

Mobile CrowdSensing

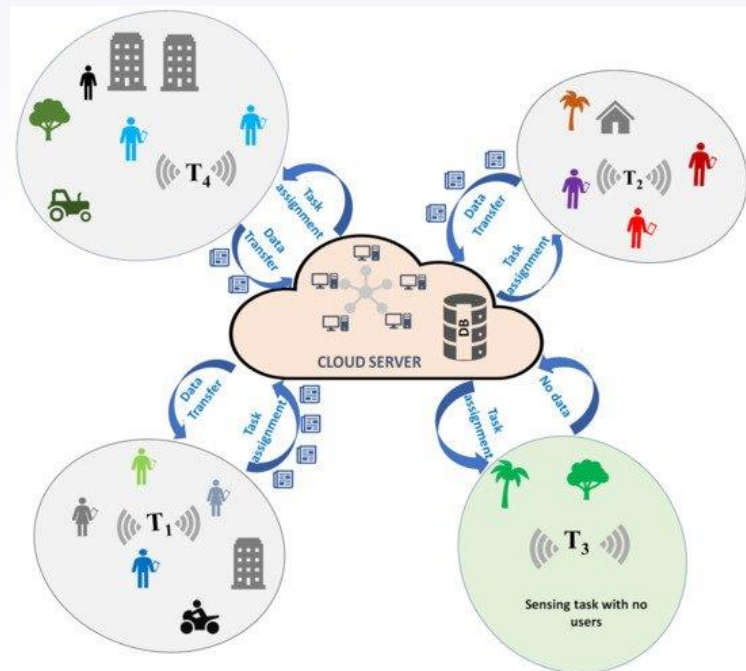


▶ Agenda

- ▶ Introduction
- ▶ Dataset Distribution
- ▶ Algorithms
- ▶ Over/Under sampling and PCA
- ▶ MLP classifier along with clustering approach.
- ▶ Conclusion

introduction

- ▶ mobile crowdsensing (MCS) is captivating growing attention
- ▶ modern smartphones are equipped with unprecedented sensing, computing, and communication capabilities that allow them to perform more complex tasks besides their inherent calling features

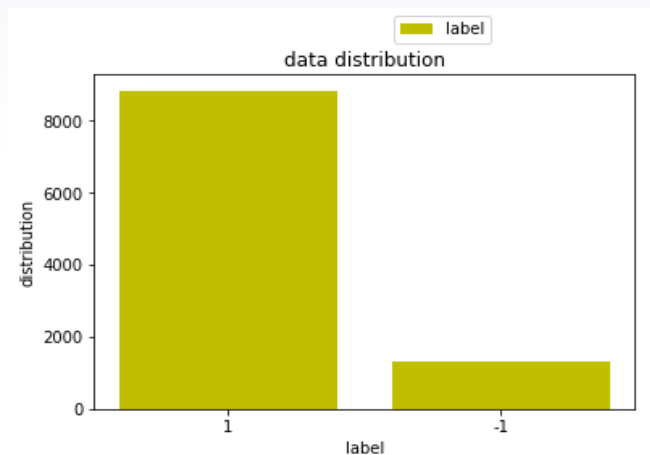


Dataset Distribution

```
1    8839
```

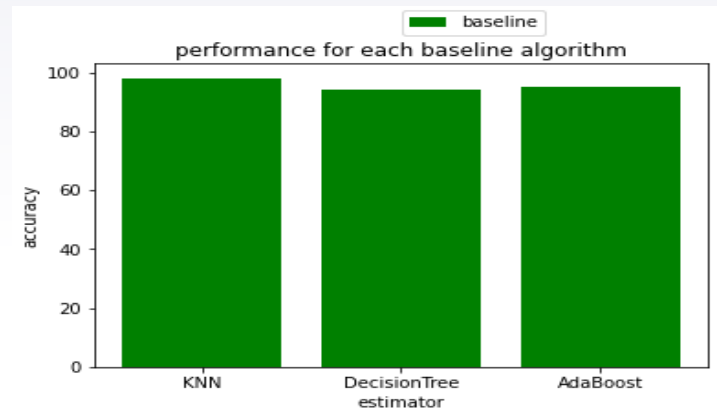
```
-1   1300
```

```
Name: Legitimacy, dtype: int64
```



