**Fraud Detection: Self Organizing Map**

**Business Problem:** The dataset consists of fictional data of customers who filled an application for advanced credit card. The bank has approved the credit card to the customers as per their system and now they want to figure out those fraudulent customers who lied on their forms to get approved.

**Dataset:** I am going to use Statlog Australian Credit Approval dataset from UCI Machine Learning Repository. The dataset consists of 690 customers and 16. However, I am going to remove the variable “Class” that tells us whether the customer was approved by the bank or not and put it in a separate dataset so that we can know in the end the customers who are frauds and were approved for the credit card by using a reverse mapping function. The dataset consists of a good mix of categorical as well as numerical data. The dataset consists of no null values and there is no correlation within the attributes. So, we can move forward with applying the dataset to SOM model after normalizing the dataset.

A close up of text on a white background

Description automatically generated

Fig.1 Null values in the dataset

A screenshot of a cell phone

Description automatically generated

Fig.2 Correlation Matrix

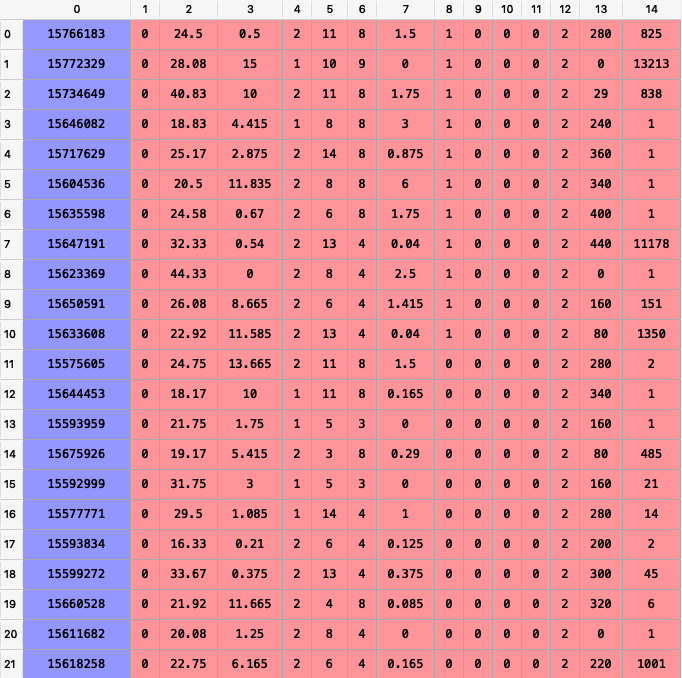
**Visualising the results:** I have used Minisom 1.0 package based on Numpy for implementing Self Organizing Map. Here, self-organizing maps is used for anomaly detection. We are going to use MID (Mean Interneuron Distance) to detect outliers. The higher the MID, the more the BMU will be separated from all other neurons and hence will be an outlier. We are going to use colors in the visualization to detect outliers. Here, we have assumed that most of the applications filled by the customer are correct and so the applications that are wrong will be outliers and that is how we will be able to capture fraud customers. As seen from the below map, we will select highest MIDs (shown in white) and that will be our fraud customer. Then we will use reverse mapping function to know the cutomer ID of those customers in the BMU with highest MID.

**Fraudulent Customers**

**A screenshot of a cell phone

Description automatically generated**

**Fraud Customer List**

****