CORRELATION MODEL IN THE ADOPTION OF E-PAYMENT SERVICES: A MACHINE LEARNING APPROACH

By

Tan Xi En

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2022

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: ______CORRELATION MODEL IN THE ADOPTION OF__________

_______E-PAYMENT SERVICES : A MACHINE LEARNING_________

________APPROACH___________________________ __________

Academic Session: ___JAN 2022____

I ____________________TAN XI EN__________________

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

_________________________ ______________________

(Author's signature) (Supervisor's signature)

Address:

____No.12, Jalan Tasik Damai 6______

____Taman Tasik Damai, 57100____

_____Sungai Besi, Kuala Lumpur______

____ TONG DONG LING______

Supervisor's name

Date: _____22/4/2022_____

Date: _____22/4/2022_____

FACULTY OF INFORMATION, SCIENCE AND TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ___22/4/2022_____

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that ______***Tan Xi En*** ____________________ (ID No: __***19ACB04098***__) has completed this final year project entitled "__**CORRELATION MODEL IN THE ADOPTION OF E-PAYMENT SERVICES: A MACHINE LEARNING APPROACH** ____ _" under the supervision of ________Dr. Tong Dong Ling_________ (Supervisor) from the Department of _____Computer Science________, Faculty of ____Information, Science and Technology______ , and __ (Co-Supervisor)* from the Department of ________, Faculty of __________

I understand that University will upload softcopy of my final year project in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

DECLARATION OF ORIGINALITY

I declare that this report entitled “**CORRELATION MODEL IN THE ADOPTION OF E-PAYMENT SERVICES: A MACHINE LEARNING APPROACH**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ______________________________

Name : ____Tan Xi En___________

Date : _____22/4/2022___________

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Dr Tong Dong Ling who has given me this bright opportunity to engage in a Machine Learning Project. It is my first step to establish a career in Data Analytics Field. A million thanks to you.

To a very special person in my life, Fong Wai Fung, for her patience, unconditional support, and love, and for standing by my side during hard times. Finally, I must say thanks to my parents and my family for their love, support, and continuous encouragement throughout the course.

ABSTRACT

The purpose of this study was to understand the highly correlated factors that influence user intention to adopt e-payment into their daily lives. The research framework incorporates the 4 main constructs of the UTAUT model and additional determinants of the extended UTAUT model to model the relationship. Previous research has shown that the UTAUT model is able to model the relationship of factors that influence user intention to adopt e-payment. To analyse results, a questionnaire was developed based on the UTAUT model, and was distributed among university graduates, researchers, and students. In total, 286 samples were recorded and analysed. Previously, identifying highly correlated user behaviours was done through a lot of statistical research and hypothesis testing. The main goal of the project is to automate this process, by using machine learning to identify the important features. To extract highly correlated features, several pairwise correlation methods were used, such as Pearson's Correlation, and Spearman's Correlation. Then, by using Correlation Based Feature selection algorithm, we select the best subset of features out of the highly correlated features to do predictive modelling. This is a novel method, as we do not need to rely on statistical analysis, rather we can automate the process of identifying important features using machine learning models. The end goal of the project is to develop a model that identifies the important features that affect user intention to adopt e-payment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

UTAUT	Unified Theory of Acceptance and Use of Technology
PE	Performance Expectancy
EE	Effort Expectancy
SI	Social Influence
FC	Facilitating Conditions
AT	Attitude Towards Using Technology
SE	Self-Efficacy
AX	Anxiety
T	Trust
E-Payment	Electronic Payments
QR	Quick Response
CFS	Correlation Based Feature Selection
ICT	Information and Communication Technology

1 Chapter 1: Introduction

This section will be discussing about problem statements, motivation, objectives and project scope of this project. We will include the report organization at the end section as well.

1.1 Problem Statement and Motivation

Lack of E-payment users in Malaysia

In the age of a cashless society, Mobile Technology and Financial Services have taken the next great step in terms of technology, giving birth to a new alternative of e-commerce transactions, that is e-payment services. However, as e-payment services are still relatively novel, there is still a huge majority of Malaysians who are unwilling to open to the idea of adopting e-payment into their daily lives. A survey conducted by oppotus.com states that only 16% of the older age generation (age 45 or above) are e-wallet users [1]. Therefore, most of the older age generation in Malaysia are unwilling to make the shift to e-payments.

Therefore, there still exists a lack of understanding of what features, facilities, influence an e-payment should have to boost the usage intention of e-payment. This project will aim to provide insights on how to encourage new users to adopt e-payment systems.

Lack of understanding of E-payment Features

Stakeholders such as merchants, NFC device owners and mobile application developers have a lack of understanding of important features. When it comes to building e-payment apps, although the functionality of e-payment apps is the same, some apps gain more popularity compared to the others. This leaves a lot of the stakeholders confused as they are uncertain what features allude to new users to adopt the e-payment services. Furthermore, due to a lack of User, there are a lack of e-payment facilities in Malaysia. Combined with the lack of understanding by stakeholders, there seem to be a disconnect between e-payment developers and e-payment users. Therefore, to resolve this, this project aims to visualize the relationship between the features and user intention to adopt e-payment.

Complicated Research Procedure for UTAUT Framework

When it comes to adoption of digital products, there already exist many related works that tries to understand what features influence new users to adopt digital products. E-payment is not an exception. There already exist many previous works that explains how behavioural intention to adopt digital products works. However, these researchers are incredibly tedious. To conduct such research, researchers have to put in a lot of effort into researching a theoretical framework, e.g., UTAUT model and have a deep understanding of features needed. To get the results required, they would have to compute the results by hand or by another software such as PLS-SEM and analyse it. Normally, researchers would use many statistical methods, such as Cronbach's Alpha, average variance, fit of goodness to understand the results. This project aims to solve the issue by using machine learning algorithms, Decision tree, to do the analysis instead. This reduces the tedious statistical analysis methods and may help to visualize new relationships between the factors.

1.2 Project Objectives

- To identify set of significant user behaviors that help influence the behavioral intention to use E-payment
- To research the underlying relationship between the correlated features and its effect towards the acceptance of e-payment systems.
- To develop a generalized machine learning model that is applicable for any theoretical framework.

1.3 Project Scope

In this project, a questionnaire based on User Intention to adopt e-payment features is created and distributed among university students. The survey data was collected over a period of 12 weeks with a total of 286 respondents. The data set contains UTAUT factors, user history and intention with e-payment and moderating factors. The goal of the project is to conduct analysis on these factors using machine learning approach and correlation analysis to gain a deeper understanding of the features.

The project will aim to create a new correlation-based algorithm, where based on correlation algorithm used, produce three types of features, intersected features, non-intersected features and union features. These features will then be used on our decision tree analysis and correlation network graph analysis. From the analysis, we will then understand the important attributes that help influence user intention to adopt e-payment.

The project also aims to produce a generalized solution for adoption of digital products for any technological framework. The solution will be a machine learning model, that can select important attributes and provide good accuracy score with deep analysis. We will evaluate the model using precision, recall, f1-score and Matthew correlation coefficient as our evaluation metric. The model will be train and tested using 10-fold cross validation.

1.4 Impact and Contributions

Firstly, this project helps stakeholders gain an understanding on important determinants that affect user to adopt e-payment services. This research will be of great interest to stakeholders such as merchants, NFC device owners and mobile application developers as it helps determine what determinants affect customer intention to adopt e-payment. With the results of the project, there will be huge improvement for the e-payment features as mobile application developers understand the determinants that allure user to adopt e-payment.

Moreover, the increase in the number of users will also help to improve technological infrastructure. More e-payment machines will be installed at every retail shop and become widely available. This will help to improve the payment process.

Furthermore, this project helps to model the relationship between the determinants of the UTAUT factor. Historically, UTAUT factors are always computed using complex statistical methods such as multiple linear regression, Cronbach analysis. This requires a lot of research for researchers or developers who are unfamiliar with the technological model. The project provides an automated and less subjective method for analyzing human user behaviors. We no longer need to use complicated statistical methods to compute highly correlated user behaviors. Rather, we can discover important subset of features by using machine learning model to identify important features.

1.5 Report Organization

In Chapter 1, we introduce the problem statement faced and motivations that influenced us to take up this project. The project objective and scope are also listed in Chapter 1.

We also discuss about the background information of e-payment system in Malaysia and UTAUT model. Lastly, we also talk about the impact and contribution of the proposed method.

In Chapter 2, We do a lot of literature review on UTAUT model. We investigate the factors that influence customer to adopt e-payment into their daily lives and come up with hypothesis. We then summarize other literature review of UTAUT factors that influence customer to adopt e-payment system and use it as reference for our project. We also do a performance review on Tree-Based algorithm that we plan to use for our experimentation.

In Chapter 3, We discuss about the system methodology, how we carry out our project, how we perform data collection and data analysis. We discuss about our machine learning pipeline and the steps we carried out throughout the entire project. We also show our Gantt Chart timeline for how we distribute our project task.

In Chapter 4, We talk more detail about our correlation-based feature selection. We discuss a variety of feature selection technique, that is, CFS, Forward Selection and Backward selection technique. We show the flowchart of our algorithm. We also show a new way of analysing relationship between features using Correlation Network Graph.

In Chapter 5, we show how we set up the necessary software and hardware to perform our experimentation. We show each step by step and our coding line by line. We also discuss about expected output after we execute the python code.

In chapter 6, we discuss about our testing setup and experimentation results. We do decision tree pathway analysis to analyse how decision tree select UTAUT factor items and come up with hypothesis for why these features were selected. We also do a performance comparison of each model algorithm. Lastly, we do another performance comparison of features selected using Correlation Based Feature selection algorithm.

In chapter 7, we discuss about the conclusion of our project. We outline possible future work for our project and recommendations to improve the workflow of our project. We also discuss about the novelties and the contribution of the project. We discuss in detail about the significance and impact the project helps other researchers.

2 Chapter 2: Literature Review

2.1 Introduction

The increased use of e-payments has prompted the question of what factors influence user intention to use the e-payment procedure. Many methodologies have been employed and investigated, in which researchers use a theoretical framework to try to find factors that influence new users' intention to adopt e-payment services. We will look at a few similar studies to see what factors influence people's willingness to use electronic payments. From the similar studies, we aim to incorporate the factors selected in existing work to help evaluate User intention to adopt e-payment services in their daily lives.

2.2 UTAUT Model Framework

The motivation behind UTAUT Model

There has been much emphasis place on research towards User's acceptance of technology since business organizations have started to integrate information technology. One of the main motivations behind the development of the UTAUT model is to understand what factors affects user's beliefs, their understanding, and their resistance towards new technology acceptance. Over the past two decades, there have been many studies that are dedicated towards the subject of digital adoption. These studies have been conducted using numerous theoretical frameworks. Other than UTAUT model, there are numerous theoretical framework that we can reference, such as Theory of Planned Behaviour (TPB), Social Cognitive Theory (SCT) and Innovation Diffusion Theory (IDT).

Below is a diagram of UTAUT model:

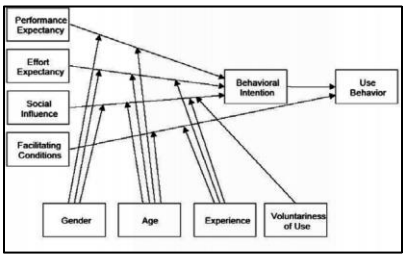


Figure 2.2.1 Original UTAUT Model

Limitations Behind Other Model

Although these theoretical models have contributed significant impact towards the studies of technology adoption, there are still some limitations. Firstly, there is no unified terminology for the models. Each theoretical models have their own unique terms, which may confuse some other researchers who are not familiar. Secondly, behavioural research or social science is very complex and therefore a single theoretical model is not reasonably able to cover all plausible factors. Each models have their own strength and weakness; however, they do not complement each other. In 2003, Venkatesh et al reviewed the existing models and created a unified model we now know as UTAUT model [2].

2.3 UTAUT Model Attributes

The 4 basic Determinants of UTAUT

The four determinants of UTAUT are performance expectancy, effort expectancy, Social Influence and Facilitating Conditions. These 4 determinants are essential in determining user acceptance and user behaviour. The key independent variables on behaviour intention and uses of information technology are normally influenced by age, gender, education and voluntariness.

Performance Expectancy

Performance expectancy (PE) is defined as “the degree to which an individual believes that using the system will help him or her attain gains in job performance”. Performance expectancy combines constructs from five different models, which are, TAM / TAM2, C-TAM-TPB, MM, MPCU, IDT and SCT [2].In general, performance expectancy is the belief that adopting new technology will help to improve work efficiency. The new technology provides extrinsic motivation and adds value to user work performance when it is adopted. User will feel new sense of accomplishment after adoption of the new technology.

In previous studies, it is observed that Performance Expectancy has positive relationship towards behavioural intention when it comes to digital adoption of electronic payments. Therefore, the study proposes the following hypotheses:

H1: Performance Expectancy (PE) has a positive effect on behavioural Intention.

Effort Expectancy

Effort Expectancy (EE) is defined as "the degree of ease associated with the use of the system" [2]. Three constructs from three existing models explains the concept of effort expectancy, namely, perceived ease of use from TAM/TAM2, complexity from MPCU, and ease of use from IDT. According to previous studies, it is also observed that effort expectancy has a positive relationship towards behavioural intention when it comes to the adoption of electronic payment system.

H2: Effort expectancy does not have a significant relationship towards behavioural intention.

Social Influence

Social Influence is defined as "the degree to which an individual perceives that important individual believes he or she should use the new system" [2]. Important individuals here, for example, are usually referred to as colleagues, peers or celebrities. Numerous studies have highlighted the significance of social influence, which refers to effects bought about by young celebrities and role models. Therefore, user believes that adopting new technology will improve his's social image and social relation. User also gauge social factors to adopt new technology, that is based on the culture and social norms of the new technology adoption.

H3: Social influence has a positive relationship on behavioural intention.

Facilitating Conditions

Facilitating conditions are defined as "the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system" [2]. Facilitating conditions refers to perceived behavioural control and compatibility to technology. For example, NFC tag scan readers. The significance of facilities conditions is that they are relevant to improving the value and experience of users when adopting new technology.

According to previous studies, Facilitating Conditions does not have a significant relationship on behavioural intention. In fact, previous studies show that facilitating conditions do not influence behavioural intention to adopt digital payment. We can then formulate such hypotheses:

H4: Facilitating Conditions does not influence behavioural intention.

Extended Determinants of UTAUT Model

Among the empirical studies that were conducted based on the UTAUT framework, they assumed that positive behavioural are essential in determining actual usage behaviour. Under this framework, the statement above is acknowledged, and behavioural intention does directly affect usage behaviour. Behavioural intention, is also, indirectly affected by the four main determinants of the UTAUT behaviour. This research proposes an extended UTAUT model that combines 4 more extra determinants that have an impact on behavioural intention. They are attitude towards technology, self-efficacy, Trust, and anxiety.

Attitude Towards Technology

Attitude Towards Using Technology (AT) are defined as "the degree of an individual's overall affective reaction to using a system". Overall affective reaction refers to an individual's liking, enjoyment, joy and pleasure associated with technology use. Previous studies suggest that there is a positive relationship between attitude towards using technology and behavioural intention. For example, taking an example based on online shopping, in the context of mobile banking adoption, non-cash system usage, smartphone adoption, etc., online or internet shopping behaviour has been found to be influenced by the individual attitudes of the shopping system. This means that, online shopping behaviour supports the research that was carried out based on the usage of e-payments and identical topics. According to Ambigai Rajendran, a customer is more willing to adopt new technology if he has a positive attitude towards the e-payment services [3]. Thus, we can propose the following hypotheses:

H5: Attitude towards technology have a positive relationship on behavioural intention.

Self-Efficacy

According to the SCT proposed by Bandura, self-efficacy refers to the evaluation of people for their efficiency and ability to perform a particular task well [4]. It is not about individual skills, but about how individuals utilize these technologies. In this case, self-efficacy is the trust of individuals who possess the aptitude and skills to be successful when performing m-technology-related tasks. Mobile self-efficacy has been defined by

Nikou and Economides as recognition of an individual ability to utilize mobile device to work (e.g. Internet Search) [5]. We can then infer such hypotheses:

H6: Self-efficacy has a positive relationship on behavioural intention.

Anxiety

Computer anxiety refers to the fear of acceptance of new technological system. Research has shown that perceived anxiety is negatively related to Perceived Ease of Use Furthermore, previous research revealed a negative correlation between test anxiety and behavioural intention. According to Kathryn, it is noted that anxiety has a strong negative effect on behavioural intention to adopt digital adoption [6].The p-value is below 0.05, therefore there is a significant relationship with behavioural intention. Thus, we can propose the following hypotheses:

H7: Anxiety has a negative relationship on behavioural intention.

Trust

Trust is defined as willingness to obey other party's action as long as its slightly beneficial for them. Trust becomes one of the factors for new users to adopt new technology system. Trust plays a central role in a situation where a person is dependent on another party under a risky and uncomfortable situation. Trust provides a subjective guarantee that consumers obtain a positive experience about the ability, honesty, and goodwill of e-payment service providers [7].Gefen identifies trust and security as major factors in the acceptance of e-commerce technology. If there are safety mechanisms in the technology system, it will increase the factor of trust. Trust and transparency are the key factors in transforming how business transactions will be conducted in the future.

We can then formulate such hypotheses:

H8: Trust has a positive relationship on behavioural intention.

Behavioural Intention

The definition of Behavioural Intention is defined as “the degree to which e-payment service user’s motivations intend to accept and use the system”. This is also the goal of the model. Venkatesh et al assumes that BI will have a significant positive influence on technology usage [2]. Thus, this research proposes the following hypothesis:

H9: Behavioural intention will have a significant positive influence on adoption of e-payment systems.

2.4 Summary of UTAUT Determinants

The following is the definition of the 8 determinants of UTAUT Factors.

Table 2.4-1 Summary of UTAUT Determinants

Component	Description
Performance Expectancy	The degree to which users believe that using e-payment service will improve their efficiency of work or life
Effort Expectancy	The degree to which consumers perceive e-payment service as easy to understand and use.
Social Influence	Users' perceived support level by those who are important to them or those who are influential in their choice of e-payment service
Facilitating Conditions	The degree of support of technical infrastructures and resources required towards using e-payment service apps.
Attitude Towards Using Technology	The degree of an individual overall affective reaction to using e-payment system.
Self-Efficacy	What user believe himself or herself is capable of using e-payment system
Anxiety	The unwillingness of user to use e-payment system
Trust	The willingness of a user to obey the other party's action.
Behaviour Intention	The degree to user intends to accept and use the e-payment system

2.5 Review and Analysis of Existing Work

Integrating Cognitive Antecedents to UTAUT Model to Explain Adoption of Blockchain Technology Among Malaysian SMEs

Figure 2.5.1 Proposed UTAUT Model for Adoption Among Malaysian SMEs

The above paper, written by Hamad Khazaeri, proposes a new conceptual framework for adoption of blockchain technology among SMEs in Malaysia using Unified Theory of Acceptance and Use of Technology as theoretical framework [8]. Using Venkatesh UTAUT model as the base model, the paper considers the following constructs that is relevant in predicting acceptance behaviour towards blockchain adoption. There are altogether 8 determinants proposed by the paper, that is Behavioural Intention, Personal Innovativeness, Perceived Trust, Perceived Security, Performance Expectancy, Effort Expectancy, Social Influence and Technology Awareness. The author believes that these 8 determinants are essential in predicting User behaviour towards adopting blockchain.

The author uses a variety of method to test the validity of the data. He first performs Cronbach-alpha to perform a reliability measure, to ensure that all questions are relevant with the topic. Then, he performs a validity measure. He then uses Confirmatory Factor Analysis (CHA) to perform a goodness-of-fit test. Goodness-of-fit test is a statistical hypothesis test to see how well sample data fit a distribution from a population with a normal distribution.

Overall, the results of the data analysis was that perceived innovativeness (PI) was the significant predictor towards adopting blockchain technology. Perceived trust and perceived security was shown to have significant influence in adopting blockchain technology. Performance Expectancy, Effort Expectancy and Social Influence had only a mild positive impact in respondents adopting blockchain technology. The hypothesis of Technology awareness, however, is rejected, and therefore does not influence user in adopting blockchain Technology.

E-Money Payment: Customers' Adopting Factors and the Implication for Open Innovation

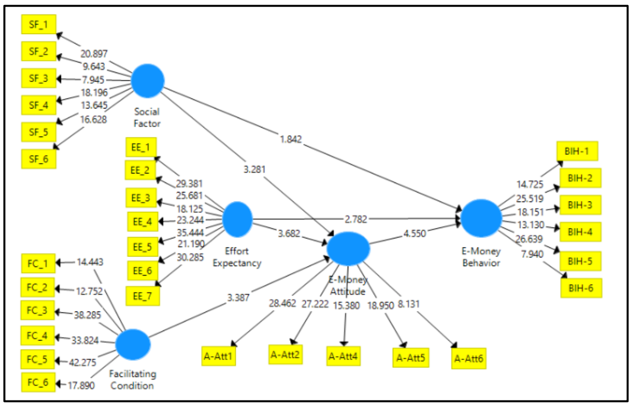


Figure 2.5.2 Proposed UTAUT Model for Adoption of E-Money Payment

The above paper, written by Widayat, Ilyas and Novita in 2020, was about investigating the adoption of e-money payment among Indonesian Youth Generation. The study was conducted and distributed around University of Muhammadiyah Malang Campus [9]. Each factor was measured by valid indicators. For this experiment, the factors used to determined factors that influence the adoption of e-money payment model are facilitating conditions, effort expectancy, social factors, e-money attitude and e-money intention behaviour.

The authors used AVE, CR, Rho-A, and Cronbach's alpha for latent variables (e-money attitude, e-money behaviour), effort expectancy, facilitating conditions, and social factors to test the validity of the data. The constructive validity of all latent variables was discovered to be constructively valid. The composite reliability test also revealed that the values were greater than 0.7.

The Heterotrait-Monotrait Ratio was used to assess discriminant validity (HTMT criteria). The result of the data analysis shows that the intention to adopt e-payment are significantly influenced by attitudes towards e-money. A positive attitude towards e-money infers that user are more likely to adopt e-payment. Besides being influenced by attitude, the behavioural intention to adopt e-payment is also influenced by effort expectancy and social factors. This indicates that clients will continue to use e-payment, making this their first choice of payment as a result of external influences such as stores they tour and close friends or family as socioeconomic aspects.

Young Generation's Mobile Payment Adoption Behavior: Analysis Based on an Extended UTAUT Model

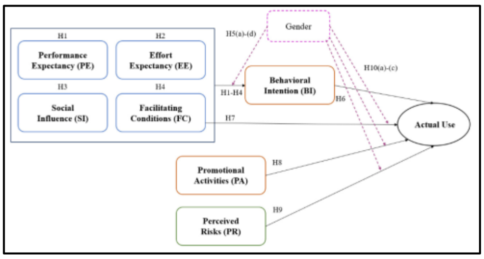


Figure 2.5.3 Proposed UTAUT Model for Taiwanese Youth

The above model, proposed by Min-Fang Wei, Yir-Hueih Luh, Yu-Hsin Huang and Yun-Chi Chang aims to understand Taiwanese young generation's mobile payment adoption behaviour analysis based on an extended UTAUT model [10]. For their research, they used the basic determinants of UTAUT factors, that is, Performance expectancy, Effort expectancy, Social Influence and Facilitating conditions. They used gender as moderating factors. As they are using an extended UTAUT model, they have also added 2 new factors, that is promotional activities and perceived risks.

