|  |  |  |
| --- | --- | --- |
| **1** | **Name** | **Yong Zong Wei** |
| **2** | **Email Address** | [soyayong@gmail.com](mailto:soyayong@gmail.com) |

**Thinking like a Data Scientist**

**In this Assignment, you will demonstrate your understanding of the data science methodology by applying it to a given problem. Pick one of the following topics to apply the data science methodology to:**

1. **Fake News Detection**
2. **Sentiment Analysis (Detection)**
3. **Emails (Detecting spam emails)**
4. **Hospitals (Providing better health care)**
5. **Credit Cards (Best spender)**
6. **Credit Cards Fraud Detection**
7. **Food Recipes (Best seller meal)**
8. **MySejahtera (Detecting crowded areas, new clusters)**
9. **Loan Payment**
10. **Car Price**
11. **Stock Market**
12. **Any other topic you are interesting in.**

**You will have to play the role of a data scientist, you have to come up with a problem that is more specific but related to one of these topics.**

**Data Science Methodology:**

1. **Business Understanding**

Trying to understand what the problem is, and what you are trying to solve.

1. **Analytic Approach**

Selecting the right analytic approach (machine learning algorithm) depends on the main problem and objective.

1. **Data Requirements**

Listing down the necessary data content, the formats of the data.

1. **Data Collection**

Data scientist takes place to determine whether or not they have what they need.

1. **Data Understanding**

Study the dataset, read the dataset, looking at the number of rows and columns.

1. **Data Preparation**

In this stage Data Scientist will focus on Data Wrangling, and Exploratory Data Analysis.

1. **Modelling**

Applying the chosen analytic approach.

1. **Evaluation**

Checking the results.

1. **Deployment**

Trying to apply the model on the new coming data.

1. **Feedback**

Getting comments.

**Topic:**

**Credit Card Fraud Detection**

I'm selecting the above topic for my assignment 3 for research and understand the fraud issues for credit cards and can use data science skills to prevent these issues.

**Methodology:**

1. **Business Understanding**

Research and Understanding

Base on some data search from U.S. Federal Trade Commission (FTC) resource, credit card fraud was the second top in FTC that most reported in 2019. A total of 53,763 credit card frauds were reported with a total loss of $135M.

In April 2020, early in the pandemic, The Wall Street Journal reported that fraud losses – including losses linked to credit and debit cards – cost banks, merchants, and in some cases, cardholders $16.9 billion in 2019, according to Javelin Research. Note the dramatically higher number here; $16.9 billion is more than 125 times greater than the $135 million, credit card-only figure from the Sentinel report.

Main Problem:

* Enormous credit card transactions are processed every day. Out of that, only less than 2% of transactions are fraudulent, which are need to sort out.
* The scammers always used adaptive techniques.
* There are some unclassified data because many of the fraudulent transactions are not reported or caught.
* Fraudulent detection model has to be the fastest possible process and most accurate.

Objective:

Improve the credit card fraud detection with 99% accuracy.

1. **Analytic Approach**

Base on the main problem and objective that have been defined, there are some parts that focus on classification models and anomaly detection.

Classification model

* Decision Tree
* Random Forest

Anomaly detection

* PCA Transformation
* Mahala Nobis Distance
* Local Outlier Factor

1. **Data Requirements**

Data for fraudulent credit card transactions:

* Type of purchase record
* Transaction’s record
  + Method of transactions
  + Sequence of transactions
  + Amount of transactions
  + Location of transactions
  + History of transactions
* User detail
* Data source in structure data

1. **Data Collection**

The data was collected from data source as structure data. Detail data will must include user, transaction, history, date data (daily, weekly, monthly, and yearly) transaction record and location.

1. **Data Understanding**

Study the dataset, check the missing data, formats for the dataset, clean and remove unnecessary data and read the dataset.

1. **Data Preparation**

EDA, importing libraries/modules, read dataset rows and columns, values, method for apply in dataset.

1. **Modelling**

Developing the model base on the chosen machine learning algorithm, training and testing the dataset and then implement classification model.

1. **Evaluation**

The results of the built model, compare the results, improve the results by tunning the parameters.

1. **Deployment**

Deploying the model on the new data.

1. **Feedback**

Getting feedback and comments from people in charge.

**References:**

1. <https://towardsdatascience.com/machine-learning-for-credit-card-fraud-detection-a-jupyter-book-for-reproducible-research-8ca5edad7b5d>
2. <https://castle.io/use-cases/transaction-fraud/?utm_source=google&utm_medium=cpc&utm_campaign=search_tf&gclid=Cj0KCQjwnJaKBhDgARIsAHmvz6f4JoaHJR6XnVqPj5kSsXHP0xx2MSDeP-N1OR1YjXWqo2HaY_s_fGUaAtQUEALw_wcB>
3. <https://towardsdatascience.com/credit-card-fraud-detection-9bc8db79b956>
4. <https://www.datasciencecentral.com/profiles/blogs/credit-card-fraud-detection-case-study-improving-safety-and>
5. <https://www.projectpro.io/project-use-case/credit-card-fraud-detection-classification-problem>
6. <https://www.researchgate.net/publication/344788652_CREDIT_CARD_FRAUD_DETECTION_USING_DATA_MINING_TECHNIQUES>
7. <https://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=4457&context=thesesdissertations>
8. <https://spd.group/machine-learning/credit-card-fraud-detection/>
9. <https://www.ijert.org/credit-card-fraud-detection-using-machine-learning-and-data-science>
10. <https://www.kaggle.com/c/1056lab-credit-card-fraud-detection>
11. <https://www.kaggle.com/merryyundi/credit-card-fraud-detection>
12. <https://www.kaggle.com/gpreda/credit-card-fraud-detection-predictive-models>
13. <https://www.kaggle.com/aniruddhachoudhury/credit-card-fraud-detection-99-accuracy>
14. <https://www.fico.com/blogs/credit-card-fraud-its-still-thing-and-big-ever>