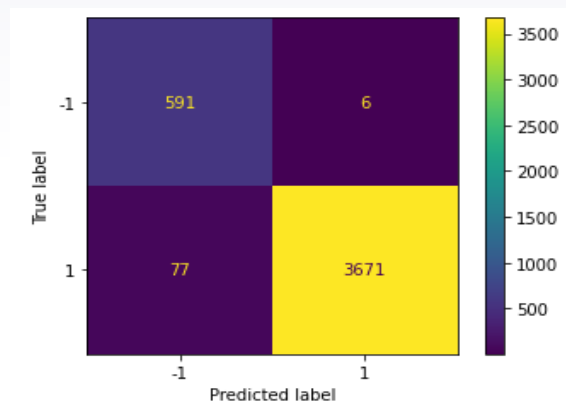
ML algorithms

- ▶ Applying three of the ML algorithms:
KNN, Decision Tree, AdaBoost
- ▶ Accuracy of three of the ML algorithms:

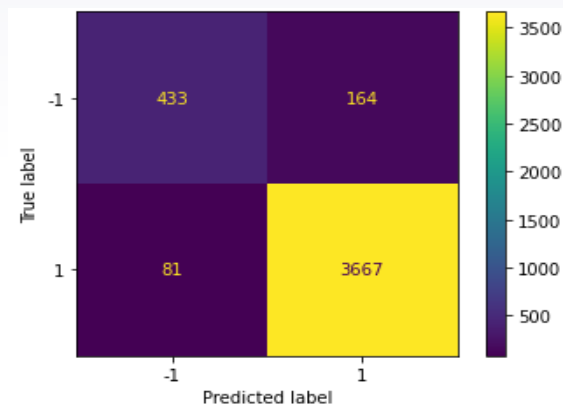


Algorithms	KNN	DecisionTreeClassifier	AdaBoostClassifier
Accuracy	98.1%	94.4%	95.3%

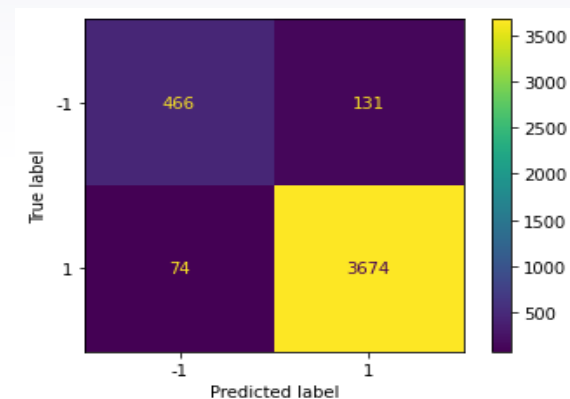
ML algorithms (confusion matrix)



KNN



Decision Tree



AdaBoost

Over/under-sampling with ML algorithms

Applying over/under-sampling with three algorithms:

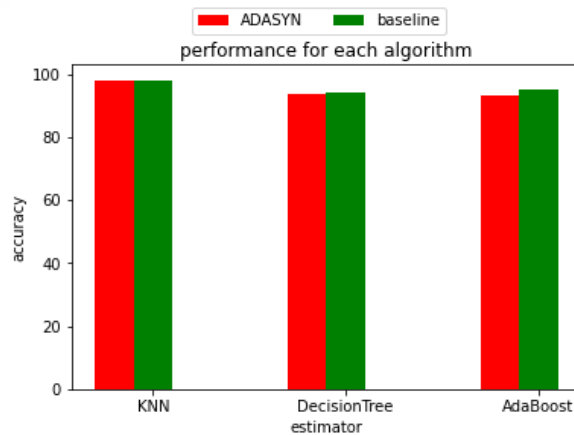
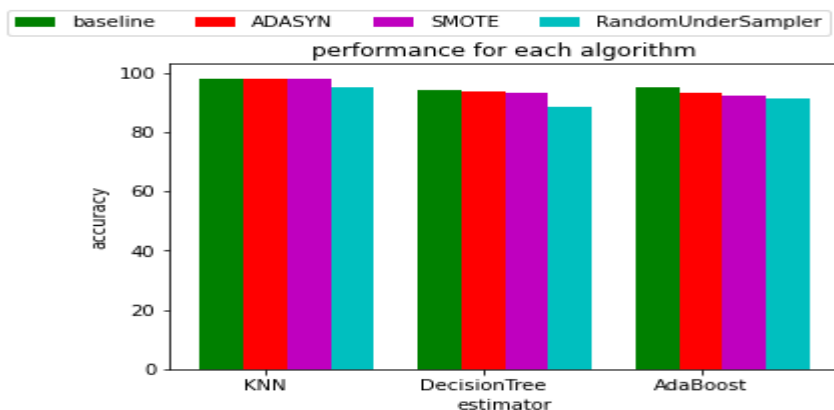
- ADASYN
- SMOTE
- Random Undersampling