To test the validity of their data, the researchers used PLS-SEM with bootstrap approach to validate the proposed model. The measurement model was initially analysed by measuring the validity and reliability of construct measures. They use Cronbach's alpha and average variance extracted as a form of measurement metric. In this study, all latent constructs are greater than -.7, satisfying the rule of thumb for validity. In their discriminant validity test, the highest value of inter-correlation estimates was less than 0.557, which is less than the maximum cut-off of 0.85. After the assessment of model reliability and validity, the researchers continue to evaluate the hypothesized relationships of the inner model using bootstrap-t-test to validate its significance.

From their analysis, they have discovered among the 4 UTAUT constructs only social influence construct has a significantly positive effect on young generation's behavioural intention to adopt mobile payment with a p-value of below 0.05. The moderating role of gender on behavioural intention indicates a strong social influence impact on men's intention to adopt mobile payment service. Effort expectancy, performance expectancy and facilitating conditions have little to none influence as their p-value were above 0.05, indicating the null hypotheses was accepted.

Determinants Of Mobile Payment Usage And The Moderating Effect Of Gender: Extending The UTAUT Model With Privacy Risk

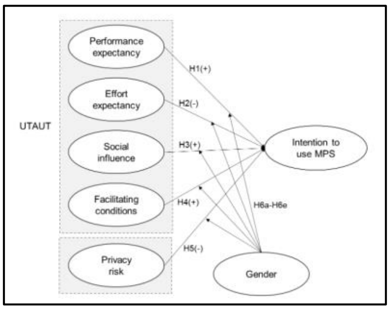


Figure 2.5.4 Proposed UTAUT Model for Adoption of MPS

The above research was conducted by Jin-Myong Lee, Bohan Lee and Jong-Youn Rha, where they investigate mobile payment influence adoption among south korea's youth generation using extended UTAUT model [11]. The motivation behind their research was that they observed that South Korea's MPS adoption rate is quite lacking when compared to United states. The frequency of using MPS is just two times a month on average for each south Koreans. To conduct their research, they proposed 4 basic determinants of UTAUT, that is Performance expectancy, Effort Expectancy, Social Influence, Facilitating Conditions. Privacy Risk was an extended determinant and Gender is used as moderating factor.

To test the validity of their results, they used Cronbach's alpha to test the internal consistency of each construct. They then employed using Confirmatory factor analysis to do factor loading around UTAUT constructs. Lastly, they do discriminant validity to analyse the overall correlation between the constructs and squared root of Average variance. To test their hypothesis, the researchers did Structure Model Testing using AMOS 20.0 software. They recorded the results and test if their hypothesis was accepted.

The summary of their analysis was that the significant factor that affects South Korean user to adopt mobile payment service are Performance Expectancy, Social Influence and Perceived Risk. Effort Expectancy and Facilitating Conditions were not supported because they had a p-value of above 0.05. Facilitating conditions was shown to have significant positive effect on Male Gendered Users, whereas privacy risk was a major concern for female users.

2.6 Comparison of Existing Works

Table 2.6-1 Comparison of Existing Works

Literary Works	Determinants Used	Significant Determinants	Insignificant Determinants
Determinants Of Mobile Payment Usage And The Moderating Effect Of Gender: Extending The UTAUT Model With Privacy Risk	• Performance expectancy • Effort expectancy • Social Influence • Facilitating Conditions • Privacy Risk	• Performance Expectancy • Social Influence • Effort Expectancy • Facilitating Conditions	• Privacy Risk
Integrating Cognitive Antecedents to UTAUT Model to Explain Adoption of Blockchain Technology Among Malaysian SMEs	• Personal Innovativeness • Perceived Trust • Perceived Security • Performance Expectancy • Effort Expectancy • Social Influence • Technology Awareness	• Trust • Security • Effort Expectancy • Performance Expectancy • Social influence • Personal Innovativeness	• Technology Awareness
Young Generation's Mobile Payment Adoption Behavior: Analysis Based on an Extended UTAUT Model	• Performance expectancy • Effort expectancy • Social Influence • Facilitating Conditions • Perceived Risk • Promotional Activities	• Social Influence • Perceived Risk	• Facilitating Conditions • Performance Expectancy • Effort Expectancy
E-Money Payment: Customers' Adopting Factors and the Implication for Open Innovation	• Facilitating Conditions • Effort Expectancy • Social Factors • E-money attitude	• Effort Expectancy • Social Factors • E-money attitude	• Facilitating Conditions

2.7 Review of Predictive Model Algorithms

There were many algorithms there were considered when building our machine learning pipeline. However, they remained an issue of multicollinearity between the factors due to the nature of our project. Therefore, we can only consider tree-based algorithm. The chosen algorithm for this project are Decision Tree algorithm, Random Forest algorithm and XGBoost algorithm.

Decision Tree

Decision Trees are constructed from a given set of attributes. As there are many possible trees to be built from a given set of attributes, finding the optimal tree is difficult because of the exponential size of the search space. Therefore, efficient algorithms have been developed to help build an optimal decision tree in a reasonable amount of time [12].

The measures for selecting best splits are normally based on the features that will produce the most homogenous resulting datasets. The most common method is to minimize entropy and maximize information gain [13].

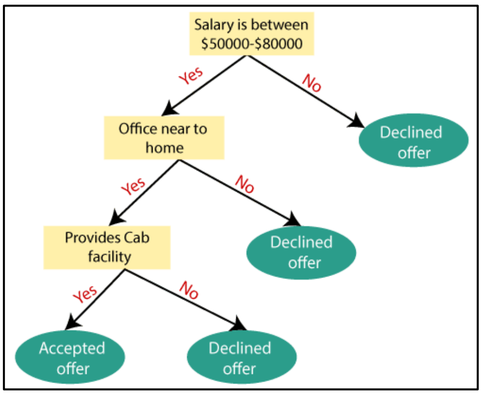


Figure 2.7.1 Example of Decision Tree

Entropy is a measure of randomness within a dataset. It measures the homogeneity of a node.

The formula of entropy is as follows:

$$Entropy(t) = -\sum p(j|t) \log p(j|$$

Figure 2.7.2 Formula of Decision Tree Entropy

Information gain is a gain measure that is used in the ID3 and C4.5 algorithms. The information gain is defined as the reduction in entropy from splitting on a given feature. Only splits with maximized reduction are selected.

The formula of information gain is as follow:

$$Entropy(p) - \left(\sum (ni/n)Entropy(i)\right)$$

Figure 2.7.3 Formula of Decision Tree Information Gain

Random Forest

Random Forest Model is an improvement from decision tree model. A single decision tree model will not be able to make accurate predictions on its own. This may be because the single decision tree is not the most optimal tree for the features. Bagging is a technique that builds many decisions tree models by randomly sampling from the original dataset. This ensures variety in decision trees, which also helps to prevent overfitting. Each node is allowed to randomly split based on the selection of the model's feature. As a result, we can build many trees and combine the predictions to improve the machine learning model's predictive power [13].

The figure Below shows an example of random forest model.

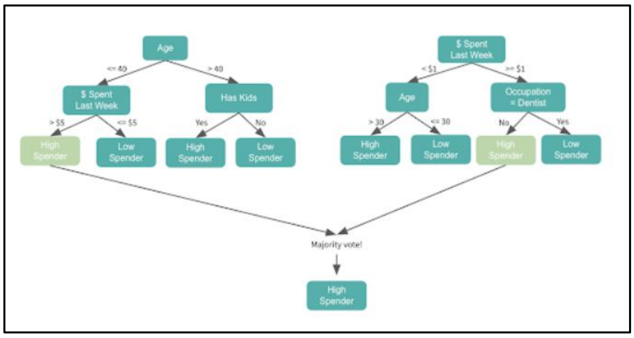


Figure 2.7.4 Diagram of Random Forest Model

As the number of trees in a forest grows large, the generalization error converges to a limit. The generalization error of a forest of tree classifiers is determined by the strength of the individual trees in the forest as well as their correlation.

XGBoost

XGBoost is a scalable machine learning system for tree boosting [14]. Boosting is a general method for improving the accuracy of any given learning algorithm Boosting is an ensemble tree method that builds consecutive small trees. Each tree focused on correcting the net error from the previous tree. The final forecast is a weighted average of all individual forecasts. Gradient boosting is the most well-known application of boosting. It optimises using the gradient descent algorithm.

The most important learning step for Tree based algorithm is the determination of split. XGBoost uses exact greedy algorithm for splitting node. To split data efficiently, XGBoost arranges data according to feature values and visits data in sorted order to accumulate gradient statistics for structure score.

The formula for rank function of XGBoost is provided below:

$$\boxed{r_k(z) = \frac{1}{\sum_{(x,h)\in\mathcal{D}_k} h} \sum_{(x,h)\in\mathcal{D}_k, x<z} h,}$$

Figure 2.7.5 Rank Function of XGBoost

The approximation factor algorithm for XGBoost is provided below:

$$\boxed{\sum_{i=1}^{n} \frac{1}{2} h_i (f_t(\mathbf{x}_i) - g_i/h_i)^2 + \Omega(f_t) + constant}$$

Figure 2.7.6 Approximation Factor Algorithm

3 Chapter 3: System Methodology

3.1 System Block Diagram

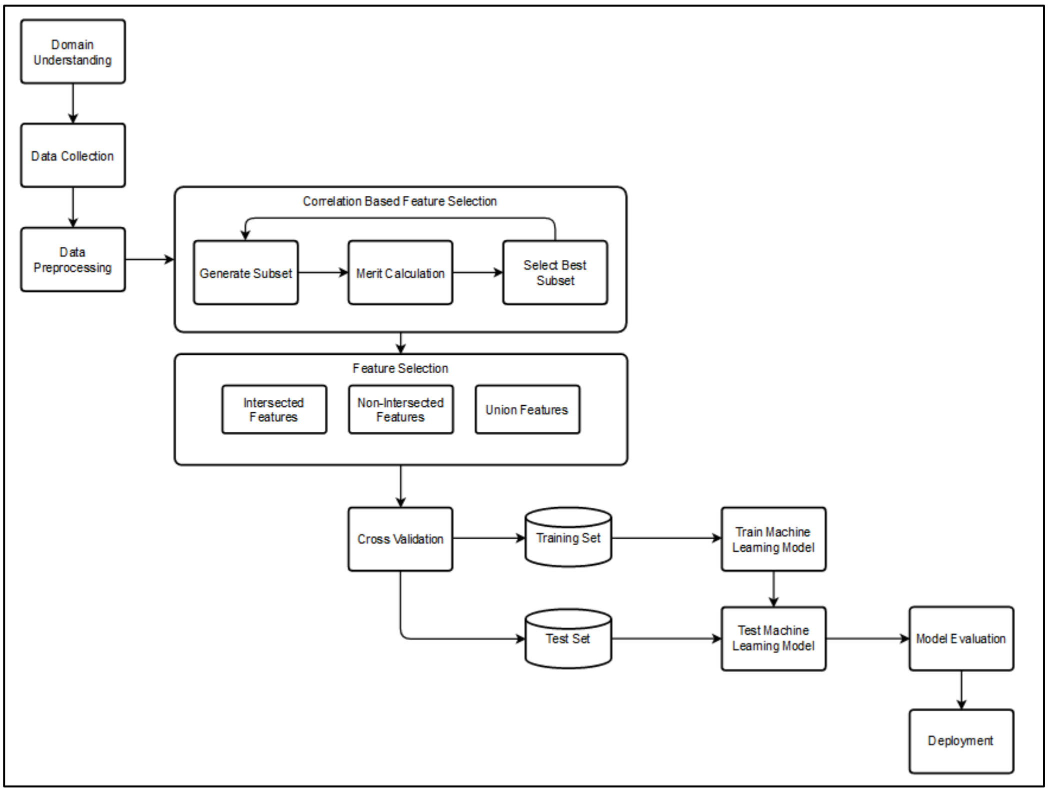


Figure 3.1.1 System Block Diagram

The above block diagram is the proposed machine learning pipeline designed as our project solution. The steps are as followed.

Domain Understanding

Before we begin any machine learning problem, we must first understand the nature of our problem. Our project objective is to understand features that affect customer intention to adopt an e-payment system. To achieve better understanding, we have done literature review on previous related work to identify features selected by related works. Due to the nature of using correlation analysis, the feature selected will be highly correlated. To circumvent this issue, we use tree-based algorithm as they can avoid multicollinearity issue. The machine learning algorithms used will be Decision Tree algorithm, Random Forest algorithm and XGBoost algorithm. Random Forest is a bagging method, whereas XGBoost is a boosting method, extended from Decision Tree algorithm.

Data Collection

A google form questionnaire was designed to be used as the data collection instrument for this study. It was designed based on the four main constructs of the UTAUT model defined by Venkatesh et al, and additional constructs proposed in this study. The questionnaire was distributed mainly at UTAR university, at both graduate and undergraduate levels. It was distributed mainly via email or word-or-mouth. Majority of the respondents are involved in the education field, either as students or researchers. The data was collected from May 2021 to August 2021.

The questionnaire had a total of 11 sections. Section 1 of the questionnaire was about socio-demographic questions that were developed for the intention of understanding the social status of the respondents. Section 2 contains questions that provide insights on the familiarity that respondents have towards blockchain, electronic payments and cryptocurrency. They were all yes / no questions and respondents can select more than one item. Section 3 – 11 consisted of five-point Likert scales statements ranging from strongly agree to strongly disagree. They were concerning the UTAUT factors, which were PE, EE, SI, FC, AT, SE, AX, T and BI. A total of 286 questionnaires were collected. The table below presents the descriptive statistics of the respondents.

Table 3.1-1 Descriptive Statistics of Respondents

Description	Frequency
Age	< 25 Years: 83 (29.02%)
	26 – 40 Years: 95 (33.22%)
	41 – 55 Years: 86 (30.07%)
	Above 55 years: 22 (7.69%)
Gender	Male: 170 (59.44%)
	Female: 116 (40.56%)
Education Level	College / University: (66.79%)
	Graduate School: 46 (16.43%)
	Primary School: 6 (2.14%)
	Secondary / High School: 41 (14.64%)
Work Industry	Banking / Finance: 34 (22.67%)
	Education: 62 (41.33%)
	Healthcare: 6 (4.0%)
	Manufacturing: 31 (20.67%)
	Retail / Hypermarket: 17 (11.33%)

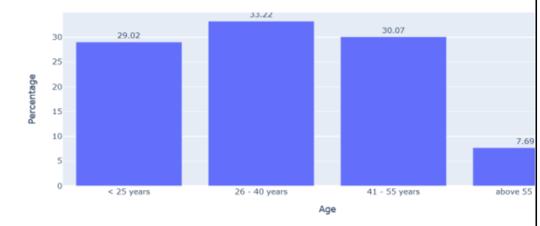
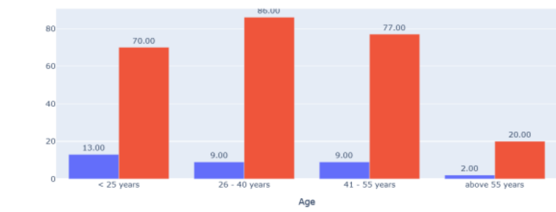
Data Understanding

To understand the Data collected, we have categorized our features into 3 groups of factors, mainly, targeted factors, moderated factors and UTAUT Factors. Targeted factors refer to the features we wish to study as our target group. For example, age, gender, education level is our targeted features, as we wish to understand how UTAUT features affect this targeted groups. The definition of moderated factors refers to factors that occur when relationship between two variables depends on another factor. In this project example, it would refer to “Have you made any electronic payments in the past 12 months?” where we want to understand whether have the targeted user made any electronic payments in the last 3 months. Lastly, UTAUT features refers to the UTAUT item constructs used in UTAUT Framework proposed by Venkatesh et al. These features include PE1, PE2, AT1, AT2 and many more.

To gain a better understanding of our attributes, we have plotted several bar graphs to help visualize the relationship between our targeted features and moderated features. We construct these bar chars using plotly, a python library, to plot these visualizations. These bar charts show the distribution of User group based on our targeted variable. Some interesting pattern may be observed from the distribution of the user group. The findings for the graph are discussed below.

Age

Table 3.1-2 Age Graph

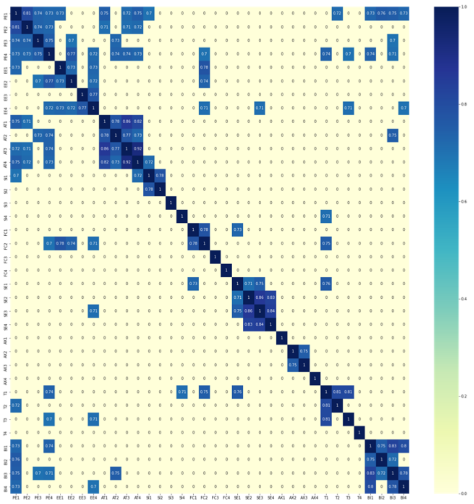
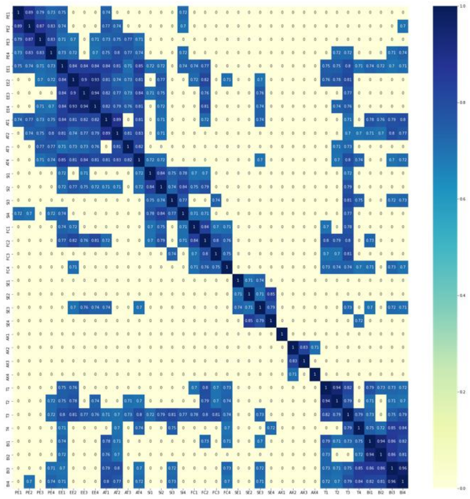
<table>
<tr><th>Distribution of Age Group</th><th>Comparison with Targeted Variable</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

The preliminary analysis done for the “Age” Variable indicates a fairly even distribution between the age range of “< 25 years”, “26 – 40 years”, “41 – 55 years” and “above 55 years”. This is primarily because the survey were targeted for university students and adults who have working experience. The descriptive statistics indicates that 29% of respondents were below 25 years old, 33.22% were around 26 – 40 years and 30.07% were around 41- 55 years. Based on the percentage result, it is safe to assume that there won’t be any bias based on age group, as the percentage for each age group is roughly the same and is evenly distributed.

When comparing age groups with the targeted variable of “Have you made any electronic payments in the past 12 months?”, the graph indicates that majority of each respondent in each age group are highly likely to adopt e-payment.

Is there any difference between respondent group “< 25 years” and “41 – 55 years”?

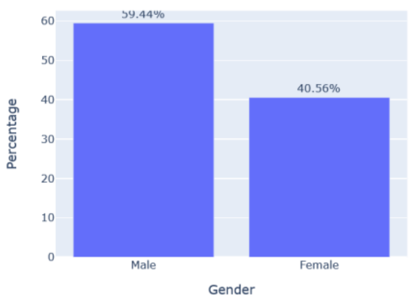
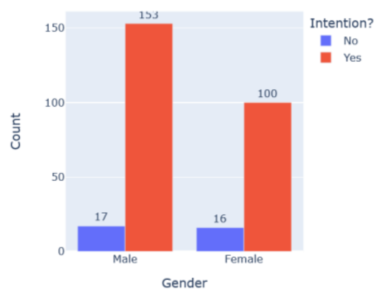
Table 3.1-3 Age Correlation Heatmap

<table>
<tr><th>< 25 Years</th><th>41 – 55 years</th></tr>
<tr><td></td><td></td></tr>
</table>

For the analysis of age group, we want to verify is there any difference between age groups “< 25 years” and “41 – 55 years”. This is based on our hypothesis that different age groups will be influenced by different types of factors in the UTAUT model. To test our hypotheses, we filter our results to just “< 25 years” respondent and “41 – 55 years” respectively. We then construct a correlation heatmap to visualize the relationship between the UTAUT factor items. As we can see, both heatmaps are not visually the same, therefore, it proves our hypotheses that there is a difference in significant factors between respondent group aged “< 25 years” and “41 – 55 years”.

Gender

Table 3.1-4 Gender Graph

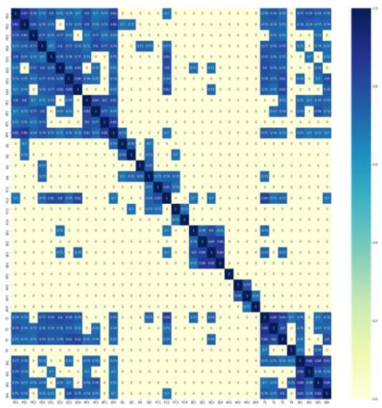
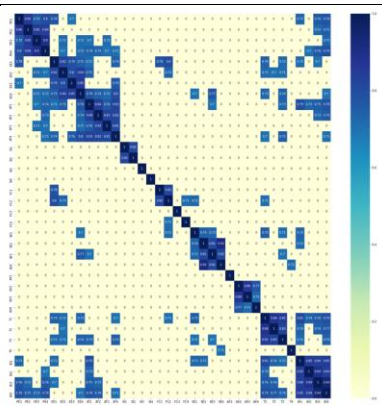
<table>
<tr><th>Distribution Among Gender Groups</th><th>Comparison with Targeted Variable</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

The preliminary analysis done for “Gender” Variable indicates an even distribution between the gender groups “Male” and “Female”. The results are shown to be a bit fairly biased towards “Male” respondents, as “Male” has a much higher distribution compared to “Female” respondents. This is hypothesized because there are much more men in FICT, Faculty of Information, Science and Technology compared to female. The descriptive statistics indicate that there are 59.44% Male and 40.56% female among 286 respondents.

When comparing gender groups to our targeted variable, Have you made any electronic payments in the past 12 months?”, the graph indicates that majority for each gender group are more inclined to have used an E-payment before in the past 12 months.

Is there any difference between respondent group “Male” and “Female”?

Table 3.1-5 Gender Correlation Heatmap

Male	Female
	

For the analysis of gender group, we want to verify is there any response difference between gender groups “Male” and “Female”. This is based on our hypothesis that different gender groups will be influenced by different types of factors in the UTAUT model. To test our hypotheses, we filter our results to just “Male” respondent and “Female” respondents respectively. We then construct a correlation heatmap to visualize the relationship between the UTAUT factor items. As we can see, both heatmaps are not visually the same, therefore, it proves our hypotheses that there is a difference in significant factors between respondent group gendered “Male” and “Female”.

Data Pre-processing

Once the data have been collected, we may begin to train our model. Before training our decision tree model, we must first pre-process the data so that our decision tree model is able to learn from the features successfully.

The first step for data pre-processing is to check for any null values in our data. To handle missing values of categorical data, the standard procedure is to replace it with values such as “Not Available” or “NA”. Otherwise, the alternative is to replace it manually. We have found no missing values using python to check the data frame.

From a brief analysis of our 63 attributes, we have a mixture of categorical and numerical attributes. For the moderated factors sections, taking “Age” as an example, we observe values such as “< 25 years”, “26 – 40 years”, “41 – 55 years” and “above 55 years”. These values are considered as categorical data and Decision Tree, Random Forest or XGBoost algorithm are unable to process these results. Therefore we have to convert these values into numerical values. This process is also known as label binarize.

