H_{map} = argmaxP(C|x_1, x_2, \ldots, x_n) = argmaxP(C) \prod_{i=1}^{n} P(x_i|C) \quad (2.4)$$

In Naïve Bayes Classifier, by means of Equation can forecast which classes can be used in Naïve Bayes Model. But, if the attribute X in Equation has quantitative types, then the probability will be very minor such that the value cannot be used to find the value. So, we essential to use other technique such as normal (Gaussian) distribution.

$$P = \left(X_i = x_i \middle| C = c_j\right) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} exp\left(-\frac{\left(X_i - \mu_{ij}\right)^2}{2\sigma_{ij}^2}\right) \quad (2.5)$$

Where:

: Opportunity

: The attribute

: The attribute value

: Class

: The sub class

: The mean of all attributes

: Standard deviation

2.6.1 Using Naïve Bayes Algorithm

Naive Bayes classifier makes statistical estimation and is based on Bayes Theorem.

A and B are random numbers;

$$P(A \mid B) = P(B \mid A)\,P(A) / P(B) \quad (2.6)$$

P (A): The independent probability of event A is the primitive probability.

P (B): Independent probability of event B.

P (A | B): Probability of B event (conditional probability) when it is known that A event occurs.

P (A | B): Probability of event A (aftershock probability) when event B is known.

The Naive Bayes concept can be explained as follows: X is considered an instance of unknown class membership. Example X = {x1, x2,…, xn} consists of attribute values. In this example class, it is assumed to be n class. C1, C2,…, Cn class values are accepted. The following possibilities are calculated for the sample that will determine the class.

P (〚X|C〛_i) = (P(X | C_i) P (C_i))/(P(X)) (2.7)

Simplification is made for the probability of P (X | Ci) to reduce the processing load in the calculation. Assuming that the Xi values of the sample are independent of each other, the following relation is used.

P (X|C_i) = ∏_(k=1)^n 〖P(X_k ┃ C_i) 〗 (2.8)

In order to classify the unknown example X, it is sufficient to compare only the numerator values since the denominators in P (Ci | X) are equal to each other. The class of the unknown instance is considered to be the same as the class of the largest of these values.

(_C_i^argmax) {P (X|C_i) P (C_i)}

The above equation, post-probabilities, is also known as the largest post-classification method (MAP).

C_MAP=(_C_i^ argma {(k=1)^n 〖P(X_k ┃C_i) 〗} (2.9)

2.6.2 Confusion Matrix

Confusion matrix is used to portion the presentation of a classification algorithm. The terminology connected to the confusion matrix can be rather confusing, but the matrix itself is simple to understand in Table 2.1.

Table 2.1 Confusion Matrix

Actual True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

In Table 2.1, 'True' or 'False' indicates if the class is properly predicted or not, while 'Positive' or 'Negative' indicates the prediction of the class of people causes mental illness or not. From the confusion matrix, accuracy of the system's performance. Where the formula of each of these things are:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (2.10)$$

CHAPTER 3

CLASSIFICATION OF PSYCHOLOGICAL ILLNESS

In this section, Naive Bayes classification is definitely described. Creating tuple and their attributes of classification system is explained using Naïve Bayes Classifier. Statistical Machine Translation system is presented in detail in this chapter.

3.1 Classification and Prediction

Classification is the process of finding a set of models that describe and distinguish data classes or concepts, and for the purpose being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known). Data classification is a two-step process.

In the first step, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, called the class label attribute. The data tuples analyzed to build a model collectively from the training dataset. The class label of each training sample is provided, and this step is known as the supervised machine learning. Typically, the learned model is represented in the form of classification rules, decision trees, or mathematical formulae. The rules can be used to categorize data samples, and provide a better understanding of database contents.

In the second step, the model is used for classification. The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae and neural networks. Classification can be used for predicting the class label of data objects. However, in many applications users may wish to predict some missing and unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data and is often specifically referred to as prediction. Although prediction may refer to both data value prediction and, thus, is distinct from classification. Prediction also encompasses the identification of distribution trends based on the available data.

3.2 Comparing Classification Methods

Classification methods can be compared and evaluated according to the criteria:

- Predictive accuracy: It refers to the ability of the model to correctly predict the class label of new or previously unseen data.
- Speed : It refers to the computation costs involved in generating that are used in the model.
- Robustness : It is the ability of the model to make correct predictions that involve noisy data or data with missing values.
- Scalability : It refers to the ability to construct the model efficiently given large amounts of data.
- Interpretability : It refers to the level of understanding and insight that is provided by the model.

3.3 Classification Methods

- Artificial Neural Network: Non-linear predictive models that learnt through training and resemble biological neural networks in structures.
- Bayesian Classification: Bayesian Classifier are statistical classifiers. They can predict class membership probabilities, such as probability that a given sample belongs to a particular class.
- Genetic Algorithms: Optimization techniques that uses processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
- Rough Set Approach: Routh set theory can be used for classification to discover structural relationships within imprecise or noisy data. It applies to discrete-valued attributes. Continuous-valued attributes must, therefore, be discretized prior to its use.
- Decision Trees: Tree-shaded structures that represent sets of decision. These decisions generate rules for the classification of dataset.
- Nearest Neighbor Method: A technique that classifies each record in a dataset on a combination of the classes of the k record (s) most similar to it in a historical dataset. Sometimes called the k- nearest neighbor technique.

- Rule Induction: The extraction of useful if-then rule from data based on statistical significance.

3.4 Data Collection of Proposed System

There are two datasets in the system. They are Dataset 1 and Dataset 2 and are collected from Kaggle.com data free site. In dataset 1, there are 158 instances with 27 attributes and 1154 instances with 22 attributes in dataset 2.

Table 3.1 Sample 21 instances in Dataset 1

Gender	Are you above 30 years of age?	Employment Status	How are you feeling today?	Eating and sleeping	(If sad)have you been in the same mental state for the past few days?	Is your sadness momentarily or has it been constant for a long time?	At what time of the day are you extremely low?	Has there been a sudden and huge change in your life?	Your stress is related to which of the following areas?
Female	No	Student	Fine	Yes	No	For some time	Evening	No	Personal
Male	No	Student	Fine	No	Yes	For some time	Morning	Yes	Personal
Male	No	Student	Fine	No	Maybe	Significant time	Evening	No	Personal
Female	No	Student	Fine	No	Maybe	Significant time	Evening	Not sure	Home
Female	No	Student	Good	Yes	No	Not sad	Afternoon	Yes	Personal
Female	No	Student	Fine	No	Maybe	Significant time	Evening	Not sure	Work
Female	No	Student	Sad	No	Maybe	For some time	Afternoon	Yes	Personal
Female	No	Student	Sad	Maybe	Yes	For some time	Afternoon	Yes	Personal
Female	No	Student	Fine	No	No	Significant time	Evening	Yes	Personal
Female	No	Unemployed	Fine	No	Yes	For some time	Morning	Yes	Personal
Male	No	Student	Fine	No	No	Not sad	Evening	No	Home
Male	No	Employed	Good	Yes	No	For some time	Afternoon	Yes	None
Female	No	Student	Fine	No	No	Not sad	Afternoon	No	Work
Female	No	Student	Fine	Yes	No	For some time	Afternoon	Not sure	Work
Male	No	Student	Good	Yes	No	Not sad	Morning	No	None
Female	No	Student	Sad	Yes	No	Long time	Evening	Yes	Home
Male	No	Student	Fine	Yes	No	Long time	Morning	Yes	Personal
Female	Yes	Unemployed	Sad	Maybe	No	For some time	Evening	Yes	Personal
Female	No	Student	Sad	No	Yes	For some time	Morning	Not sure	Personal
Female	No	Student	Good	Yes	No	For some time	Evening	Not sure	Financial
Male	No	Student	Fine	Yes	No	Not sad	Morning	No	Work

3.5 Data Preprocessing

Real world data are generally incomplete, noisy, and inconsistent. Data preprocessing is an important issue for both data warehousing and data mining as real-world data tend to be incomplete, noisy, and inconsistent. It includes data cleaning, data integration, data transformation, and data reduction.

Data cleaning routines can be used to fill in missing values, smooth noisy data, identify outliers and correct data inconsistencies. Data integration combines data from multiple sources form a coherent data store. Metadata, correlation analysis, data conflict detection, and the resolution of semantic heterogeneity contribute towards smooth data integration. Data transformation routines convert the data into appropriate forms for mining. Data reduction techniques such as data cube aggregation, dimension reduction, data compression, numerosity reduction, and discretization can be used to obtain a reduce representation of the data while minimizing the loss of information content.

CHAPTER 4
SYSTEM DESIGN AND IMPLEMENTATION

The proposed system's implementation is presented details in this chapter. The system design and, structure of the system is described in this chapter. Two datasets of the system are explained with Dataset 1 and Dataset 2 and implementation of the system containing components are also explained in this chapter. Finally, a graphical user interface of the system is presented step-by-step detailed explanation with figures and experimental results of the system are displayed with table and bar chat.

4.1 Overview Design of the System

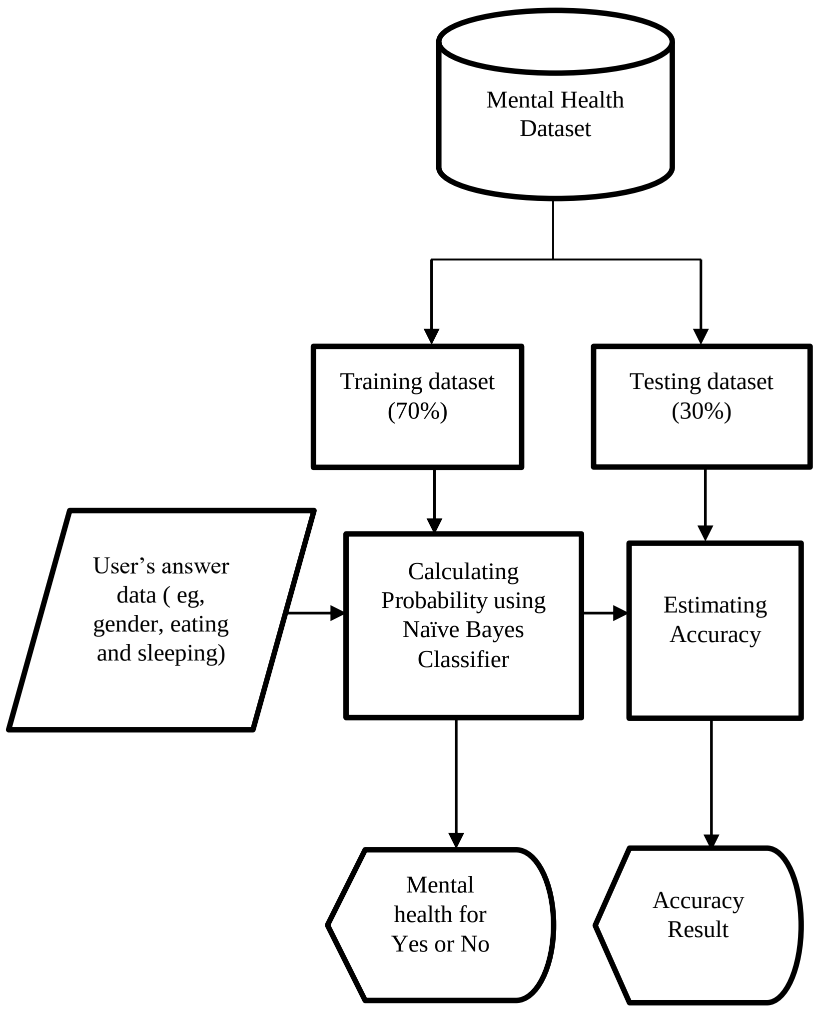


Figure 4.1 Overview Design of the System

The flow diagram of the proposed system is shown in Figure 4.1. The proposed system is implemented to classify where the patients or users causes psychological illness or not using Naïve Bayes classifier. Mental health dataset 159 instances are used for this system from Kaggle.com. There are 27 attributes, Prediction status (Yes) or (No) are defined as Class.

While the system is started, the 159 users' answers for 28 questions assigned attributes for the system are collected as dataset. Dataset is divided into two datasets: training (70%) and testing (30%).

70% training dataset is calculated to get probability result by using Naïve Bayes algorithm. And then the user is classified where he or she causes mental illness problems or not according to these probability result.

