**Introduction and Discovery**

Ransomware Bitcoin dataset is related to the ransomware attacks and bitcoin transaction related to them. Ransomware is a type of malware which encrypts (lock) the victim file and then they will demand ransom payment in bitcoin to decrypt files back. The dataset contains information about various ransomware attacks on various addresses.

**Framing the problem:**

1. Identify patterns and trends: We can identify patterns and trends in ransomware attacks such as some may use looped transaction while demanding payment which means it involves same address as both sender and receiver.
2. Cluster Analysis: group similar ransomware attacks on based of attack patterns and other vectors.
3. Classification Analysis: Performing Classification and telling about which group attack pattern is this.

**Initial Hypothesis:**

1. Certain types of ransomware attacks may follow specific pattern such as higher value of length, more weight, and more looped transactions.
2. Specific pattern result in higher income transaction. So, developing machine learning models to learn more about attacks to detect and prevent these attacks.
3. Ransomware attacks are more in certain industries such as finance and health sectors due to high income and flow of data.

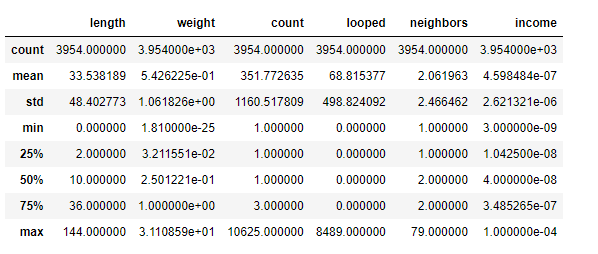
**Data Preparation**

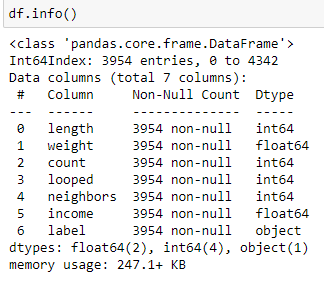
**Data Inventory :** The dataset used for creating machine learning models is information on bitcoin transfer from specific time period. Data is obtained from UCI Bitcoin Heist Ransomware dataset <https://archive.ics.uci.edu/dataset/526/bitcoinheistransomwareaddressdataset> It has following features.

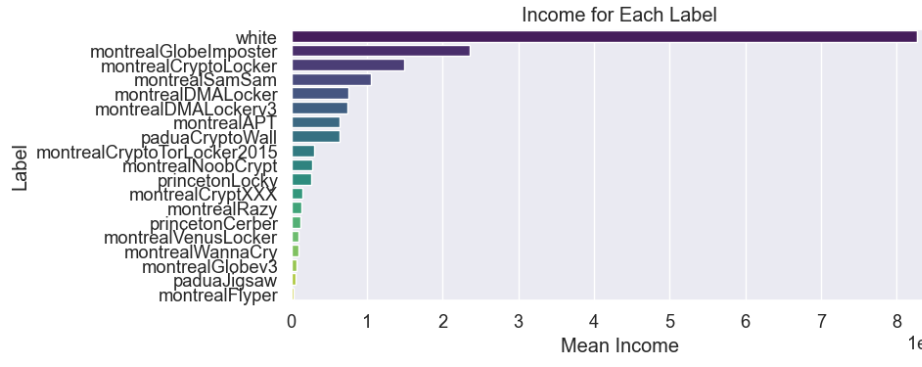
1. address: String. Bitcoin address.
2. year: Integer. Year.
3. day: Integer. Day of the year.
4. length: Integer.
5. weight: Float.
6. count: Integer.
7. looped: Integer.
8. neighbours’: Integer.
9. income: Integer. Satoshi amount (1 bitcoin = 100 mill satoshis).
10. label: Category String. Name of the ransomware family (e.g., Cryptxxx, crypto locker etc) or white (i.e., not known to be ransomware).

**Data Processing:**

1. There are no duplicated and no null values in dataset.
2. The income feature is represented in Satoshi (1 bitcoin = 100 million Satoshi’s). So value of income is divided by 100 million to get bitcoin value.
3. Dummy values for field label is added to handle categorical value of different types of labels there are 12 different types of labels.
4. The address and day column is also removed from the dataset.