	ADASYN	SMOTE	Random Under-sampling
KNN	97.79%	97.76%	95.30%
Decision Tree	93.78	93.23%	88.28%
AdaBoost	93.18%	92.17	91.30%

Plotting bar chart

Plotting bar chart to compare between baseline and over/under-sampling.

over-sampling (ADASYN) is selected for each ML method.



Dataset before and after over/under-sampling

Before over/under-sampling

```
1    8839  
-1   1300  
Name: Legitimacy, dtype: int64
```

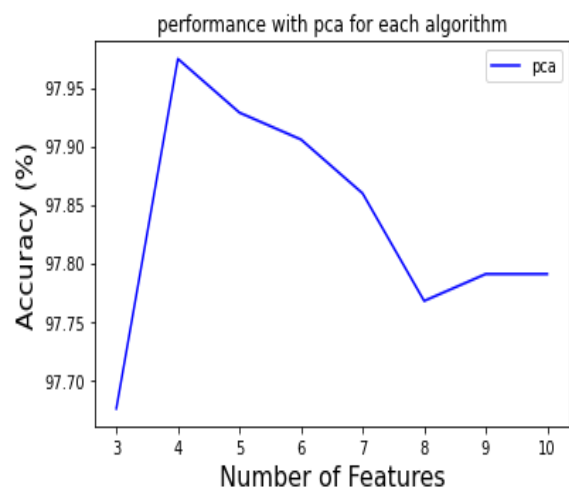


After over-sampling

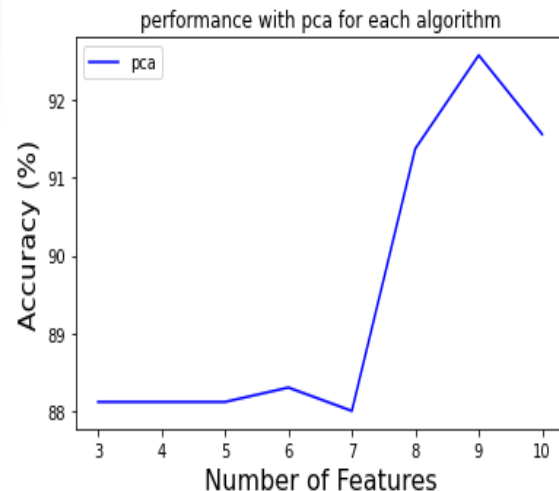
```
-1    8879  
1     8839  
Name: Legitimacy, dtype: int64
```

PCA with ML algorithms

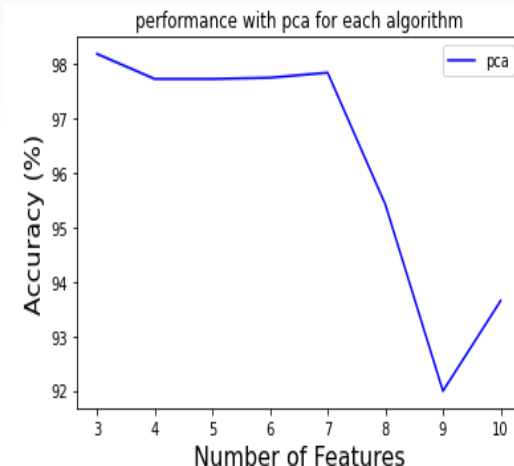
Applying feature selection and dimension reduction methods by PCA