And the performance of the system is evaluated with accuracy result by using Random Split Test, Folds Test and Unlabeled Test. Finally, the system displays the user causes mental illness class yes or no and evaluation results with table and bar chat.

4.2 Two Datasets with different Attributes

In this system, 27 and 21 different attributes of dataset1 and 2 are shown. In dataset1, there are 158 instances are used and 500 instances are used in dataset2.

Table 4.1 Attributes for Dataset 1

No	Attributes Name	Attributes Value
1	Gender	Female, Male
2	Above 30 years?	Yes, No
3	Employment	Student, Unemployed, Employed
4	Today feeling ?	Fine, Good, Sad, Depressed
5	Eating and Sleeping	Yes, No, Maybe
6	(If sad)for the past few days	Yes, No, Maybe
7	Is your sadness for a long time?	For some time, Significant time, Not sad, Long time
8	At what time of the day low?	Evening, Morning, Afternoon
9	Sudden and huge change in your life?	Yes, No, Not sure
10	Your stress?	Personal, , None, Financial, Work
11	Little pleasure or interest in the activities usually enjoy?	Very Often, Sometimes, Never, Often
12	Confident you feel in your capabilities recently.	1,2,3,4,5

13	Supported you feel by others around your friends, family.	Highly supportive, Little bit, Satisfactory, Not at all
14	Frequently have you been doing things?	Very Often, Sometimes, Never, Often
15	Mental Health Condition	Yes, No, Maybe
16	How easy is it for you to take medical leave for a mental health condition?	Not so easy, Very easy, Difficult, Easy
17	You use substance abuse(e.g. smoking, alcohol)?	Never, Often, Sometimes, Very Often
18	You take medication in the near past for mental health?	Yes, No, Maybe
19	Having trouble concentrating on things?	Yes, No, Maybe
20	Do you feel bad about yourself or your family down?	Yes, No, Maybe
21	Hours you spend per day on watching mobile phone?	1-2 hours, 5-10 hours, More than 10 hours, 2-5 hours
22	Appointment with a psychologist for your current mental state?	1, 2, 3, 4, 5
23	COVID-19 pandemic affected your mental ?	Yes, No, Not sure
24	How often do you get offended or angry or start crying?	Never, Sometimes, Often, Very often
25	Feel yourself vulnerable or lonely?	1, 2, 3, 4, 5
26	Comfortable about your mental health?	1, 2, 3, 4, 5

27 attributes of dataset 1 are described in Table 4.1.

In Table 4.1, 22 attributes of dataset2 are shown. Some attributes of two datasets are the same and some are different.

Table 4.2 Attributes for Dataset 2

No	Attributes Name	Attributes Value
1	Does employer provide mental health benefits as healthcare coverage?	Yes, No, don’t know, not eligible
2	Do the options for mental health care available under employer-provided coverage?	Yes, No, not sure
3	Has employer ever formally discussed mental health?	Yes, No, don’t know
4	Does employer offer resources to learn more about mental health concerns and options for seeking help?	Yes, No, don’t know

5	If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:	Very easy, very difficult, somewhat easy, somewhat difficult
6	Would you feel comfortable discussing a mental health disorder with your coworkers?	Yes, No, Maybe
7	Do you feel that your employer takes mental health as seriously as physical health?	Yes, No, don't know
8	Do you know local or online resources to seek help for a mental health disorder?	Yes, No, don't know
9	Do you believe your productivity is ever affected by a mental health issue?	Yes, No, Unsure
10	If yes, what percentage of your work time (time performing primary or secondary job functions) is affected by a mental health issue?	1-25%, 26-50%, 51-75%, 76-100%
11	How willing would you be to share with friends and family that you have a mental illness?	Very open, somewhat open, neutral
12	Do you have a family history of mental illness?	Yes, No, don't know
13	Have you had a mental health disorder in the past?	Yes, No, Maybe
14	Do you currently have a mental health disorder?	Yes, No, Maybe
15	Have you been diagnosed with a mental health condition by a medical professional?	Yes, No
16	If so, what condition(s) were you diagnosed with?	Anxiety, Mood
17	Have you ever sought treatment for a mental health issue from a mental health professional?	1,0
18	If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?	Sometimes, Often, Never
19	If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?	Sometimes, Often, Never
20	Age	Above 20 years
21	Gender	Male, Female
22	Do you work remotely?	Sometimes, Often, Never

4.3 Implementation of the system by using Naïve Bayesian Classification

When the system is started, Home Page is displayed in Figure 4.2. In Home Page, Thesis's Title "Classification of Psychological Illness using Naïve Bayes", the author's name and Roll No "Ma Kay Khaing Soe, MCS-19", and Supervisor's name and rank "Dr. Win Lae Hnin, Associate Professor".

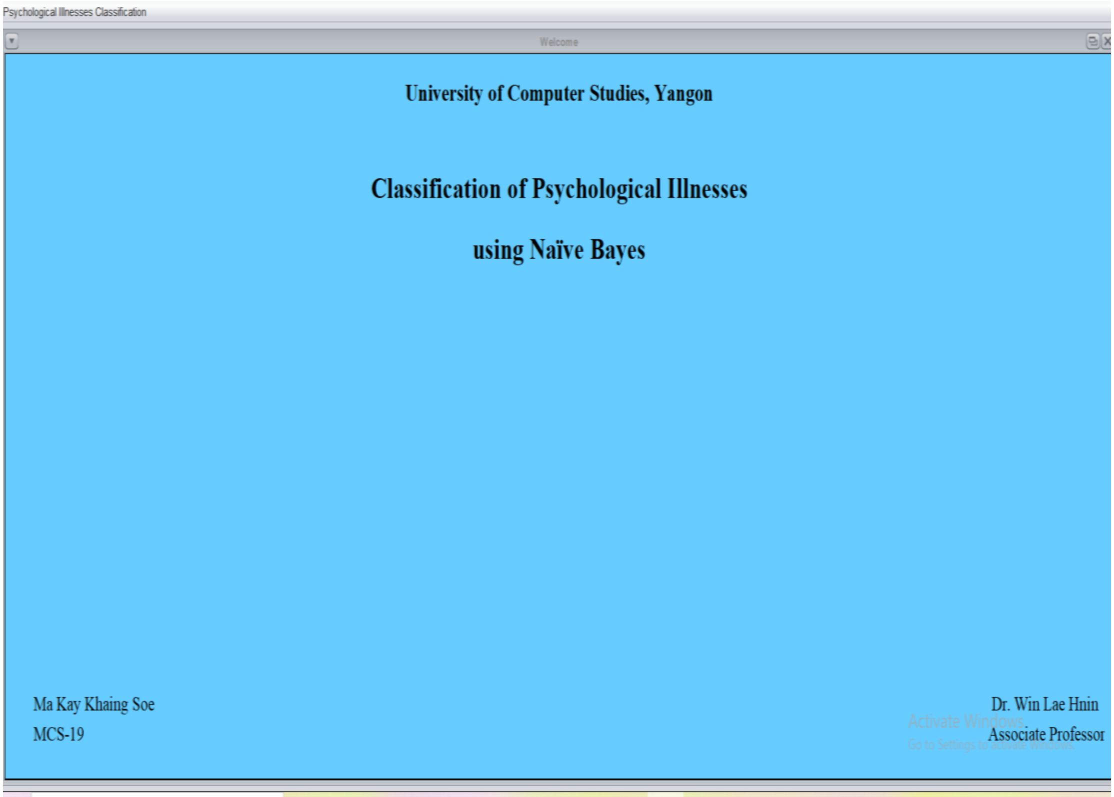


Figure 4.2 Home Page of the System

In this system's menu, there are three main tasks: Preprocessing, Train Naïve Bayes Classifier, and Test Naïve Bayes Classifier as shown in Figure 4.3.

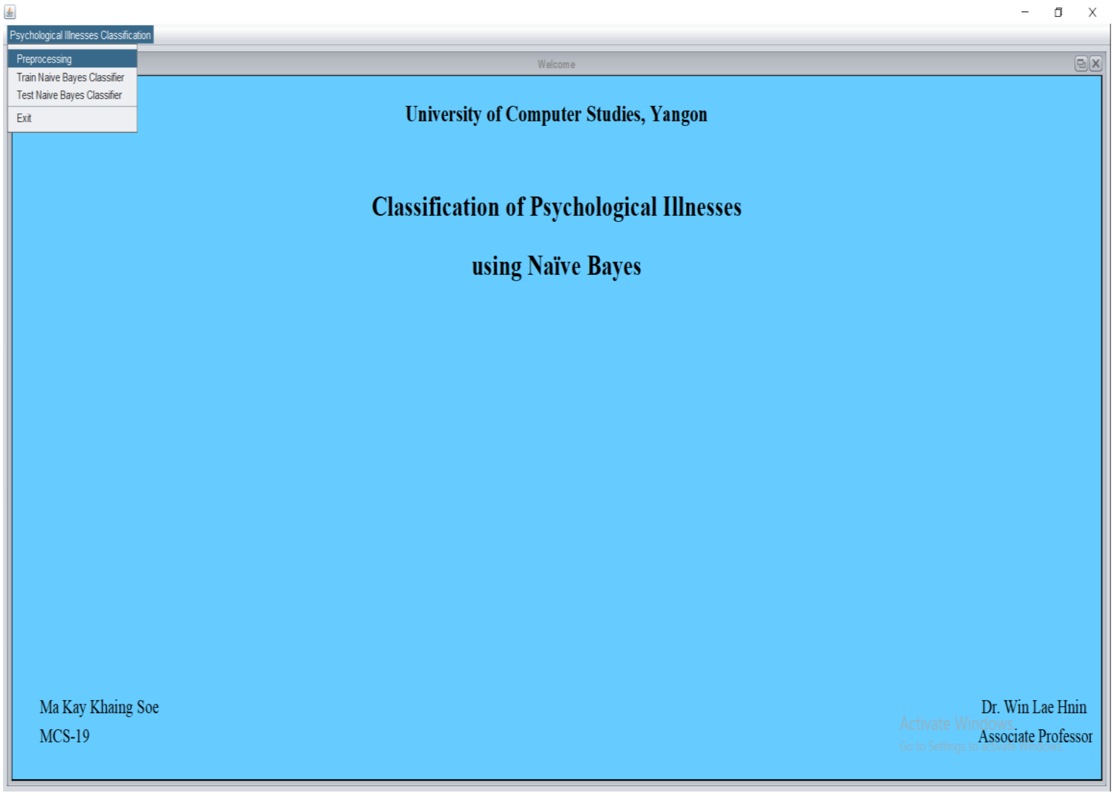


Figure 4.3 Menu Page of the System

4.3.1. Preprocessing

First task is "Preprocessing". In "Preprocessing", there are four parts: Input Dataset (xlsx), Input Dataset (csv), Normalization, and Split Training or Testing. In Figure 4.4 and Figure 4.5, "Input Dataset (xlsx)" button is clicked, one of two datasets is selected by clicking "Select Excel Dataset" button, and then "Open" button is clicked. In these Figures, Dataset 1 is selected.