Table 3.1-6 Feature Conversion

No	Attributes	Values
1	Age	'< 25 years': Class 1 '26 - 40 years': Class 2 '41 - 55 years': Class 3 'above 55 years': Class 4
2	Gender	'Male': Class 1 'Female': Class 2
3	Marital Status	'Single': Class 1 'Married': Class 2 'Other': Class 3
4	Education Level	'Primary school': Class 1 'Secondary/High school': Class 2 'College/university': Class 3 'Graduate school': Class 4 'Other': Class 5
5	Work Industry	'Banking / Finance': Class 1 'Education': Class 2 'Healthcare': Class 3 'Manufacturing': Class 4 'Retail / Hypermarket': Class 5 'Other': Class 6
6	Work Position	'Junior management': Class 1 'Middle management': Class 2 'Top management': Class 3 'Professional': Class 4 'Other': Class 5

After we have converted our categorical attributes into numerical values, we can construct our dataset to test and train model. One of the objectives of our project is to understand features that affect customer intention to adopt E-Payment. For the nature of this project, we have to filter out our data by users whom have experience in using E-Payment. Therefore, using "Have you ever purchased anything using the E-payment mode?" as our filter variable, we filter only results where the respondent answered “Yes” for our following questions. We then finalize our dataset and prepare for feature selection process.

Our finalized Dataset used for training and testing the model is shown below:

Table 3.1-7 Attribute Name and Types After Conversion

No	Attributes	Values
1	Age	Categorical: 0, 1, 2, 3
2	Gender	Categorical: 0, 1
3	Marital Status	Categorical: 0, 1, 2
4	Education Level	Categorical: 0, 1, 2, 3, 4
5	Work Industry	Categorical: 0, 1, 2, 3, 4, 5
6	Work Position	Categorical: 0, 1, 2, 3, 4
7	PE1	Numerical: 1 to 5
8	PE2	Numerical: 1 to 5
9	PE3	Numerical: 1 to 5
10	PE4	Numerical: 1 to 5
11	EE1	Numerical: 1 to 5
12	EE2	Numerical: 1 to 5
13	EE3	Numerical: 1 to 5
14	EE4	Numerical: 1 to 5
15	AT1	Numerical: 1 to 5
16	AT2	Numerical: 1 to 5
17	AT3	Numerical: 1 to 5
18	AT4	Numerical: 1 to 5
19	SI1	Numerical: 1 to 5
20	SI2	Numerical: 1 to 5
21	SI3	Numerical: 1 to 5
22	SI4	Numerical: 1 to 5
23	FC1	Numerical: 1 to 5
24	FC2	Numerical: 1 to 5
25	FC3	Numerical: 1 to 5
26	FC4	Numerical: 1 to 5
27	SE1	Numerical: 1 to 5

28	SE2	Numerical: 1 to 5
29	SE3	Numerical: 1 to 5
30	SE4	Numerical: 1 to 5
31	AX1	Numerical: 1 to 5
32	AX2	Numerical: 1 to 5
33	AX3	Numerical: 1 to 5
34	AX4	Numerical: 1 to 5
35	T1	Numerical: 1 to 5
36	T2	Numerical: 1 to 5
37	T3	Numerical: 1 to 5
38	T4	Numerical: 1 to 5
39	BI1	Numerical: 1 to 5
40	BI2	Numerical: 1 to 5
41	BI3	Numerical: 1 to 5
42	BI4	Numerical: 1 to 5

Feature Selection

Once our dataset have been finalized, we can now use Correlation analysis to do feature selection. Correlation analysis is a technique of analyzing the linear relationship between two variables. The two variables can be independent or dependent based on the strength of the relationship computed. We define the strength of the correlation as correlation coefficient. Correlation coefficient may be derived from various formula such as Pearson's Correlation coefficient, Spearman's Correlation coefficient and many more.

In this Project, we explore two correlation analysis technique, that is Pearson's Correlation Coefficient and Spearman's Correlation Coefficient.

Pearson's r is a measure that assesses the association between two continuous (or metrical) variables.

Below is the formula for Pearson's Correlation

$$r = \frac{\sum\left[\left(\frac{x_i-\bar{x}}{s_X}\right)\left(\frac{y_i-\bar{y}}{s_Y}\right)\right]}{n}$$

Figure 3.1.2 Pearson's Correlation Formula

The correlation values for Pearson's correlation is between -1 and 1. The closer the value to -1 or 1, the stronger the association. Positive values represents positive relationship, whereas negative values represents negative relationship.

Other than Pearson's Correlation, Spearman Correlation is encouraged to be used for Likert-Scale Response item. This is because Likert-Scale are ordinal factors, therefore, we should perform non-parametric measure to compute the correlation of the likert-scale factor items, in which case, the UTAUT items.

Spearman's q (a special case of Pearson's r) is a nonparametric measure that outputs the correlation using the ranked scores of the two variables. It is commonly used for ordinal variables.

Spearman's q s calculated based on this equation:

$$\boxed{\rho = 1 - \frac{6\sum d_i^2}{n^3 - n}}$$

Figure 3.1.3 Spearman's Correlation Formula

The output of the values of Spearman Correlation lies between -1 and 1. Spearman's q can be interpreted as the difference of normality and proportion of validity between the two ordinal variables.

Setting the Correlation threshold at 0.7, we will be able to filter out and select meaningful attributes from our UTAUT Factors. As we are using two different correlation techniques, we then split our features into intersected features, non-intersected features and union features. We then compare the model accuracy to prove that our correlation feature selection technique is a success.

Model Building

The models that have been built to analyze our dataset respectively are tree-based algorithm, built using python Scikit-Learn library. The tree-based algorithms used are Decision Tree algorithm, Random Forest Algorithm and XGBoost algorithm. Decision Tree algorithm is selected because it can circumvent the issues of multicollinearity in our features. Random Forest is a bagging method, whereas XGBoost is a boosting method. The models have been built with 5-fold cross validation with GridSearchCV in each fold to obtain the best parameters for each fold.

The tuned hyperparameters are shown in the table below

Table 3.1-8 Hyperparameters for Tree-Based Model

Classifiers	Hyperparameters	Value
Decision Tree Algorithm	Max_depth	2, 3, 5, 10, 20
	Min_samples_leaf	5, 10, 20, 50 ,100
	Criterion	Gini, Entropy
Random Forest Algorithm	Max_depth	2, 3, 5, 10, 20
	Min_samples_leaf	5, 10, 20, 50 ,100
	N_estimators	5, 10, 20, 50, 100
	Bootstrap	True, False
XGBoost Algorithm	Max_depth	2, 3, 5, 10, 20

	Min_samples_leaf	5, 10, 20, 50 ,100
	Learning Rate	0.01, 0.05, 0.1
	Colsample_bytree	0.3, 0.7

Model Evaluation

In this project, the system should be able to identify important features that influence user to adopt e-payment system. The system should be able to perform well on UTAUT model-based dataset and come up with highly correlated set of user behaviors. However, the accuracy may vary based on different demographic groups. Accuracy of the model may be affected due to classification model overfitting the dataset or number of records is insufficient to properly train the model. Hence, one of the evaluation metrics we may use to determine the performance of our model is confusion matrix.

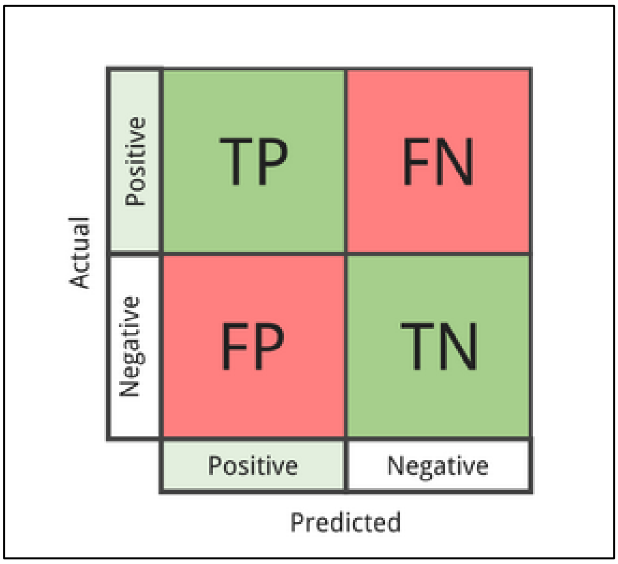


Figure 3.1.4 Confusion Matrix

A confusion matrix is a table that is used to describe the performance of a classification model. The confusion matrix is a 2x2 matrix, that has four quadrants. The values of the four quadrants are [TN FP] [FN TP] respectively. TN (True Negatives) represents the number of negative samples that are correctly predicted as negative. FP (False Positives) represents the number of negative samples that are falsely predicted as positive. TP (True Positives) represents the number of positive samples that are correctly predicted as positive and FN (False Negatives) represents the number of positive samples that are falsely predicted as negative.

To generate a confusion matrix output, we may use Python Scikit-Learn library to generate confusion matrix or classification report. From the confusion matrix, a few key constructs may be calculated, that is **precision**, **Recall** and **F1 Score**.

Precision calculates the number of positive class predictions that belong to the positive class.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Figure 3.1.5 Precision Score Formula

Recall calculates the number of positive class predictions made of all positive examples in

the dataset.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Figure 3.1.6 Recall Score Formula

F1-Score is a function of recall and precision. F1-score is used to balance the precision and recall

if there is a large number of actual negatives.

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Figure 3.1.7 F1-Score Formula

To evaluate each model, we need to observe their precision, recall and f1-score value respectively. A model with high precision value indicates that it is good at classifying the targeted class correctly. A model with high recall value indicates that it is good at classifying incorrect targeted class as negative value. F1-score balance our precision and recall value. A high F1-Score indicates that our model can classify targeted class correctly and incorrect class as negative value.

3.2 Project Timeline

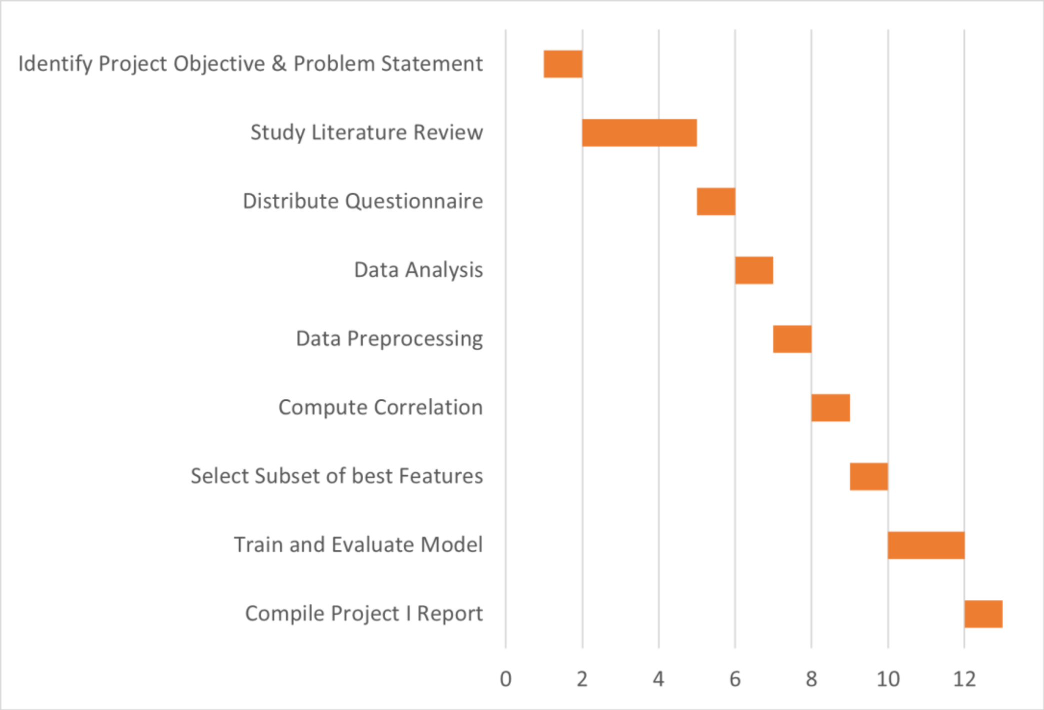


Figure 3.2.1 Project I Timeline

Figure 3.2.1 shows the Gantt Chart for the development timeline for Project I. This timeline was more focused on understanding literature review of UTAUT model and distributing questionnaire for data collection. During Project I, a lot of effort was spent on analysing the dataset and pre-processing the dataset. Then we compute the correlation of each feature and select the best subset of features using Correlation Based Feature Selection Algorithm. After selecting the best subset, we may begin to train and evaluate the model.

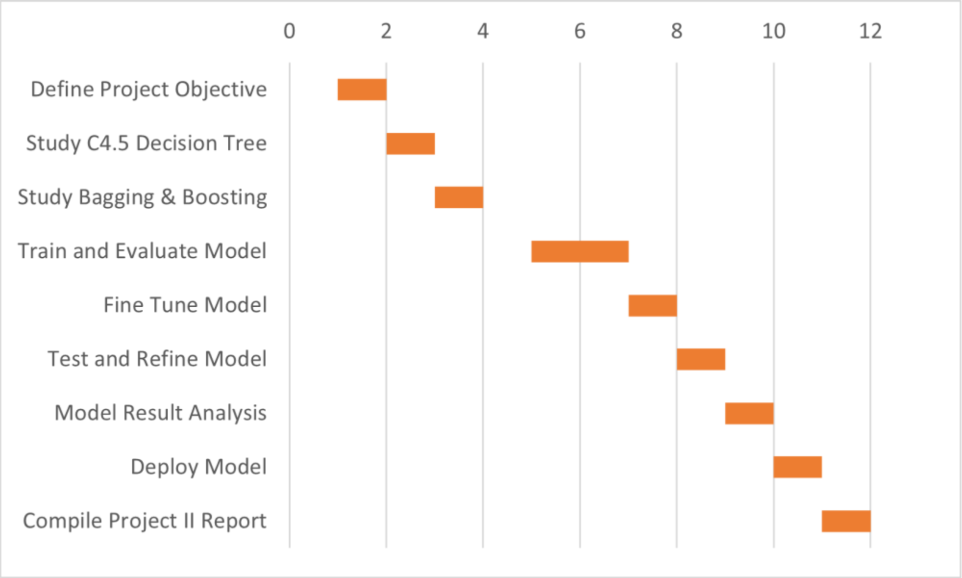


Figure 3.2.2 Project II Timeline

Figure 3.2.2 shows the timelines for the project development for FYP Project II. For this timeline, now that we understand the data, more emphasis is placed on understanding the machine learning models like c4.5 Decision Tree and Bagging and boosting Method. Then, we train and evaluate the machine learning model, measure the accuracy whether it can identify significant user behaviour that have high accuracy. Then, we identify underlying factors among the subset of features and perform result analysis. We can then deploy the model for real world performance purpose.

4 Chapter 4: Feature Selection Design

4.1 Correlation Network Graph

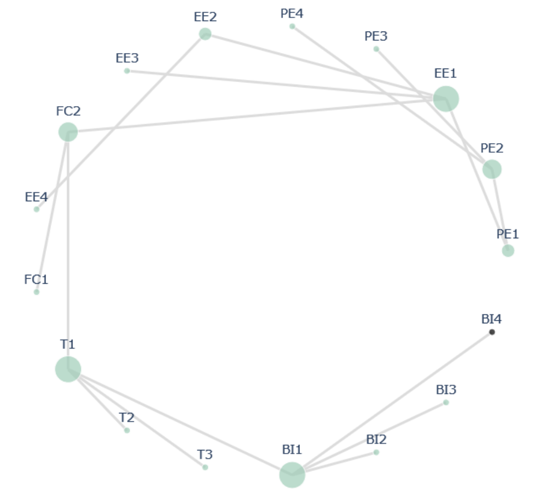
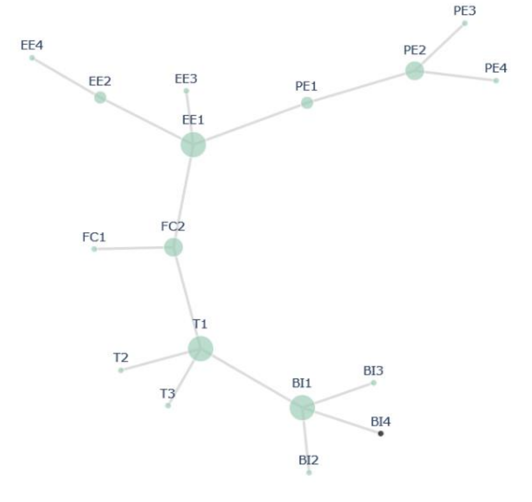
Network graphs is a mathematical structure to show relations between points in a less statistical manner. It allows us to study relationship between factors in a more aesthetically pleasing manner.

Network graphs are made up of nodes and edges. Commonly, Nodes or vertices are the discrete entities of the graph or dataset. Edges or links are used to represent relations among the nodes. Weighted edges can be used to represent correlation coefficient. In this scenario, a node represents a feature in the dataset, which is a UTAUT Factor item, whereas Edges are used to represent the strength of correlation between the two UTAUT Factors. A positive relationship between the two factors is coloured green whereas a negative relationship is coloured red. The strength of the correlation is represented by the size of the node. A node that is larger in size also represents that it has a large degree, thus it affects numerous factors in the dataset.

In network analysis, there are two types of graph used to visualize correlation matrix. Circular layout and Fruchterman-Reingold Layout. Fruchterman-Reingold layout is in the format of a minimum spanning tree. We can visualize more relationships effectively using Fruchterman-Reingold layout.

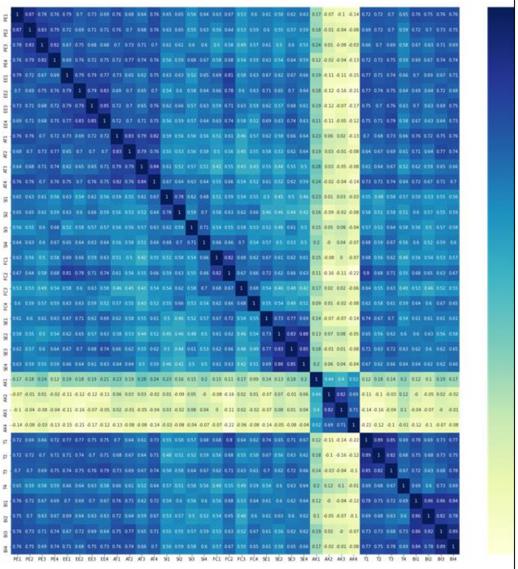
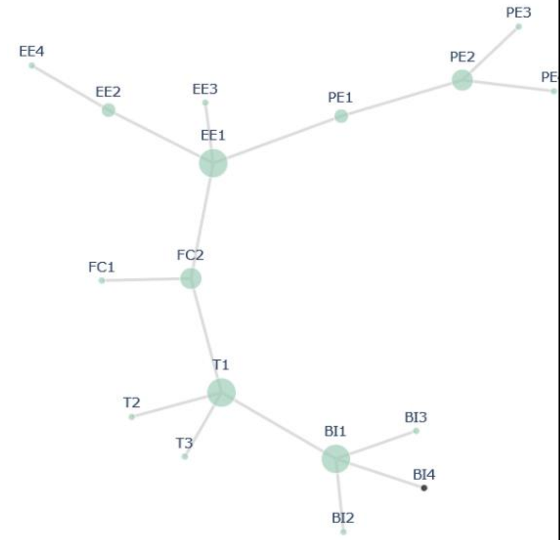
Comparison between Circular layout and Fruchterman-Reingold Layout

Table 4.1-1 Comparison of Network Layout Graph

Circular Layout	Fruchterman-Reingold Layout
	

Comparison with Correlation Heatmap Graph

Table 4.1-2 Comparison of Heatmap and Network Graph

Correlation Heatmap	Correlation Network Graph
	

As we can see from the above graph, the correlation network graph is much more visually appealing compared to the correlation heatmap. The strength of the correlation network graph becomes more apparent when dealing with large feature sets. The correlation network graph visually shows the relationship between each feature in the form of a minimum spanning tree, whereas the correlation heatmap can only show the strength of the relationship between two features.

4.2 Feature Selection

Feature selection is a technique for extracting the least number of features from a problem domain while retaining a high classification measure in representing the original features. It is a technique categorized as data pre-processing, that is a necessary step in determining important features.

When the number of features chosen is small, the likelihood of information content is low. When several features are chosen, the presence of noise for irrelevant data is more likely to occur. Therefore, feature selection should be on the right selection of subsets, avoiding too large or too small number of features. To find the most optimal features among a large group of features, the fastest solution is to generate candidate feature subsets and use a search algorithm to find the best possible combination of features.

There are four basic steps in a typical feature selection process. The process of feature selection is as below:

- A generation procedure, where we generate all possible candidate subset.
- An evaluation function to evaluate the subset to determine its relevancy towards the classification task. In this project, we use correlation merit formula as an evaluation measure.
- Stopping criteria to determine where to terminate the execution of the search algorithm. The final subset will be the most optimal feature subset.
- Validation procedure is to check whether the selected feature subset is valid.

Feature selection can be classified into two types, that is Filter based Feature selection and Wrapper based feature selection. Now, we will discuss about Filter Based approach in selecting features via correlation analysis.

4.3 Filter Based Method

Correlation Merit Based Selection

The correlation-based feature selection (CFS) method is a filter approach and therefore independent of the final classification model. It evaluates feature subsets only based on data intrinsic properties, that is the correlation of the subset.

The goal of the algorithm is to find a feature subset with low feature-feature correlation to avoid redundancy and high feature-class correlation to maintain or increase predictive power.

Correlation Merit Formula calculates the correlation value of the features in the subset using a correlation coefficient [15]. The correlation coefficient may be Pearson's coefficient or Spearman's Coefficient. The correlation merit formula loops through all the feature in the subset with the targeted value. It then generates a merit value to indicate the strength of the subset.

Below is the algorithm of the proposed filter-based method:

$$Merit = \frac{\overline{kavg\left(corr_{fc}\right)}}{\sqrt{k + k(k-1)\overline{avg\left(corr_{ff}\right)}}}$$

Figure 4.3.1 Correlation Merit Based Formula

Below is the working implementation in python

```python
# Calculate Merit
def merit_calculation(X, Y, func):
    k = X.shape[1]

    # average feature-class correlation
    rcf_all = [func(X.iloc[:, i], Y) for i in range(k)]
    rcf_all = list(map(lambda x : abs(x[0]), rcf_all))
    rcf = np.mean(rcf_all)

    # average feature-feature correlation
    corr = X.corr()
    if corr.shape != (1, 1):
        corr.values[np.tril_indices_from(corr.values)] = np.nan
    corr = abs(corr)
    rff = corr.unstack().mean()

    return (k * rcf) / math.sqrt(k + k * (k-1) * rff)
```

Figure 4.3.2 Python Implementation of CFS algorithm

The flowchart of the feature selection algorithm is provided below:

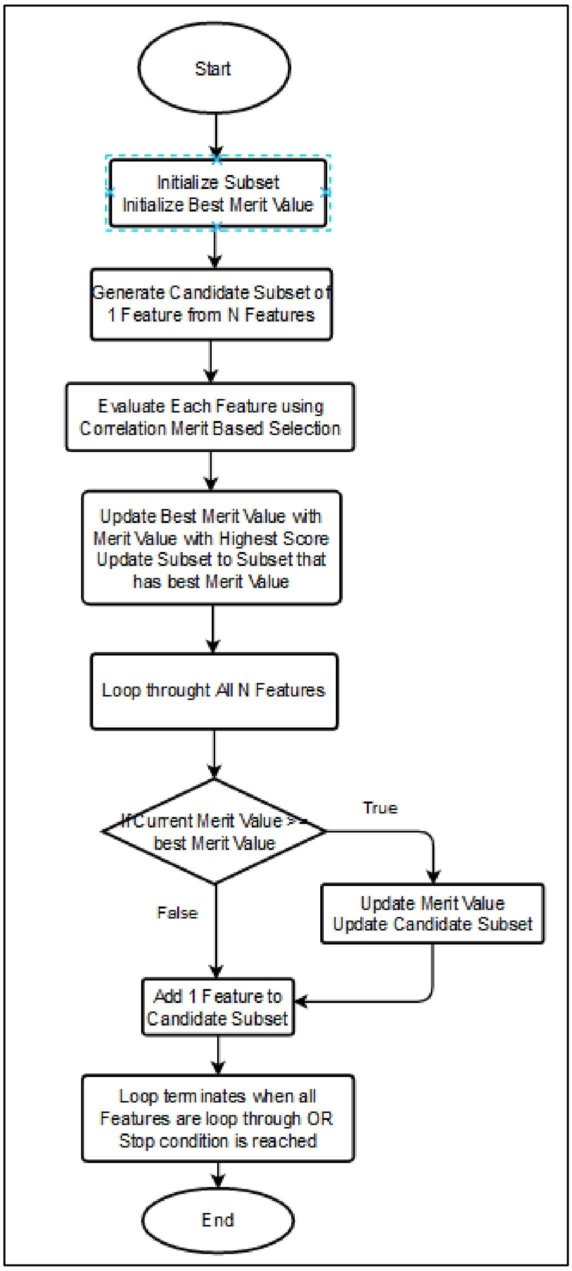


Figure 4.3.3 Flowchart of CFS Algorithm

When the algorithm is executed, it will first generate a list of candidate subsets from 1 to n feature size. The algorithm will then calculate the merit value of each subset. It will use the subset with the highest merit value as the initial search merit. Then the algorithm will constantly loop itself, slowly adding new features each iteration. Once the features are added, it will calculate the merit value. If the current merit value is smaller than the merit of the subset, the subset will be replaced and the merit value will be updated. The algorithm will backtrack for a maximum of 5 backtracks across all iteration. This approach is a heuristic search method.