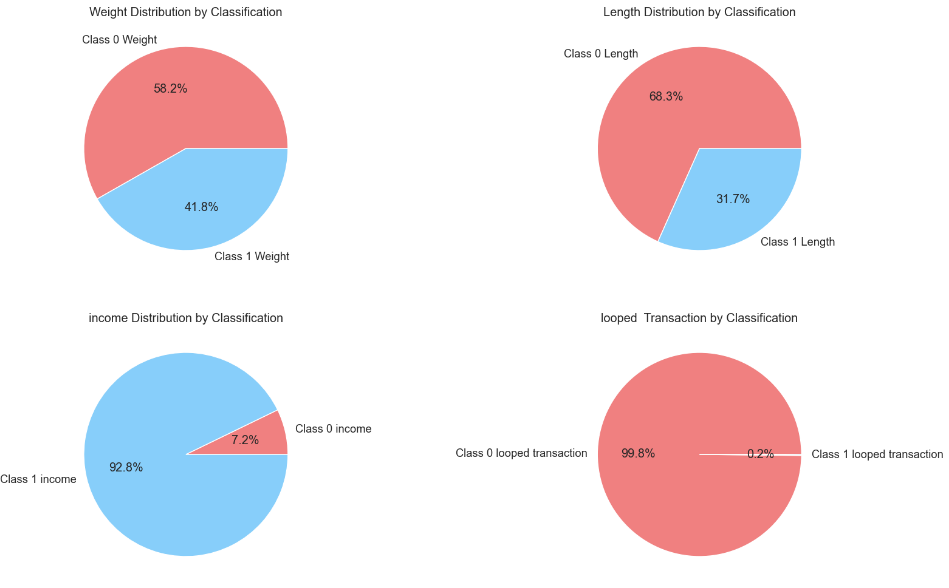


**Model Implementation:**

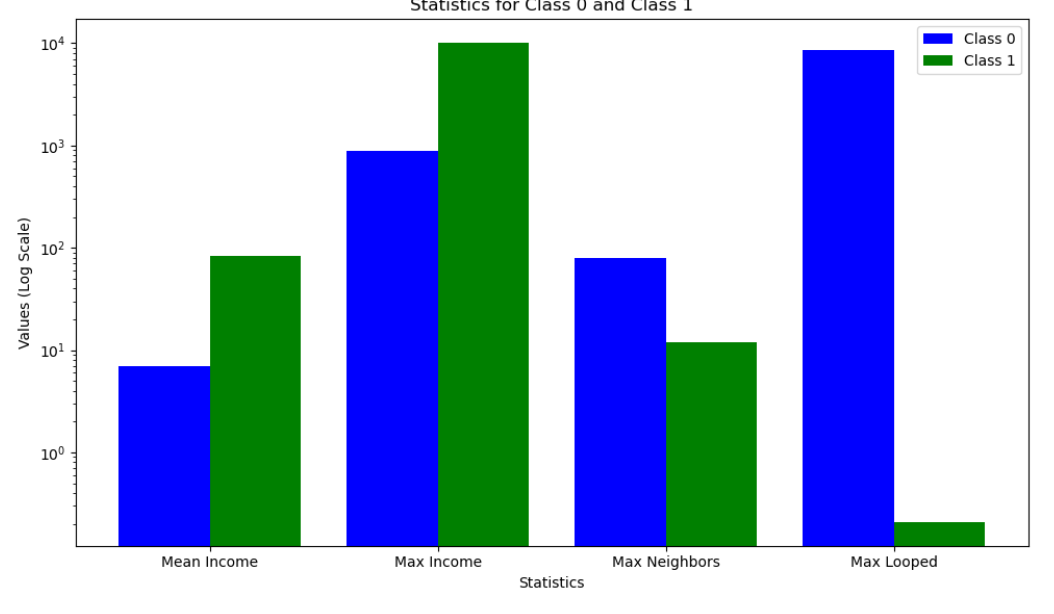
1. **Clustering Techniques**:
   1. **KMeans =** KMeans is a popular unsupervised clustering model which main aim is to divide data points into K clusters. In this case K = 2.
   2. **Agglomerative Clustering =** This is Hierarchical Clustering that starts with each datapoint and then sub merges the closest one until the stated number of cluster matches
   3. **DBSCAN =** Density based spatial clustering with noise. It groups datapoint together based on their respective densities.



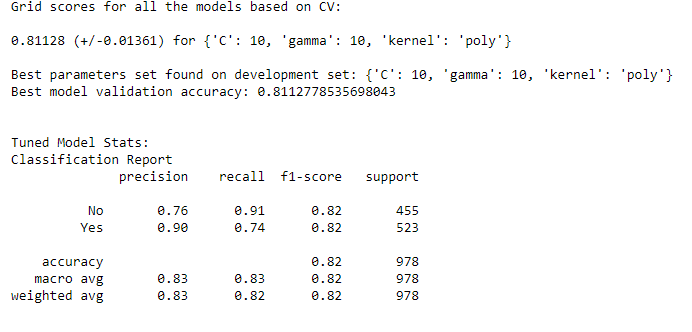
**Justification** = Since our dataset have following features, we can determine different attack patterns and groups. Like shown picture we can see different attack patterns by two different classes. The clustering process helps us to determine different clusters of transaction based on similar attack patterns like in this case Class 0 uses more weight of bitcoin, more length and one thing to note is they uses looped transaction very much which is used for fraudulent activities and indicates more sophisticated money laundering activities as compared to class 1 which contains white transactions mostly.



