



Assignment Cover Sheet

Pattern Recognition & Machine Learning UG+PG 11482, 11512

You must keep a photocopy or electronic copy of your assignment.

Name of Tutor:	Dr Xing Wang		
Student ID:	U3284513	Unit/Subject Code:	11482
Assignment No.:	1A	Number of pages: (including this cover sheet)	3
Date and Time Due:	29 th of August, 23:59pm		
Date and Time Submitted:	10:00pm 31/08/2025		

I declare that this assignment is solely my own work, except where due acknowledgements are made. I acknowledge that the assessor of this assignment may provide a copy of this assignment to another member of the University, and/or to a plagiarism checking service whilst assessing this assignment. I have read and understood the University Policies in respect of Student Academic Honesty.

Date: 23/08/2025

INTRODUCTION

This assignment proposes a machine learning model to classify whether a credit card transaction is fraudulent or genuine. Credit card fraud is becoming increasingly prevalent in our society and could cause damage in terms of trust, financial and security aspects. This model will analyze transactions behaviors, such as time between transactions, amount and the time made. As of such, the proposed machine learning model demonstrates one of the core characteristics of machine learning, classification which in this case refers to fraudulent credit card activities. This is also one of the core characteristics in machine learning and also proves to be one of the reasons why a machine learning solution is favored. Due to the high volume of card transactions a day, manual reviews of each transaction are not a possible solution. This method is also prone to human errors and takes a large amount of time to review. With a machine learning model, it can flag a transaction as fraudulent in a quicker manner and with better accuracy, thus reducing errors and time constraints. This will also help financial institutions deal with credit card transactions in a more efficient manner and reduce the risk of raising a false transaction which can inconvenience their customers.

MODEL

The proposed main goal of the project is to design, train and predict a learning model whether a credit card transaction is fraudulent or not based on training data. Another goal includes reducing the number of false positives and negatives raised. Due to the sensitive nature of the dataset, the model will only classify the charges as either 1 – being fraudulent and 0 otherwise with not much more information. This method is referred to as supervised learning, due to the datasets having labels for the transactions. In this case, the labels are either fraudulent or non-fraudulent. This, however, lies with the main goal of the model itself, as it will not be trained in fraudulent techniques used. The main patterns that the model will extract from the data are the timing between each transaction and the first transaction measured in seconds, and the amount of funds moved during these transactions. The model will then output the result in binary classification, with 1 being fraudulent and 0 being otherwise. This output can then be used to automate fraud checks in a bank and improve the manual fraud verification process. In order to prove that model can predict accurately, several validation techniques will be employed:

- Dividing the model into training, validation and testing sets. The model will be built upon the training set and validated using the validation set to verify its prediction capabilities. It will then be compared against the testing set to ensure the model has the same performance.

- Using cross validation techniques such as K-Fold cross validation, where the data is divided into “folds” of equal size, and the model will iteratively train itself over the folds to achieve the best results.
- Evaluate the model using various techniques such as accuracy score, recall value, AUC and ROC.

This problem qualifies as a machine learning problem because it has to learn from an existing dataset of bank transactions with different labels, which can then accurately make its predictions. This dataset has to have well-defined input-output results, in this case whether a charge is fraudulent or not, for the model to work appropriately. For the model to make accurate predictions, it must learn the patterns behind fraudulent transactions and utilize these patterns to predict on unseen dataset. Thus, with all reasons listed above, this problem qualifies for a machine learning project.