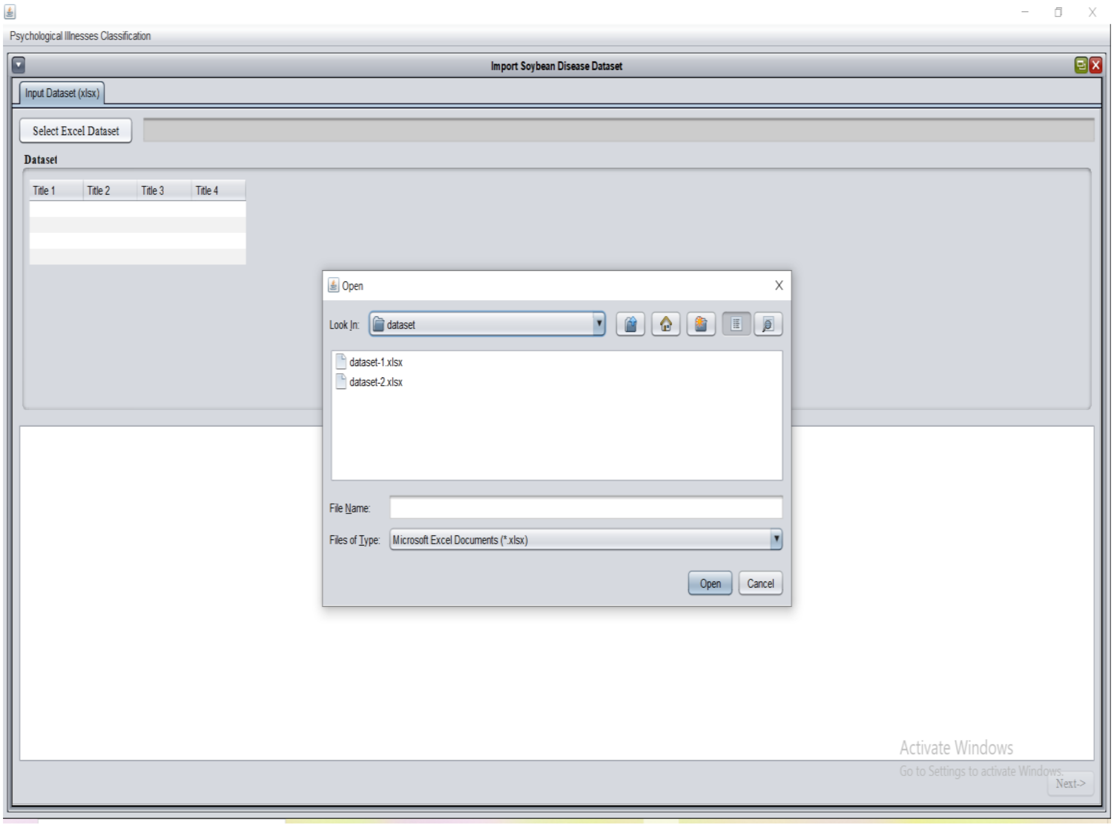


Figure 4.4 Inserting Dataset(xlsx) Page

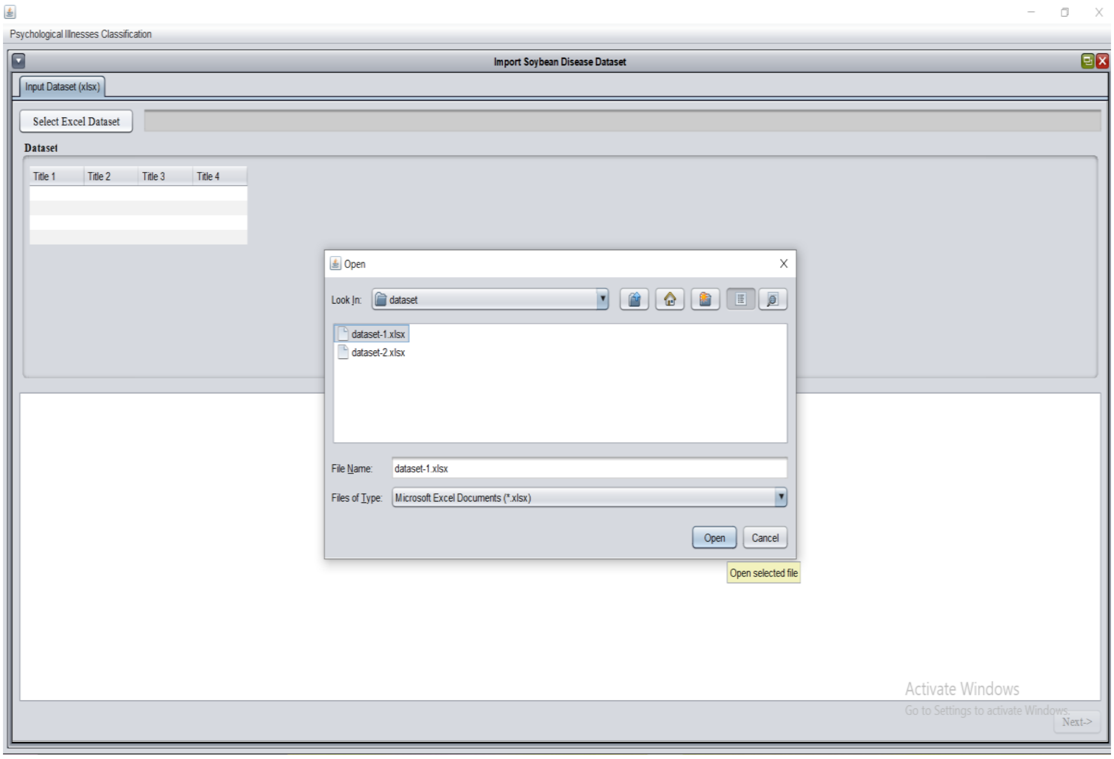


Figure 4.5 Selecting Dataset(xlsx) Page

After selecting Dataset 1, Dataset 1 containing 157 instances and 27 attributes with excel file (xlsx) is seen in Figure 4.6. And attribute names with letters in Figure 4.7 is shown when “Change Attribute Names” button is clicked.

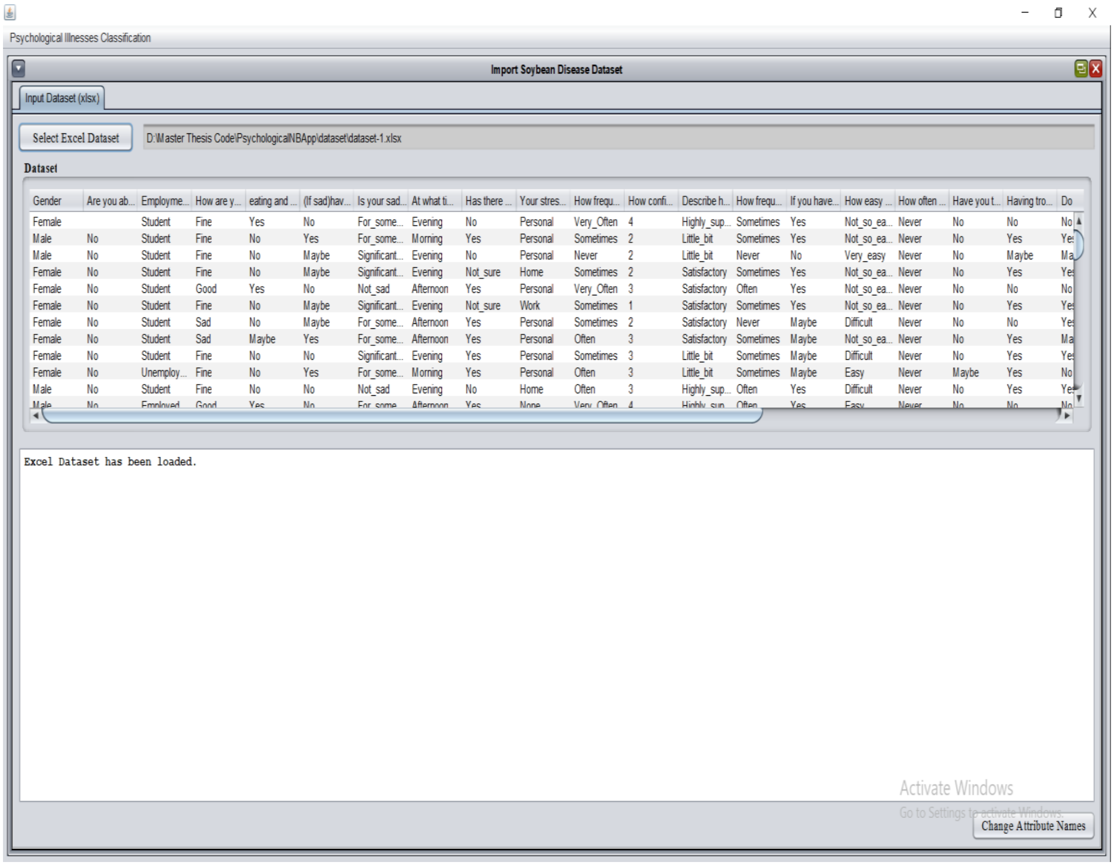


Figure 4.6 Dataset 1 (xlsx) Page with Attribute Names

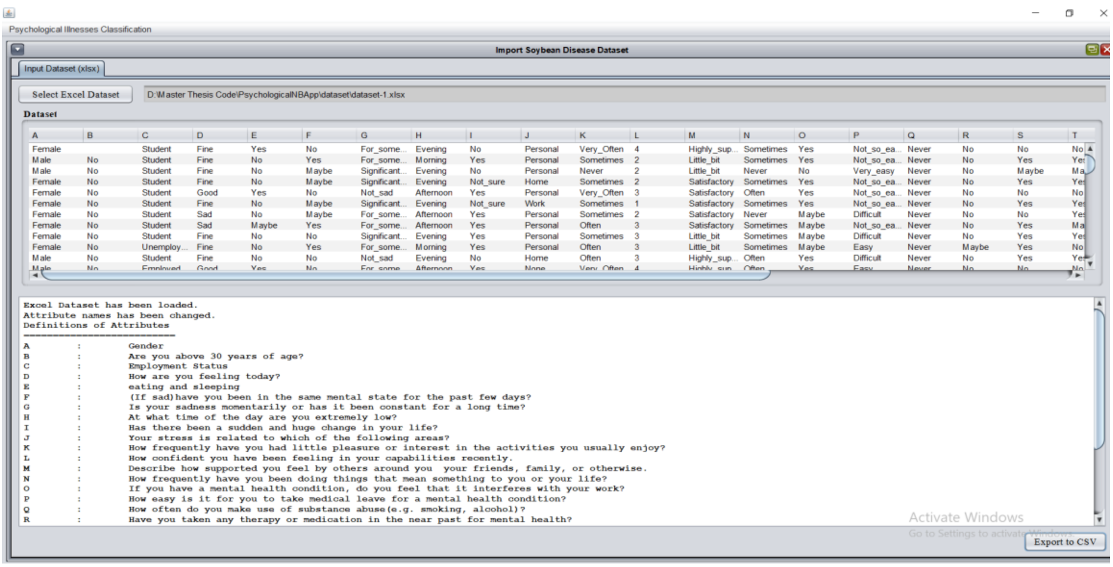


Figure 4.7 Dataset 1 (xlsx) Page changed Attribute Names

Dataset 1 with attribute names changed is saved with (csv) by clicking "Export to CSV" and "Next" button is clicked to go next part in Figure 4.8.

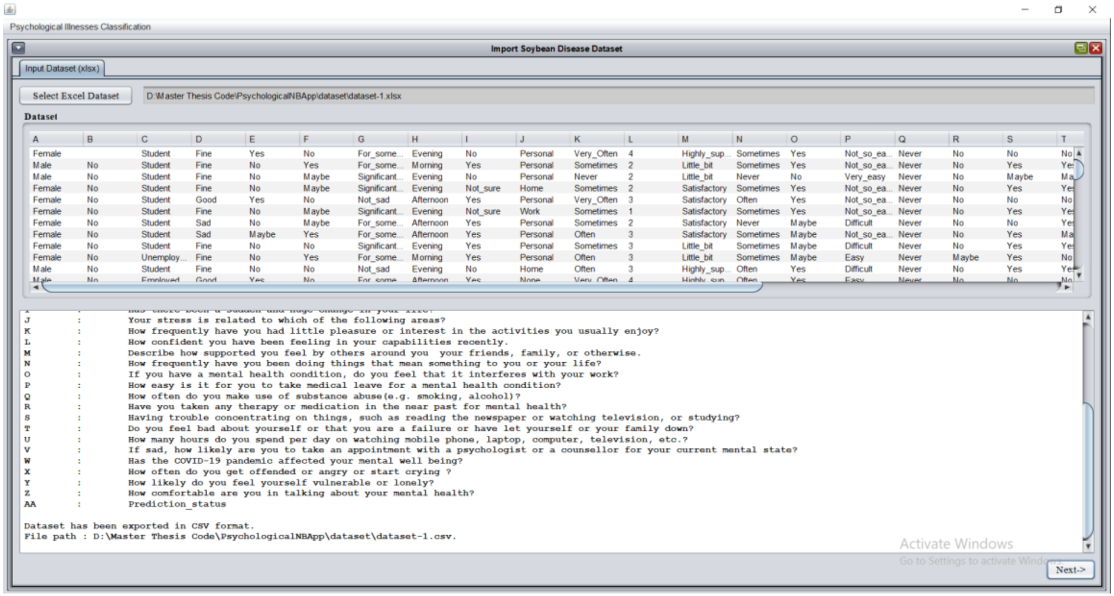


Figure 4.8 Dataset 1 (xlsx) Page changed Attribute Names to Next Step

In the second part, Input Dataset (csv) is clicked and then Dataset 1 (csv) is selected by using "Select Dataset (CSV)" button and "Open" button as shown in Figure 4.9 and Figure 4.10 because the system processes CSV file.

