

# School of Computing

## CA326 Year 3 Project Proposal Form

### SECTION A

Project Title : Atlas

Student 1 Name : Jordan Ukawoko

ID Number : 20353751

Student 2 Name : Luke Whelan

ID Number : 20463814

*(A third team member is exceptional and requires detailed justification.)*

Staff Member Consulted \_\_\_\_\_

### Project Description (1-2 pages)

#### Purpose

Atlas is software that can indicate if there has been fraudulent activity on a given bank account. Atlas has the ability to analyse transaction patterns and determine if activities are abnormal. This activity could be online spending, ATM withdrawals or in-store purchases. If Atlas finds any abnormal patterns then it will flag the transaction as suspicious. The aim is to use Machine Learning algorithms through Python to help us identify abnormal or unusual transactions and using R to confirm our mathematical calculations and formulas. Atlas will work with a CSV uploaded by the user. The main idea is the user will export their transactions from their online banking app into a CSV file format and Atlas will go through all the transactions for any unusual activity. The software will list the location of purchase (if applicable) of the fraudulent transaction flagged and the date of the fraudulent purchase etc. For example:

**(28/06/22,"Nike Store8741\*",400.00 - Fraudulent Purchase Detected).\*\***

Atlas has the ability to:

- Compute the total amount of transactions from the CSV file
- Compute the number of usual transaction from the CSV file
- Compute the number of fraudulent transactions from the CSV file
- Find the percentage of fraudulent transactions against usual transactions from the CSV file with and without duplicates transactions

example of a duplicate transaction →

**(30/06/22,"Amazon8741\*",80.00).\*\***

**(30/06/22,"Amazon8741\*",80.00)\*\***

- Find the location (if applicable) of the fraudulent activity and the date of purchase
- 

## How will we get strong Accuracy?

Firstly, we plan on importing all the necessary libraries in Python to aid us in creating a strong model to accurately detect fraudulent transactions. We will be using the Pandas module in Python to help us import our dataset. Following on, we plan on creating a transaction distribution to help us find the total transactions, usual transactions and fraudulent transactions and creating a percentage using this formula

$$\text{UsualTransactions} / \text{FraudulentTransactions} * 100$$

This formula will give us a percentage of the fraudulent activity detected in the CSV file. etc “16% of transactions are fraudulent”. We will be using Machine Learning algorithms to build our accuracy model. We will be testing different models and comparing and contrasting between the models to make sure we find the different accuracy scores of each model. We will then combine all the models for 99% accuracy (for example the Decision Tree Model might miss detecting a fraudulent transaction while the K-Nearest model might detect all the fraudulent transactions etc).

Other ways we plan on creating an accurate fraud detection software is to have the software look for duplicate transactions and remove them to scale the data.

For graphical purposes we plan on having the software create two different plots. Fraud detection with duplicate transactions vs Fraud Detection without duplicate transactions and coming up with two different percentages using a modification of the formula above

$$\text{UsualTransactionswDups} / \text{FraudulentTransactions} * 100$$

## Why Atlas?

Credit / Debit Card is the main way of purchasing goods both in store and online for many people today. However, sometimes we don't know if our card has been used for purchases without us knowing. To tackle this issue this is where Atlas comes in. Atlas is able to track the pattern of all transactions and flag any abnormalities. Machine Learning algorithms today can help us classify and identify abnormalities.

## **Division of Work**

Atlas will use Machine Learning to conclude results. If either of us have difficulty, we will help each other and try to tackle the problem together. We both plan on developing the GUI (Graphical User Interface). We also both plan to develop the core of the software together.

However because there are a lot of machine learning models to research and understand we plan on focusing on 3 different machine learning models each. For example, Jordan will focus on research of the Decision Tree, K-Nearest Neighbour and Logistic Regression Models while Luke will focus on the Support Vector Machine, Random Forest the XGBoost Models.

## **Programming Languages**

- Python3
- R (for our own calculations)

The software will be written in Python as the main programming language and will use different machine learning algorithms. We will be using python libraries for developing the software. QT Designer will be used as the GUI and will be integrated into the Python Code. R will be used purely for our own calculations

## **Programming Tools**

- Pandas (for CSV processing)
- Machine Learning Libraries
- QtDesigner (GUI )
- Numpy
- Matplotlib
- Visual Studio Code
- Visual Studio Online

These programming tools will aid us in developing our software. We will be working with different machine learning libraries and different python libraries. Visual Studio Code will be used for code writing and testing. Visual Studio Online will be used for project management and tracking ideas

## **Learning Challenges**

Learning the idea of Machine Learning is the main learning challenge and considering this is the focus of our third year project, it will be very important that we research this technique. The different Machine Learning algorithms we are planning on researching and learning more about are The Decision Tree model, K-Nearest Neighbours model, Logistic Regression model, Support Vector Machines model, XGBoost and the Random Forest Model. We plan on researching on how to check the accuracy of each model for our project.

### **Hardware/ Software Platform**

- Latest Windows / Preferably Windows 11 or Macintosh - Preferably Monterey
- A Functioning Computer
- CSV file containing purchase history (You can get this for example by going to AIB internet banking, clicking on account, purchase history then clicking convert purchase history to CSV)

### **Special Hardware / Software Requirements**

- Not Applicable