Charles Dilger
AIM460
Professor Matthew Fried

Isolation Forests as a Solution to Credit Card Fraud

Isolation Forests compared to alternative options for classifying credit card fraud

Related Works - Comparison

	AUC			
	iForest	ORCA	LOF	RF
Http (KDDCUP99)	**1.00**	0.36	NA	NA
ForestCover	0.88	0.83	NA	NA
Mulcross	0.97	0.33	NA	NA
Smtp (KDDCUP99)	0.88	0.80	NA	NA
Shuttle	**1.00**	0.60	0.55	NA
Mammography	0.86	0.77	0.67	NA
Annthyroid	0.82	0.68	0.72	NA
Satellite	0.71	0.65	0.52	NA
Pima	0.67	0.71	0.49	0.65
Breastw	0.99	0.98	0.37	0.97
Arrhythmia	0.80	0.78	0.73	0.60
Ionosphere	0.85	**0.92**	0.89	0.85

Figure I

Isolation Forests

Isolation Forests dominate against ORCA, Local Outlier Factor, and Random Forests

Model	Detection Rate (%)	AUC-ROC
Isolation Forest	95.3	0.964
One-Class SVM	95.0	0.958
Autoencoder (AE)	94.7	0.971

Figure II

Credit Card Fraud

In detecting credit card fraud, Isolation Forests outperform One-Class SVM and compete with an Autoencoder

Algorithm	Accuracy
Logistic Regression	90.0%
Decision Tree	94.3%
Random Forest Classifier	95.5%
Isolation Forest	71%
Local Outlier Factor	97%

Figure III

Credit Card Fraud

One study found Isolation Forests ineffective when compared to Logistic Regression, Random Forest, LOF, Decision Trees.

Methods	F1 Score	Accuracy	AUC Score
OCSVM	0.0033	0.5088	0.5154
LOF	0.0027	0.8901	0.4970
K-means	0.0054	0.9012	0.5191
Isolation Forest	0.0544	0.9512	0.9168

Figure IV

Credit Card Fraud

Isolation Forests for credit card fraud once again defeat One-Class SVM, Local Outlier Factor, and K-Means

Model	Country A						Country B					
	F1 (↑)		AUROC (↑)		AUPRC (↑)		F1 (↑)		AUROC (↑)		AUPRC (↑)	
	2018	2020	2018	2020	2018	2020	2018	2020	2018	2020	2018	2020
LightGBM	21.52	0.7	89.98	66.49	18.74	0.31	17.15	0.48	93.5	75.51	34.84	2.68
	1.6	0.3	0.1	2.38	1.78	0.05	1.93	0.39	0.31	1.24	1.48	0.29
ECOD	0.48	0.2	62.2	<u>62.49</u>	0.4	0.25	0.57	1.04	54.02	51.59	1.09	0.76
	0.2	0.22	0.64	1.02	0.02	0.01	0.26	0.38	0.54	0.87	0.06	0.04
COPOD	0.34	0.16	64.77	<u>62.64</u>	0.43	0.25	0.5	1.12	51.7	50.15	1.0	0.73
	0.21	0.15	0.51	0.98	0.02	0.01	0.2	0.44	0.59	0.91	0.05	0.04
Isolation Forest	0.16	0.19	64.14	60.55	0.43	0.23	0.71	1.3	60.52	46.86	1.35	0.67
	0.12	0.19	0.75	0.76	0.02	0.01	0.23	0.34	0.52	0.84	0.07	0.03
KNN	0.34	0.01	68.87	55.6	0.51	0.18	0.78	0.38	65.92	49.28	1.58	0.63
	0.12	0.04	0.76	0.85	0.02	0.01	0.21	0.27	0.64	0.67	0.08	0.02
GOAD	0.14	0.19	53.72	52.73	0.17	0.17	0.7	0.69	50.36	<u>64.45</u>	0.67	1.03
	0.09	0.13	1.41	1.69	0.01	0.01	0.36	0.34	2.35	1.25	0.05	0.06
NeuTraL-AD	0.6	0.02	59.12	51.52	0.35	0.15	1.45	0.38	53.23	45.19	1.08	0.58
	0.22	0.08	3.56	1.15	0.05	0.01	0.44	0.21	1.9	1.75	0.07	0.03
Internal Cont.	0.64	0.0	39.43	46.7	0.18	0.13	1.21	0.23	45.63	50.66	0.87	0.68
	0.08	0.0	1.05	0.17	0.0	0.0	0.16	0.07	2.46	0.9	0.1	0.03
NPT-AD	<u>0.97</u>	<u>0.66</u>	67.21	53.2	<u>0.81</u>	0.18	<u>1.67</u>	0.58	66.21	53.45	1.28	0.61
	0.07	0.06	1.25	0.65	0.01	0.03	0.11	0.03	1.14	0.43	0.11	0.06

LightGBM

LightGBM is found to be the only model performing consistently better than Isolation Forests, and by a wide margin. The goal of the remainder of the presentation is to determine if LightGBM is truly a better algorithm for detecting credit card fraud than Isolation Forests, as the new top competitor.

