



Credit Card Fraud Detection

Tópicos de Aprendizagem
Automática 2023/2024

André Oliveira - 107637
Duarte Cruz - 107359

Table of Contents

01

Overview

Credit Card Fraud in our daily life

02

Objectives

Goals
State of Art

03

Machine Learning

The importance of ML in fraud detection

04

Model Selection

Chosen models and evaluation

05

Results & Analysis

Comparison of results between models

06

Conclusion

Future Directions



01

Overview

Credit Card Fraud in our daily life

Overview

Credit card fraud is the **unauthorized use** of a **debit or credit card** to make purchases or withdraw cash.



In 2021, **389,845 credit card fraud reports** in the US were the most common identity fraud for people aged 20-39, per the FTC.



02

Objectives

Goals and State of Art

Objectives

Goals



Effective credit card
fraud detection
system.

State of Art



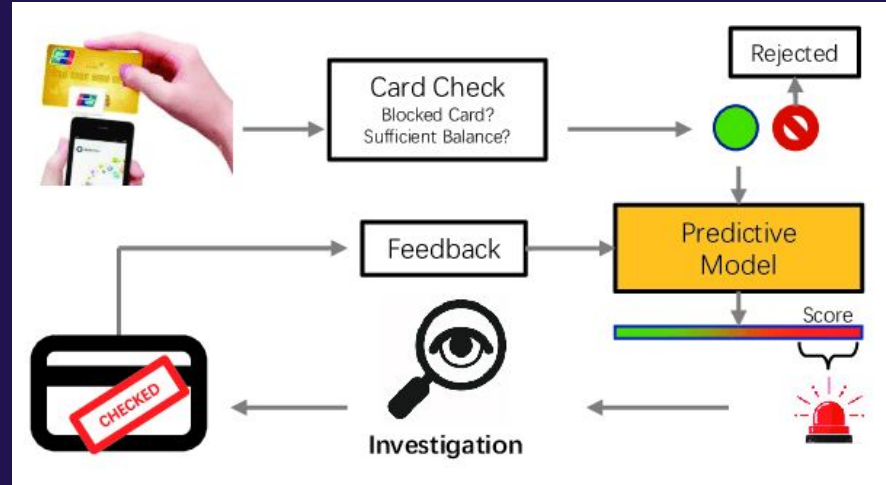
Subsampling
Associated Cost

03

Machine Learning

The importance of ML in fraud detection

Machine Learning



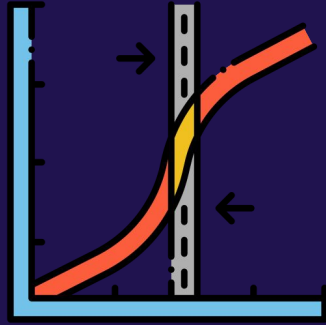
Source: The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20) Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection

04

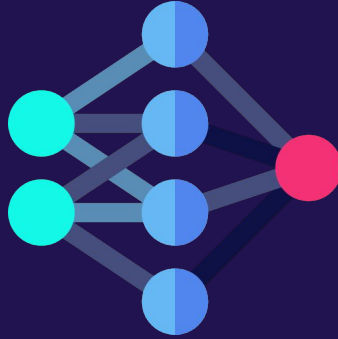
Model Selection

Chosen models and evaluation

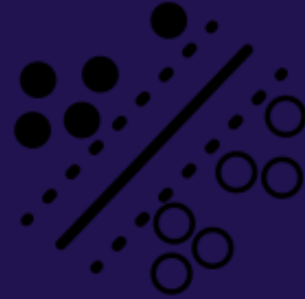
Model Selection



Logistic
Regression



Neural
Networks



Support Vector
Machine

05

Results & Analysis

Comparison of results between models

Results & Analysis

First Results

	F1 Score	Accuracy Score
Train		
Logistic Regression	0.9580	0.9580
Logistic Regression with "L2"	0.9525	0.9526
SVC	0.9525	0.9526
NN	1.0000	1.0000
Test		
Logistic Regression	0.9268	0.9268
Logistic Regression with "L2"	0.9430	0.9431
SVC	0.9430	0.9431
NN	0.9350	0.9350

TABLE V: F1 Score and Accuracy Score Values

	Precision Score	Recall Score
Logistic Regression	0.9339	0.9187
Logistic Regression with "L2"	0.9739	0.9106
SVC	0.9820	0.8862
Neural Network (NN)	0.9421	0.9268