Maximum accuracy: 98.0%
Best number of features: 4



Maximum accuracy: 92.6%
Best number of features: 9

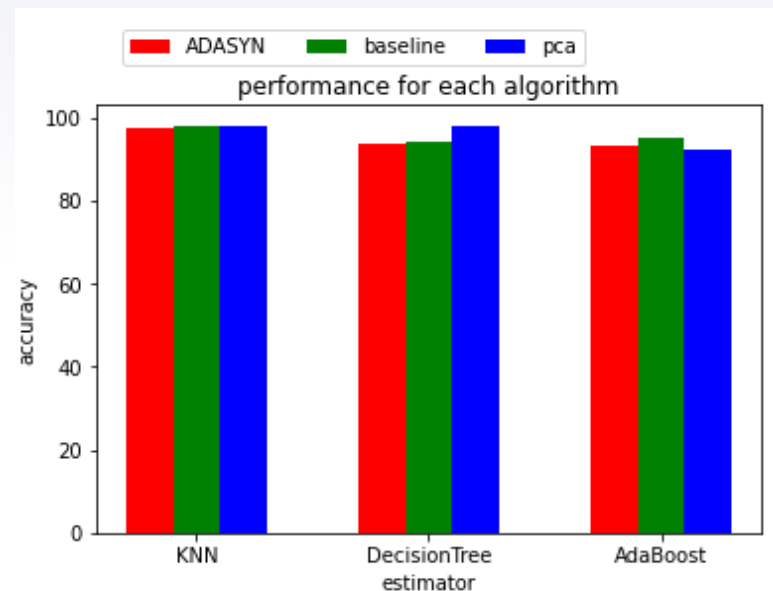


Maximum accuracy: 98.2%
Best number of features: 3

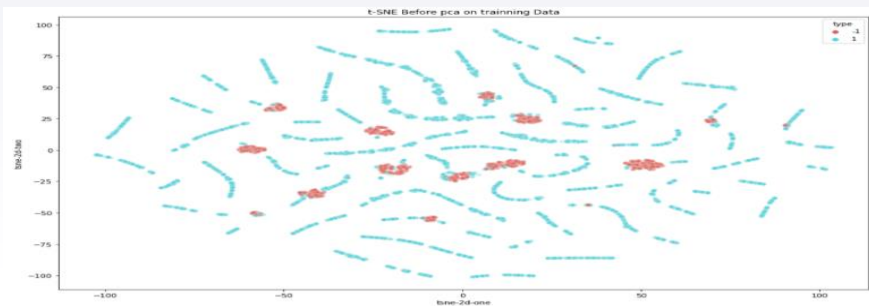
bar-chart

Comparing between baseline, PCA, ADASYN.

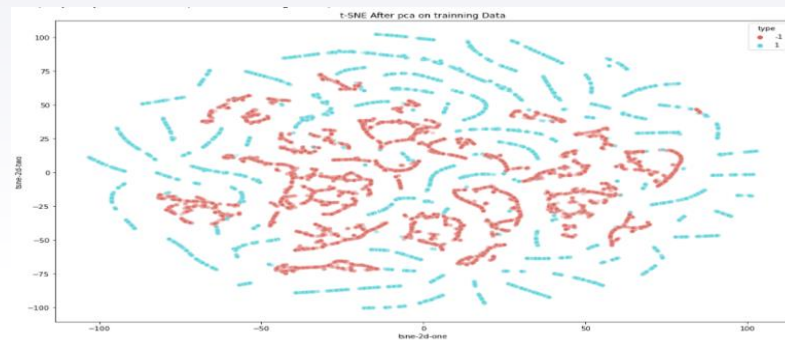
Obtaining PCA with Decision Tree is the highest accuracy.



