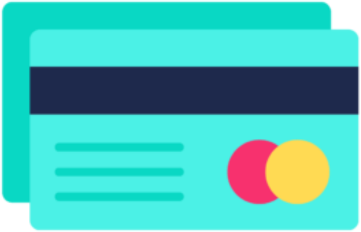
Classifying Fraudulent Transactions on Credit Cards

Ryan Gascoigne-Jones

December 2023

Introduction



- Credit card companies want to detect fraudulent transactions before they are validated.
- The easiest way to do this is to have a model automatically detect if something seems fraudulent.
- The cardholder can then be asked for extra verification.

Dataset

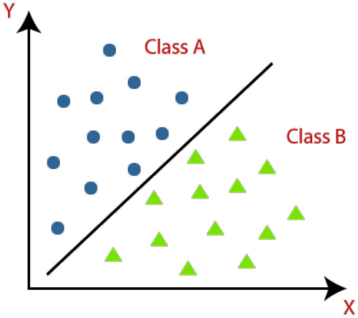
	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0

- Transactions made in 2 days by Europeans during September 2013.
- Contains 284,807 anonymised transactions.
- Contains 492 fraudulent transactions.

Includes:

- Monetary value of transaction.
- Time of transaction.
- Whether the transaction was fraud or genuine (Class).
- 28 other abstract features which have been generated by PCA.

Classification



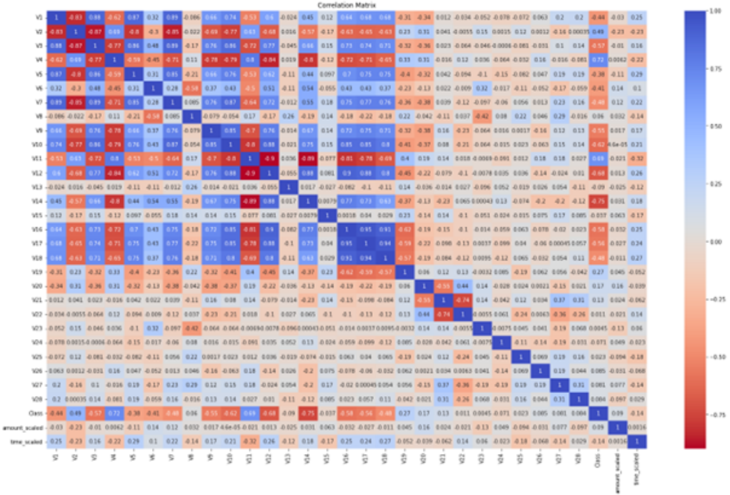
- Only need to predict whether a transaction is fraud or not.
- Model will aim predict what class a transaction is.

Data preprocessing

- First I randomly undersampled the genuine transactions so the dataset had a class balance.
- This gave a sub-sample dataset of 984 samples, 50% genuine, 50% fraudulent.
- I also duplicated this sub-sample and cleaned the extreme outliers.

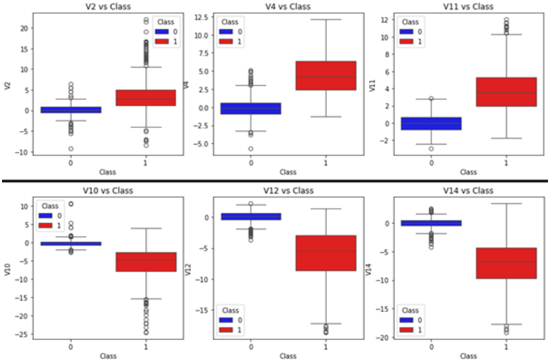
Feature Selection

Feature Correlation matrix



- V2, V4, V11 were the most strongly positively correlated.
- V10, V12, V14 were the most strongly negatively correlated.

Feature Selection



- These features relationships and distributions were then analyses with respect to the class of each transaction.
- V4, V11 and V14 were found to be the best fitting features to use in models.