TABLE VI: Precision Score and Recall Score Values

Results & Analysis

Gradient Boosting Regressor

	F1 Score	Accuracy Score
Train		
Logistic Regression	0.9157	0.9160
SVC	0.9115	0.9119
NN	0.9185	0.9187
Test		
Logistic Regression	0.9226	0.9228
SVC	0.9224	0.9228
NN	0.9267	0.9268

TABLE VII: F1 Score and Accuracy Score Values with Feature 14

	Precision Score	Recall Score
Logistic Regression	0.9643	0.8780
SVC	0.9906	0.8537
Neural Network (NN)	0.9267	0.9268

TABLE IX: Precision Score and Recall Score Values with Feature 14

	F1 Score	Accuracy Score
Train		
Logistic Regression	0.9430	0.9431
SVC	0.9457	0.9458
NN	0.9498	0.9499
Test		
Logistic Regression	0.9146	0.9146
SVC	0.9105	0.9106
NN	0.9228	0.9228

TABLE VIII: F1 Score and Accuracy Score Values with Top 5 Features

	Precision Score	Recall Score
Logistic Regression	0.9397	0.8862
SVC	0.9391	0.8780
Neural Network (NN)	0.9262	0.9187

TABLE X: Precision Score and Recall Score Values with Top 5 Features

Results & Analysis

Tuning Hyper-Parameter

Solver	Max Iterations	Class Weight	Penalty	C
liblinear	5000	balanced	11	[0.001, 0.01, 0.1, 1, 10, 100, 1000]
liblinear	5000	balanced	12	[0.001, 0.01, 0.1, 1, 10, 100, 1000]

TABLE XI: Possible Values for Logistic Regression Parameters

	Precision Score	Recall Score	F1 Score	Support
0	0.88	0.96	0.92	123
1	0.96	0.87	0.91	123
Accuracy			0.91	246
Macro Avg	0.92	0.91	0.91	246
Weighted Avg	0.92	0.91	0.91	246

TABLE XII: Logistic Regression Report

kernel	C	Gamma
rbf	[0.001, 0.01, 0.1, 1, 10, 100, 1000]	[0.0001, 0.001, 0.01, 0.1, 1]

TABLE XIII: Possible Values for SVC Parameters

	Precision Score	Recall Score	F1 Score	Support
0	0.88	0.97	0.92	123
1	0.96	0.87	0.91	123
Accuracy			0.92	246
Macro Avg	0.92	0.92	0.92	246
Weighted Avg	0.92	0.92	0.92	246

TABLE XIV: SVC Report

Solver	Max Iterations	Hidden Layer Sizes	Activation	Alpha	Learning Rate	Initial Learning Rate
adam	5000	[(12,12), (12,12,12)]	['tanh', 'relu']	[1e-3,1e-4]	constant	[0.001, 0.01]
adam	5000	[(12,12), (12,12,12)]	['tanh', 'relu']	[1e-3,1e-4]	invscaling	[0.001, 0.01]

TABLE XV: Possible Values for NN Parameters

	Precision Score	Recall Score	F1 Score	Support
0	0.91	0.94	0.93	123
1	0.94	0.91	0.93	123
Accuracy			0.93	246
Macro Avg	0.93	0.93	0.93	246
Weighted Avg	0.93	0.93	0.93	246