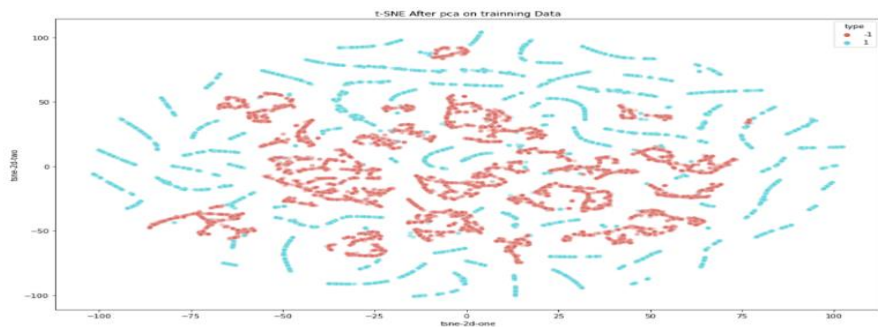
t-SNE



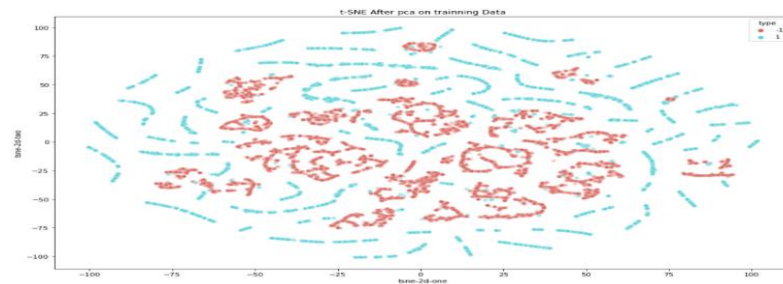
Before pca



After pca=3



After pca=4

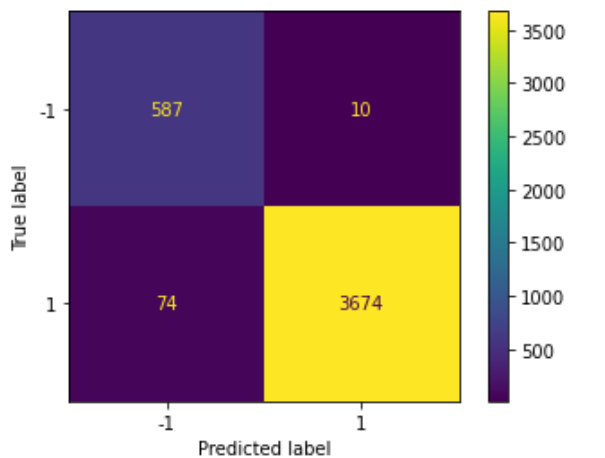


After pca=9

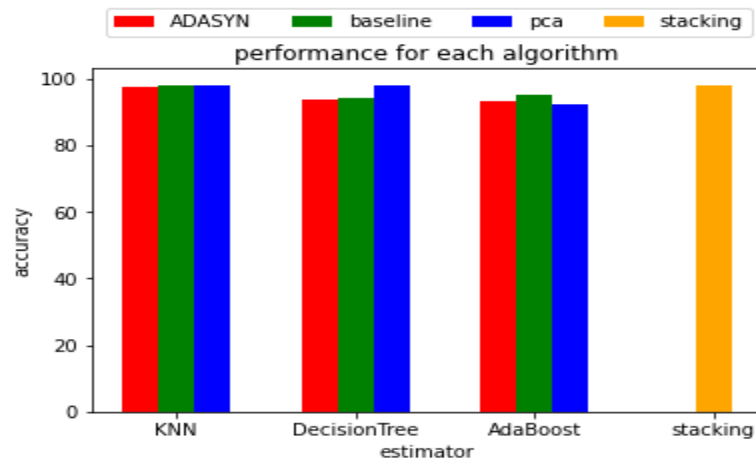
Stacking

Applying pipeline between three models with PCA .

Applying stacking to aggregate between three models by pipeline.