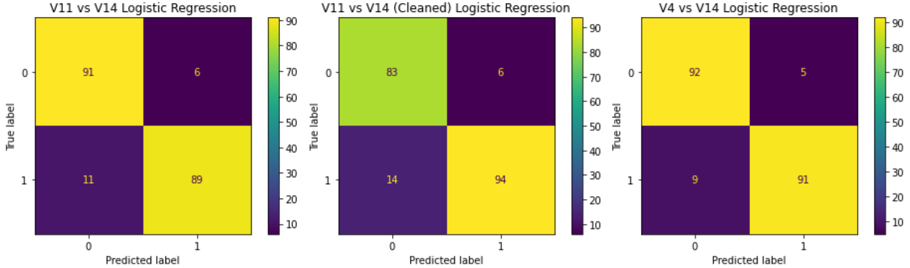
Logistical Regression

- Finds relationship between 2 features by using a sigmoid function to predict what class a datapoint belongs to.
- I created logistic regression models using V14 and V11, V4 and V14, and a cleaned version of V14 and V11.

```
Precision:  0.9368421052631579
Recall:  0.89
F1:  0.9128205128205129
Accuracy:  0.9137055837563451
```

```
Precision:  0.94
Recall:  0.8703703703703703
F1:  0.9038461538461539
Accuracy:  0.8984771573604061
```

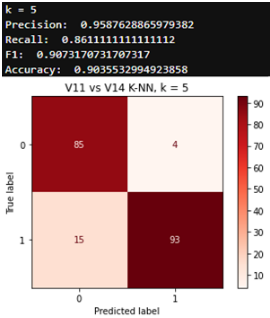
```
Precision:  0.9479166666666666
Recall:  0.91
F1:  0.9285714285714285
Accuracy:  0.9289340101522843
```

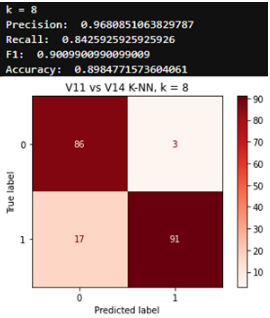


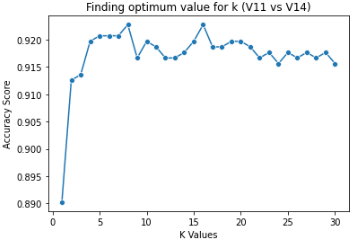
- The best model was the one using V4 and V14.

K-Nearest Neighbour

- K-NN aims to predict what class a specific datapoint is by looking at the class of the datapoints within its neighbourhood.
- The value of k determines how many reference points are in a points neighbourhood.



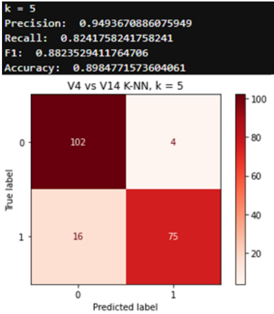


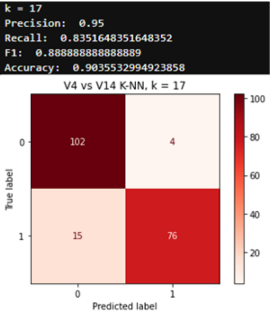


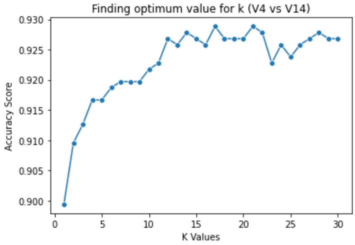
- First I used V11 and V14 with an initial k value of 5.
- I then ran a cross-validation optimiser to find the optimum value of k.

K-Nearest Neighbour

- I started the K-NN model using V4 and V14 with the same k value.
- In this case there was a bigger improvement when the same k optimiser was ran and found the best k value to be 17.
- This created a model that had better precision, accuracy and recall.



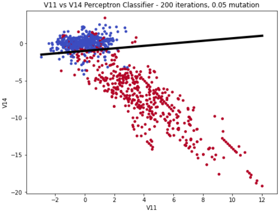




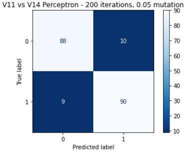
- However this was still a worse model than the one using V11 and V14.