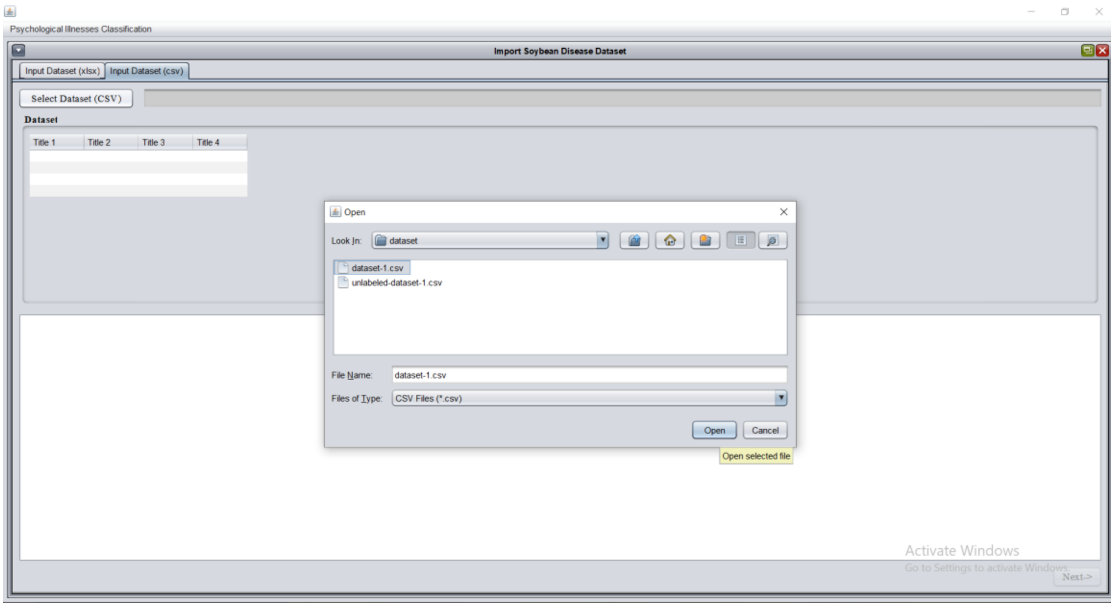


Figure 4.9 Inserting Dataset 1 (csv) Page

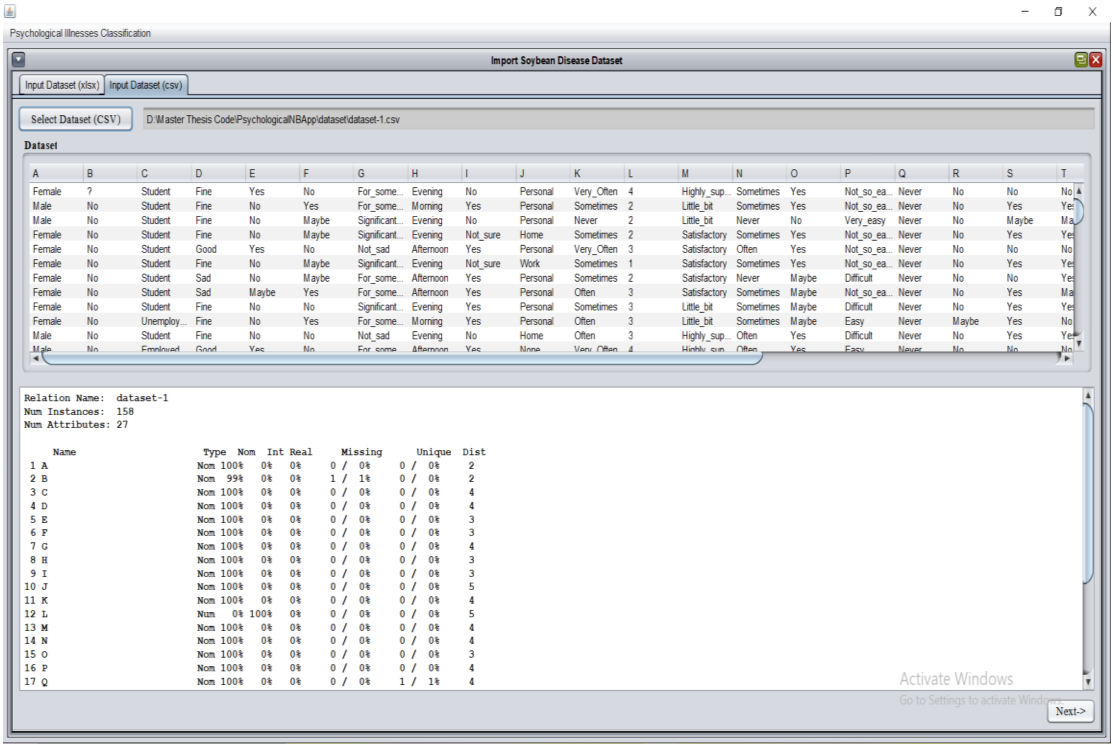


Figure 4.10 Dataset 1 (csv) Page

The third part is “Normalization”. Dataset 1 (csv) is normalized as shown in Figure 4.11 and Figure 4.12.

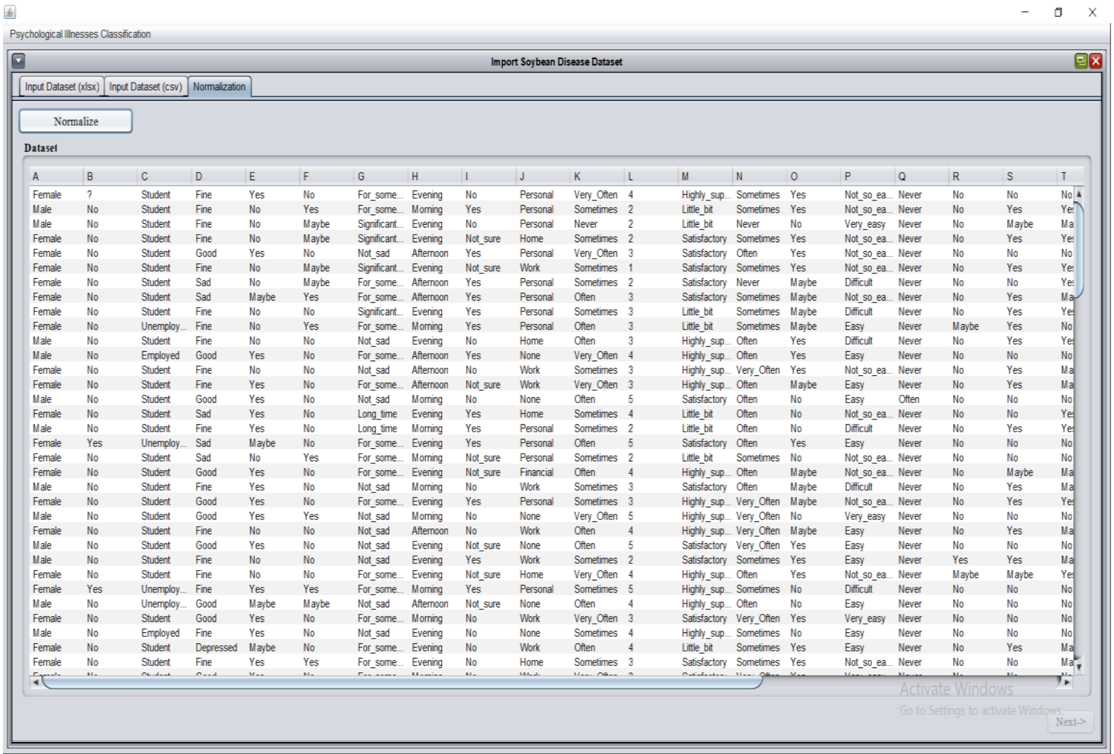


Figure 4.11 Normalization Page

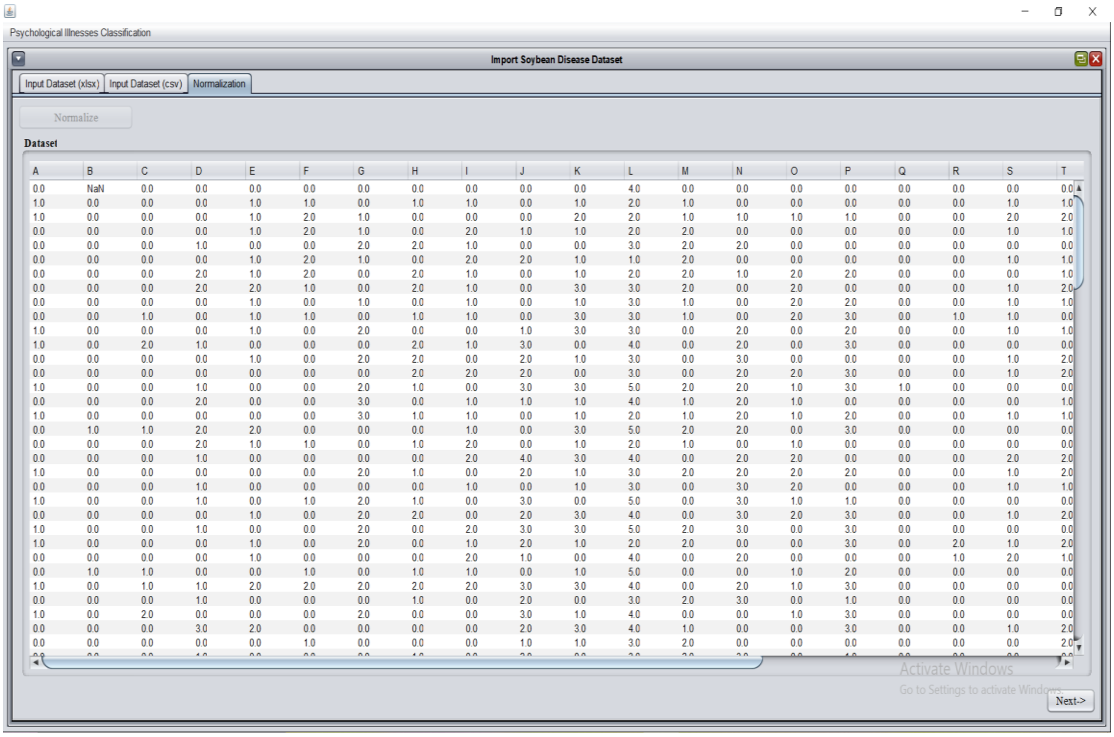


Figure 4.12 Normalization Page to Next Step

In the forth part, "(%) in Taining" text box is typed after clicking "Split Training/Testing" button. Now training percent is typed With 70% and then testing is 30%. As there are 158 instances in Dataset 1, training 70% and testing 30% are training 111 instances and testing 47 instances shown in Figure 4.13.

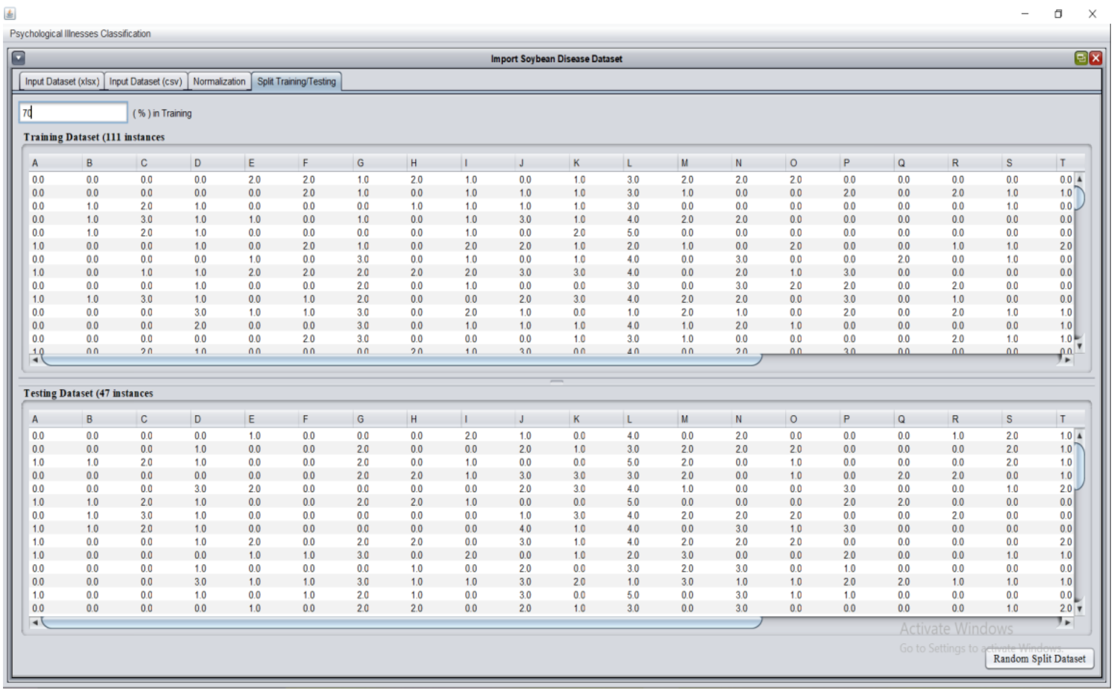


Figure 4.13 Spliting Training or Testing Page

4.3.2. Training Naïve Bayes Classifier

Second task of the system is processed by clicking “Train Naïve Bayes Classifier” shown Figure 4.14.