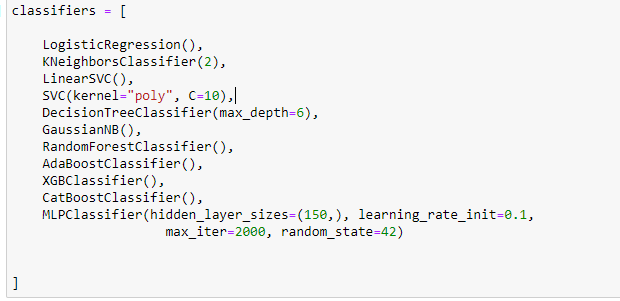
1. **Classification Techniques:** After clustering the data , each cluster is different subset to used further for classification and analysis. Now after clustering we have new feature or target which we called Classification. We have two classes 0 and 1 (0 have all the fraudulent activities) and class 1 have not fraudulent activities.



* 1. **Logistic Regression =** Binary classifier that models the probability using activation function to tell probability which class model belongs.
  2. **K-Nearest Neighbors (KNN) =** non parametric classifier which gives label based on majority.
  3. **Support Vector Machines (SVM) =** Used GridSearchCV for finding the best hyperparameters. As it is exhaustive search so I have commented out the code for now to perform fast computation and execution. Below are the best parameters.



* 1. **Decision Tree**.
  2. **Gaussian Naive Bayes =** Probability Classifierdoes not give good result.
  3. **Random Forest =** ensemble method to combine multiple models
  4. **AdaBoost, XGBoost (Most Efficient) , CatBoost**
  5. **Multilayer Perceptron (MLP)** = With given below hyperparameters and hidden layer it gives the good result.



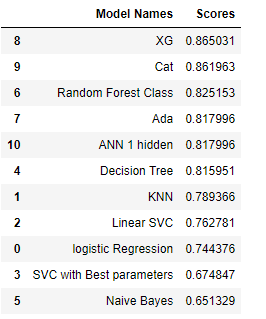
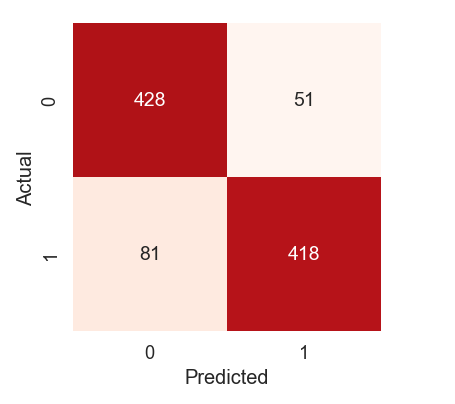
**Hypothesis Testing =**

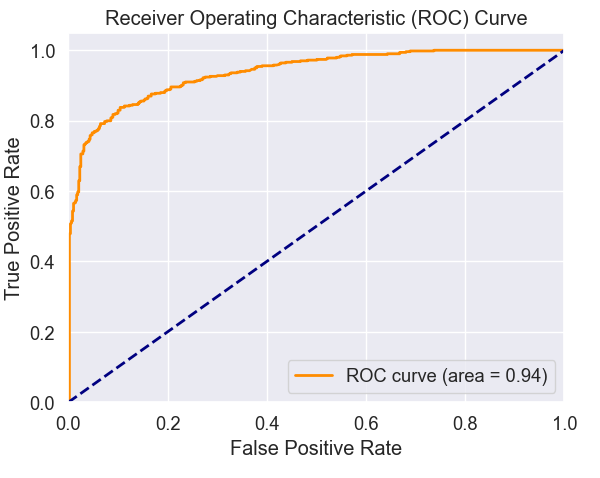
1. Transaction within specific cluster have different attack vectors and pattern related to labels in dataset.
2. Using Clustering Algorithm on following transaction can reveal new patterns and typologies in the network.
3. Further using Classification after doing clustering can help perform better due to underline patter in each of the clusters present.

**Result Interpretation and Implication**

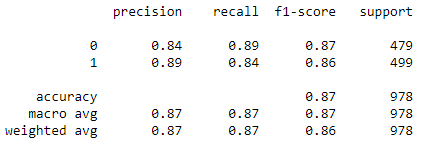
XGBoost model achieved the highest accuracy of 86.5%. Where as CatBoost is also like XGBoost. Moreover, Naïve Bayes performs the worst with accuracy of 65%. Random Forest, MLP neural network and Ada Boost also show a good score of around 83%. For the **confusion matrix** of our best model which is XGBoost we can say that

1. True Positive (TP) = 418 are correctly predicted positive class
2. False Positive (FP) = 51 Incorrectly predicted positive class
3. True Negative (TN) = 428 Correctly predicted negative class
4. False Negative (FN)= 81 are incorrectly predicted.