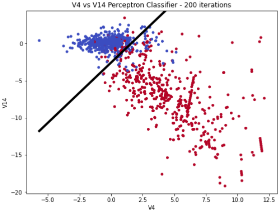
Perceptron

- I tried using a perceptron, which is a single layer neural network.
- It works by trial and error and attempts to increase the accuracy of a classification by mutating it slightly every iteration.
- I tested this using V11 and V14 as well as V4 and V14. 200 iterations. 0.05 mutation value.

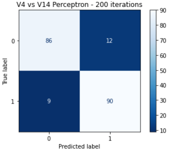



```
Precision:  0.9
Recall:  0.9090909090909091
F1:  0.9045226130653266
Accuracy:  0.9035532994923858
```



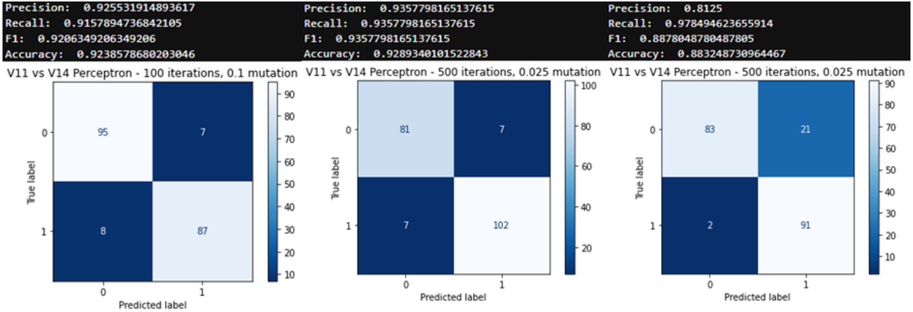



```
Precision:  0.8823529411764706
Recall:  0.9090909090909091
F1:  0.8955223880597014
Accuracy:  0.8934010152284264
```



Perceptron

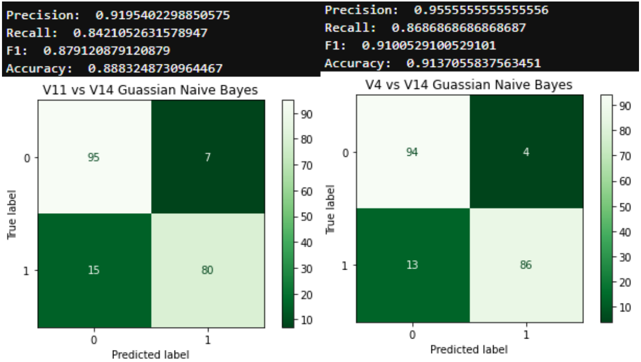
- The perceptron using V11 and V14 seemed to create the best model.
- I then attempted to optimise this perceptron.
- This gave some models with very high metrics.
- But other times it was ran, it created some very poor models.
- One model gave a recall of 97.8%



- Although this could result in models better than others I have created so far it is still too unreliable with a very high model variance.

Naive Bayes

- Naive Bayes is more suited to categorical input values, but they can also prove efficient with relatively small training datasets.
- I applied this model to the pairs V11 and V14 as well as V4 and V14.



- V4 and V14 was the best with a very high precision but the recall is still lower than previously generated models.

Conclusion: Models

- Best model is the logistic regression model using V4 and V14.
- It has the best recall of all the reliable models at 91%.
- Recall is perhaps the most important for these models as we want to reduce the number of false negatives.
- Perhaps other models could prioritise a much higher recall at the expense of precision and accuracy.
- The perceptron can create better models, but it is very inconsistent and unreliable.
- K-NN and Naive Bayes failed to create models that were any better than those previously created.
- V4 and V14 did prove to have the most representative relationship when it came to predicting the class of a transaction.

Conclusion: Dataset

- Most models have a high variance.
- This is because of the massive undersampling required to create a class balance.

How the dataset could be improved to improve the models

- A larger dataset with many more fraudulent transactions.
- A more recent dataset.
  - As there is now more transactions in total as well as the proportion of them being online changing considerably.