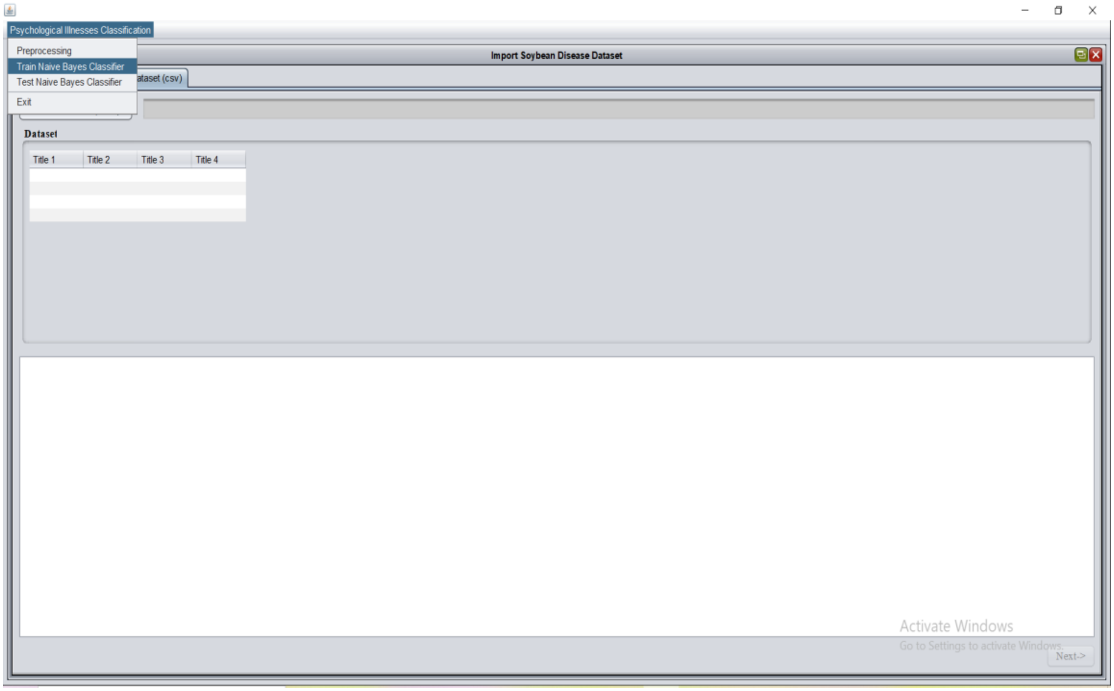


Figure 4.14 Training Page with Naïve Bayes Classifier

111 instances of 70% training dataset are shown in Figure 4.15 and are often refreshed by clicking “Refresh” if “Build Classifier” button is clicked, the system displays the probabilities value of the whole trained instances, and each instance based on Class Label “Yes” and “No” shown in Figure 4.16.

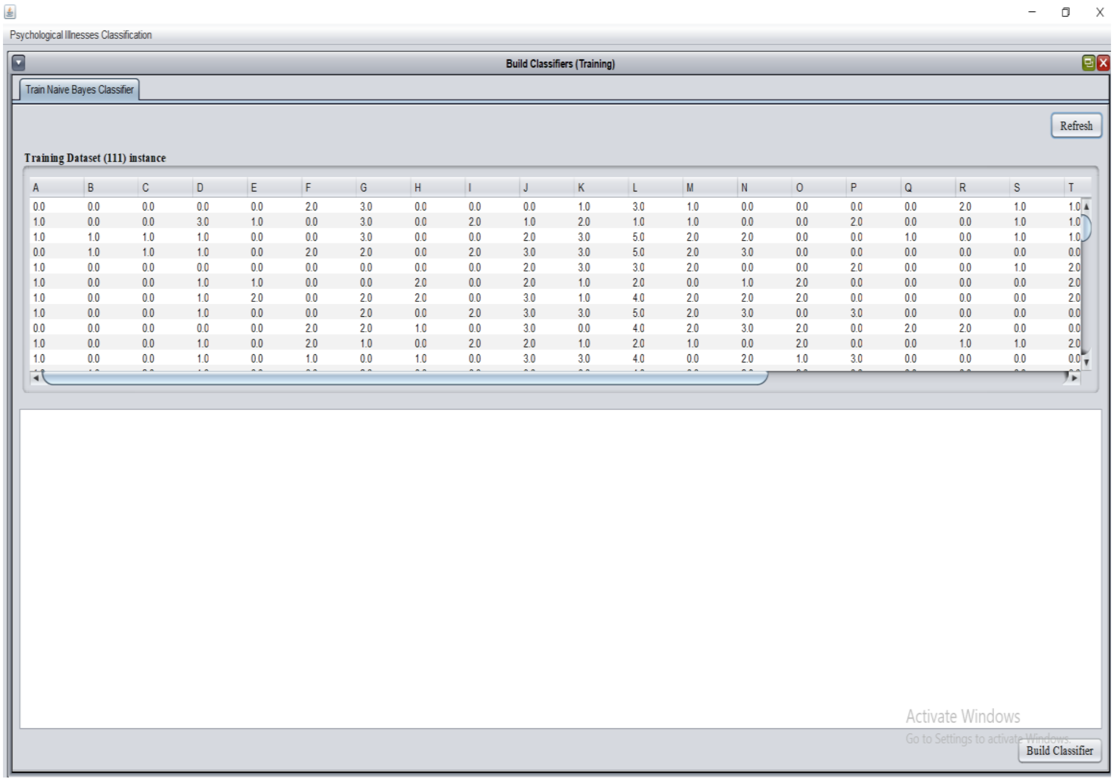


Figure 4.15 Spliting Training Page

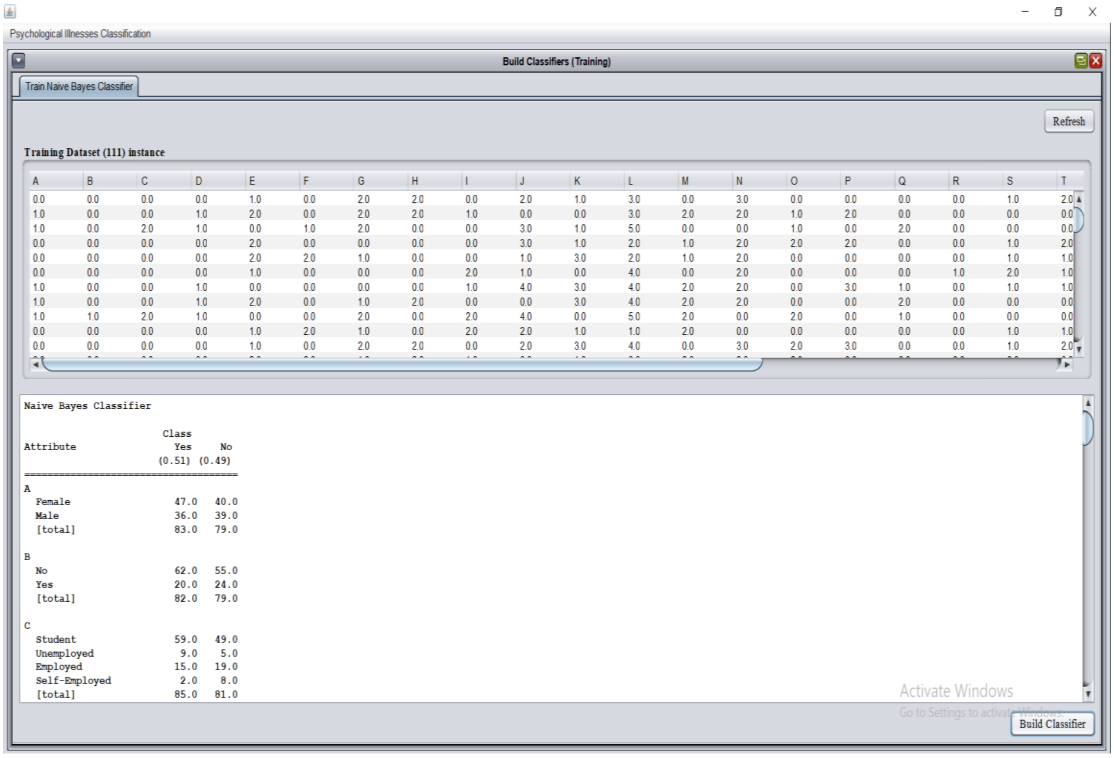


Figure 4.16 Training Dataset 1 and their Probability Values Page

4.3.3. Testing Naïve Bayes Classifier

The third main task is testing of 30% of Dataset 1. If “Test Naïve Bayes Classifier” button is clicked, five menus is appeared shown in Figure 4.17 and Figure 4.18. In the testing, three testing: Random Split Test, K-Folds Test, and Unlabeled test.

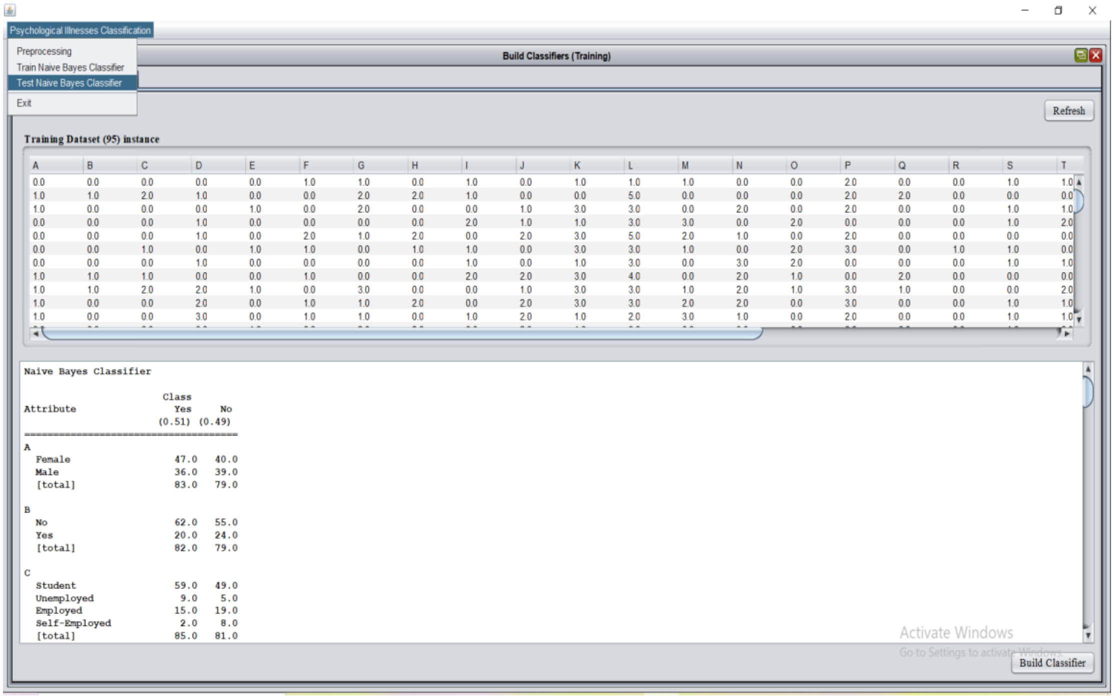


Figure 4.17 Testing Menu of the System

“Random Split Test” is a method of conducting controlled, randomized experiments with the goal of improving metric. In this system, random split (70% training and 30% testing) is evaluated. Moreover 50%:50%,60%:40, and customized and its test result and result with graph are shown in Figure 4.17, Figure 4.18 and Figure 4.19.