Algorithms compared

Algorithm 1 Isolation Forest

Input: Dataset $\mathbf{D}$, trees t, subsample ψ
Initialize forest $F \leftarrow \emptyset$
for $i = 1$ to t **do**
  Sample ψ transactions from $\mathbf{D}$ to form $\mathbf{D}'$
  Build tree $T_i \leftarrow$ iTree($\mathbf{D}'$, 0, limit)
  $F \leftarrow F \cup \{T_i\}$
end for
for each $\mathbf{x} \in \mathbf{D}$ **do**
  Compute $E[h(\mathbf{x})] \leftarrow$ avg PathLength($\mathbf{x}, T_i$) $\forall T_i \in F$
  Calculate $s(\mathbf{x}, \psi) = 2^{-E[h(\mathbf{x})]/c(\psi)}$
end for
return anomaly scores $s(\mathbf{x}, \psi)$

Isolation Forest

Algorithm 2 LightGBM Classifier

Input: Dataset $\mathbf{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, learning rate η, trees M
Initialize $\hat{y}_i^{(0)} = 0$ for all i
for $m = 1$ to M **do**
  Compute gradients $g_i = \frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}$ for all i
  Compute Hessians $h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(m-1)})}{\partial (\hat{y}_i^{(m-1)})^2}$ for all i
  Build histogram-based tree f_m fitting $-g_i$ using leaf-wise growth
  Update predictions $\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta \cdot f_m(\mathbf{x}_i)$ for all i
end for
return final predictions $\hat{y}(\mathbf{x}) = \text{sigmoid}(\hat{y}^{(M)})$

LightGBM

3 credit card fraud datasets were collected from Kaggle.com. The data went through preprocessing including data balancing, and was run through both algorithms.

Metric	Isolation Forest	LightGBM
Basic Performance Metrics		
Accuracy	0.9655	**0.9990**
Precision	0.6465	**0.9800**
Recall	0.6531	**1.0000**
F1-Score	0.6497	**0.9899**
ROC-AUC	0.9652	**1.0000**
MCC	0.6316	**0.9894**
Cross-Validation Analysis (5-Fold)		
Accuracy	0.9626 ± 0.0043 (95% CI: [0.9633, 0.9752])	0.9997 ± 0.0002 (95% CI: [0.9994, 1.0001])
Precision	0.6879 ± 0.0497 (95% CI: [0.6189, 0.7569])	0.9996 ± 0.0003 (95% CI: [0.9992, 1.0001])
Recall	0.6924 ± 0.0597 (95% CI: [0.6099, 0.7758])	0.9999 ± 0.0003 (95% CI: [0.9993, 1.0006])
F1-Score	0.6890 ± 0.0444 (95% CI: [0.6274, 0.7506])	0.9997 ± 0.0002 (95% CI: [0.9994, 1.0001])
ROC-AUC	0.9750 ± 0.0050 (95% CI: [0.9680, 0.9821])	1.0000 ± 0.0000 (95% CI: [1.0000, 1.0000])
Statistical Significance Tests		
T-test (Fraud vs Normal)	t-statistic: 38.5888 p-value: 6.4478e-244	t-statistic: 401.8760 p-value: 0.0000e+00
Significance at $\alpha = 0.05$	Yes	Yes
Cohen's d (Effect Size) Interpretation	3.9972 large effect	**41.6283** large effect

SIDE-BY-SIDE COMPARISON – DATASET 2

	Isolation Forest	LightGBM
Accuracy	0.445250	0.998000
Precision	0.445387	0.996100
Recall	0.446500	0.999500
F1-Score	0.445943	0.998003
ROC-AUC	0.484863	0.999982
MCC	-0.109500	0.996004

SIDE-BY-SIDE COMPARISON – DATASET 3

	Isolation Forest	LightGBM
Accuracy	0.591250	0.999500
Precision	0.592781	0.999500
Recall	0.583000	0.999500
F1-Score	0.587850	0.999500
ROC-AUC	0.646721	0.999999
MCC	0.182525	0.999000

Ablation Study

Isolation Forests are an anomaly detection algorithm, so we tried again with a significant class imbalance to see if Isolation Forests would start improving over LightGBM

SIDE-BY-SIDE COMPARISON (WITHOUT SMOTE)		
	Isolation Forest	LightGBM (Class Weights)
Accuracy	0.965500	0.999000
Precision	0.646465	1.000000
Recall	0.653061	0.979592
F1-Score	0.649746	0.989691
ROC-AUC	0.965176	0.999938
MCC	0.631612	0.989223

Dataset 1 re-tested without SMOTE, with a large class imbalance.

Isolation Forest (10% fraud)		
Metric	Dataset 2	Dataset 3
Accuracy	0.8438	0.9170
Precision	0.2137	0.5867
Recall	0.2100	0.6058
F1-Score	0.2119	0.5808
ROC-AUC	0.7276	0.9064
MCC	0.1251	0.5348

Isolation Forests used on Dataset 2 and 3 with a larger class imbalance, in order to test if Isolation Forests will become competitive in this case

The conclusion is that Isolation Forests are a great method for detecting credit card fraud. However, LightGBM is an even stronger method, providing even more accuracy to the classification of credit card transactions as fraudulent or safe.

Datasets used

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

https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud

https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023