DAB 402 – CAPSTONE PROJECT

ASSESSEMENT 2 – Formulation of analytics problem

Cyber Space Security in Banking (CSS)

Credit Card Fraud Detection



GROUP 13

|  |  |
| --- | --- |
| **NAMES** | **STUDENT ID:** |
| Jai Surya | 0731608 |
| Rajit Kumar | 0730468 |
| Amita Mehta | 0730478 |

***Date – 02/06/2020***

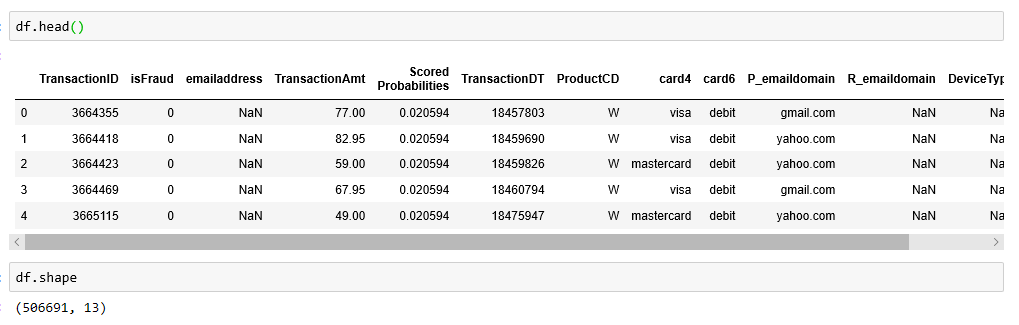
**Title:** Credit Card Fraud Detection

**Problem Statement:**

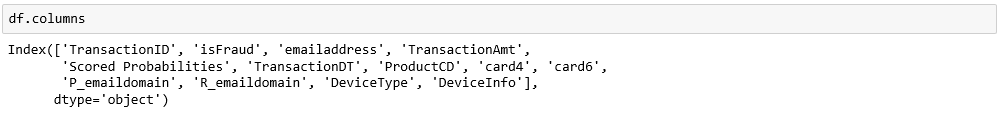
Electronic computing and communication pose some of the most complex challenges engineering has ever faced. They range from protecting the confidentiality and integrity of transmitted information and deterring identity theft to preventing the scenario recently dramatized in the Bruce Willis movie *"Live Free or Die Hard,"* in which hackers take down the transportation system, then communications, and finally the power grid. The most complex challenge engineering has ever faced is electronic communication and computing. **Cyber Crime** is the most serious problem in the present era. According to U.S.A. government more than 600 billion amount of fraud happened annually and it is increasing by 1.2 billion. And more than 200 billion amounts of fraud happened in the banking sector. South Africa has recently been afflicted by fraud in credit and banking information from online banking subscribers.

**Analytical Problem:**

* We got the dataset name hiwott-cyber-security-dataset. This dataset contains 50,6691 rows and 13 columns. We have more data for test and train the model to predicting the accuracy.



* The dataset contains the following columns



* According to above outputs the dataset is data driven problem and data-driven is a part of the analytical problem. To resolve this problem, we must use the analytical thinking. Analytical thinking helps to observe and identify this problem to develop more ideas about it. Gathering Information and developing solutions about the problem is also a part of the analytical problem. Analytical thinking is the ability to quickly identify cause and effect of this problem. Identify and examine the cause and effects while solving this problem. To gather the knowledge, we must do more research and information analysis.
* This dataset contains unnecessary columns and null values. So, we must do some cleaning and use pre- processing techniques for further analytics. And pre-procession is used to resolve the analytic problem. So, our topic relates to the analytical problem. This pre-processing and the cleaning of data will be done with the help of Excel tools.
* This dataset contains the huge number of rows. So, it is hard to analyses this data. And to create the co-relation between the feature we must use some analytics tools. To visualize the data, we can use the Tableau. As tableau helps to visualize the things in proper way and good manner. By plotting the graph, we can create and find the correlation between the features.
* Now we step into the python tool. After the visualization and pre-processing we must do the feature selection. There number of columns in each dataset so its little difficult to select the feature. We can use feature selection based on predicting our model.It is the process of reducing the number of input variables when developing a predictive model. And it helps us to improve performance of the model. In our project the main and measure feature are card number, account Number and card types. And use different modeling techniques. And consider the better one for further work.
* After feature selection we partition the data into test and train form for building different models. And select the best model which predict the best accuracy. We may use the regression and linear model.
* We have another dataset for credit card and bank fraud detection. So, we do clustering based on the feature. And create a column which tells whether the fraud is happened or not while the transaction. Which help us to find the location where the more fraud cases happened by correlating the fraud column with the region column. So, we can find some solution to prevent the peoples from fraud cases.

**Reference**

* Pdfs.semanticscholar.org

<https://pdfs.semanticscholar.org/f920/211b1e8a3d2c3e3ac89caea3ba3caba98fd9.pdf>

* Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier

Credit Card Fraud Detection

[https://www.kaggle.com/mlg-ulb/creditcardfraud](javascript:openWebLink('https://www.kaggle.com/mlg-ulb/creditcardfraud'))

* Detecting credit card fraud by genetic algorithm and scatter search

#### Detecting credit card fraud by decision trees and support vector machines

* [https://openaccess.dogus.edu.tr/xmlui/handle/11376/2366#sthash.sDqj3beI.dpbs](javascript:openWebLink('https://openaccess.dogus.edu.tr/xmlui/handle/11376/2366#sthash.sDqj3beI.dpbs'))
* Dataset got from data.world
* <https://data.world/hiwott/cyber-security-dataset>