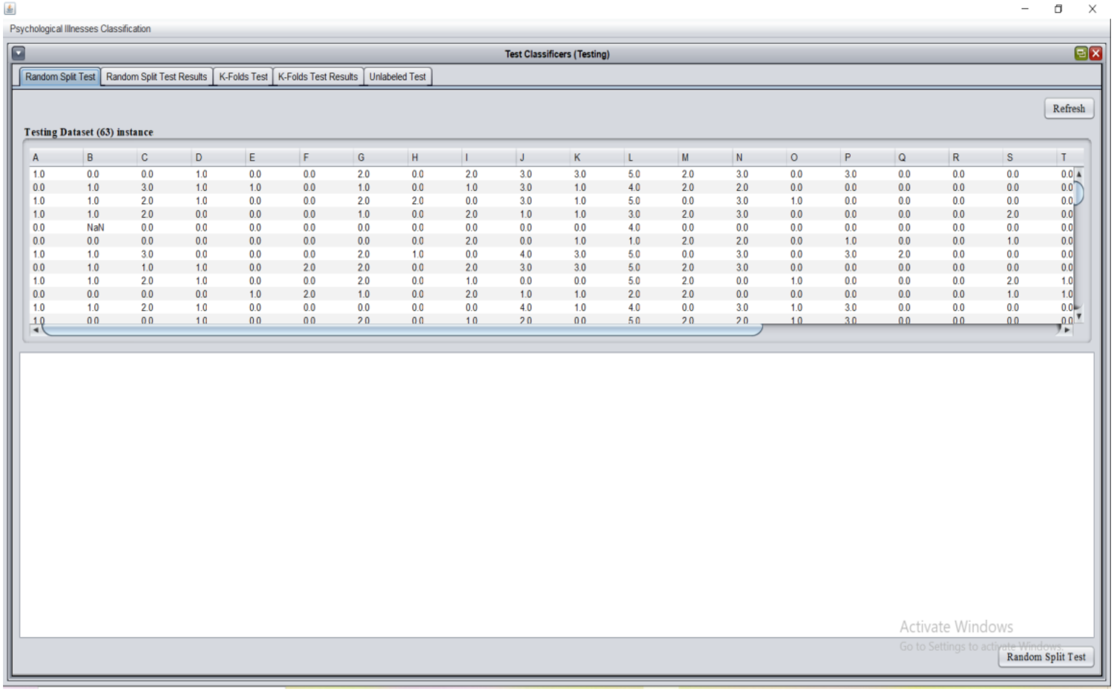


Figure 4.18 Testing Page

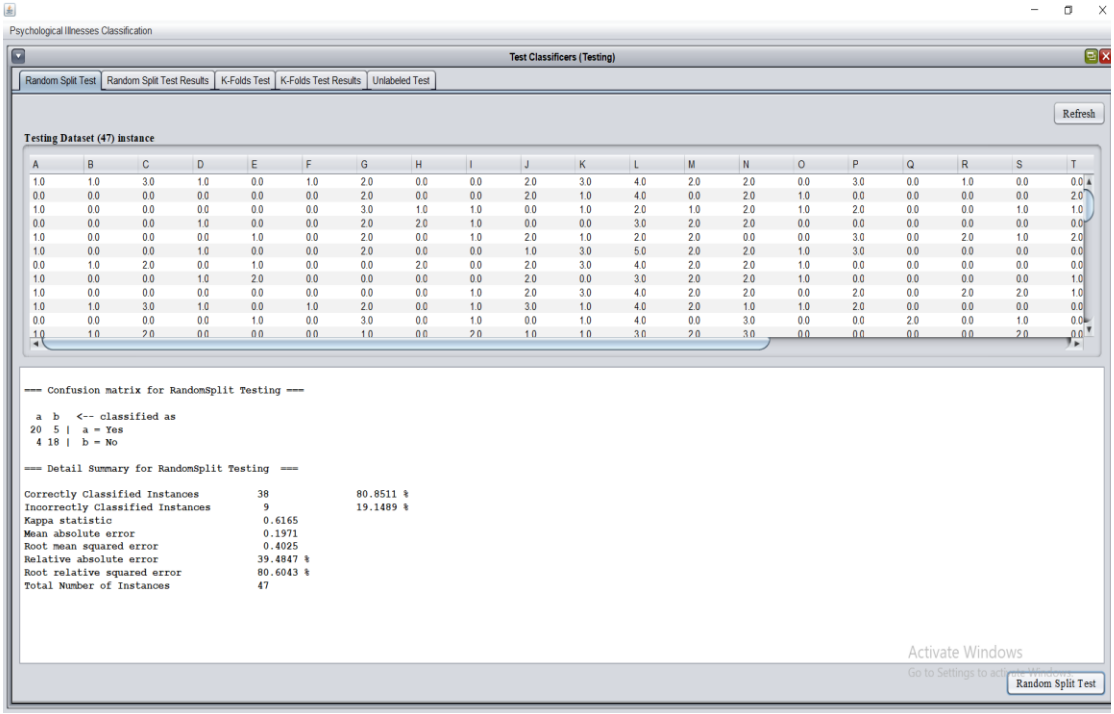


Figure 4.19 Random Split Testing Page

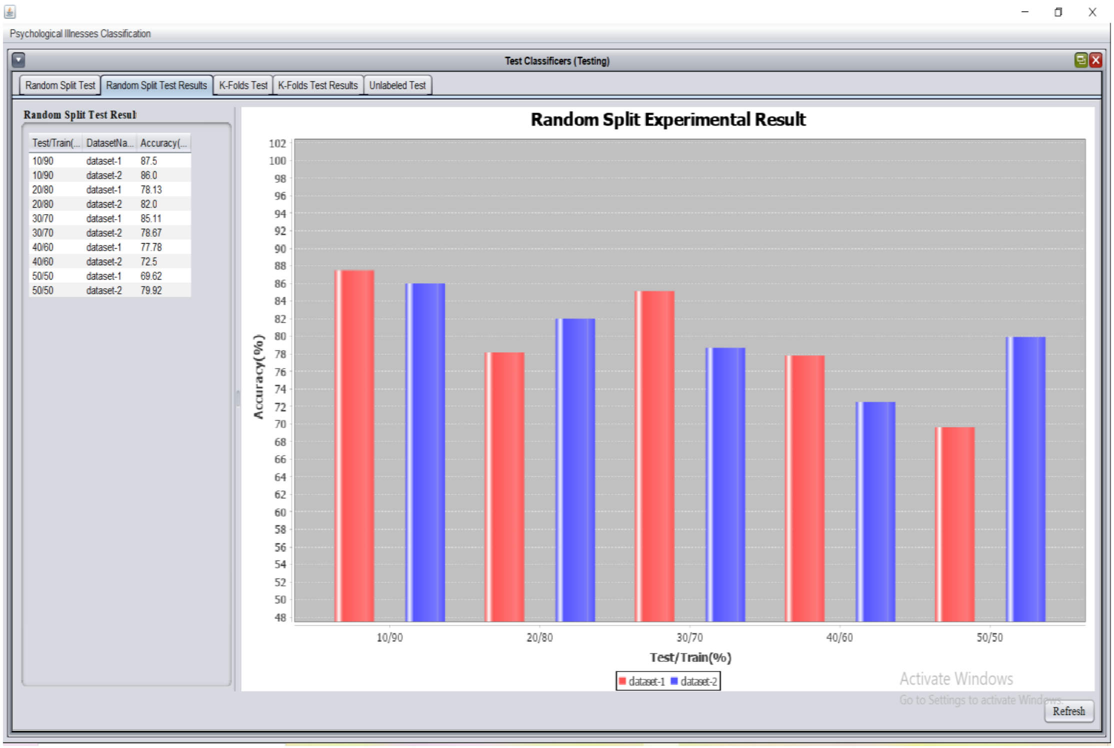


Figure 4.20 Random Split Testing Result Page

"K-Folds Test" is a validation technique in which data is split into k-subsets and the holdout method is repeated k-times where each of the k-subsets are used test set and other k-1 subsets are used for the training purpose and its test result and result with graph are shown in Figure 4.21 and Figure 4.22.

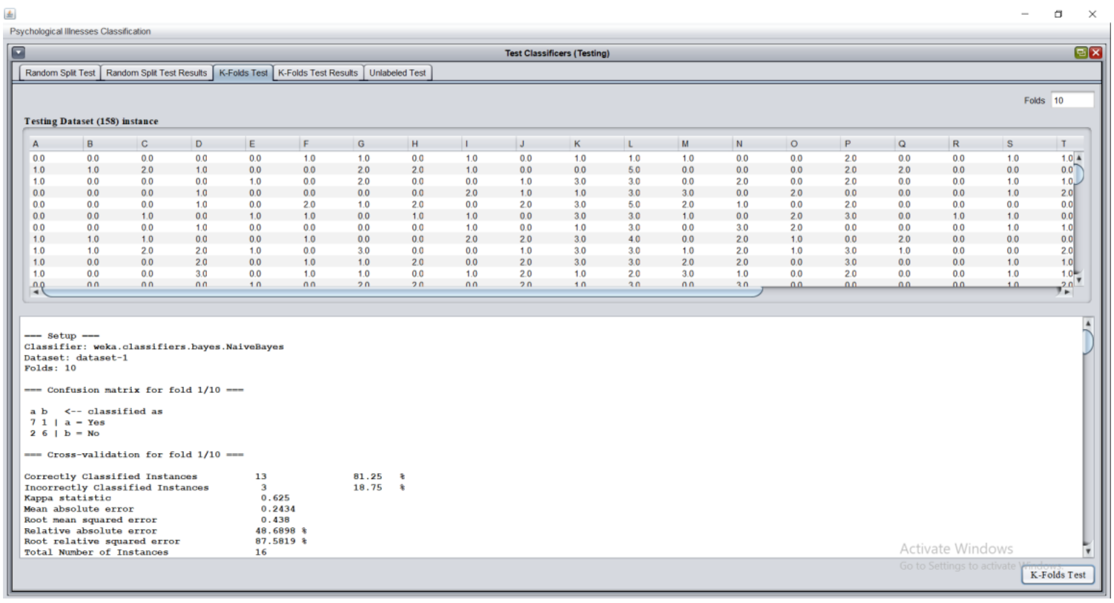


Figure 4.21 K-Folds Testing Page

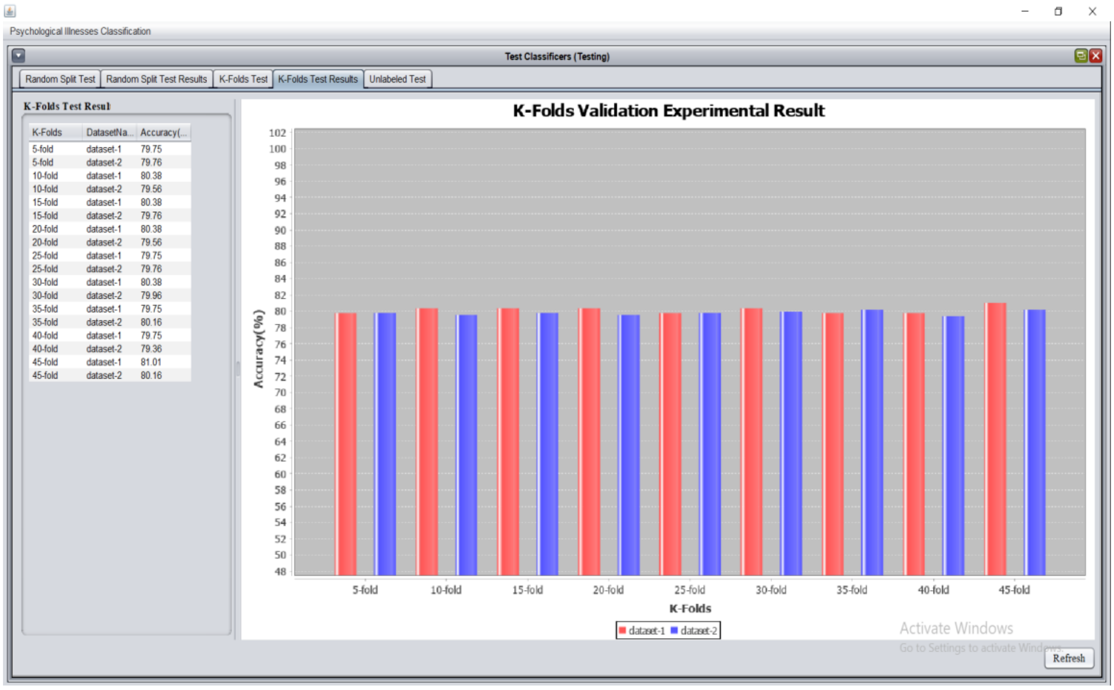


Figure 4.22 K-Folds Testing Result Page

"Unlabeled Test" is to test for the new dataset shown in Figure 4.23.

