SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	SOH JOEN SHIUAN
Metric Number:	MCS241028
Student Email & Phone:	sohjoenshiuan@graduate.utm.my
Project Title:	Customer Churn Prediction in E-Commerce Industry using Classification Approach
Supervisor 1:	-
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

In the fast-paced growing environment, retail sales transform business model into online sales to speed up

transaction procedure to keep up with the digital trend. According to Ikhlass Boukrouh (2025), this is known as

electronic commerce (e-commerce) in which business transactions are conducted on the internet. This transformation

brings general challenges such as high competition between e-commerce platforms, high return and refund rate from

customers. This leads to customer churn and the trend keeps growing. Based on the statistics from Jack M. Germain

(2023), the 42% if B2C companies are churning 3% or more, followed by another 16% that churns 4% or more.

Customer churn happens when customer has decided to stop using the service. This reduces the brand loyalty, increasing

future customer acquisition cost and revenue down, leading to business loss. (Daniyal Asif, 2025)

Churn prediction requires supervised learning method to train. The Traditional ML Models under supervised learning

method include Logistic Regression, Decision Trees, Random Forest, Gradient Boosting (XGBoost, LightGBM,

CatBoost), Support Vector Machines (SVM), and k-Nearest Neighbours (k-NN). Random Forest algorithm would be

used to identify the patterns in customer churn. Following the pattern would reveal the root cause that affects the

customer churn rate in e-commerce platform.

Problem Background:

Churn prediction allows the e-commerce management to identify whether the customers are likely to stop using the e-commerce platform or not, allowing for preventive intervention. Without addressing customer churn, e-commerce has higher tendency to lose the current customers, leading to lesser commission from transaction between customers and sellers. At the same time, it would be increasing future customer acquisition costs through marketing, advertisements, and promotions. Therefore, conducting research is necessary to reveal reasons that caused customers lose interest in future engagement on the platform.

Problem Statement:
E-commerce Customer daily active time has been gradually reducing. Reducing active time indicates that the customer might be leaving the e-commerce platform in the future. The number of customers would be reducing gradually, leading to lesser transactions to be made in the e-commerce platform. The seller would be directly influenced followed by the e-commerce platform. Seller unable to sell the products to the customer. E-commerce platform would gain lesser commission from overall transactions. By increasing e-commerce customer active time, customer would have higher tendency to make more transactions in e-commerce platform. This requires customer daily active period data, total time spent on the platform, purchase history data to train a model that predicts the tendency of customer churn.
For the current solution, management predict the users that is frequently active in the platform would be purchasing more items. In the fact that customer is active but without purchasing items. This gives a contradiction point that customer active time does not directly influence the number of customer transactions.

Aim of the Project:
To develop and enhance a predictive model that uses Random Forest algorithm that labels the customer as churn or not churn, provide actionable insights in improving the sales revenue and customer active time.

Objectives of the Project:
    1)   To preprocess the data by applying one-hot encoding technique on existing dataset.

    2)   To develop a predictive model that able to predict the customers that is going to churn

    3)   To evaluate the model accuracy on predicting the churn status of customer

....
....

Scopes of the Project:
1) Data source would be obtained from public datasets such as Kaggle.

2) Python would be used to process and clean the Excel file data.

3) MySQL would be used to store data.

4) SQL would be used to translate the data into summarized form.

5) Power BI would be used to visualize the data from database.

Expected Contribution of the Project:

Current method would use manual check to identify whether the customer is churn or not. The moment to identify

whether the customer is churn or not, the customer already churn. With the use of predictive model, marketing able to

find out a list of customers that has sign to stop using the platform early. Marketing able to implement early prevention

strategies such as free shipping/discount vouchers, lucky draw contest to keep these customers from leaving the

platform. E-commerce revenue would be increasing in proportion to increased transaction.

Project Requirements:

Software:	Visual Studio Code, Python, SQL, Excel, MySQL, Power BI
Hardware:	Windows 11 PC, 16 GB RAM, 500 GB Memory
Technology/Technique/ Methodology/Algorithm:	Random Forest Algorithm

Type of Project (Focusing on Data Science):

[/] Data Preparation and Modeling

[/] Data Analysis and Visualization

[/] Business Intelligence and Analytics

[/] Machine Learning and Prediction

[/] Data Science Application in Business Domain

Status of Project:

[/] New ....

[ ] Continued ....

If continued, what is
the previous title? ....

SECTION C: Declaration

I declare that this project is proposed by:

[/] Myself ....

[ ] Supervisor/Industry Advisor (   )

Student Name: SOH JOEN SHIUAN ....

SJS ....   5/4/2025 ....
Signature   **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:   ....

                                 Signature   **Date**

Name of Supervisor 2 (if any):   ....

                                 Signature   **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL   [ ] CONDITIONAL APPROVAL (Major)*
[ ] CONDITIONAL APPROVAL (Minor)   [ ] FAIL*
* Student has to submit new proposal form considering the evaluators' comments.

Comments:

....

....

....

....

....

....

....

....

....

....

Name of Evaluator 1:

Project1 Proposal Form MSc (Data Science)

Signature   **Date**

Name of Evaluator 2:

Signature   **Date**