4.4 Wrapper Based Methods

Wrapper Based feature selection is a feature selection technique that selects feature by using classification algorithm, rather than using a statistic method, unlike its counter part, filter based feature selection. The general consensus is that filter based feature selection are much faster than wrapper based feature selection. However, in terms of classification accuracy, wrapper based feature selection normally produce better results. The general argument is that the classifier that will be built from the feature subset should provide a better estimate of accuracy than a separate measure that may have an entirely different classification bias.

Below is an example of Stepwise Feature Selection Graph:

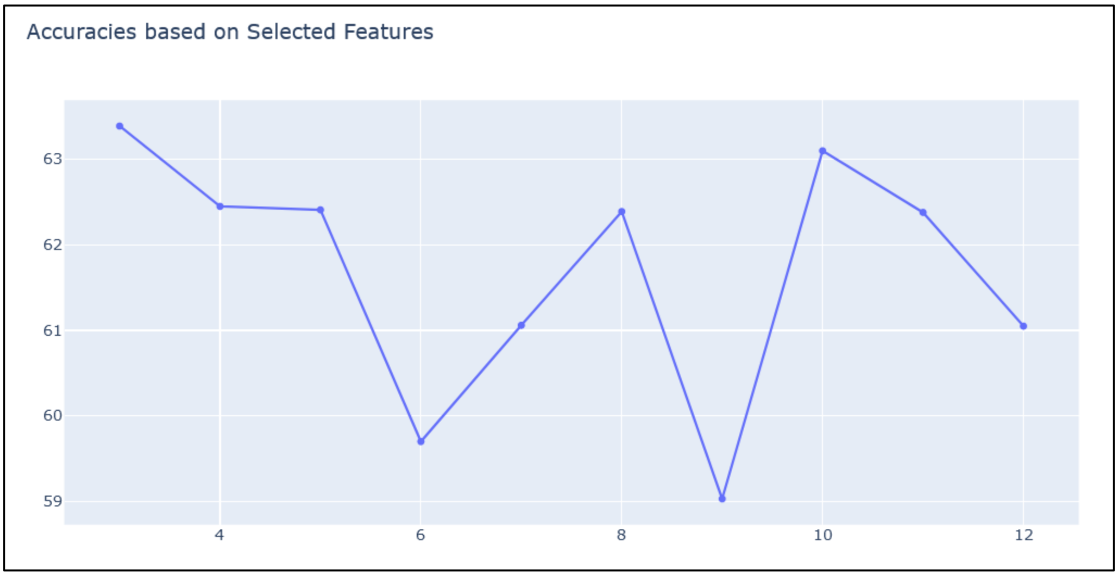


Figure 4.4.1 Stepwise Feature Selection Graph

The main disadvantage of wrapper approaches is that during the feature selection process, the classifier must be repeatedly called to evaluate a subset. Therefore, wrapper-based feature selection will be much more computationally expensive, compared to filter-based feature selection, however, it will provide better classification accuracy.

Wrapper based feature selection can be classified into two types, that is, forward feature selection and backward feature selection.

Forward Feature selection.

Below is a flowchart of the entire algorithm:

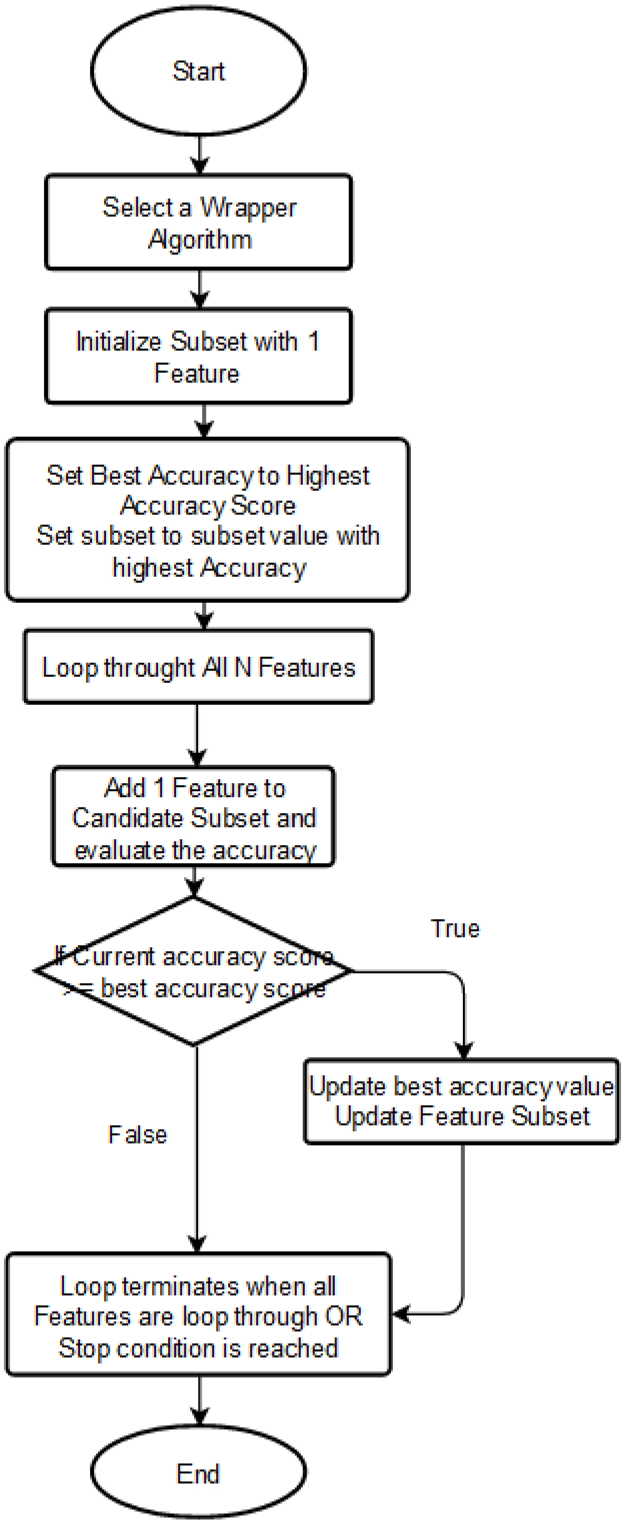


Figure 4.4.2 Forward Selection Algorithm Flowchart

To do forward feature selection, we must select our classification algorithm. For this purpose, we will use a decision tree as our classification algorithm. We set our backtrack to a maximum of five iterations. Our functions are executed by first having a base number of features of 1. We then use the features in the feature subset to calculate its classification accuracy. Over each iteration, we slowly add a feature and update the best classification accuracy. The function constantly loops iteratively until it discovers a subset that has better classification accuracy than the current one. When the accuracy value is updated, the best subset is also updated to the corresponding subset. The function executes iteratively until it reaches the stop termination, or it has gone through all possible features.

Here is the working implementation in python

```python
def forward_selection(X, Y, clf, n_selected_features = 10):
    # Number of Features
    n_features = X.shape[1]

    # selected feature set, initialized to be empty
    count, feature_set = 0, []

    while count < n_selected_features:

        f_loop = [i for i in range(n_features) if i not in feature_set]

        # 1. Get All Accuracy Score for All Subset in Feature Set
        acc_arr = []

        for i in f_loop:
            feature_set.append(i)
            acc_arr.append(get_acc_score_kcv(X.iloc[:, feature_set], Y, clf))
            feature_set.pop()

        # 2. Convert Python List to Series
        acc_arr = pd.Series(acc_arr, index = f_loop)

        # 3. Get the largest Feature
        idx = acc_arr.idxmax()

        # 4. add the feature which results in the largest accuracy
        feature_set.append(idx)
        count += 1

    # Sort Feature Set
    feature_set = sorted(feature_set)

    # Get Feature from X.Columns
    feature_set = list(map(lambda x: X.columns[x], feature_set))

    return feature_set
```

Figure 4.4.3 Python Implementation of Forward Selection Algorithm

Backward Selection

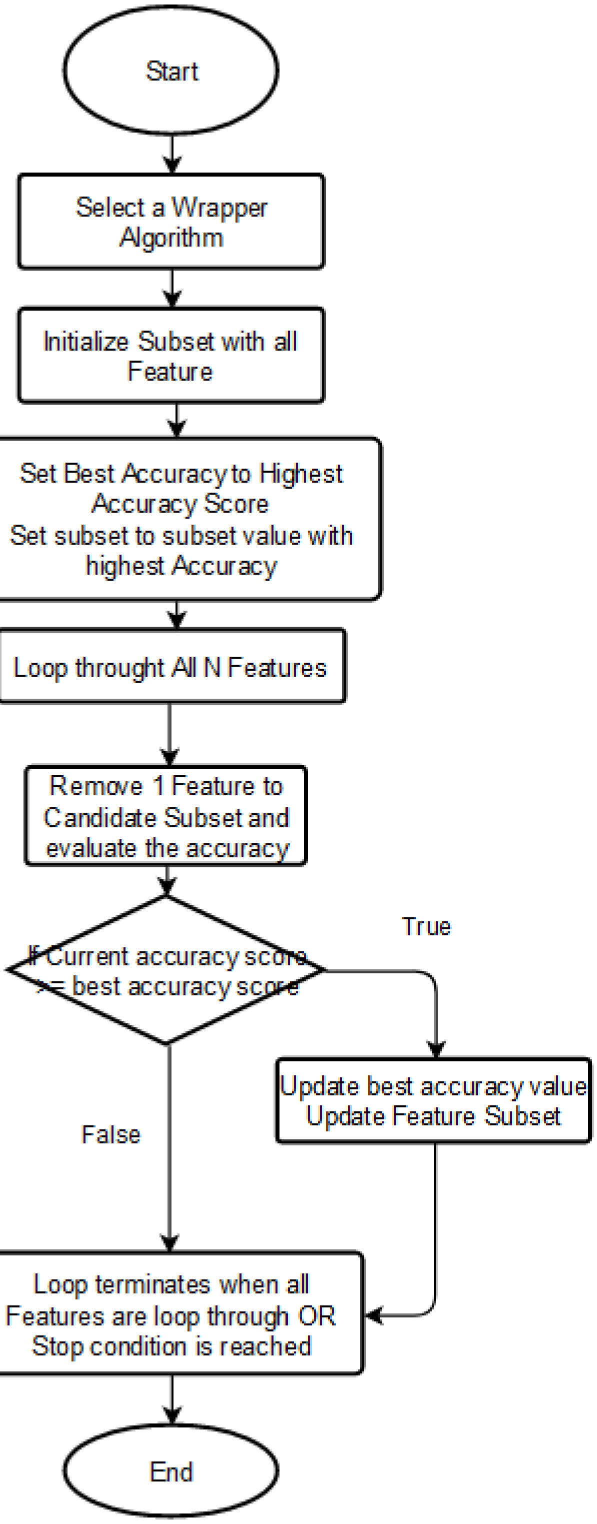


Figure 4.4.4 Flowchart of Backward Selection Algorithm

For backward selection, rather than starting from one feature, we start off with all the features in our subset. We then update our classification accuracy by using our classification algorithm to train and get an accuracy value. The function then iteratively removes each feature, one by one. When one feature is removed, the model is then trained and tested on the current feature subset to get classification accuracy. If the current accuracy is better than the classification accuracy, we update the value and the subset.

Compared to forward selection, backward selection will be more time-consuming and computationally expensive as, rather than slowly adding one feature, we test all possible feature sets backwards. However, the classification accuracy will be much better.

Below is the working implementation in python.

```
def backward_selection(X, Y, clf, n_selected_features = 10):
    # Number of Features
    n_features = X.shape[1]

    # selected feature set, initialized to be empty
    count, feature_set = n_features, [*range(n_features)]

    while count > n_selected_features:

        f_loop = [i for i in range(n_features) if i in feature_set]

        # 1. Get All Accuracy Score for All Subset in Feature Set
        acc_arr = []

        for i in f_loop:
            feature_set.remove(i)
            acc_arr.append(get_acc_score_kcv(X.iloc[:, feature_set], Y, clf))
            feature_set.append(i)

        # 2. Convert Python List to Series
        acc_arr = pd.Series(acc_arr, index = f_loop)

        # 3. Get The Largest idx
        idx = acc_arr.idxmax()

        # 4. delete the feature which results in the largest accuracy
        feature_set.remove(idx)

        count -= 1

    # Sort Feature Set
    feature_set = sorted(feature_set)

    # Get Feature from X.Columns
    feature_set = list(map(lambda x: X.columns[x], feature_set))

    return feature_set
```

Figure 4.4.5 Python Implementation of Backward Selection Algorithm

5 Chapter 5: Experiment / Simulation

5.1 Hardware Setup

The hardware involved in this project is just a laptop. The laptop is used to design the machine learning model and analyze the dataset. The laptop is also used to develop the GUI implementation of our feature selection tool.

Table 5.1-1 Hardware Specifications

	Minimum Requirement
Processor	Intel(R) Core(TM) i5- 8250U CPU @ 1.60GHz 1.80 GHz
Memory (RAM)	8GB
Disk Space	4GB
Display	1280 x 800

5.2 Software Setup

The programming languages that have been used in this project are:

Python Language

Python is a high-level programming language, that is actively used in the Machine Learning community. Python is an interpreted language, therefore, instead of compiling the entire code for syntax error, it can run code cell by cell. This unique ability allows machine learning engineer to debug code quickly without compiling the entire code again.

Figure 5.2.1 Python Language

Javascript Language

Javascript is a high-level programming language that is one of the core technologies of the world-wide web. It is an interpreted language, similar to python. This mean it runs code cell-by-cell rather than compiling the entire code block. Javascript is favoured by the full stack community, used to create web apps and server using Node JS and express library. We use Javascript to code our GUI Implementation.

Figure 5.2.2 Javascript Language

Visual Studio Code

Visual studio code is a source code editor, developed by Microsoft Windows. It is a lightweight code editor that can be used to edit files fast and efficient. Visual studio code can run any language as long as we have the language installed in our path. Visual studio code also comes with many extension, its ability does not lose out to official IDE. We use visual studio code to implement our GUI and debug errors.

Figure 5.2.3 Visual Studio Code IDE

Jupyter Notebook IDE

Jupyter Notebook IDE is a web-based interactive development environment to execute python code cell-by-cell. Its cell by cell output is a favorite among machine learning developers as developers only have to debug error in the executing cell. Jupyter notebook IDE is extensively used for machine learning development and provides many functionalities suited for that purpose.

Figure 5.2.4 Jupyter Notebook IDE

Weka

Weka is a collection of machine learning algorithms for data mining tasks. It is a visualization tool that displays machine learning data analysis. Weka contains tools for data pre-processing, classification, regression, clustering, association rules and many more. It is developed using Java programming language making it fully portable. For the machine learning algorithm analysis, we will be using Weka to analyse and visualize the results.

Figure 5.2.5 Weka Tool

5.3 Settings and Configuration

To run our jupyter notebook, Navigate to Train-Test Folder

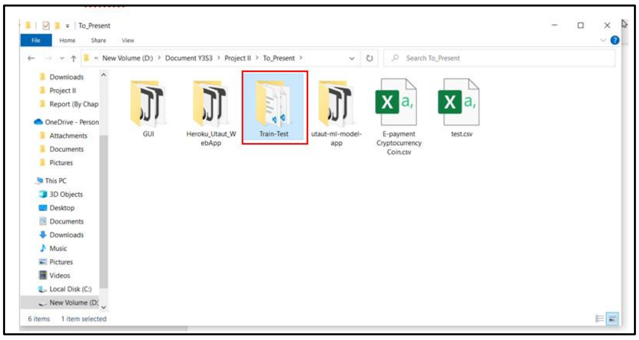


Figure 5.3.1 Navigate to Train-Test Folder

Shift + Right click on any empty space. Select “Open PowerShell window here”

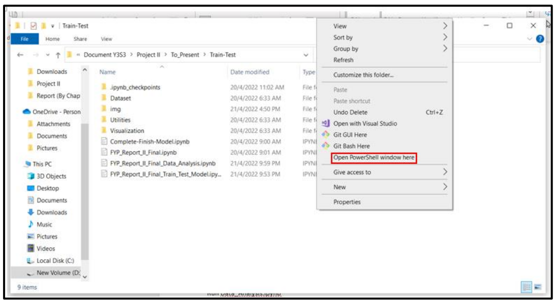


Figure 5.3.2 Open PowerWindow Shell

In powerShell window, type the command “jupyter notebook”. This will launch jupyter notebook ide.

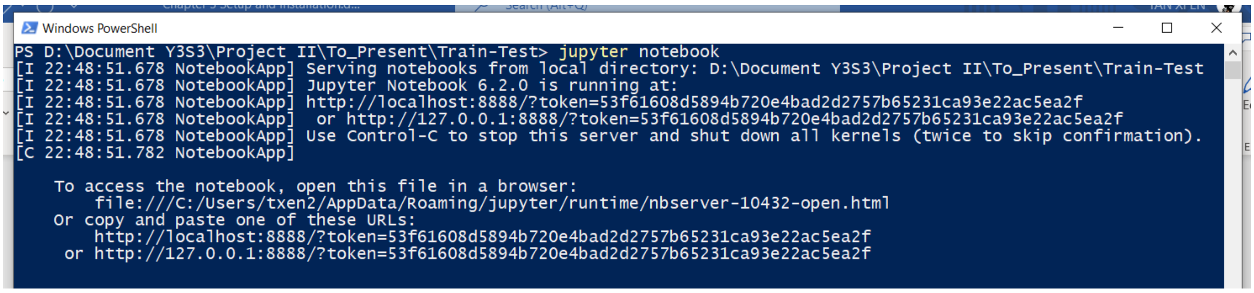


Figure 5.3.3 PowerWindow Shell

On any web browser, in the URL bar, enter “localhost:8888/tree”

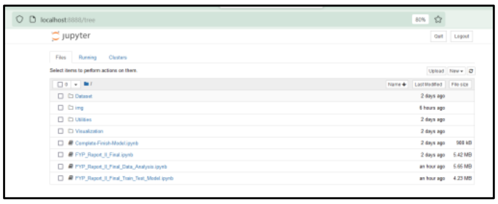


Figure 5.3.4 Jupyter Notebook IDE

You can now run our Data_Analysis.ipynb and Train_Test_Model.ipynb.

Run GUI

To run our GUI, Navigate to GUI Folder

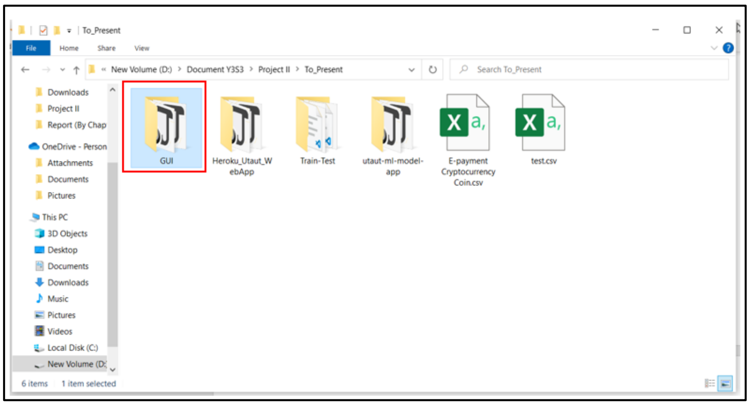


Figure 5.3.5 Navigate To GUI

Shift + Right click on any empty space. Select “Open PowerShell window here”

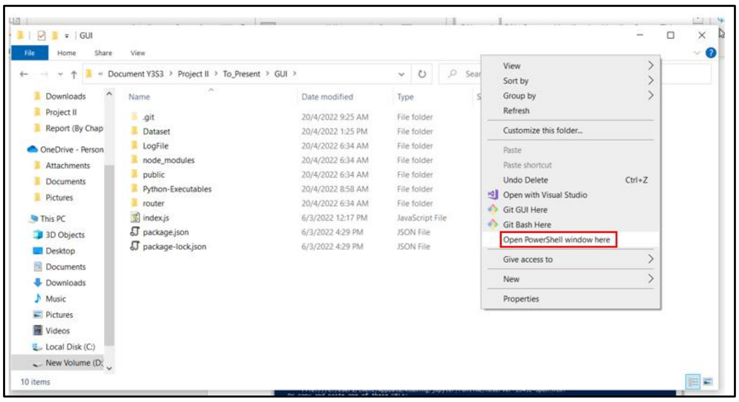


Figure 5.3.6 Open PowerWindowShell

In powerShell window, type the command “code .”. This will open visual studio code window.