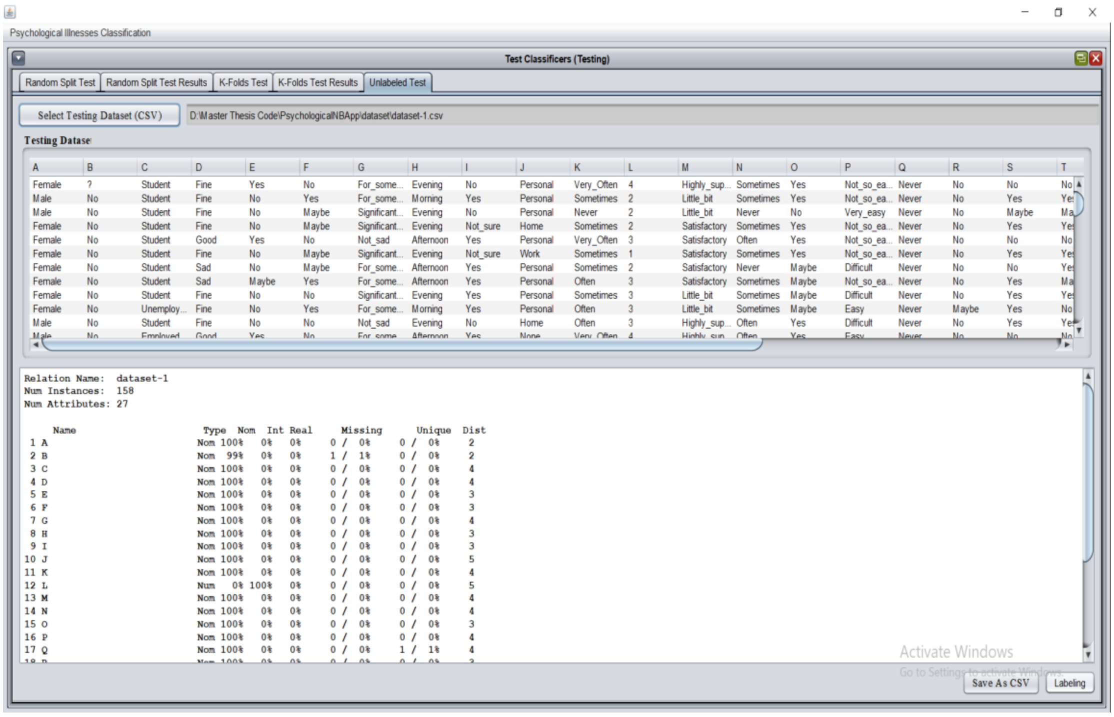


Figure 4.23 Unlabeled Test

4.4 Predicting a class label by using Naïve Bayesian Classification

Let X be a data sample whose class label is unknown. Each data sample is represented by an n dimensional feature vector, X= (x_1, x_2, …,x_n).

In this sample Dataset1, data sample X=(age=)

Class- C1, Yes

Class- C2, No

P(C1) = 11/20 = 0.55

P(C2) = 9/20 = 0.45

P(Q1 = female|C1) = 9/11 = 0.818

P(Q1 = female|C2) = 5/9 = 0.555

P(Q2 = no|C1) = 10/11 = 0.909

P(Q2 = no|C2) = 9/9 = 1

P(Q3 = student|C1) = 9/11 = 0.818

P(Q3 = student|C2) = 8/9 = 0.888

P(Q4 = fine|C1) = 7/11 = 0.636

P(Q4 = fine|C2) = 4/9 = 0.444

P(Q5 = yes|C1) = 3/11 = 0.272

P(Q5 = yes|C2) = 4/9 = 0.444

P(Q6 = no|C1) = 6/11 = 0.545

P(Q6 = no|C2) = 6/9 = 0.666

P(Q7 = for some time|C1) = 5/11 = 0.454

P(Q7 = for some time|C2) = 5/9 = 0.555

P(Q8 = evening|C1) = 7/11 = 0.636

P(Q8 = evening|C2) = 2/9 = 0.222

P(Q9 = no|C1) = 2/11 = 0.181

P(Q9 = no|C2) = 3/9 = 0.333

P(Q10 = personal|C1) = 7/11 = 0.636

P(Q10 = personal|C2) = 4/9 = 0.444

P(Q11 = very often|C1) = 1/11 = 0.090

P(Q11 = very often|C2) = 3/9 = 0.333

P(Q12 = 4|C1) = 2/11 = 0.181

P(Q12 = 4|C2) = 2/9 = 0.222

P(Q13 = highly supportive|C1) = 2/11 = 0.181

P(Q13 = highly supportive|C2) = 4/9 = 0.444

P(Q14 = sometime|C1) = 6/11 = 0.545

P(Q14 = sometime|C2) = 2/9 = 0.222

P(Q15 = yes|C1) = 5/11 = 0.454

P(Q15 = yes|C2) = 4/9 = 0.444

P(Q16 = not so easy|C1) = 5/11 = 0.454

P(Q16 = not so easy|C2) = 5/9 = 0.555

P(Q17 = never|C1) = 11/11 = 1

P(Q17 = never|C2) = 8/9 = 0.888

P(Q18 = no|C1) = 10/11 = 0.909

P(Q18 = no|C2) = 9/9 = 1

P(Q19 = no|C1) = 4/11 = 0.363

P(Q19 = noC2) = 4/9 = 0.444

P(Q20 = no|C1) = 3/11 = 0.272

P(Q20 = no|C2) = 4/9 = 0.444

P(Q21 = 1-2 hours|C1) = 2/11 = 0.181

P(Q21 = 1-2 hours|C2) = 1/9 = 0.111

P(Q22 = 1|C1) = 7/11 = 0.636

P(Q22 = 1|C2) = 8/9 = 0.888

P(Q23 = yes|C1) = 6/11 = 0.545

P(Q23 = yes|C2) = 5/9 = 0.555

P(Q24 = never|C1) = 1/11 = 0.090

P(Q24 = never|C2) = 1/9 = 0.111

P(Q25 = 1|C1) = 2/11 = 0.181

P(Q25 = 1|C2) = 1/9 = 0.111

P(Q26 = 1|C1) = 5/11 = 0.454

P(Q26 = 1|C2) = 1/9 = 0.111

P(X|Class=C1)= (P(Q1=female|C1) * P(Q2=no|C1) * P(Q3=student|C1) * P(Q4=fine|C1) * P(Q5=yes|C1) * P(Q6=no|C1) * P(Q7= for some time |C1) * P(Q8=evening|C1) * P(Q9=no|C1) * P(Q10=personal|C1) * P(Q11=very often|C1) * P(Q12=4|C1) * P(Q13= highly supportive |C1) * P(Q14=sometime|C1) * P(Q15=yes|C1) * P(Q16= not so easy |C1) * P(Q17=never|C1) *

P(Q18=no|C1) * P(Q19=no|C1) * P(Q20=no|C1) * P(Q21=1-2 hours|C1) * P(Q22=1|C1) * P(Q23=yes|C1) *
P(Q24=never|C1) * P(Q25=1|C1) * P(Q26=1|C1))
=0.818*0.909*0.818*0.636*0.272*0.545*0.454*0.636*0.181*0.636*0.090*
0.181*0.181*0.545*0.454*0.454*1*0.909*0.363*0.272*0.181*0.636*0.545*0.090*0.181*0.454
=2.629

P(X|Class=C2)= (P(Q1=female|C2) *P(Q2=no|C2)* P(Q3=student|C2) *P(Q4=fine|C2) * P(Q5=yes|C2) * P(Q6=no|C2) * P(Q7= for some time |C2) * P(Q8=evening|C2) * P(Q9=no|C2) * P(Q10=personal|C2) * P(Q11=very often|C2) * P(Q12=4|C2) * P(Q13= highly supportive |C2) * P(Q14=sometime|C2) * P(Q15=yes|C2) * P(Q16= not so easy |C2) * P(Q17=never|C2) * P(Q18=no|C2) * P(Q19=no|C2) *
P(Q20=no|C2) * P(Q21=1-2 hours|C2) * P(Q22=1|C2) * P(Q23=yes|C2) * P(Q24=never|C2) * P(Q25=1|C2) * P(Q26=1|C2))
=0.555*1*0.888*0.444*0.444*0.666*0.555*0.222*0.333*0.444*0.333*
0.222*0.444*0.222*0.444*0.555*0.888*1*0.444*0.444*0.111*0.888*0.555*0.111*0.111*0.111
=2.772

And then the user outputs maximum possible predicted results by comparing such probabilities.

P(X|Ci)*P(Ci)

P(X|Class=C1)*P(C1)

=2.629*0.55 =1.445

P(X|Class=C2)*P(C2)

=2.772*0.45 =1.247

Therefore, the Naïve Bayesian Classifier predicts class label =" Yes" for sample tuple X. In above sample calculation, probabilities values of Class 1 (Yes) and Class 2 (No) in each attribute, in all attributes, and the product of probability values of tuple x and probability of each class. Finally, the system classifies yes or no of each tuple depending upon the calculated probabilities of training dataset, testing data is process and classify class label Yes or No, and then unlabeled dataset is tested with answers of one patient, many patients and some patients and classified Class label.

4.5 Performance Result of the Proposed System

The system can help users to reduce the consuming time for confirming their conditions they feel. As the system's performance, Random Split Test's result and K-Folds Test's result are displayed with Table 5.1 and with bar chat in Figure 4.24.

Table 5.1 Accuracy Results

	Random Split Test (30%, 70%)	K-Folds Test 10-Folds
Dataset 1	80.38	85.11
Dataset 2	79.56	78.67

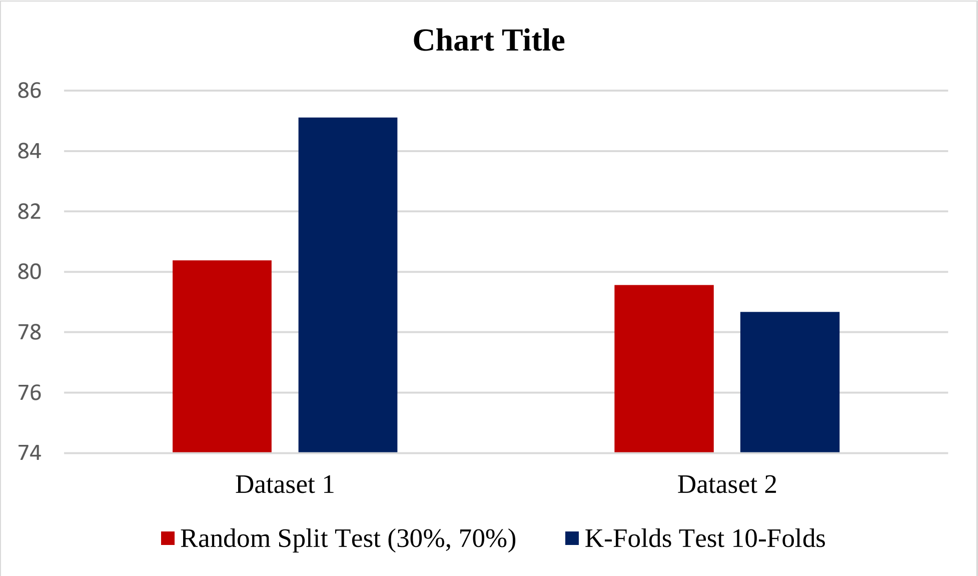


Figure 4.24 Accuracy Result with Chart

CHAPTER 5
CONCLUSION AND FURTHER EXTENSIONS

This thesis is developed to classify psychological illness disease from two datasets by using Naïve Bayes Classifier. The needed components of the system constructed and investigated and their overall performance of the system are analyzed. In this chapter, the thesis are concluded with advantages of the system, and future work are presented.

