1. Fraud detection

   Analysis using public datasets from:

   https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

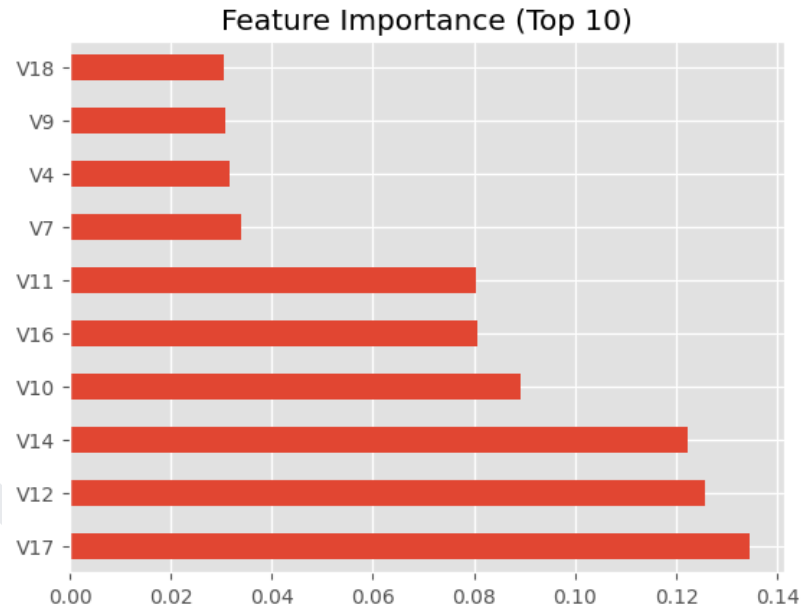
   Dataset preview:

	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

   Modelling: Random forest

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56864
1	0.94	0.83	0.88	98
accuracy			1.00	56962
macro avg	0.97	0.91	0.94	56962
weighted avg	1.00	1.00	1.00	56962

   Feature Importance Visualization:



   1. Most Influential Features:

      V17, V12, and V14 are the most important features in fraud detection, each contributing the highest to the model's predictions.

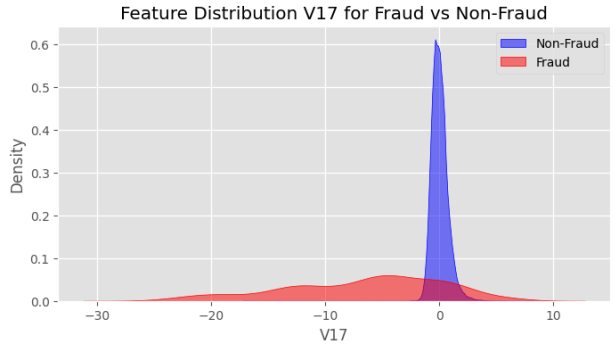
      This means that the values of these variables statistically best differentiate fraudulent and non-fraudulent transactions.

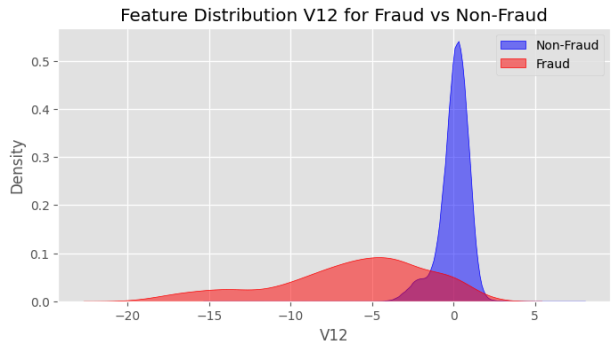
   2. Low-Influence Features:

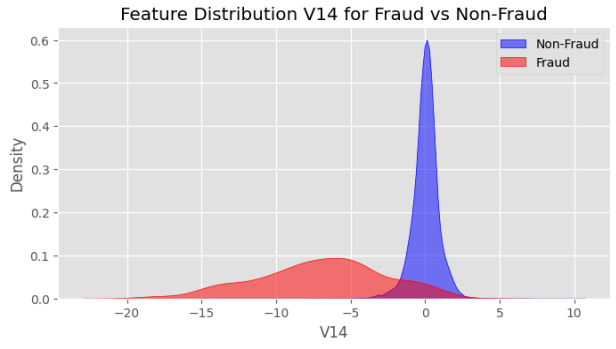
      Features such as V7, V4, V9, and V18 have relatively low importance values among the top 10, but are still more significant than other features that do not appear in the graph.

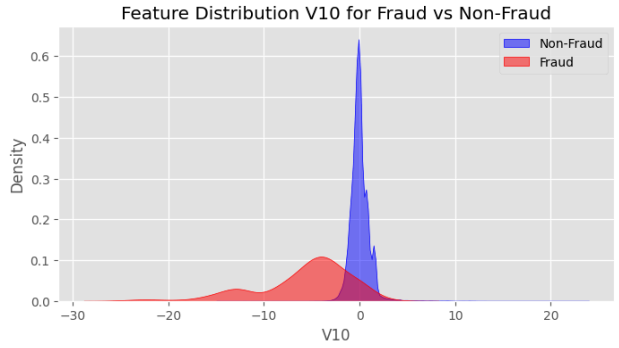
      This means that their contribution to the model's decisions is still significant, but not dominant.

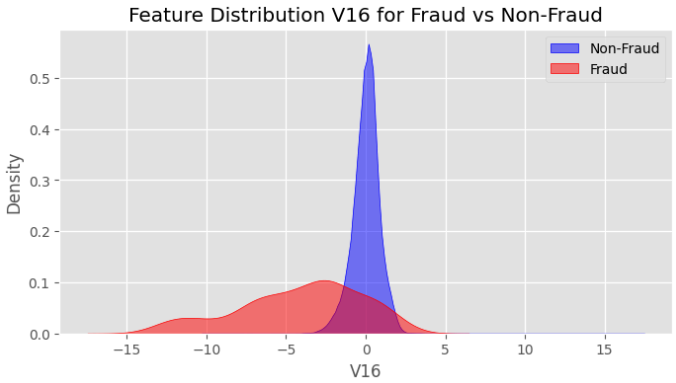
Visualization of the distribution of important features for fraud (1) vs non-fraud (0) classes:











If the V17 plot shows that: Fraud values tend to be lower/extremely negative, while non-fraud values are spread out in the middle, Then the model can utilize this for early detection. A similar approach can be applied to other features such as V12, V14, etc.