Figure 5.3.7 PowerWindow Shell

Ctrl + ` to open up terminal. Type the command “node index” to launch the GUI

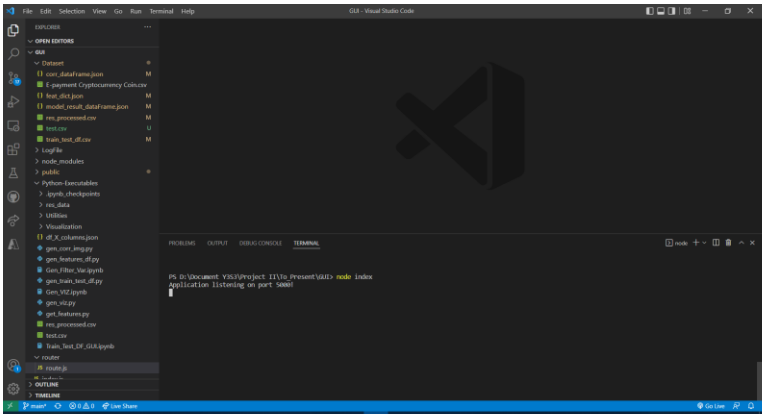


Figure 5.3.8 Visual Studio Code

Open up any web browser. At the URL bar, enter “localhost:5000”

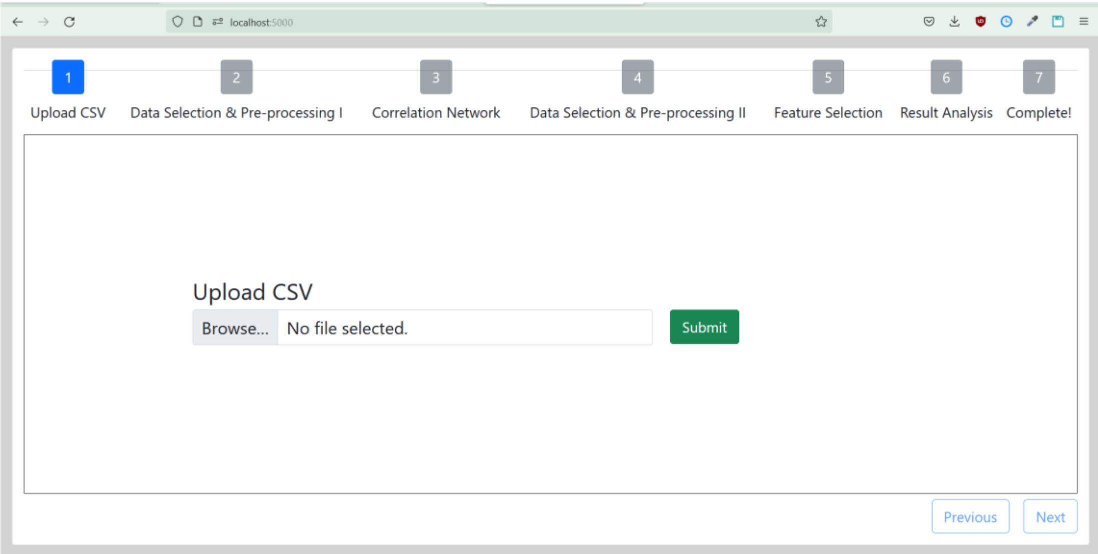


Figure 5.3.9 GUI Implementation

You can now use our GUI to run your experiment Simulation.

5.4 System Operation

The following screenshots shows how our system operate to get the experimentation data and result.

Load Libraries

This steps loads the necessary libraries to execute our Jupyter Notebook cells.

Load Libraries

```
In [1]: import time
        import numpy as np
        import pandas as pd
        import seaborn as sns
        import ipywidgets as ipw
        import plotly.express as px
        import scipy.stats as stats
        import matplotlib.pyplot as plt
        import plotly.graph_objects as go

        from plotly.subplots import make_subplots

        from sklearn import svm, tree
        from sklearn.tree import DecisionTreeClassifier
        from sklearn.ensemble import RandomForestClassifier

        from sklearn.model_selection import train_test_split, KFold, GridSearchCV
        from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, matthews_corrcoef, precision_recall_curve, r

        from scipy.stats import pearsonr, spearmanr, kendalltau, pointbiserialr, chi2, chi2_contingency

        from imblearn.over_sampling import SMOTE

        from xgboost import XGBClassifier

        from dtreeviz.trees import *

        %matplotlib inline
```

Figure 5.4.1 Load Libraries

Read Data

This step reads the data from our “E-payment Cryptocurrency Coin.csv” dataset using pandas library. These steps transfer the data from the dataset into a pandas dataframe object.

Load Dataset

```
In [3]: import pandas as pd
        df = pd.read_csv("Dataset/E-payment Cryptocurrency Coin.csv")
```

Figure 5.4.2 Read Dataset

Data Pre-processing

This step pre-processed the data from our dataset. We first check for null value in our dataset, to ensure that there are no missing or incorrect data.

Check for null value

```
In [16]: all_df.isnull().sum()
```

Figure 5.4.3 Check Null

Expected Output

Age	0	AT1	0	SE1	0	BI1	0
Gender	0	AT2	0	SE2	0	BI2	0
Marital Status	0	AT3	0	SE3	0	BI3	0
Education Level	0	AT4	0	SE4	0	BI4	0
Work Industry	0	SI1	0	AX1	0	dtype: int64	
Work Position	0	SI2	0	AX2	0		
PE1	0	SI3	0	AX3	0		
PE2	0	SI4	0	AX4	0		
PE3	0	FC1	0	T1	0		
PE4	0	FC2	0	T2	0		
EE1	0	FC3	0	T3	0		
EE2	0	FC4	0	T4	0		
EE3	0						
EE4	0						

We also label binarize some columns. This is to ensure that we do not have any data that is not in numerical format.

Label Binarizer

```
In [17]: def convert_nominal(arr, term_arr):
             tmp_dict = {val:ind for (ind, val) in enumerate(term_arr)}
             return arr.map(lambda x : tmp_dict[x])
```

Figure 5.4.4 Label Binarizer

Expected Result:

Before

	Age	Gender	Marital Status	Education Level	Work Industry	Work Position
0	< 25 years	Female	Single	College/university	Banking / Finance	Other
1	< 25 years	Female	Single	College/university	Other	Other
2	41 - 55 years	Female	Single	College/university	Manufacturing	Middle management
3	< 25 years	Male	Single	College/university	Education	Other
4	< 25 years	Female	Single	College/university	Other	Other
...	...	...	...	...	...	...
280	41 - 55 years	Male	Married	Secondary/High school	Other	Other
281	above 55 years	Male	Married	Graduate school	Education	Top management
282	above 55 years	Female	Married	College/university	Other	Other
283	41 - 55 years	Male	Married	Graduate school	Education	Professional
284	< 25 years	Male	Single	College/university	Other	Other

After

	Age	Gender	Marital Status	Education Level	Work Industry	Work Position
0	0	1	0	2	0	4
1	0	1	0	2	5	4
2	2	1	0	2	3	1
4	0	1	0	2	5	4
5	1	0	1	2	2	3
...	...	...	...	...	...	...
280	2	0	1	1	5	4
281	3	0	1	3	1	2
282	3	1	1	2	5	4
283	2	0	1	3	1	3
284	0	0	0	2	5	4

Generate Bar Graph

The bar graph indicates a distribution of the User Group. To execute the bar graph function, we need to set the targeted variable. For example, to observe the distribution of Age group in our dataset, we set col to "Age".

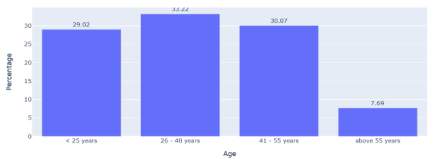
Display Bar Chart Side by Side

```
In [16]: def bar_graph_side(mod_var, filter_var, col_list, graph_df):
             fig_arr = []

             tmp_df = pd.DataFrame(graph_df[mod_var].value_counts()).T
             tmp_df = tmp_df[col_list]
             tmp_df.iloc[0, :] = tmp_df.iloc[0, :] / tmp_df.iloc[0, :].sum() * 100.0
             fig_arr.append(show_bar_graph_percentage(tmp_df, "", mod_var, "", "Percentage"))

             tmp_df = graph_df.loc[:, [mod_var]]
             tmp_df["Target"] = graph_df.loc[:, [filter_var]]

             tmp_df = pd.crosstab(tmp_df.iloc[:, 1], tmp_df.iloc[:, 0])

             tmp_df = tmp_df[col_list]

             fig_arr.append(show_bar_graph(tmp_df, "", mod_var, "Legend", "Count"))

             h_arr = [go.FigureWidget(fig) for fig in fig_arr]

             display(ipw.HBox(h_arr))
```

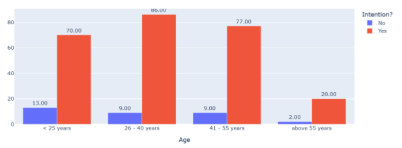
Figure 5.4.5 Generate Bar Graph

Expected Result:

Distribution of Age Group



Comparison with Targeted Variable



Generate Correlation Network Graph

To generate the correlation network graph, we have to first create a correlation matrix. We then load our own custom made network graph, created using plotly libraries. We set the graph layout to "fruchterman-reingold layout".

```
In [29]: # Fruchterman Reingold Layout
         fruFig = network_graph(
             corr_df,
             "Important Features - Fruchterman Reingold Layout",
             nx.fruchterman_reingold_layout,
             0.75
         )
         layout = dict(
             width = 800,
             height = 800,
             hovermode = "closest",
             plot_bgcolor = "#fff",
         )
         fruFig.update_layout(layout)
```

Figure 5.4.6 Generate Correlation Network Graph

Expected Result

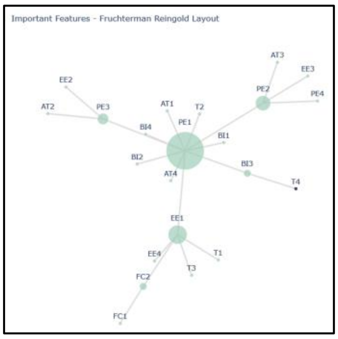


Figure 5.4.7 Correlation Layout Graph

SMOTE Imbalanced

```
tmp_Y = df_Y

imb_df = pd.DataFrame(index = index_dict[tmp_Y.name])

imb_df["Count"] = tmp_Y.value_counts().sort_index().set_axis(index_dict[tmp_Y.name])

imb_df["Count (%)"] = round(imb_df["Count"] / imb_df["Count"].sum() * 100.0, 2)

if min(imb_df["Count"]) > len(index_dict[tmp_Y.name]) * 2:
    oversample = SMOTE()
    df_X, df_Y = oversample.fit_resample(df_X, df_Y)
```

Figure 5.4.8 SMOTE Imbalanced

Expected Result:

Before

	Count	Count (%)
Male	149	59.13
Female	103	40.87

After

	Count	Count (%)
Male	149	50.0
Female	149	50.0

Decision Tree

```
Decision Tree

tmp_X = df_X
tmp_Y = df_Y

model = DecisionTreeClassifier(max_depth = 3)
name = "Decision Tree"

acc_score = get_acc_score_kcv(tmp_X, tmp_Y, model)

print(f"Accuracy Score: {round(acc_score, 2)}%")

model_dict[name] = create_ModelObj(model, name, tmp_X, tmp_Y, index_dict[tmp_Y.name])
```

Figure 5.4.9 Decision Tree

Decision Tree HyperTuning

```
Decision Tree (Hyperparameter Tuning)

params = {
    'max_depth': [2, 3, 5, 10, 20],
    'min_samples_leaf': [5, 10, 20, 50, 100],
    'criterion': ["gini", "entropy"]
}

model = DecisionTreeClassifier()
grid_search = GridSearchCV(estimator=model, param_grid=params, cv=5, n_jobs=-1, verbose=1, scoring = "accuracy")

grid_search.fit(tmp_X, tmp_Y)

Fitting 5 folds for each of 50 candidates, totalling 250 fits

GridSearchCV(cv=5, estimator=DecisionTreeClassifier(), n_jobs=-1,
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [2, 3, 5, 10, 20],
                         'min_samples_leaf': [5, 10, 20, 50, 100]},
             scoring='accuracy', verbose=1)

grid_search.best_estimator_

DecisionTreeClassifier(criterion='entropy', max_depth=10, min_samples_leaf=5)

tmp_X = df_X
tmp_Y = df_Y

model = grid_search.best_estimator_
name = "Decision Tree (Hyperparameter Tuned)"

acc_score = get_acc_score_kcv(tmp_X, tmp_Y, model)

print(f"Accuracy Score: {round(acc_score, 2)}%")

model_dict[name] = create_ModelObj(model, name, tmp_X, tmp_Y, index_dict[tmp_Y.name])

Accuracy Score: 45.74%
```

Figure 5.4.10 Decision Tree HyperTuned

Random Forest

Random Forest (Bagging Method)

```
tmp_X = df_X
tmp_Y = df_Y

model = RandomForestClassifier()
name = "Random Forest"

acc_score = get_acc_score_kcv(tmp_X, tmp_Y, model)

print(f"Accuracy Score: {round(acc_score, 2)}%")

model_dict[name] = create_ModelObj(model, name, tmp_X, tmp_Y, index_dict[tmp_Y.name])
```

```
Accuracy Score: 53.34%
```

Figure 5.4.11 Random Forest

Random Forest HyperTuning

Random Forest (Hyperparameter Tuned)

```
params = {
    'max_depth': [2, 3, 5, 10, 20],
    'min_samples_leaf': [5, 10, 20, 50, 100],
    "n_estimators": [5, 10, 20, 50, 100],
    "bootstrap": [True, False]
}
```

```
model = RandomForestClassifier()
grid_search = GridSearchCV(estimator=model, param_grid=params, cv = 5, n_jobs=-1, verbose=1, scoring = "accuracy")

grid_search.fit(tmp_X, tmp_Y)
```

```
Fitting 5 folds for each of 250 candidates, totalling 1250 fits

GridSearchCV(cv=5, estimator=RandomForestClassifier(), n_jobs=-1,
             param_grid={'bootstrap': [True, False],
                         'max_depth': [2, 3, 5, 10, 20],
                         'min_samples_leaf': [5, 10, 20, 50, 100],
                         'n_estimators': [5, 10, 20, 50, 100]},
             scoring='accuracy', verbose=1)
```

```
grid_search.best_estimator_
```

```
RandomForestClassifier(bootstrap=False, max_depth=5, min_samples_leaf=5,
                       n_estimators=20)
```

Figure 5.4.12 Random Forest HyperTuned

XGBoost

XGBoost (Boosting Method)

```
tmp_X = df_X
tmp_Y = df_Y

model = XGBClassifier(eval_metric='error', use_label_encoder=False)
name = "XGBoost"

acc_score = get_acc_score_kcv(tmp_X, tmp_Y, model)

print(f"Accuracy Score: {round(acc_score, 2)}%")

model_dict[name] = create_ModelObj(model, name, tmp_X, tmp_Y, index_dict[tmp_Y.name])
```

```
Accuracy Score: 51.77%
```

Figure 5.4.13 XGBoost

Decision Tree Visualization

Decision Tree Visualization

```
viz = dtreeviz(
    model_dict[name].model,
    tmp_X,
    tmp_Y,
    feature_names = tmp_X.columns,
    class_names = index_dict[tmp_Y.name],
    fancy = False)

viz
```

Figure 5.4.14 Generate Decision Tree Visualization

Expected Result:

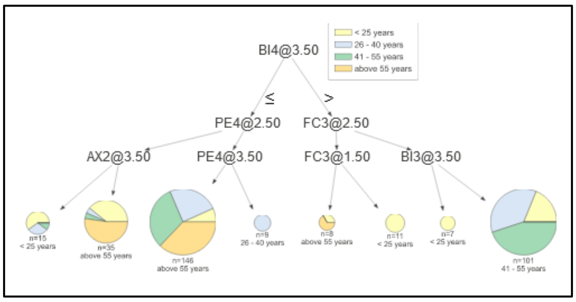


Figure 5.4.15 Decision Tree Visualiztion

Precision Graph

Precision

```
clf_report_arr = [(name, model_dict[name].clf_report) for name in model_arr]
tmp_df = get_df_type(clf_report_arr, "Precision")
pfr_graph(tmp_df, "Model", "Score", "Precison Comparison")
```

Figure 5.4.16 Generate Precision Bar Graph

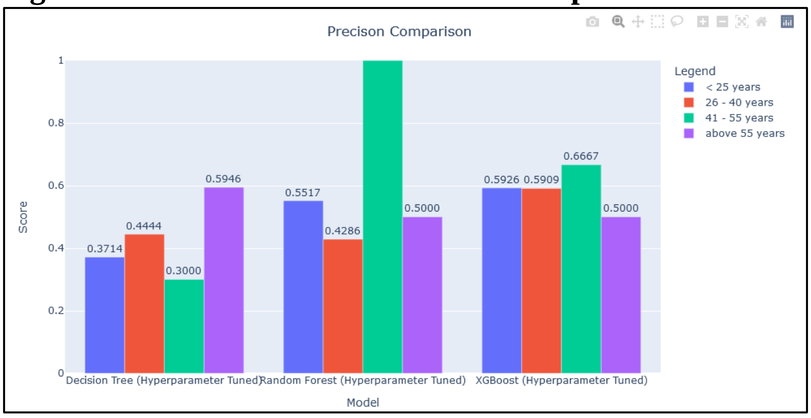


Figure 5.4.17 Precision Bar Graph

Recall Graph

Recall

```
clf_report_arr = [(name, model_dict[name].clf_report) for name in model_arr]
tmp_df = get_df_type(clf_report_arr, "Recall")
pfr_graph(tmp_df, "Model", "Score", "Recall Comparison")
```

Figure 5.4.18 Generate Recall Graph

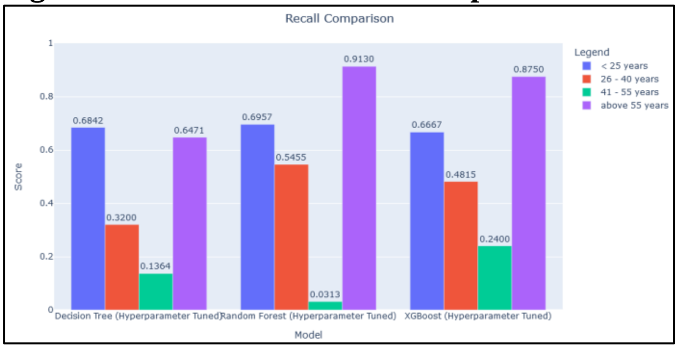


Figure 5.4.19 Recall Bar Graph

F1-Score Graph

F1-Score

```
clf_report_arr = [(name, model_dict[name].clf_report) for name in model_arr]
tmp_df = get_df_type(clf_report_arr, "F1-Score")
pfr_graph(tmp_df, "Model", "Score", "F1-Score Comparison")
```