TABLE XVI: NN Report

Results & Analysis

Tuning Hyper-Parameter

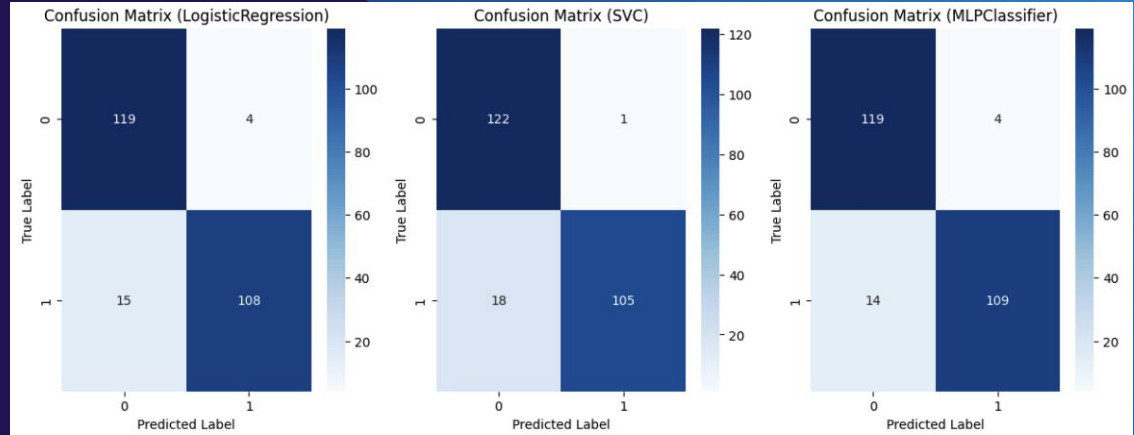


Fig. 34: Confusion Matrix for All Algorithms

	F1 Score	Accuracy Score	Precision Score	Recall Score
Logistic Regression	0.9552	0.9553	0.9828	0.9268
SVC	0.9389	0.9390	0.9821	0.8943
NN	0.9553	0.9553	0.9667	0.9431

TABLE XVII: Results of all algorithms after the Tuning Hyper-Parameter

Results & Analysis

K-Fold Results

	Average Accuracy
Logistic Regression	0.94
SVC	0.93
NN	0.92

TABLE XVIII: K-Fold Results

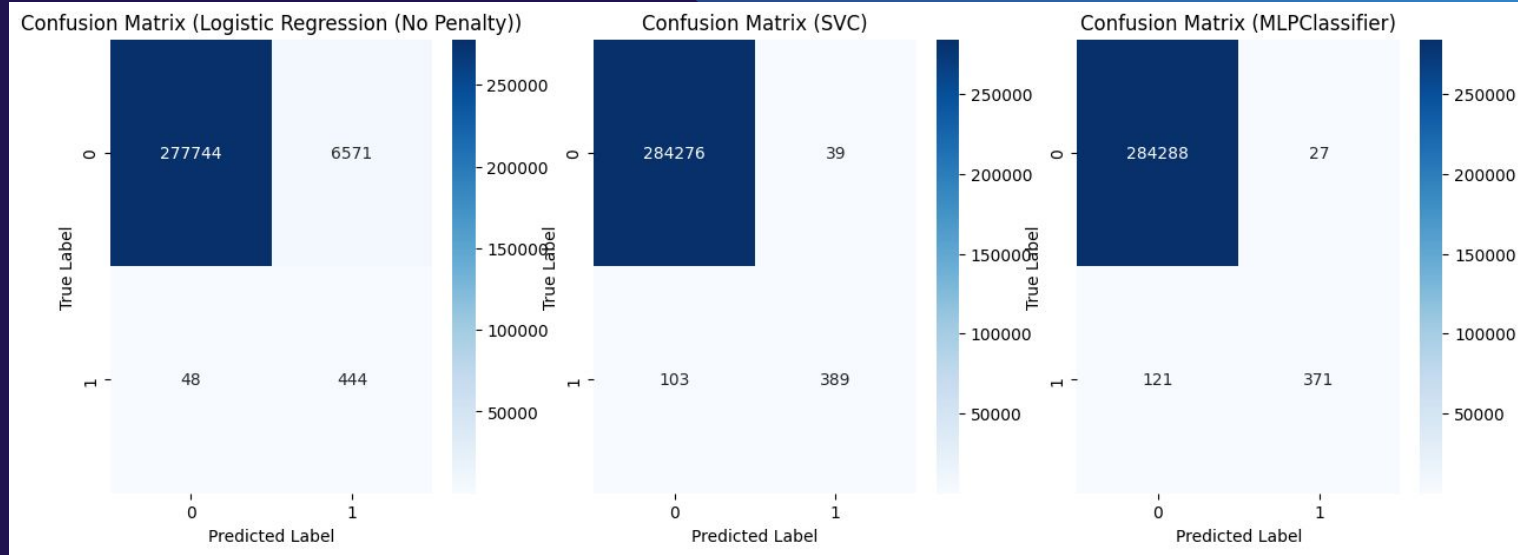


06

Conclusion

Future Directions

Results & Analysis



	F1 Score	Accuracy Score	Precision Score	Recall Score
Logistic Regression	0.9867	0.9768	0.0633	0.9024
SVC	0.9995	0.9995	0.9089	0.7907
NN	0.9995	0.9995	0.9322	0.7541

TABLE XIX: Final Results

Thank You

“Without big data, you are blind and deaf in the middle of a freeway.”

— Geoffrey Moore