Accuracy=98.1%

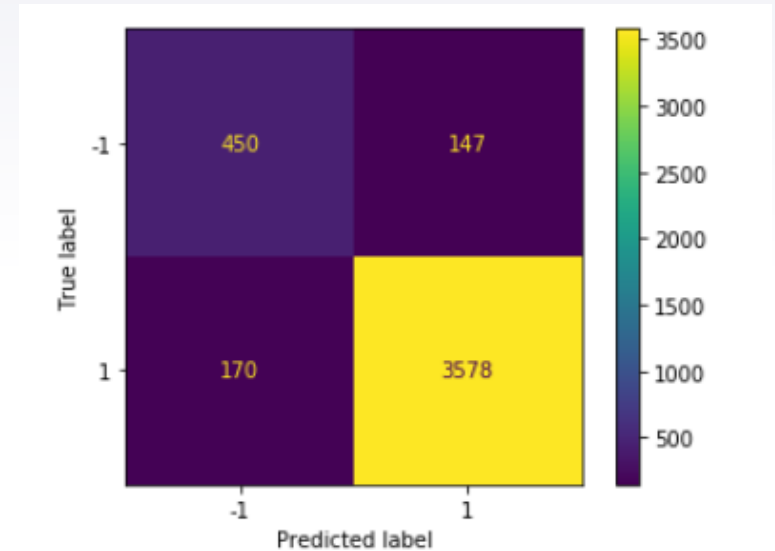
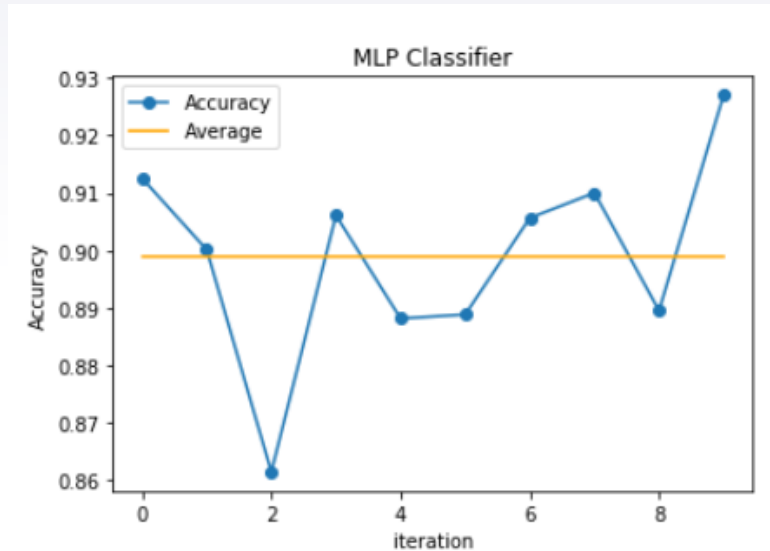


Comparing between baseline, pca, over-sampling and stacking



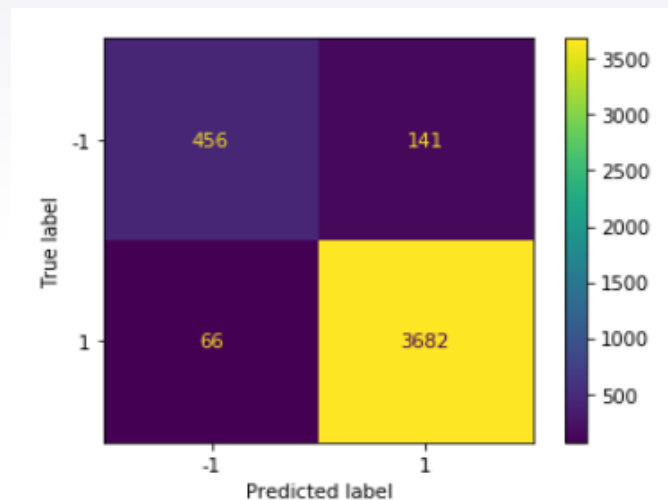
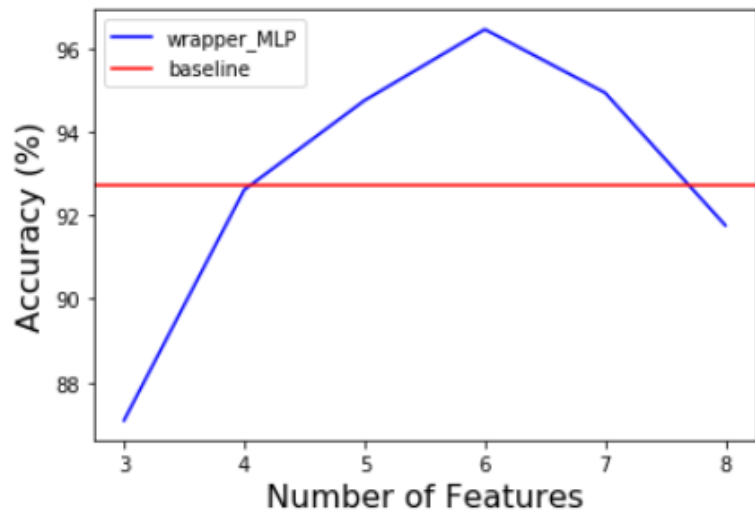
MLP classifier along with
clustering approach.