Figure 5.4.20 Generate F1-Score Graph

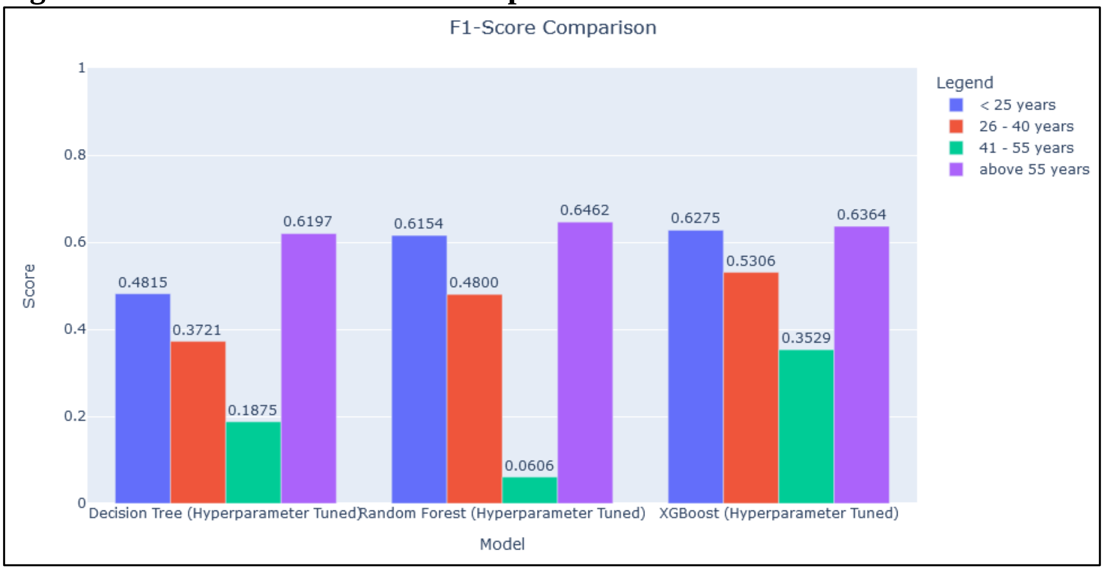


Figure 5.4.21 F1-Score Bar Graph

GUI Steps

Launch our GUI

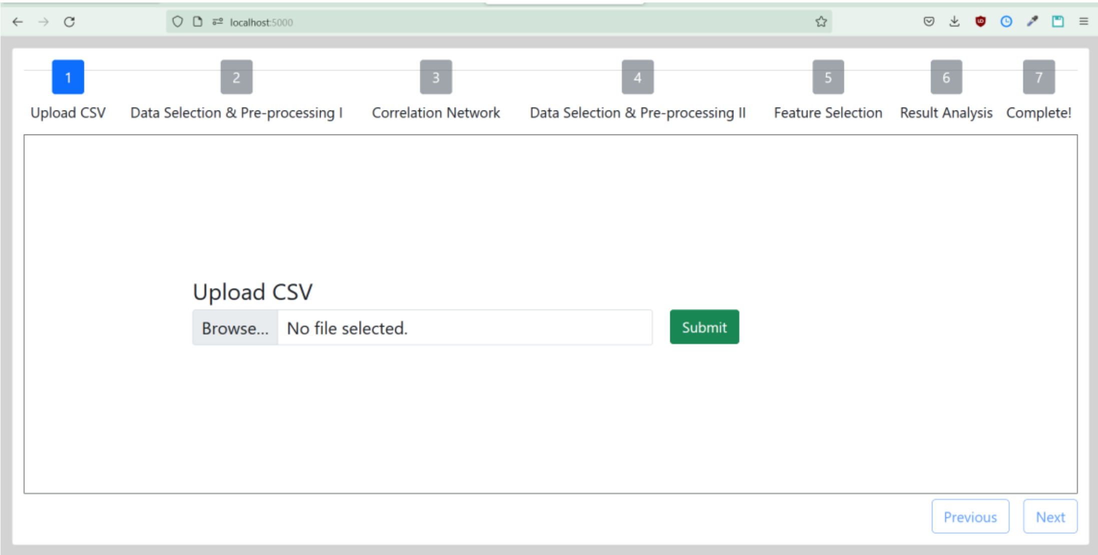


Figure 5.4.22 GUI Homepage

Upload any Dataset

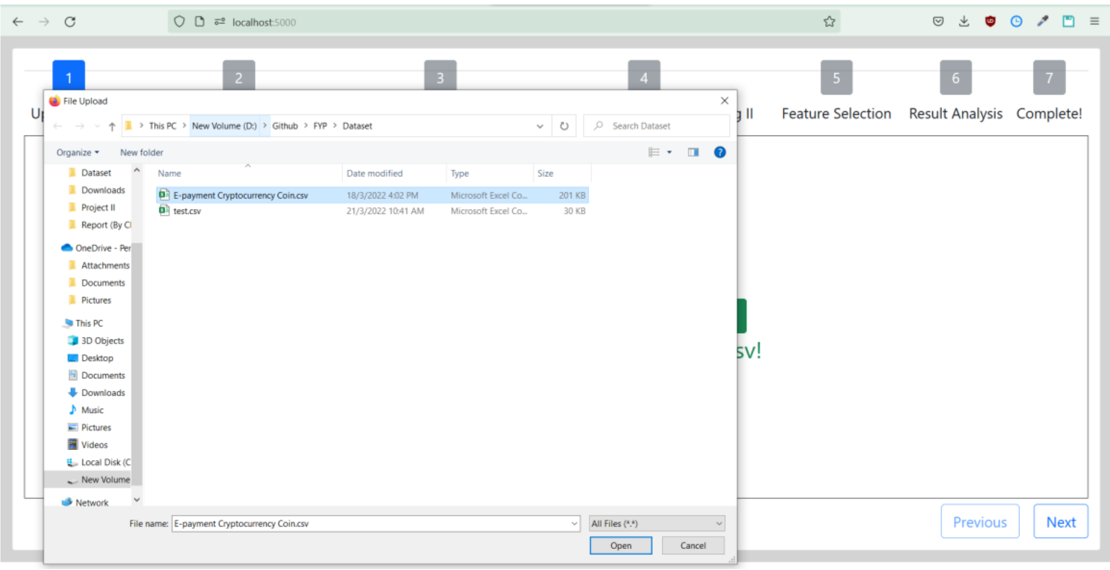


Figure 5.4.23 Upload Dataset

Select the Features to Preprocess and train test model

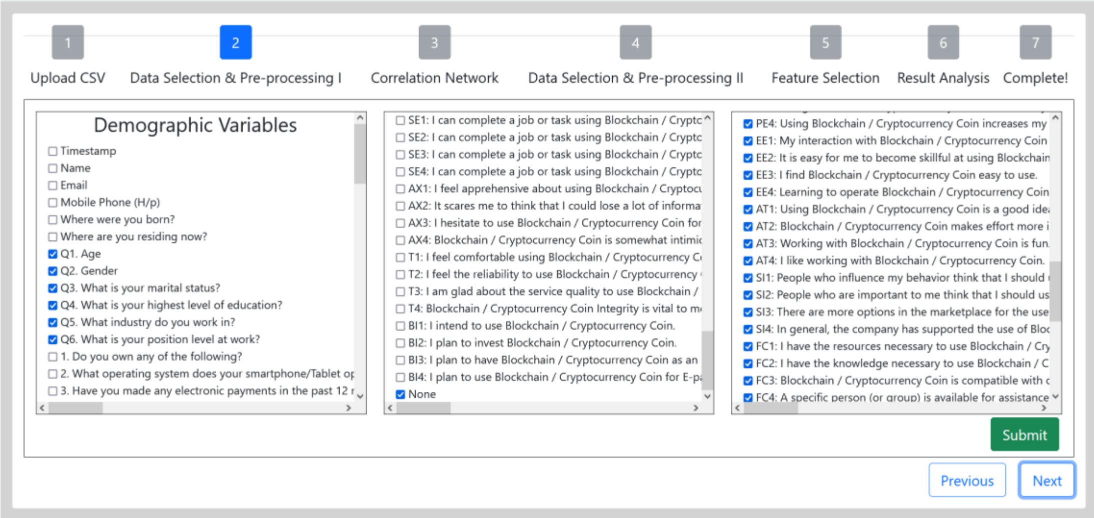


Figure 5.4.24 Select Features

Generate Correlation Network graph

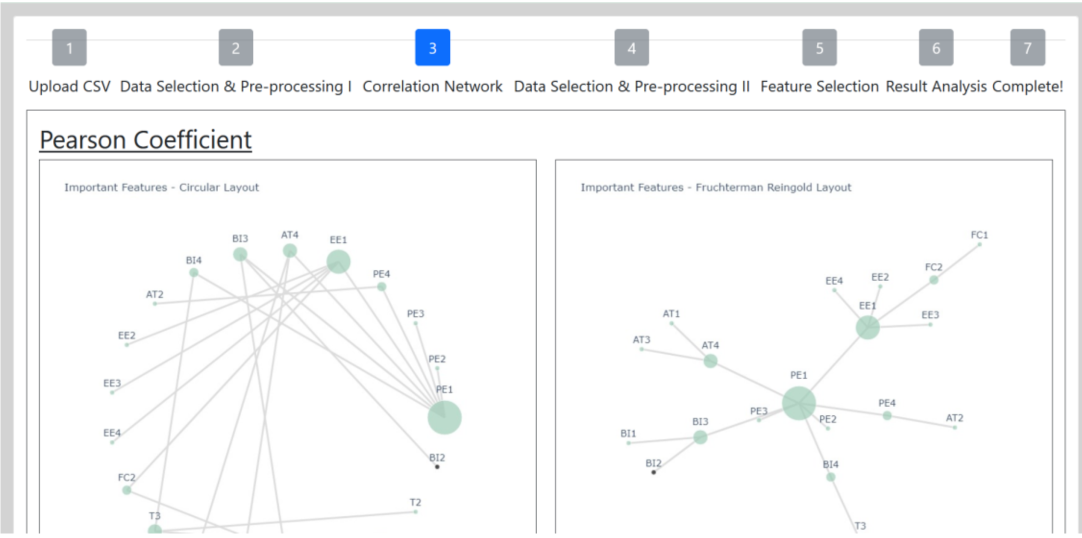


Figure 5.4.25 Generate Corr Network graph

Filter Targeted Variable

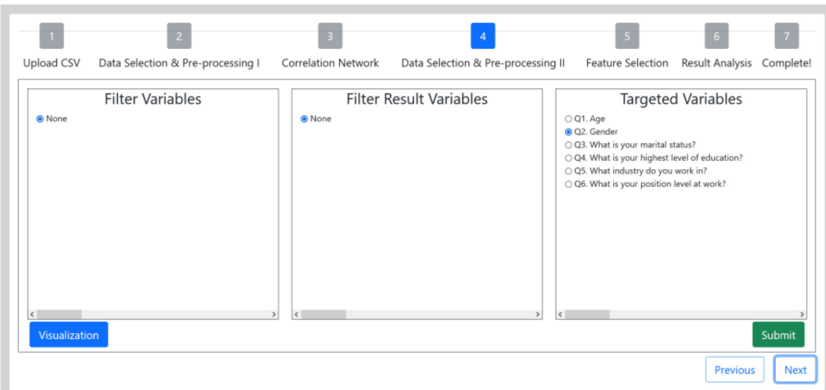


Figure 5.4.26 Select Target Variable

Generate Intersected, Non-Intersected and Union Features from Correlation Feature Selection

Intersected Feature Set

- EE3: I find Blockchain / Cryptocurrency Coin easy to use.
- EE1: My interaction with Blockchain / Cryptocurrency Coin is clear and understandable.
- FC3: Blockchain / Cryptocurrency Coin is compatible with other systems I use.

	EE3	EE1	FC3
EE3	1.00		
EE1	0.79	1.00	
FC3	0.64	0.60	1.00

Non-Intersected Feature Set

Union Feature Set

- EE1: My interaction with Blockchain / Cryptocurrency Coin is clear and understandable.
- FC3: Blockchain / Cryptocurrency Coin is compatible with other systems I use.
- EE3: I find Blockchain / Cryptocurrency Coin easy to use.

	EE1	FC3	EE3
EE1	1.00		
FC3	0.60	1.00	
EE3	0.79	0.64	1.00

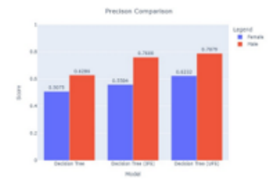
Figure 5.4.27 Generate Intersected Features

Result Analysis

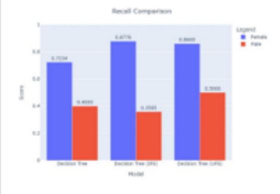
Model Result

	Precision	Recall	F1-Score	Accuracy Score	MCC	Time Taken
Decision Tree	0.57	0.55	0.54	68.04	0.13	0.05
Decision Tree (IFS)	0.66	0.61	0.58	67.16	0.27	0.05
Decision Tree (UFS)	0.71	0.68	0.67	60.47	0.38	0.03

Precision Graph



Recall Graph



F1-Score Graph

Accuracy Score Graph

Figure 5.4.28 Result Analysis

6 Chapter 6: System Evaluation and Discussion

6.1 System Testing

For this section, we aim to do analysis of attributes of customer intention to adopt E-payment using Decision Tree Modelling. These models are built using Jupyter notebook as our IDE and using Python language. We specifically select Python to build these prediction models because python contains Scikit-library, which contains all the necessary functions to do machine learning modelling.

The selected machine learning classifiers for this project are all tree-based algorithms. The reason why we only use tree-based algorithm is because Decision tree can avoid multicollinearity issues where the attributes are highly correlated. If the attributes are collinear, it is very likely to produce highly biased and inaccurate result. We select random Forest and XGBoost algorithm because random forest is a bagging method, whereas XGBoost is a boosting method.

The dataset were preprocessed and analyzed according to the above steps. Using our CFS Subset based evaluator, we split our attributes into intersected features, non-intersected features and union features. Thus, we will be building 4 classifier models and evaluating the result. To ensure the accuracy is valid and all data are used in building the model, we will be using 10-fold cross validation as well. The dataset is split 80-20, 80% as the training set and 20% for the testing set.

To evaluate the performance of our model, we will be using confusion matrix and classification report to compare the scores of each model. Specifically, we will be evaluating the model using accuracy, Matthew correlation coefficient, precision, recall and f1-score.

6.2 Performance of Tree-Based Classification Algorithm

We first hypertuned each model to get their best performing parameter. To conduct hypertuning, each model have been built using 5-fold cross validation with GridSearchCV. Each model have their own hyperparamters setting tuned to obtain the best parameters. The hypertuned parameters for each model are shown below.

Decision Tree Hyperparameter Tuning

Table 6.2-1 Decision Tree Hyperparameters

	Parameters	Value
Decision Tree Model	Max_Depth	10
	Min_Sample_Leaf	5
	Criterion	Entropy

Random Forest Hyperparameter Tuning

Table 6.2-2 Random Forest Hyperparameters

	Parameters	Value
Random Forest Model	Max_Depth	5
	Min_Samples_Leaf	5
	N_estimators	20
	Bootstrap	False

XGBoost Hyperparameter Tuning

Table 6.2-3 XGBoost Hyperparameters

	Parameter	Value
XGBoost Model	Max_Depth	5
	Learning_Rate	0.1
	N_Estimators	20
	Colsample_ByTree	0.3

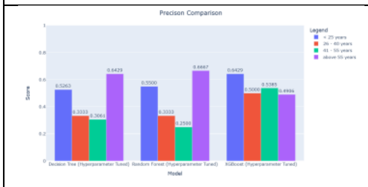
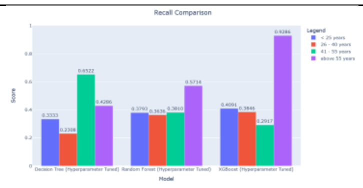
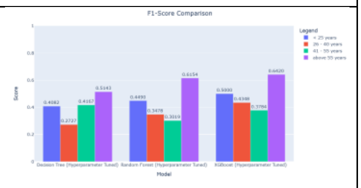
Overall Results

Table 6.2-4 Overall Results among Three Model

	Precision	**Recall**	**F1-Score**	**Accuracy Score**	**MCC**
Decision Tree (Hyperparameter Tuned)	0.449970	0.40	0.397191	40.579710	0.217765
Random Forest (Hyperparameter Tuned)	0.472000	0.43	0.442430	52.590580	0.248488
XGBoost (Hyperparameter Tuned)	0.538018	0.52	0.493607	47.318841	0.371556

From an overall perspective, we can observe that among the three models that are hypertuned, the model with the highest accuracy is Random Forest model. From here, it is indicated that the best performing model is via XGBoost algorithm. This is observed from the Matthew Correlation Coefficient score, as the scoring metric is the highest among the 3 models. Furthermore, the F1-Score of XGBoost model is the best performing one. Although Random Forest Model did not have the highest MCC Score, it has the highest precision score. This indicates that our Random forest model is great at classifying targeted class correctly. Random Forest model also has the best accuracy, however its MCC score is too low compared to our XGBoost model. Overall, the best performing tree-based algorithm is XGBoost model.

Table 6.2-5 Precision, Recall, F1-Score of Three Algorithm

Precision	Recall	F1-Score
		

From the graphs above, particularly the recall graph, we can observe that in terms of Predicting class group "< 55 years" as incorrect class, XGBoost have an extremely high recall value. For the F1-Score metric, we can see that the bar graph for XGBoost is distributed much more fairly. Although Random Forest model has the highest accuracy, we can see that the results are not consistent when we compare precision, recall and F1-Score. It has extremely low precision for class group "26 – 40 years" and "41 – 55 years". XGBoost has a much more all rounded distribution therefore XGBoost is considered to be the best performing model among these 3 tree-based model.

6.3 Differences between Age Group

To conduct this experiment, we take "Age" as our target group and "Have you ever purchased anything using the E-payment mode?" as our filter variable. After we have filtered our dataset, we use CFS to generate the intersected features, non-intersected features and union features. These features are selected using Pearson's Correlation method and Spearman's Correlation method using a threshold of 0.7.

Features selected using CFS

Table 6.3-1 Features for Age Selected using CFS

<table>
<tr><td>Union Features</td><td>Intersected Features</td><td>Non-intersected features</td></tr>
<tr><td>'AT3', 'AX2', 'SI2'</td><td>'AT3', 'SI2'</td><td>'AX2'</td></tr>
<tr>
<td><table><tr><th></th><th>AX2</th><th>AT3</th><th>SI2</th></tr><tr><th>AX2</th><td>1.000000</td><td>NaN</td><td>NaN</td></tr><tr><th>AT3</th><td>0.026229</td><td>1.000000</td><td>NaN</td></tr><tr><th>SI2</th><td>0.090854</td><td>0.522027</td><td>1.0</td></tr></table></td>
<td><table><tr><th></th><th>AT3</th><th>SI2</th></tr><tr><th>AT3</th><td>1.000000</td><td>NaN</td></tr><tr><th>SI2</th><td>0.522027</td><td>1.0</td></tr></table></td>
<td><table><tr><th></th><th>AX2</th></tr><tr><th>AX2</th><td>1.0</td></tr></table></td>
</tr>
</table>

Overall Results

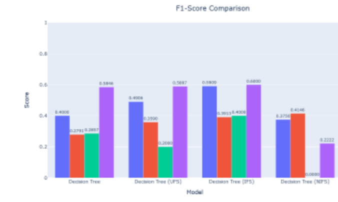
Table 6.3-2 Overall result for Age Group Features Model

	Precision	**Recall**	**F1-Score**	**Accuracy Score**	**MCC**
Decision Tree	0.398788	0.41	0.378822	43.913043	0.234201
Decision Tree (UFS)	0.505738	0.46	0.432365	40.942029	0.297747
Decision Tree (IFS)	0.549258	0.51	0.495905	41.503623	0.362424
Decision Tree (NIFS)	0.264606	0.31	0.254831	32.355072	0.113628

From an overall perspective, we can observe that among the three models that uses features selected based on CFS, the model with the highest accuracy is the Decision Tree model that uses intersected features. This proves that our algorithm is working as intended, as we can observe the difference in accuracy between features non-intersected and features intersected. Using CFS, we are able to gain a performance that is comparable to Decision Tree model that uses all features.

For age group, we can observe that all 4 models perform mediocrely. The accuracies for all 4 models do not exceed 45%. We can observe that particularly for Decision Tree that uses non-intersected features, it has very low precision and recall score, indicating that it has a hard time differentiating users in their age group. All models have low precision and low recall. This indicates that for age, it may not cause much difference as everyone had the same amount of exposure to age group.

Table 6.3-3 Precision, Recall, F1-Score for Age Group

Precision	Recall	F1-Score
		

From the graphs above, particularly the recall graph, we can observe that it is much easier for the models to differentiate users that are not “above 55 years old”. This indicates that user above 55 years old may have different factors affecting their intention to adopt e-payment. Using Decision Tree model train on Intersected features as our base model, we can observe that it has a higher precision for user group “< 25 years” and user group “41 – 55 years”. This means that for young adults and adults in retirement, Acceptance towards technology and Social influence affects their intention to adopt e-payment. Overall, our decision tree model trained using Intersected features have the best precision and recall score.

To gain a better understanding on what factors affect age groups, we may consider using Decision Tree visualization. Using Weka as our machine learning tool, we plot the decision tree visualization graph.

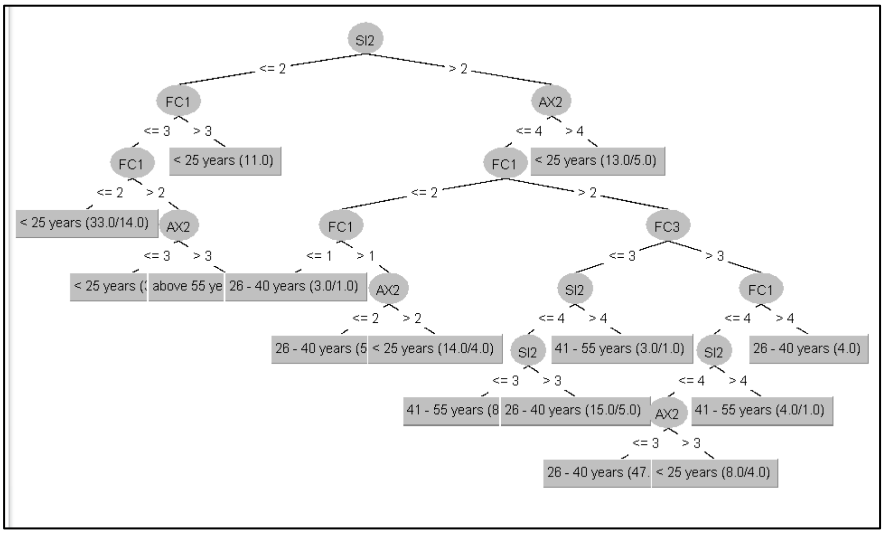


Figure 6.3.1 Decision Tree Visualization for Age

Table 6.3-4 Decision Tree Rules for Age Graph

Model	Number of Rules					Accuracy
J48 Classifier	< 25 years	26 – 40 years	41 – 55 years	Above 55 years	Total	45.23%
	6	5	3	1	15	

From the results above, we can observe that the Decision tree from Weka Modelling identifies more user "< 25 years" compared to other age group. From the tree above, we can have several inference:

- The factors that affect user intention to adopt e-payment, particularly age group, are FC1, SI2, AX2 and FC3. These attributes are Resource necessary to use Bitcoin, Afraid of Bitcoin, Blockchain is compatible with other system and people who are important encourage users.
- For Users who don't care about resources to use bitcoin, they are more likely to be the < 25 years.
- Users above 55 years old have fear that all their e-payment will be lost if they press the wrong key.
- Users among 26 – 40 years, and users among 41 – 55 years are dependent on whether someone important to them encourages them to use Bitcoin.

6.4 Differences between Gender Group

To conduct this experiment, we take "Gender" as our target group and "Have you ever purchased anything using the E-payment mode?" as our filter variable. After we have filtered our dataset, we use CFS to generate the intersected features, non-intersected features and union features. These features are selected using Pearson's Correlation method and Spearman's Correlation method using a threshold of 0.7.

Features selected using CFS

Table 6.4-1 Feature Set for Gender Group

Union Features	Intersected Features	Non-intersected Features
'AX1', 'EE1', 'EE3', 'FC2', 'FC3', 'SE1'	'EE1', 'FC3', 'SE1', 'AX1'	'FC2', 'EE3'

	EE3	AX1	EE1	SE1	FC3	FC2
EE3	1.000000	NaN	NaN	NaN	NaN	NaN
AX1	0.187271	1.000000	NaN	NaN	NaN	NaN
EE1	0.793750	0.193314	1.000000	NaN	NaN	NaN
SE1	0.619711	0.140300	0.671106	1.000000	NaN	NaN
FC3	0.625142	0.165338	0.583730	0.536561	1.000000	NaN
FC2	0.705653	0.113573	0.812177	0.716698	0.673631	1.0

	EE1	FC3	SE1	AX1
EE1	1.000000	NaN	NaN	NaN
FC3	0.583730	1.000000	NaN	NaN
SE1	0.671106	0.536561	1.0000	NaN
AX1	0.193314	0.165338	0.1403	1.0

	FC2	EE3
FC2	1.000000	NaN
EE3	0.705653	1.0

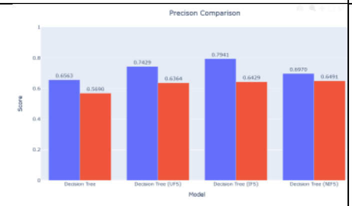
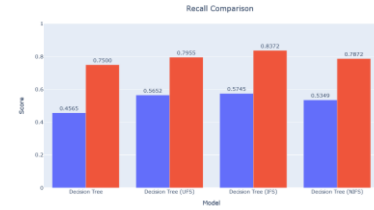
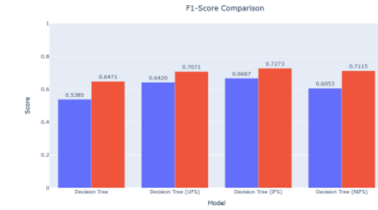
Overall Result

Table 6.4-2 Overall Result for Gender Group

	Precision	Recall	F1-Score	Accuracy Score	MCC
Decision Tree	0.613578	0.600000	0.591554	58.071429	0.215666
Decision Tree (UFS)	0.690794	0.677778	0.673800	62.833333	0.369830
Decision Tree (IFS)	0.721849	0.700000	0.695623	61.595238	0.424138
Decision Tree (NIFS)	0.671983	0.666667	0.660762	55.333333	0.333890

Among the 4 models, we can observe that the model with the highest accuracy is Decision Tree Model trained using Union Feature set. It performs even better than Decision Tree that is trained on all features, proving that our CFS method is capable of extracting meaningful features. However, if we use Matthew Correlation Coefficient as a comparison metric, we can see that the better model is still Decision Tree trained using intersected feature set. Compared to age group, the 4 models here have better accuracy, achieving an accuracy score of over 60%, and better precision, recall score performance. The precision and recall score of Decision Tree trained on intersected feature set are above 70%, this indicates that the model is capable of sorting user based on their gender. It also provides the hypotheses that the features that affects user intention based on gender are much more apparent.

Table 6.4-3 Precision, Recall, F1-Score for Gender Model

Precision	Recall	F1-Score
		

Based on the precision, recall and f1-score above, we can see that for decision tree models for Gender Group are much better at identifying genders between respondent. All models have a higher score for Male gender in the precision graph. This indicates that they are much better at predicting male gender correctly. All models have higher score for female gender in the recall graph. This indicates that they are much better at classifying incorrect female gender respondent. In the f1-score graph, we can observe that Decision Tree model trained using all features have worse f1-score, whereas the decision tree model trained using intersected features have better f1-score. This indicates that the best model among the 4 models for classification is the decision tree model trained using Intersected features. Although its accuracy is not the highest, it has a better matthew correlation coefficient and f1-score.

Visualization of Decision Tree Graph using Weka Tool

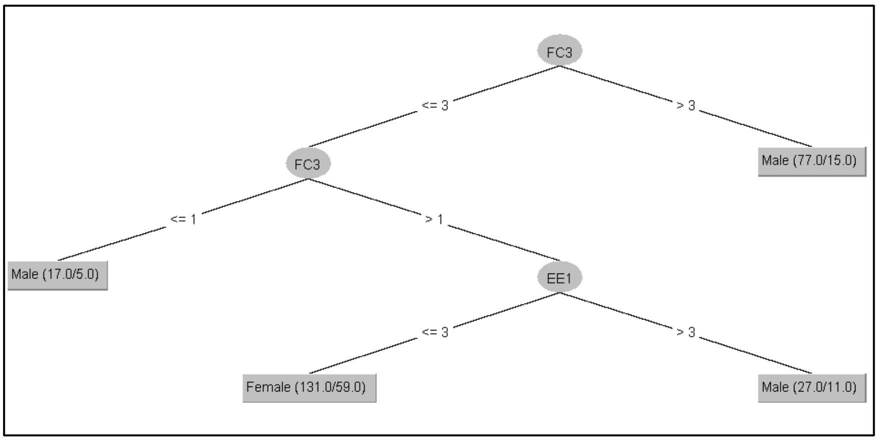


Figure 6.4.1 Decision Tree Visualization for Gender

Table 6.4-4 Decision Tree Rules For Gender

Model	Number of Rules			Accuracy
J48 Classifier	**Male**	**Female**	**Total**	**58.33%**
	3	**1**	**4**	

From the above tree graph, we can conclude that the decision tree model trained using J48 classifier in Weka Tool identifies Male respondent more than female Respondent. From the tree above, we have the following inference:

- The factors that affect Users, based according to Gender, intention to adopt E-payment are FC3, Blockchain is compatible with other systems I use, and EE1, User Interaction with Blockchain is Clear.
- Male Users are more likely to adopt E-Payment if it is compatible with other system they use. This is observed from the graph above, as users whom score 3 or higher on FC3 are more likely to be Male Respondents.
- Female users are less Likely to adopt E-Payment if the interaction with Blockchain isn't clear. This is observed from the graph above, as users whom score 3 or lower on EE1 are Female respondents.