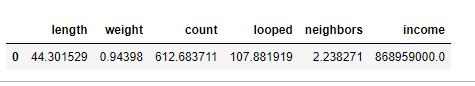
1. **Does the model appear valid and accurate on the test data**? Yes, the model appears valid and shown accuracy of 86.5% for classification.
2. **Does the model output/behaviour make sense to the domain experts?** Yes, model is classifying according to specific pattern or behaviour of attackers and telling us whether it is a white (non-fraud) transaction or fraud transaction.
3. **Do the parameter values make sense in the context of the domain?** GridSearchCV it is a exhaustive search to find the optimal hyperparameters for SVM models. But it was commented out due to its high computation and search.
4. **Is the model sufficiently accurate to meet the goal?** Yes, I think model is sufficiently accurate providing accuracy of 86.5% to meet the goal of classifying transactions based on the specific pattern used by hackers.
5. **Does the model avoid intolerable mistakes?** The precision for class 0 is 0.84 while for class 1 is 0.89. More high precision show that model make few false positive predictions. And higher recall also shows it correctly classify large proportion of data.



1. **Are more data or inputs needed?** No large data is not required. This ransomware dataset is itself very large of 10 lakh rows I resampled it to ~ 4k. And received satisfactory result.
2. **Is a different form of the model required to address the problem?** By looking and analysing at different performance factors of model such as accuracy, precisions, recall and f1-score new model is not needed to address the problem. Our model is meeting the project requirement or our hypothesis and it does not show any sign of overfitting and underfitting

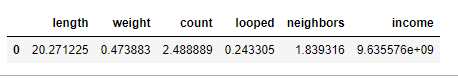
**key findings and major insights derived from the analysis =** The main key finding from the analysis is specific attack patterns used by the hackers such as high looped transaction and high length of encryption. First doing clustering according to their specific behaviour and after performing clustering doing classification on model to detect whether it is white transaction (no fraud) or fraud transaction performed by various labels.

**Out of sample prediction =** Taking data from both classes mean value and testing the best model which is XGBoost on it. Below sample dataset from class 0 mean values. If we look at its result its predicting right. It belongs to class 0





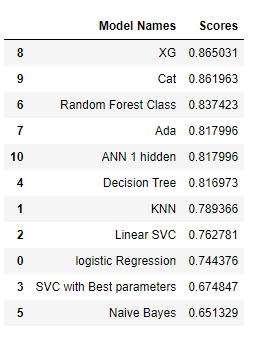
The next sample is generating from class 1 mean values. If we look at result its predicting correct for class 1 by looking at following features and patterns which create from them.





**Concluding remarks**

1. **Analytics process** 
   1. data processing = For data processing missing values were handled and income features was converted from Satoshi to bitcoin (1 bitcoin = 100 million Satoshi).
   2. Model selection and evaluation = For clustering techniques we used 3 algorithms KMeans, DB Scan and Agglomerative clustering from them selecting the best out of them on basis of silhouette score, calinski harabasz score and davies bouldin score. We selected Agglomerative on basis of its score and divide the dataset into two clusters 0 and 1. After clustering performing classification on them to classify its cluster whether it is class 0 (fraud transactions) or class 1 (no fraud transactions or white transactions) we used various models for classification such as XGBoost, CatBoost, Random forest classifier , Ada Boost , ANN with one hidden layer, Decision tree, KNN, Linear SVC , logistic regression , SVC with best parameters which are determined by GridSearchCV and Naïve Bayes below are the result of following classifier models in descending order of score.



Deep learning model with 3 hidden layer is also tried for classification to check the accuracy and difference between other classifier models it gives accuracy of 81.2% which is not bad. We can further increase hidden layers and tune it. I just tried this for leaning.



**Managerial Implications =** This model highlight need of using machine learning in cyber security field to detect ransomware detection. Managers can use this information to boost their cybersecurity strategies and prevent people from fraud transactions. They can also easily classify whether transaction is fraud or white transactions and if its fraud we can further detect patterns which group was responsible for this seeing their various patterns. Managers can also fine tune their incident response. And they can prioritize their response efforts based on model predictions.

**Video presentation link =** <https://www.youtube.com/watch?v=PYQjNt8PhEU>

**Extra Part outside project =** Integrating Machine learning models to frontend so non-technical people can also use feature of ML.

* 1. **Joblib model , Created Flask Server post Api.**
  2. **React App**