5.1 Conclusion

The proposed system presents the implementation of classification of psychological illness using Naïve Bayes. This system helps the users or patients to know whether user diagnosing the psychological illness causes or not. The questions what conditions they suffer are the attributes of the system and the most conditions which can cause psychological illness. The system displays the user who causes psychological illness or does not cause. The system is proved that the performance of the system is very good and improves the accuracy is over 80%. Also

5.2 Advantages of the System

The proposed system serves, user-friendly, and high-performance to the patients by diagnosing on their thoughts and conditions he/she feels. As a result, the classification system is more accurate in answering on the questions (attributes). The exactness and completeness of the proposed system are proved by the average value of accuracy which obtains over 80%.

5.3 Further Extensions

The proposed system is tested by using only two datasets from kaggle.com (data free site). The dataset can be extended with other resources, such as professionals or psychologists' patient records, others, by using Naïve Bayes Classifier. Obtaining a better result in this comparison is a motivation for further research by comparing Naïve Bayes Classifier with other rules.

REFERENCES

[1] Abiyoga, Arya Wicaksana and Ni Made Satvika Iswari, "Decision Support System for Choosing an Elective Course Using Naive Bayes Classifier", pp. 97-110, 2020. https://www.researchgate.net/publication/335366488_Decision_Support_System_for_Choosing_an_Elective_Course_Using_Naive_Bayes_Classifier

[2] Aditiya Hermawan, "Implementation of Naïve Bayes Algorithm for Classification of Mental Health of Social Media Users", Vol.4, No.2, December 2021. https://jurnal.kdi.or.id/index.php/bt/article/view/282

[3] Bandana Garg, "Design and Development of Naïve Bayes Classifier", 2013. https://library.ndsu.edu/ir/bitstream/handle/10365/23048/Grag_Design%20and%20Development%20of%20Naive%20Bayes%20Classifier.pdf?sequence=1

[4] B.Padmapriya, T.Velmurugan, "Classification Algorithm Based Analysis of Breast Cancer Data", International Journal of Data Mining Techniques and Applications, 2016, pp. 43-49. www.hindex.org/2016/article.php?page=1161

[5] C.W.Chein, C. C. Ming, L. C. Huang, B.H. Bernd, "A Bayesian Expert System for Clinical Detecting Coronary Artery Disease", JMed Sci, 2009. A Bayesian Expert System for Clinical Detecting Coronary Artery Disease - Google Search

[6] Daniel Berrar, "Bayes' Theorem and Naive Bayes Classifier", Vol 1, 2018, pp. 403-412. https://www.researchgate.net/publication/324933572_Bayes'_Theorem_and_Navie_Bayes_Classifier

[7] Diana Du Mtru, "Prediction of recurrent event in breast cancer using Naïve Bayes Classification", Annals of University Cralova, Math. Comp. Sei. Ser. Prediction of recurrent event in breast cancer using Naïve Bayes Classification - Google Search

[8] Dimitrios Galiatsatos, Georgia Konstantopoulou, George Anastassopoulos, "Classification of the most Significant Psychological Symptoms in Mental Patients with Depression using Bayesian Network", 2015.

https://dl.acm.org/doi/10.1145/2797143.2797159

[9] E. Chandra Blessie , Bindu George, "A Survey on Data Mining Algorithms in Prediction of Psychiatric Disorders", International Journal of Science and Research (IJSR), pp. 783-786.
https://www.ijsr.net/get_count.php?paper_id=SR21817180245

[10] Iheme, Nicholas Omoregbe, Sanjay Misra, Davies Adeloye, Adewumi, "Mobile-Bayesian Diagnosis System for Childhood Infectious Diseases, 2017.
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiLrcKL97T6AhV3TWwGHaS5CiYQFnoECB0QAQ&url=http%3A%2F%2Feprints.covenantuniversity.edu.ng%2F9450%2F1%2FFAIA295-0109.pdf&usg=AOvVaw2CvGMge-IFt3fKfJTMpZYz

[11] Maria Morgan, Carla Blank, Raed Seetan, "Plant disease prediction using classification algorithms", IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 10, No. 1, March 2021, pp. 257-264.
https://www.researchgate.net/publication/349889417_Plant_disease_prediction_using_classification_algorithms

[12] Md Shaoaib Ahmed, Tanjim TaharatAurpa, Md. Abul Kalam Azad, "Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture", Vol 34, Issue 8, Journal of King Saud University- Computer and Information Sciences, 2022, pp. 5170-5182.
https://www.sciencedirect.com/science/article/pii/S1319157821001063

[13] Megha Rathi, Arun Kumar Singh, "Breast Cancer Prediction using Naïve Bayes Classifier", International Journal of Information Technology & Systems, Vol. 1; No. 2: ISSN: 2277-9825 (July-Dec. 2012), pp. 77-80.
https://www.researchgate.net/profile/Megha-Rathi/publication/308934053_Breast_Cancer_Prediction_using_Naive_Bayes_Classifier/links/5bda7f9da6fdcc 3a8db5062f/Breast-Cancer-Prediction-using-Naive-Bayes-Classifier.pdf

[14] Mhambe Priscilla Dooshima, Egejuru Ngozi Chidozie, Balogun Jeremiah Ademola, Olusanya Olayinka Sekoni, Idowu Peter Adebayo, "A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining", International Journal of Immunology 2018;6(1), pp. 5-16.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3561661

[15] Mrs.G.Subbalakshmi (M.Tech), Mr. K. Ramesh, Mr. M. Chinna Rao, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2 No. 2 Apr-May 2011, pp. 170-176.
http://ijcse.com/docs/IJCSE11-02-02-56.pdf

[16] Mubarik Ahmad, Vitri Tundjungsari, Dini Widianti, Peny Amalia, Ummi Azizah Rachmawati, "Diagnostic Decision Support System of Chronic Kidney Disease Using Support Vector Machine", 2017.
https://www.researchgate.net/publication/322945479_Diagnostic_decision_support_system_of_chronic_kidney_disease_using_support_vector_machine

[17] Muhammad Arif, Ashjan Basri, Ghufran Melibari, Taghreed Sindi, Nada Alghamdi, Nada Altalhi, Maryam Arif, "Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review", Vol 4, Issue 1, pp. 95-110, 2020.
https://scholars.direct/Articles/biomedical-research/ibr-4-020.php? Jid = biomedical -research

[18] Nilda N. dela Cruz, Ricardo Q. Camungao," Decision Support System for Predicting Cardiovascular Diseases Using Naïve Bayesian Algorithm", nced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3178 – 3183 3178 Figure 1: Conceptual Framework of the Research ISSN 2278-3091 Volume 9, No.3, May - June 2020 International Journal of Advanced Trends in Computer Science and Engineering, pp. 3178-3183.
https://www.academia.edu/43610885/Decision_Support_System_for_Predicting_Cardiovascular_Diseases_Using_Na%C3%AFve_Bayesian_Algorithm?from_sitemaps=true&version=2

[19] Ojaswi Borate, Snehal Nibe, Sayali Manikwar, Shivani Gund, "Psychological Illnesses Revealing Method Using Machine Learning", 2020 JETIR May 2020, Volume 7, Issue 5, pp. 48-52.
https://www.jetir.org/view?paper=JETIR2005310

[20] Pierre Miasnikof, Vasily Giannakeas, Mireille Gomes, Lukasz Aleksandrowicz, Alexander Y. Shestopaloff, Dewan Alam, Stephen Tollman, Akram Samarikhalaj, Prabhat Jha, " Naïve Bayes classifiers for verbal autopsies: comparison to physician-based classification for 21,000 child deaths, 2015.
https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-015-0521-2

[21] P. Bhasker, P. Kumud, K. R. Pardasani, “Naïve Bayes Classifier for Classification of Plant and Animal miRNA”, International Journal of Computer Theory and Engineering, Vol.2, No.3, June, 2010.
P. Bhasker, P. Kumud, K. R. Pardasani, “Naïve Bayes Classifier for Classification of Plant and Animal miRNA”, International Journal of Computer Theory and Engineering, Vol.2, No.3, June, 2010. - Google Search

[22] P. Kour, S. Shastri, A.S. Bhadwal, S. Kumar, K. Singh, M. Kumari, A. Sharma, V. Mansotra, “Classification of Maternal Healthcare Data using Naïve Bayes”, International Journal of Computer Sciences and Engineering JCSE, Vol.-7, Issue-3, March 2019, pp. 388-394.
https://www.ijcseonline.org/full_paper_view.php?paper_id=3850

[23] Ria Arafiyah, Fariani Hermin, “Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method”, IOP Conf. Series: Journal of Physics: Conf. Series 948 (2018).
Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method - IOPscience

[24] Rosa Devie, Device Rosa Anamisa, Aeri Rachmad, “Classification of Disease for Rice Plant Based on Naïve Bayes Classifier with a Combination of Promethee>, 2021.
https://www.researchgate.net/publication/358019693_CLASSIFICATION_OF_DISEASES_FOR_RICE_PLANT_BASED_ON_NAIVE_BAYES_CLASSIFIER_WITH_A_COMBINATION_OF_PROMETHEE

[25] R Arafiyah, F Hermin, I R Kartika, A Alimuddin, I Saraswati, “Classification of Dengue Haemorrhagic Fever (DHF) using SVM, naive bayes and random forest”, 3rd Annual Applied Science and Engineering Conference (AASEC 2018).https://iopscience.iop.org/article/10.1088/1757-899X/434/1/012070/meta

[26] Sakshi Kapoor, Rabina Verma, Surya Narayan Panda, “Detecting Kidney Disease using Naïve Bayes and Decision Tree in Machine Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019, pp. 498-501.
https://www.ijitee.org/wp-content/uploads/papers/v9i1/A4377119119.pdf

[27] Sanjeev Kumar Yadav, Arvind Kumar Tiwari, “Classification of Enzymes Using Machine Learning Based Approaches: A Review”, Machine Learning

and Applications: An International Journal (MLAIJ) Vol.2, No.3/4, December 2015, pp.30-49.
https://www.researchgate.net/publication/315763502_Automatic_single_and_multi-label_enzymatic_function_prediction_by_machine_learning

[28] Sajana Tiruveedhula, Manda rama narasingarao, “Classification of Imbalanced Malaria Disease Using Naïve Bayesian Algorithm”, International Journal of Engineering and Technology, Vol 7, Issue 2.7, 2018.
https://www.researchgate.net/publication/325117424_Classification_of_Imbalanced_Malaria_Disease_Using_Naive_Bayesian_Algorithm

[29] Sriram Selvaraj, Bommannaraja Kanakara, “Naïve Bayesian Classifier for Acute Lymphocytic Leukemia Detection”, ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 16, September 2015, pp. 6888-6892.
https://www.researchgate.net/publication/288827919_Naive_Bayesian_classifier_for_Acute_Lymphocytic_Leukemia_detection

[30] S Vijayarani, S Deepa, “Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences”, International Journal of Computational Intelligence and Informatics, Vol. 3: No. 4, January - March 2014, pp. 278-283.
https://www.semanticscholar.org/paper/Naive-Bayes-Classification-for-Predicting-Diseases-Vijayarani/b33bb0c66edbfb5c83bcdd10e05d9bcd82d022ea

[31] T.Sajana, M.R.Narasingarao, “Classification of Imbalanced Malaria Disease Using Naïve Bayesian Algorithm”, International Journal of Engineering & Technology, 7 (2.7) (2018) , pp. 786-790.
https://www.sciencepubco/index.php/ijet/article/view/10978

[32] Vishal Deshwal, Mukta Sharma, “Breast Cancer Detection using SVM Classifier with Grid Search Technique”, International Journal of Computer Applications (0975 – 8887) Volume 178, No. 31, July 2019, pp.18-23.
https://www.researchgate.net/publication/334522256_Breast_Cancer_Detection_using_SVM_Classifier_with_Grid_Search_Technique

[33] Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi, “Early Detection of Breast Cancer Using SVM Classifier Technique”, International Journal on Computer Science and Engineering Vol.1(3), 2009, pp. 127-130.

[Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi, “Early Detection of Breast Cancer Using SVM Classifier Technique”, International Journal on Computer Science and Engineering Vol.1(3), 2009, pp. 127-130. - Google Search]