6.5 Project Challenges

Among our project challenges, one of the major issues is that we have a very imbalanced dataset due to lack of non e-payment users. As you can see among Age Graph and Gender graph, majority of our respondents are E-payment users. Due to the lack of non-e-payment users, some of the analysis may be a bit biased. To circumvent this issue, we use SMOTE imbalanced to balance out the non e-payment users. Due to SMOTE being synthetic data, the result analysis may not entirely be accurate. To improve the result analysis, we may consider getting more non e-payment users in the future.

Secondly, it was hard to do decision tree pathway analysis for our factors. This project is quite a novel approach. It was hard to find existing works that extensively research how decision tree select each factor and to hypothesize the relationship between the factor items. Due to lack of existing work, it is hard to evaluate our analysis and hypothesis. We hope that this project will be used as a base model for analysing how decision tree makes its selection process.

Lastly, another issue with our project is majority of our respondents are university students or university staff. Take "Work industry" for example. We can observe that "Education" industry has an alarmingly high number of respondents. There is a lack of distribution of survey outside of university campus. This project is not diverse enough, especially the fact that, if we categorize the respondents by education group, majority would be within IT industry. The results of this project might be more suited for users in IT industry rather than all industry. To solve this issue, we should branch out our survey outside of university campus and distribute to the real world.

6.6 Objective Evaluation

Among our three objectives, we have securely achieved all of the proposed objectives of this project. Our first objective, which was to create a simplified feature selection process was achieved via our correlation-based feature selection algorithm. Our CFS algorithm works as intended from our experimentation result above. We are able to select meaningful features and make hypothesis and conclusion. This correlation-based feature selection is a novel approach to extracting meaningful features from UTAUT model, as existing UTAUT model works require extensive statistical knowledge and background. Our correlation-based feature selection algorithm also helps to reduce the feature subset.

Our second objective, which was the analysis of relationship between UTAUT factor items, was also achieved. This was achieved from our interpretation based on WEKA results. We construct the decision tree, count the number of decision tree rules, analyse which factor items were selected and form our hypothesis based on the concluding class group. We can analyse efficiently and come to same conclusion with some of the existing work I have previewed. Decision tree pathway analysis is quite a novel approach in terms of approaching UTAUT model problems. We hope that this method will get popularized in theoretical framework field.

Our last objective to create a generalize machine learning model was also achieved from our GUI implementation. We have tested two different dataset, one which is UTAUT model, another which is TAM model. Both datasets are able to have meaningful features extracted from our CFS model. We are also able to construct decision tree visualization and use precision, recall and f1-score evaluation. Therefore, this machine learning pipeline works not just for e-payment adoption of UTAUT model, it also works for other digital adoption via any theoretical framework.

7 Chapter 7: Conclusion and Recommendation

7.1 Conclusion

In conclusion, our machine learning model can identify underlying features among highly correlated user behaviours. Using our self-implemented Correlation Feature Selection, it can extract meaningful information and model a relationship network graph among the highly correlated user behaviours. Furthermore, our machine learning model can identify important features that influence customer to adopt e-payment system. This is extremely beneficial for app developers, NFC developers and market shareholders who aim to promote e-payment adoption in Malaysia.

Our machine learning model uses a correlation-based feature selection and decision tree modelling to classify the significant user behaviours. This is quite novel as previously to identify significant user behaviours, researchers have to perform a lot of statistical analysis techniques such as Component Factor Analysis or Cronbach reliability alpha to understand the underlying factors. The benefits of our machine learning model is that it is capable of learning from the dataset and identify the set of significant user behaviour. This helps to save cost and time.

Furthermore, to simplify the analysis purpose, we have developed a GUI using Node JS and express library to build a web server. The GUI connects to our Python files and does machine learning based on the features selected by the researcher. The researcher just have to upload their own dataset that is part of a theoretical framework. The machine learning model is a generalized model that is suitable for any theoretical framework task.

7.2 Recommendation

Increase the number of Non e-payment User Respondents

One of the main issue with our project is the lack of Non e-payment users. From the distribution bar graph of our respondents, we can see that majority of our user are active e-payment users, whom have used e-payment in the past 12 months. Due to the lack of e-payment users, our results are a bit biased towards users whom are e-payment users. Increasing the number of non e-payment users would allow more interesting analysis.

Use Other Models for Analyzing Respondent Intention

To analyze the intention of respondents, we use Decision Tree pathway analysis. From our model comparison, we can see other models that have better classification accuracy compared to decision tree model. In future work, we can experiment with other predictive model and do enhanced analysis of factors.

Use Larger Dataset

Our dataset only has a number of 286 respondents with 64 features. There is a lack of diversity among the respondents as majority of them are either University Student or University Lecturers. An even in-depth analysis would realize that majority of the students are from the IT industry. Widening the scope of the respondents would allow more interesting pattern of user intention to adopt e-payment system to resurface and be discovered.

8 References

[1] Oppotus, "Navigating the E-Wallet Landscape of Malaysia," 22 July 2019. [Online]. Available: https://www.oppotus.com/e-wallet-landscape-malaysia/. [Accessed 12 April 2021].

[2] Venkatesh, Morris, Davis, and Davis, “User Acceptance of Information Technology: Toward a Unified View,” MIS Quarterly, vol. 27, no. 3. JSTOR, p. 425, 2003. doi: 10.2307/30036540.

[3] Rajendran, Ambigai, Shenoy, Sandeep, Rao, Guru, Shetty, Khyati, Understanding the Customer's Attitude Towards Acceptance and Usage of Online Banking Services Using Utaut Model, 2017

[4] R. Tadayon Nabavi, Bandura's Social Learning Theory & Social Cognitive Learning Theory, 2012.

[5] S. A. Nikou and A. A. Economides, “Mobile-Based Assessment: Integrating acceptance and motivational factors into a combined model of Self-Determination Theory and Technology Acceptance,” Computers in Human Behavior, vol. 68. Elsevier BV, pp. 83–95, Mar. 2017. doi: 10.1016/j.chb.2016.11.020.

[6] K. Mac Callum, L. Jeffrey, and Kinshuk, “Comparing the role of ICT literacy and anxiety in the adoption of mobile learning,” Computers in Human Behavior, vol. 39. Elsevier BV, pp. 8–19, Oct. 2014. doi: 10.1016/j.chb.2014.05.024.

[7] D. Gefen, I. Benbasat and P. Pavlou, "A Research Agenda for Trust in Online Environments", Journal of Management Information Systems, vol. 24, no. 4, pp. 275-286, 2008. Available: 10.2753/mis0742-1222240411.

[8] H. Khazaei, “Integrating Cognitive Antecedents to UTAUT Model to Explain Adoption of Blockchain Technology Among Malaysian SMEs,” JOIV :

International Journal on Informatics Visualization, vol. 4, no. 2. Politeknik Negeri Padang, Apr. 27, 2020. doi: 10.30630/joiv.4.2.362.

[9] W. Widayat, I. Masudin, and N. R. Satiti, “E-Money Payment: Customers’ Adopting Factors and the Implication for Open Innovation,” Journal of Open Innovation: Technology, Market, and Complexity, vol. 6, no. 3. MDPI AG, p. 57, Jul. 29, 2020. doi: 10.3390/joitmc6030057.

[10] M.-F. Wei, Y.-H. Luh, Y.-H. Huang, and Y.-C. Chang, “Young Generation’s Mobile Payment Adoption Behavior: Analysis Based on an Extended UTAUT Model,” Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 1. MDPI AG, pp. 1–20, Jan. 04, 2021. doi: 10.3390/jtaer16010001

[11] J.-M. Lee, “DETERMINANTS OF MOBILE PAYMENT USAGE AND THE MODERATING EFFECT OF GENDER: EXTENDING THE UTAUT MODEL WITH PRIVACY RISK,” International Journal of Electronic Commerce Studies, vol. 10, no. 1. Academy of Taiwan Information Systems Research, pp. 43–64, Jun. 30, 2019. doi: 10.7903/ijecs.1644.

[12] B. R Patel and K. K Rana, "A Survey on Decision Tree Algorithms of Classification in Data Mining", International Journal of Science and Research (IJSR), vol. 5, no. 4, pp. 2094-2097, 2016. Available: 10.21275/v5i4.nov162954.

[13] K. Gross, "Tree-Based Models: How They Work (In Plain English!)," 6 April 2020. [Online]. Available: https://blog.dataiku.com/tree-based-models-how-they-work-in-plain-english. [Accessed 12 April 2021].

[14] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” arXiv, 2016, doi: 10.48550/ARXIV.1603.02754.

[15] M. A. Hall, “Correlation-based feature selection for machine learning,” 1999.

9 APPENDIX A

9.1 A-1 Weekly Report

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 1**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE Research Filter based Approach and Wrapper Based Approach for Feature Selection of Dataset Research Correlation Based Feature Selection Algorithm and Source Code in Python
2. WORK TO BE DONE Produce a working implementation of CFS algorithm in Python using Priority Queue and Best First Search
3. PROBLEMS ENCOUNTERED Minor issues with getting least cost subset from Priority Queue
4. SELF EVALUATION OF THE PROGRESS Will Do Research on features selected using Correlation Based Feature Selection

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 2**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE Implemented Correlation Based Feature Selection in Python Research on existing machine learning classifiers – Random Forest, Decision Tree and XGBoost
2. WORK TO BE DONE Conduct Research on Decision Tree Visualization Conduct Research on Correlation Table Visualization
3. PROBLEMS ENCOUNTERED Unable to properly explain feature selection results from Decision Tree
4. SELF EVALUATION OF THE PROGRESS Try to produce working graph of Correlation Graph and Decision Tree Visualization

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 3**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Research on Correlation Network Graph

Produced Working Visualization of Correlation network Graph

Research on DtreeViz

2. WORK TO BE DONE

Produce working Decision Tree Graph

Separate Feature Selection into Intersected Features, Non-Intersected Features and Union Features

3. PROBLEMS ENCOUNTERED

NIL

4. SELF EVALUATION OF THE PROGRESS

Compare Results from Intersected Features, Non-Intersected Features and Union Features

Finalize Machine Learning Pipeline Model by Next week

Tong

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 4**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Use weka to visualize Decision Tree

Use DTreeViz to visualize Decision Tree

Conduct Feature Selection Analysis from Decision Tree Graph

2. WORK TO BE DONE

Finalize Machine Learning Pipeline Model

Propose GUI program to visualize Machine Learning Results

3. PROBLEMS ENCOUNTERED

NIL

4. SELF EVALUATION OF THE PROGRESS

Completed Decision Tree Analysis, May start to write report on Chapter 4 and Chapter 5

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 5**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Use SMOTE Imbalance to balance classes for unbiased results

Research on Precision Recall Graph and ROC Curve for better understanding of results

Use Precision, Recall and F1-Score to visualize model performance

Finalized Machine Learning Pipeline model

2. WORK TO BE DONE

Start designing GUI Program

Separate python chunks in Jupyter notebook into individual python program

3. PROBLEMS ENCOUNTERED

NIL

4. SELF EVALUATION OF THE PROGRESS

Start working on GUI to do a proper prototype

Tong

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 6**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Design Front Page using HTML

Completed Web server using Node JS

2. WORK TO BE DONE

Use Jquery for AJAX to send post request to Node JS, to execute python script to get results

3. PROBLEMS ENCOUNTERED

AJAX is mostly async request. Due to this, the function will complete before results are gained from python script. Need to find a way to execute python script synchronously to return results to html.

4. SELF EVALUATION OF THE PROGRESS

Minor problems fixed. Hope to fix it by next week.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 7**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Completed gen_corr_image.py

In our html, users are able to select 3 groups of features, that is filter feature, targeted feature and utaut feature. Able to pass these features to Server.

Server able to generate correlated image using features selected by User

2. WORK TO BE DONE

Completed Feature Selection.py for filtering against selected feature and producing new targeted dataset for train and test model.

Generate Visualization bar graph for expected targeted feature against selected feature

3. PROBLEMS ENCOUNTERED

Hard to debug using Node JS child spawn process. Need to read more online documentation.

4. SELF EVALUATION OF THE PROGRESS

Have not converted all python cells in jupyter notebook into individual python program. Hope to catch up by next week.

Tong

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 8**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE

Completed Feature_selection.py

Now, server able to generate final dataset for training and testing model.

Able to load dataframe output onto HTML

2. WORK TO BE DONE

Complete Model train python script

Model train python script have to generate precision, recall and f1-score graph as jpeg

Generate Decision Tree Visualization

Train and test using Decision Tree model

3. PROBLEMS ENCOUNTERED

NIL

4. SELF EVALUATION OF THE PROGRESS

NIL

Tong

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 9**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE Completed Model Train py Able to visualize precision, recall, f1-score graph Able to visualize Decision Tree Able to store and visualize results from Decision Tree Model
2. WORK TO BE DONE Start work on Chapter 4 and 5 Write Experimentation result to report
3. PROBLEMS ENCOUNTERED NIL
4. SELF EVALUATION OF THE PROGRESS NIL

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(~~Project I~~ / Project II)

Trimester, Year: Y3S3	**Study week no.: 10**
Student Name & ID: Tan Xi En 1904098	
Supervisor: Dr. Tong Dong Ling	
Project Title: Correlation Model In the adoption of E-payment Services: A Machine Learning Approach	

1. WORK DONE Redesign Block Diagram for Chapter 3 Start Writing Evaluation and Result Analysis for Chapter 5
2. WORK TO BE DONE Change Literature Review Citation to IEEE Format, instead of Harvard format
3. PROBLEMS ENCOUNTERED NIL
4. SELF EVALUATION OF THE PROGRESS NIL

Supervisor's signature

Student's signature

9.2 A-2 Poster


Name: Tan Xi En Project Supervisor: Dr. Tong Dong Ling
Correlation Model In the Adoption of E-payment Services: A Machine Learning Approach
Introduction
56%
In the recent Covid Pandemic, we have seen tremendous boost to E-payment industry. E-payment systems are on its way to replace fiat currency as the main transaction tool from now onwards. However, adoption rate among malaysian consumers is still low, especailly among the older generation. What are the factors that attracts user to adopt e-payment systems?
What are e-payment systems?
Touch 'n Go eWallet
Boost
• Platform used in making payment for goods or services purchased online using internet.
• Alternative for cash-based transactions
What is UTAUT model?
• Technological framework used to study factors that influence intention to adopt digital products
• Combination of 8 Fragmented Theories
Problem Statement
• Lack of adoption among Malaysian Consumers
• Lack of understanding of E-Payment basic features
• Complicated Research Procedure for UTAUT Framework
Project Objectives
• Identify important User Behavior that promotes E-Payment Adoption
• Research underlying relationship between factors that affect behavioral intention
• Develop a generalized Model for any theoretical framework
Project Scope
Algorithm
MACHINE LEARNING
GUI
EDITABLE STROKE
Correlation Algorithm
Machine Learning Model
GUI Program
Methodology
• Domain Understanding
• Data Collection
• Data Understanding
• Data Preprocessing
• Feature Selection
• Split Features
• Train-Test model
• SMOTE Imbalanced
• Model Evaluation
Conclusion
The system is able to identify set of significant user behavior that promotes user adoption of e-payment system. We are also able to simplify the process of doing research and analysis on UTAUT framework using an AI based approach. Our correlation based feature selection algorithm is also able to identify set of significant features.

9.3 A-3 Questionnaire Sample

E-payment Cryptocurrency Coin

Dear Participant,
We're conducting a research study entitled "users' intention to use Blockchain / Cryptocurrency Coin". This study attempts to examine the factors that affect the users' intention to use the Blockchain / Cryptocurrency Coin from your viewpoint as a user. Please feel free to view the video – (https://www.youtube.com/watch?v=pmPFHog74bI).
New Cryptocurrency Coin platform: Sign up for free and invite friends to join: https://www.woth.io/auth/signup?ref=83d8416d7c
Referral Code 83d8416d7c. Make sure you confirm your email after registration. Tks
As part of the study, I would like to invite you to participate in this survey and appreciate your time spent here. Your participation will involve a survey, which takes around 15-20 minutes to complete. Your participation in this study is voluntary. Please be assured that all information will be treated confidentially. If you have any questions, please do not hesitate to email pcsurvey7@gmail.com for more information.
Thank You

* Required

Section I

Start with the socio-demographic questionnaire that is developed for purposes of gauging the social status of a specific community or society.

1. Name *

2. Email *

3. Mobile Phone (H/p)

4. Where were you born? *

Mark only one oval.

- Malaysia
- Singapore
- Brunei
- Thailand
- Indonesia
- Philippines
- Vietnam
- Cambodia
- Laos
- Myanmar
- Others

5. Where are you residing now? *

Mark only one oval.

- Malaysia
- Singapore
- Brunei
- Thailand
- Indonesia
- Philippines
- Vietnam
- Cambodia
- Laos
- Myanmar
- Others

6. Q1. Age *

Mark only one oval.

- < 25 years
- 26 - 40 years
- 41 - 55 years
- above 55 years

7. Q2. Gender *

Mark only one oval.

- Male
- Female

8. Q3. What is your marital status? *

Mark only one oval.

- Single
- Married
- Other

9. Q4. What is your highest level of education? *

Mark only one oval.

- Primary school
- Secondary/High school
- College/university
- Graduate school
- Other

10. Q5. What industry do you work in? *

Mark only one oval.

- Education
- Banking / Finance
- Retail / Hypermarket
- Manufacturing
- Healthcare
- Other

11. Q6. What is your position level at work? *

Mark only one oval.

- Top management
- Middle management
- Junior management
- Professional
- Other

Section II

Can select more than one items.

12. 1. Do you own any of the following? *

Check all that apply.

- [ ] Mobile Smartphone
- [ ] Bank Cards (Credit, Debit, Pre-paid)
- [ ] Touch n Go, EzLink etc
- [ ] Internet Services (e.g: Broadband)
- [ ] E-wallet account (E.g: MOL or PayPal)
- [ ] HealthCare Gadget (E.g: Blood pressure measure device etc)
- [ ] Internet of Things gadget (e.g: Fitbit – measure steps, etc)
- [ ] Blockchain / Cryptocurrency Coin

13. 2. What operating system does your smartphone/Tablet operate? *

Check all that apply.

- [ ] Android (Samsung, etc)
- [ ] iOS (iPhone)
- [ ] Microsoft phone

14. 3. Have you made any electronic payments in the past 12 months? *

Check all that apply.

- [ ] Yes (Mobile)
- [ ] Yes (Bank Cards (i.e. Credit, Debit, Pre-paid))
- [ ] Yes (Touch n Go, Ezlink etc)
- [ ] Yes (Internet Services (e.g: Broadband))
- [ ] Yes (E-wallet account (E.g: MOL or PayPal))
- [ ] Yes (Blockchain / Cryptocurrency Coin solutions)
- [ ] No

15. 4. Investment Portfolio (Tick what is relevant) *

Check all that apply.

- [ ] Stocks
- [ ] Indices
- [ ] Crypto
- [ ] Commodity
- [ ] Currency
- [ ] Other

16. 5. Investment Cryptocurrency Coin (Tick what is relevant) *

Check all that apply.

- [ ] No
- [ ] Bitcoin
- [ ] Ethereum
- [ ] LiteCoin
- [ ] Ripple (XPR)
- [ ] Defi Coin
- [ ] Ontime or Intime

Other: [ ] ______________________

17. 6. E-payment purchasing, Loyalty Points and Crypto Coin *

Mark only one oval per row.

	Yes	No
(1) Have you ever purchased anything using the E-payment mode?	⬭	⬭
(2) In the next six months, do you plan to purchase anything using the E-payment mode?	⬭	⬭
(3) Do you plan to purchase any gifts/tickets etc this year using the E-payment mode?	⬭	⬭
(4) Do you have the Investment in Crypto?	⬭	⬭
(5) Do you plan to convert your Crypto Investment to Crypto Coins for E-payment?	⬭	⬭
(6) Do you	⬭	⬭

(6) Do you have the loyalty points that can be used for E-payment transactions?

(7) Are you interested to turn your loyalty points into Crypto coins for E-payment?

(8) Do you like to have the credit/debit card, e-wallet, Crypto Coin on a single platform for E-payment transactions?

(9) Do you like to have the platform of a survey redemption solution for Crypto Coin E-payment transactions?

(10) Do you like the points collected from shopping that can be used for Crypto Coin

E-payment transactions? ____________

18. Type of gifts you like

Check all that apply.

- [ ] Shopping Voucher
- [ ] Cryptocurrency Coin
- [ ] Loyalty Points
- [ ] Airline / Holiday Voucher
- [ ] Food Voucher
- [ ] Taxi / Transport Voucher
- [ ] Money in my e-wallet (e.g: TnG, Boost, Grab etc)

Other: [ ] ____________

Performance Expectancy (PE)

similar to Perceived Usefulness

19. PE1: I find Blockchain / Cryptocurrency Coin useful in me. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

20. PE2: Using Blockchain / Cryptocurrency Coin enables me to accomplish tasks more quickly. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

21. PE3: Using Blockchain / Cryptocurrency Coin increases my productivity. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

22. PE4: Using Blockchain / Cryptocurrency Coin increases my chances of getting more choices. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

Effort Expectancy (EE)

similar to Ease of Use

23. EE1: My interaction with Blockchain / Cryptocurrency Coin is clear and understandable. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

24. EE2: It is easy for me to become skillful at using Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

25. EE3: I find Blockchain / Cryptocurrency Coin easy to use. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

26. EE4: Learning to operate Blockchain / Cryptocurrency Coin is easy for me. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

Attitude toward Using Technology (AT)

27. AT1: Using Blockchain / Cryptocurrency Coin is a good idea. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

28. AT2: Blockchain / Cryptocurrency Coin makes effort more interesting. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

29. AT3: Working with Blockchain / Cryptocurrency Coin is fun. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

30. AT4: I like working with Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

Social Influence (SI)

31. SI1: People who influence my behavior think that I should use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

32. SI2: People who are important to me think that I should use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

33. SI3: There are more options in the marketplace for the use of Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

34. SI4: In general, the company has supported the use of Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

Facilitating Conditions (FC)

35. FC1: I have the resources necessary to use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

36. FC2: I have the knowledge necessary to use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

37. FC3: Blockchain / Cryptocurrency Coin is compatible with other systems I use. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

38. FC4: A specific person (or group) is available for assistance with Blockchain / Cryptocurrency Coin difficulties. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

Self-Efficacy (SE)

39. SE1: I can complete a job or task using Blockchain / Cryptocurrency Coin , if there is no one around to tell me what to do. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

40. SE2: I can complete a job or task using Blockchain / Cryptocurrency Coin , if I can call someone for help if I get stuck. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

41. SE3: I can complete a job or task using Blockchain / Cryptocurrency Coin , if I have a lot of time to complete the job for which the software is provided. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

42. SE4: I can complete a job or task using Blockchain / Cryptocurrency Coin , if I have just the built-in help facility for assistance. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

Anxiety (AX)

43. AX1: I feel apprehensive about using Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

44. AX2: It scares me to think that I could lose a lot of information using Blockchain / Cryptocurrency Coin by hitting the wrong key. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

45. AX3: I hesitate to use Blockchain / Cryptocurrency Coin for fear of making mistakes I cannot correct. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

46. AX4: Blockchain / Cryptocurrency Coin is somewhat intimidating to me. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

Trust (T)

47. T1: I feel comfortable using Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

48. T2: I feel the reliability to use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

49. T3: I am glad about the service quality to use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

50. T4: Blockchain / Cryptocurrency Coin Integrity is vital to me. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	⬭	⬭	⬭	⬭	⬭	Completely agree

Behavioral Intention to Use the System (BI)

51. BI1: I intend to use Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

52. BI2: I plan to invest Blockchain / Cryptocurrency Coin. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

53. BI3: I plan to have Blockchain / Cryptocurrency Coin as an E-Wallet. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

54. BI4: I plan to use Blockchain / Cryptocurrency Coin for E-payment transaction to buy stuff. *

Mark only one oval.