MLP classifier along with clustering approach.



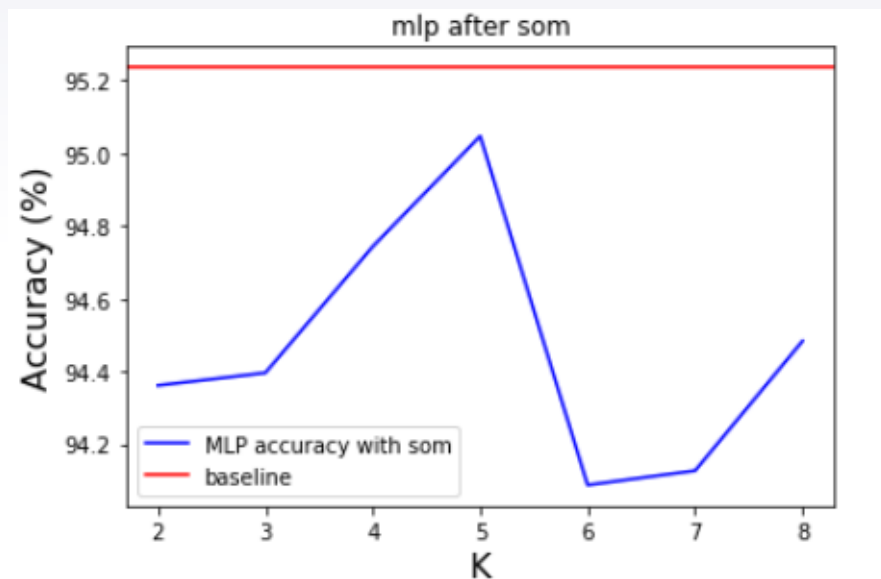
After applying MLP 10 times, we got the average accuracy (**89.8%**), and the highest accuracy equals (**92.7%**)

Feature Selection



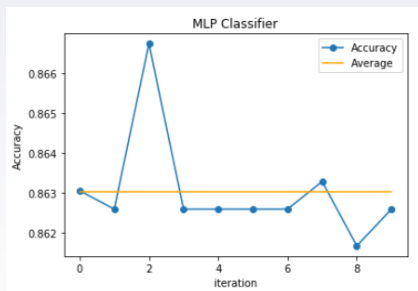
By applying the SFS we found the best accuracy equals (96.5%) with Num of feature (6)

Clustering-SOM

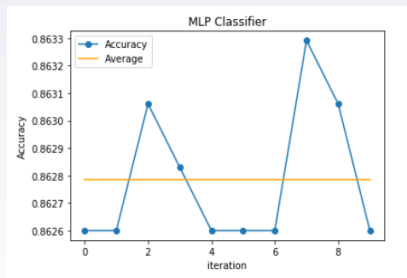


By applying the SOM, we get the heist accuracy equals (95.59%)