	1	2	3	4	5	
Completely disagree	◯	◯	◯	◯	◯	Completely agree

9.4 A-4 Sample Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Timestam	Name	Email	Mobile Ph	Where we	Where are	Q1. Age	Q2. Gende	Q3. What	Q4. What	Q5. What	Q6. What	1. Do you	2. What o	3. Have yo	4. Investm	5. Investm	6. E-paym	6. E-pa
2	2021/03/2	Chai Yin	eeuchaiyin	1.22E+08	Malaysia	Malaysia	< 25 years	Female	Single	College/ur	Baking / Fi	Other	Mobile Sm	Android (S	Yes (Mobil	Stocks	pi	Yes	Yes
3	2021/03/2	Mabellyn	mabellyn2	1.22E+08	Malaysia	Malaysia	< 25 years	Female	Single	College/ur	Other	Other	Mobile Sm	Android (S	Yes (Bank	Other	No	Yes	Yes
4	2021/03/2	Lee Beng E	leebbee70	1.25E+08	Malaysia	Malaysia	41 - 55 yea	Female	Single	College/ur	Manufactu	Middle ma	Mobile Sm	Android (S	Yes (Bank	Crypto;Ot	Ethereum	Yes	Yes
5	2021/03/2	Salman Fa	salmanfaiz	017-63746	Others	Malaysia	< 25 years	Male	Single	College/ur	Education	Other	Mobile Sm	Android (S	Yes (Mobil	Currency;(	No	No	Yes
6	2021/03/2	Darsini Ma	darsini96n	1.23E+08	Malaysia	Malaysia	< 25 years	Female	Single	College/ur	Other	Other	Mobile Sm	iOS (iPhon	Yes (Mobil	Currency	No	Yes	Yes
7	2021/03/2	Izath	drizat90@	1.12E+09	Malaysia	Malaysia	26 - 40 yea	Male	Married	College/ur	Healthcare	Profession	Mobile Sm	iOS (iPhon	Yes (Mobil	Stocks;Cry	Ethereum;	Yes	Yes
8	2021/03/2	albert tang	albertbct@	1.23E+08	Malaysia	Malaysia	above 55 y	Male	Married	College/ur	Baking / Fi	Middle ma	Mobile Sm	Android (S	Yes (Mobil	Stocks;Cry	Ethereum;	Yes	Yes
9	2021/03/2	L T Soon	corp@3qb	3.8E+08	Malaysia	Malaysia	above 55 y	Male	Married	Graduate s	Other	Top mana	Mobile Sm	Android (S	Yes (Mobil	Stocks;Cur	No	Yes	Yes
10	2021/03/2	Murali Nag	nagalingan	1.24E+08	Malaysia	Malaysia	above 55 y	Male	Married	College/ur	Baking / Fi	Profession	Mobile Sm	Android (S	Yes (Mobil	Stocks	No	Yes	Yes
11	2021/03/2	Kaung Mya	kaungmya	-4.4E+08	Myanmar	Myanmar	26 - 40 yea	Male	Married	Graduate s	Education	Top mana	Mobile Sm	Android (S	Yes (Mobil	Commodit	No	Yes	Yes
12	2021/03/3	Mohamed	riffaventur	1.25E+08	Malaysia	Malaysia	41 - 55 yea	Male	Married	Secondary	Other	Other	Mobile Sm	Android (S	Yes (Mobil	Other	No	Yes	Yes
13	2021/03/3	Nur Nabila	nabilahras	1.93E+08	Malaysia	Malaysia	< 25 years	Female	Single	Graduate s	Other	Other	Mobile Sm	Android (S	Yes (Mobil	Stocks;Cry	Bitcoin	Yes	Yes
14	2021/03/3	Rosnita M	rosnitama	1.96E+08	Malaysia	Malaysia	26 - 40 yea	Female	Single	College/ur	Other	Other	Mobile Sm	iOS (iPhon	Yes (Mobil	Other	No	Yes	Yes
15	2021/03/3	Satiamurt	satia_maa	1.64E+08	Malaysia	Malaysia	41 - 55 yea	Male	Married	College/ur	Healthcare	Profession	Mobile Sm	Android (S	Yes (Mobil	Other	No	No	No
16	2021/03/3	Putri Khair	nshputri@	1.26E+08	Malaysia	Malaysia	41 - 55 yea	Female	Other	College/ur	Other	Middle ma	Mobile Sm	iOS (iPhon	Yes (Bank	Other	No	Yes	Yes
17	2021/03/3	Mahabuya	mahabuya	1.95E+08	Malaysia	Malaysia	above 55 y	Female	Married	Primary sc	Education	Other	Mobile Sm	Android (S	Yes (Bank	Other	No	Yes	Yes
18	2021/03/3	Kasthuri	kastrms28	1.24E+08	Malaysia	Malaysia	26 - 40 yea	Female	Married	Secondary	Other	Other	Mobile Sm	Android (S	Yes (Bank	Stocks	No	No	No
19	2021/03/3	Muhamma	hafizuddin	1.11E+09	Malaysia	Malaysia	26 - 40 yea	Male	Single	College/ur	Other	Other	Mobile Sm	Android (S	Yes (Touch	Other	No	Yes	Yes
20	2021/03/3	Norihan R	norihan96	1.74E+08	Malaysia	Malaysia	26 - 40 yea	Female	Married	Secondary	Education	Other	Mobile Sm	Android (S	Yes (Bank	Other	No	Yes	Yes
21	2021/03/3	Mohamed	budin27@	1.64E+08	Malaysia	Malaysia	26 - 40 yea	Male	Married	College/ur	Other	Other	Mobile Sm	Android (S	Yes (Mobil	Other	No	Yes	Yes
22	2021/03/3	Muhamma	danialthun	6.02E+10	Malaysia	Malaysia	< 25 years	Male	Single	Primary sc	Other	Other	Mobile Sm	Android (S	No	Other	No	No	No
23	2021/03/3	Mohamad	mosaini88	1.08E+08	Malaysia	Malaysia	41 - 55 yea	Male	Married	Secondary	Other	Profession	Mobile Sm	Android (S	No	Crypto;Co	Bitcoin;Eth	No	Yes

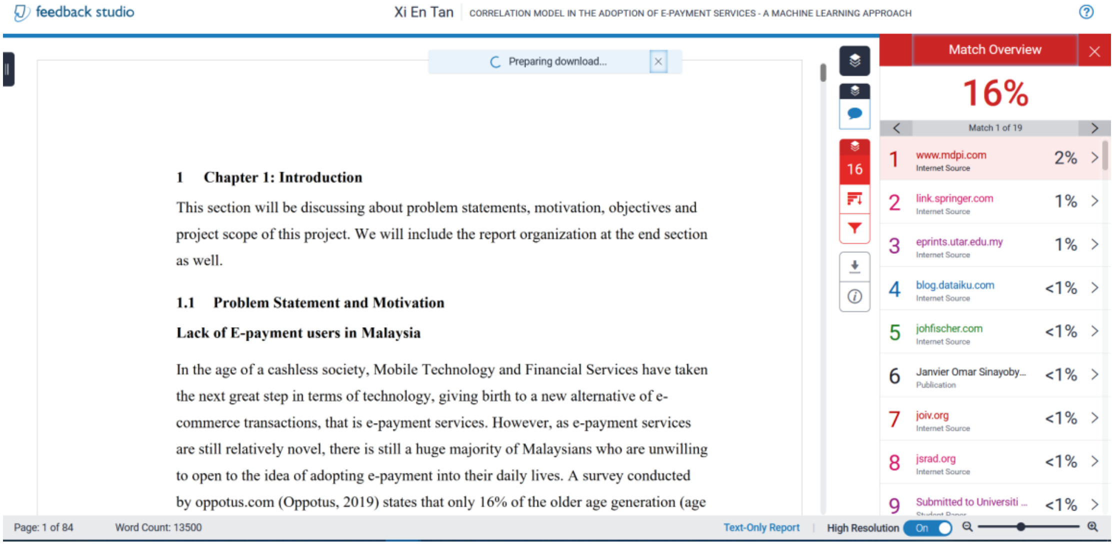
E-payment Cryptocurrency Coin

9.5 A-5 UTAUT Item Constructs

Construct	Item	Description
PE	PE1	I find Blockchain / Cryptocurrency Coin useful in me.
	PE2	Using Blockchain / Cryptocurrency Coin enables me to accomplish tasks more quickly.
	PE3	Using Blockchain / Cryptocurrency Coin increases my productivity.
	PE4	Using Blockchain / Cryptocurrency Coin increases my chances of getting more choices.
EE	EE1	My interaction with Blockchain / Cryptocurrency Coin is clear and understandable.
	EE2	It is easy for me to become skillful at using Blockchain / Cryptocurrency Coin.
	EE3	I find Blockchain / Cryptocurrency Coin easy to use.
	EE4	Learning to operate Blockchain / Cryptocurrency Coin is easy for me.
SI	SI1	People who influence my behavior think that I should use Blockchain / Cryptocurrency Coin.
	SI2	People who are important to me think that I should use Blockchain / Cryptocurrency Coin.
	SI3	There are more options in the marketplace for the use of Blockchain / Cryptocurrency Coin.
	SI4	In general, the company has supported the use of Blockchain / Cryptocurrency Coin.
FC	FC1	I have the resources necessary to use Blockchain / Cryptocurrency Coin.
	FC2	I have the knowledge necessary to use Blockchain / Cryptocurrency Coin.
	FC3	Blockchain / Cryptocurrency Coin is compatible with other systems I use.
	FC4	A specific person (or group) is available for assistance with Blockchain / Cryptocurrency Coin difficulties.
AT	AT1	Using Blockchain / Cryptocurrency Coin is a good idea.
	AT2	Blockchain / Cryptocurrency Coin makes effort more interesting.
	AT3	Working with Blockchain / Cryptocurrency Coin is fun.
	AT4	I like working with Blockchain / Cryptocurrency Coin.
SE	SE1	I can complete a job or task using Blockchain / Cryptocurrency Coin , if there is no one around to tell me what to do.
	SE2	I can complete a job or task using Blockchain / Cryptocurrency Coin , if I can call someone for help if I get stuck.
	SE3	I can complete a job or task using Blockchain / Cryptocurrency Coin , if I have a lot of time to complete the job for which the software is provided.

	SE4	I can complete a job or task using Blockchain / Cryptocurrency Coin , if I have just the built-in help facility for assistance.
AX	AX1	I feel apprehensive about using Blockchain / Cryptocurrency Coin.
	AX2	It scares me to think that I could lose a lot of information using Blockchain / Cryptocurrency Coin by hitting the wrong key.
	AX3	I hesitate to use Blockchain / Cryptocurrency Coin for fear of making mistakes I cannot correct.
	AX4	Blockchain / Cryptocurrency Coin is somewhat intimidating to me.
T	T1	I feel comfortable using Blockchain / Cryptocurrency Coin.
	T2	I feel the reliability to use Blockchain / Cryptocurrency Coin.
	T3	I am glad about the service quality to use Blockchain / Cryptocurrency Coin.
	T4	Blockchain / Cryptocurrency Coin Integrity is vital to me.
BI	BI1	I intend to use Blockchain / Cryptocurrency Coin.
	BI2	I plan to invest Blockchain / Cryptocurrency Coin.
	BI3	I plan to have Blockchain / Cryptocurrency Coin as an E-Wallet.
	BI4	I plan to use Blockchain / Cryptocurrency Coin for E-payment transaction to buy stuff.

10 Appendix B: Plagiarism Check Summary



CORRELATION MODEL IN THE ADOPTION OF E-PAYMENT SERVICES - A MACHINE LEARNING APPROACH

ORIGINALITY REPORT

16%	13%	11%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	link.springer.com Internet Source	1%
3	eprints.utar.edu.my Internet Source	1%
4	blog.dataiku.com Internet Source	<1%
5	johfischer.com Internet Source	<1%
6	Janvier Omar Sinayobye, Kyanda Swaib Kaawaase, Fred N. Kiwanuka, Richard Musabe. "Hybrid Model of Correlation Based Filter Feature Selection and Machine Learning Classifiers Applied on Smart Meter Data Set", 2019 IEEE/ACM Symposium on Software Engineering in Africa (SEiA), 2019 Publication	<1%

joiv.org

7	Internet Source		<1%
8	jsrad.org Internet Source		<1%
9	Submitted to Universiti Tunku Abdul Rahman Student Paper		<1%
10	www.coba.usf.edu Internet Source		<1%
11	www.caee.utexas.edu Internet Source		<1%
12	iosrjen.org Internet Source		<1%
13	Submitted to University of Salford Student Paper		<1%
14	Oliveira, Tiago, Manoj Thomas, Goncalo Baptista, and Filipe Campos. "Mobile payment: Understanding the determinants of customer adoption and intention to recommend the technology", Computers in Human Behavior, 2016. Publication		<1%
15	www.analyticsvidhya.com Internet Source		<1%
16	eprints.lancs.ac.uk Internet Source		<1%

17	Mousa Jaradat, Mohammed-Issa Riad, and Mamoun S. Al Rababaa. "Assessing Key Factor that Influence on the Acceptance of Mobile Commerce Based on Modified UTAUT", International Journal of Business and Management, 2013. Publication	<1 %
18	www.investopedia.com Internet Source	<1 %
19	Submitted to RDI Distance Learning Student Paper	<1 %
20	Stavros A. Nikou, Anastasios A. Economides. "Mobile-based assessment: Investigating the factors that influence behavioral intention to use", Computers & Education, 2017 Publication	<1 %
21	tessera.spandidos-publications.com Internet Source	<1 %
22	dias.library.tuc.gr Internet Source	<1 %
23	huggingface.co Internet Source	<1 %
24	journals.sagepub.com Internet Source	<1 %
25	livelig.com Internet Source	<1 %

26	hdl.handle.net Internet Source	<1 %
27	www.icommercecentral.com Internet Source	<1 %
28	Min-Fang Wei, Yir-Hueih Luh, Yu-Hsin Huang, Yun-Cih Chang. "Young Generation's Mobile Payment Adoption Behavior: Analysis Based on an Extended UTAUT Model", Journal of Theoretical and Applied Electronic Commerce Research, 2021 Publication	<1 %
29	Submitted to newcastle under lyme college Student Paper	<1 %
30	Submitted to Nottingham Trent University Student Paper	<1 %
31	Submitted to University of Wollongong Student Paper	<1 %
32	Veera Bhatiasevi. "An extended UTAUT model to explain the adoption of mobile banking", Information Development, 2016 Publication	<1 %
33	Brunner, Gerd. "Structure features for content-based image retrieval and classification problems", Universität Freiburg, 2007. Publication	<1 %

34	Darmatasia, Aniati Murni Arymurthy. "Predicting the status of water pumps using data mining approach", 2016 International Workshop on Big Data and Information Security (IWBIS), 2016 Publication	<1 %
35	Theresadevi Moodley, Irene Govender. "Factors influencing academic use of internet banking services: An empirical study", African Journal of Science, Technology, Innovation and Development, 2016 Publication	<1 %
36	Submitted to Universiti Teknologi Malaysia Student Paper	<1 %
37	Submitted to University College London Student Paper	<1 %
38	Submitted to British University in Egypt Student Paper	<1 %
39	Submitted to Hankuk University of Foreign Studies Student Paper	<1 %
40	S.P. Meharunnisa, M. Ravishankar, K. Suresh. "IMPROVEMENT IN DETECTION ACCURACY OF DIGITAL MAMMOGRAM USING POINT TRANSFORM AND DATA MINING TECHNIQUE", ICTACT Journal on Image and Video Processing, 2018	<1 %

	Publication	
41	hpcf-files.umbc.edu Internet Source	<1 %
42	www.techdroidtips.com Internet Source	<1 %
43	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
44	Submitted to The Hong Kong Institute of Education Student Paper	<1 %
45	Submitted to University of KwaZulu-Natal Student Paper	<1 %
46	Submitted to University of Liverpool Student Paper	<1 %
47	www.ecoccs.com Internet Source	<1 %
48	Submitted to International University of Malaya-Wales Student Paper	<1 %
49	Kojo Kakra Twum, Daniel Ofori, Gabriel Keney, Bright Korang-Yeboah. "Using the UTAUT, personal innovativeness and perceived financial cost to examine student's intention to use E-learning", Journal of Science and Technology Policy Management, 2021	<1 %

Publication

50	Submitted to Lebanese University Student Paper	<1 %
51	Submitted to Universidad Carlos III de Madrid Student Paper	<1 %
52	Yi Cao, Christian Ritz, Raad Raad. "Feature selection for low bit rate mobile augmented reality applications", Signal Processing: Image Communication, 2015 Publication	<1 %
53	www.idosi.org Internet Source	<1 %
54	Ken Pohronezny, Myrine Hewitt, Janice Infante. " Susceptibility of pepper fruit to infection by pv. ", Canadian Journal of Plant Pathology, 1993 Publication	<1 %
55	Submitted to Universiti Malaysia Sabah Student Paper	<1 %
56	Submitted to University of Bath Student Paper	<1 %
57	Widayat Widayat, Ilyas Masudin, Novita Ratna Satiti. "E-Money Payment: Customers' Adopting Factors and the Implication for Open Innovation", Journal of Open Innovation: Technology, Market, and Complexity, 2020 Publication	<1 %

58	ideas.repec.org Internet Source	<1 %
59	Submitted to University of London External System Student Paper	<1 %
60	etd.uum.edu.my Internet Source	<1 %
61	flipkarma.com Internet Source	<1 %
62	hackernoon.com Internet Source	<1 %
63	www.abacademies.org Internet Source	<1 %
64	Krishna Moorthy, Ooi Yin Chiang, Aufa Amalina Kamarudin, Loh Chun T'ing, Chin Yoon Mei, Sonia Johanthan. "Factors That Influence the Adoption of E-Wallets by the Malaysians", International Journal of E-Adoption, 2021 Publication	<1 %
65	Mervat Albufalasa, Madani O. Ahmed, Yasser A. Gomaa. "Virtual Learning Platforms As A Mediating Factor In Developing Oral Communication Skills: A Quantitative Study Of The Experience Of English Major Students At The University of Bahrain", 2020 Sixth	<1 %

	International Conference on e-Learning (econf), 2020 Publication	
66	Submitted to Multimedia University Student Paper	<1 %
67	eprints.soton.ac.uk Internet Source	<1 %
68	idoc.pub Internet Source	<1 %
69	researchspace.ukzn.ac.za Internet Source	<1 %
70	scholarworks.aub.edu.lb Internet Source	<1 %
71	stars.library.ucf.edu Internet Source	<1 %
72	Submitted to University of Mauritius Student Paper	<1 %
73	ir.knust.edu.gh Internet Source	<1 %
74	tanthiamhuat.files.wordpress.com Internet Source	<1 %
75	www.coursehero.com Internet Source	<1 %

76	Chanchai Phonthanukitithaworn, Carmine Sellitto, Michelle W.L. Fong. "An investigation of mobile payment (m-payment) services in Thailand", Asia-Pacific Journal of Business Administration, 2016 Publication	<1 %
77	Florian Resatsch. "Working Prototype: An NFC-based Mobile Phone Ticketing System", Ubiquitous Computing, 2010 Publication	<1 %
78	Huaxin Li, Zheyu Xu, Haojin Zhu, Di Ma, Shuai Li, Kai Xing. "Demographics inference through Wi-Fi network traffic analysis", IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, 2016 Publication	<1 %
79	Li-Bin Wang, Ling-Ling Xu, Li-Jian Chen, Kai-Kai Zhang et al. "Methamphetamine induces intestinal injury by altering gut microbiota and promoting inflammation in mice", Toxicology and Applied Pharmacology, 2022 Publication	<1 %
80	Sien Chen. "Chapter 2 Estimating Customer Lifetime Value Using Machine Learning Techniques", IntechOpen, 2018 Publication	<1 %

81	academic-pub.org Internet Source	<1 %
82	educationdocbox.com Internet Source	<1 %
83	repository.sustech.edu Internet Source	<1 %
84	researcharchive.vuw.ac.nz Internet Source	<1 %
85	www.emeraldinsight.com Internet Source	<1 %
86	www.ijraset.com Internet Source	<1 %
87	"Advances in Neural Networks – ISNN 2011", Springer Science and Business Media LLC, 2011 Publication	<1 %
88	"CS1-PC-22_HR", ActEd Publication	<1 %
89	Ashraf Elsafty, Abdulaziz Boghdady. "The Cognitive Determinants Influencing Consumer Purchase-Intention Towards Subscription Video on Demand (SVoD): Case of Egypt", International Journal of Marketing Studies, 2022 Publication	<1 %

90	Feature Extraction Construction and Selection, 1998. Publication	<1 %
91	K. Z. Mao. "The ties problem resulting from counting-based error estimators and its impact on gene selection algorithms", Bioinformatics, 07/31/2006 Publication	<1 %
92	Mohammad Alryalat, Yogesh K. Dwivedi, Michael D. Williams. "A Conceptual Model for Examining E-Government Adoption in Jordan", International Journal of Electronic Government Research, 2012 Publication	<1 %
93	Stefano Diomedi, Francesco E. Vaccari, Matteo Filippini, Patrizia Fattori, Claudio Galletti. "Mixed Selectivity in Macaque Medial Parietal Cortex during Eye-Hand Reaching", iScience, 2020 Publication	<1 %
94	Submitted to Universitas Pelita Harapan Student Paper	<1 %
95	core.ac.uk Internet Source	<1 %
96	core.kmi.open.ac.uk Internet Source	<1 %

97	napier-surface.worktribe.com Internet Source	<1 %
98	ntnuopen.ntnu.no Internet Source	<1 %
99	pure.southwales.ac.uk Internet Source	<1 %
100	repository.tudelft.nl Internet Source	<1 %
101	researchrepository.wvu.edu Internet Source	<1 %
102	so01.tci-thaijo.org Internet Source	<1 %
103	ujcontent.uj.ac.za Internet Source	<1 %
104	www.cs.waikato.ac.nz Internet Source	<1 %
105	www.igi-global.com Internet Source	<1 %
106	www.ros.hw.ac.uk Internet Source	<1 %
107	www.section.io Internet Source	<1 %

University Tunku Abdul Rahman			
Form Title: Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 22/4/2022	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Tan Xi En
ID Number(s)	19ACB04098
Programme / Course	Bachelor of Computer Science (Honours)
Title of Final Year Project	CORRELATION MODEL IN THE ADOPTION OF E-PAYMENT SERVICES: A MACHINE LEARNING

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceed the limits approved by UTAR)**
Overall similarity index: 16 **% Similarity by source** Internet Sources: 13 % Publications: 11 % Student Papers: 7 %	
Number of individual sources listed of more than 3% similarity: -	
Parameters of originality required, and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are***	

Note: Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Tong

Signature of Supervisor

Name: TONG DONG LING

Date: 22/4/20222

Signature of Co-Supervisor

Name: ____________________

Date: ____________________

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	19ACB04098
Student Name	Tan Xi En
Supervisor Name	Dr. Tong Dong Ling

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

____

(Signature of Student)
Date: 22/4/2022