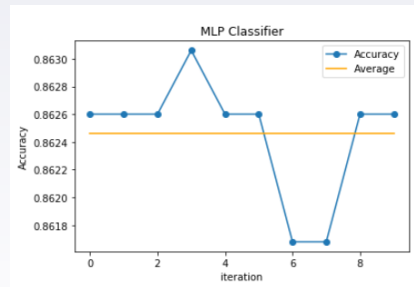
Tuning Hidden Layer



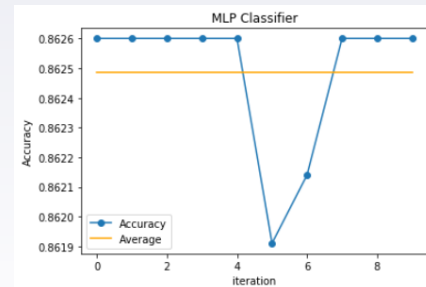
Hidden Layer = 3
Accuracy = 86.30%



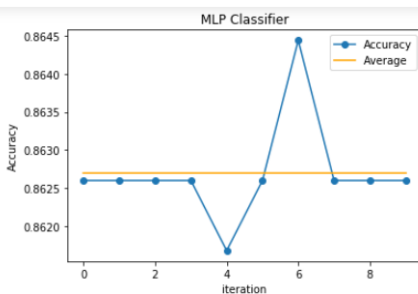
Hidden Layer = 4
Accuracy = 86.28%



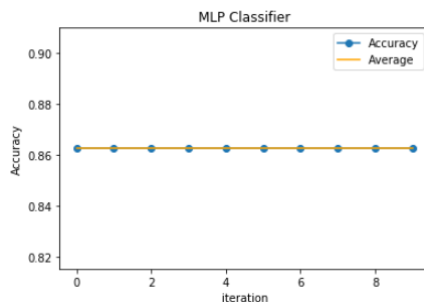
Hidden Layer = 5
Accuracy = 86.25%



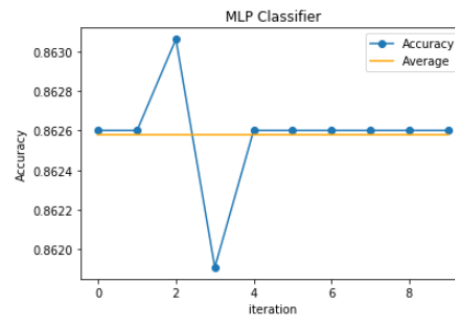
Hidden Layer = 6
Accuracy = 86.25%



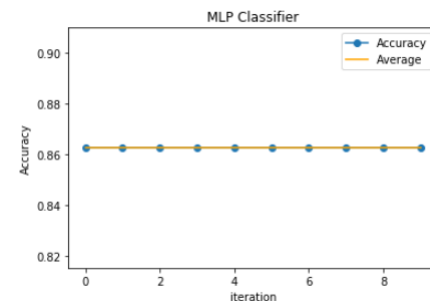
Hidden Layer = 7
Accuracy = 86.27%



Hidden Layer = 8
Accuracy = 86.26%

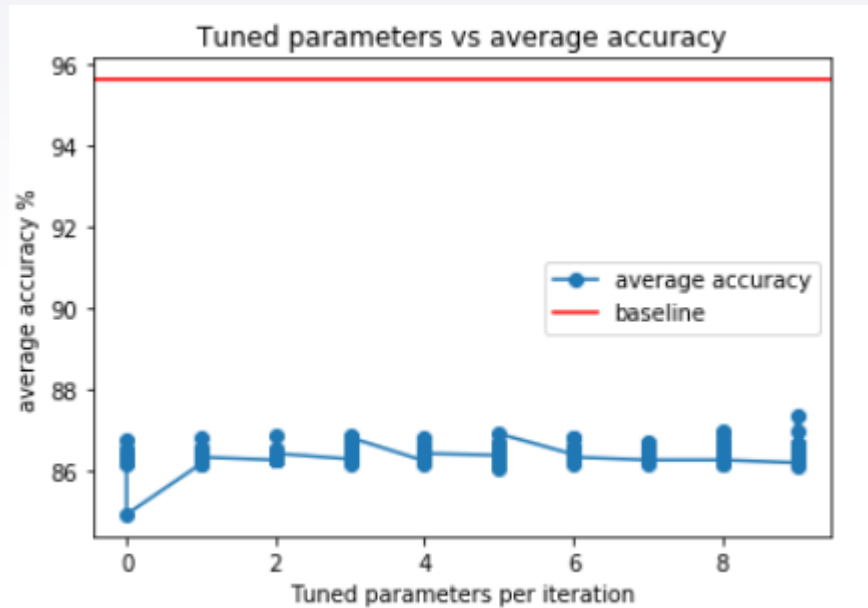


Hidden Layer = 9
Accuracy = 86.26%



Hidden Layer = 10
Accuracy = 86.26%

Tuning Num of Neurons



Accuracy: 86.26%

Conclusion

to sum up, what we did, We get KNN in baseline has the highest accuracy, the best solution to imbalance data is ADASYN, the best number of features in PCA with KNN =4, decision tree = 3 and AdaBoost =9 , the stacking and PCA with decision tree have the highest accuracy. When going through the MLP classifier we got the highest accuracy after applying the wrapper method with Sequential feature selection equals 96.5 with the number of features equals 6.

Thanks!