Credit Card Fraud Detection

Saira Faiz, Harshilkumar Pareshbhai Kansara, & Colby Jamieson

Can we predict which credit card transactions are fraudulent?

Date	Posted Transactions	Amount
December 1, 2021	U of A CAB TIMS	$4.60
	ESSO	$50.00
December 4, 2021	Etsy	$45.55
	EDM EPark	$2.00
December 5, 2021	Sephora Canada	$39.75
	UAlberta Bookstore	$55.66
	U of A CAB TIMS	$4.60

Introduction

Credit Card Fraud Statistics

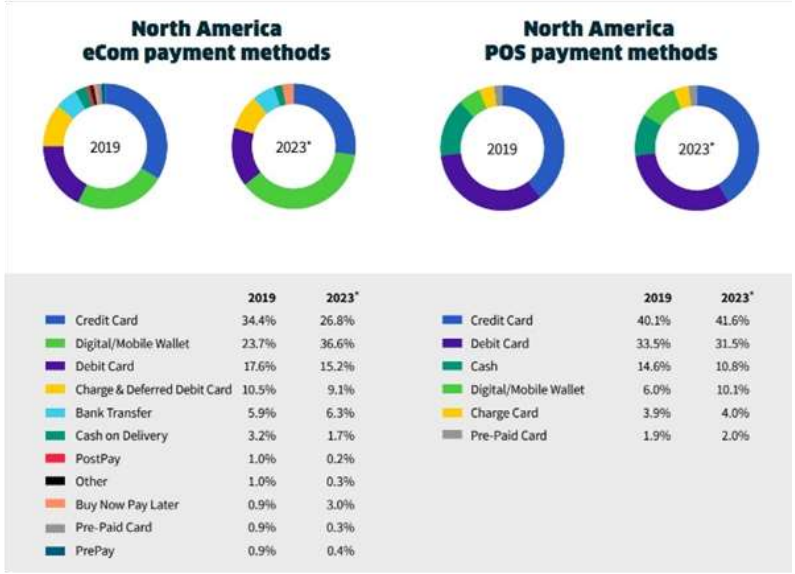
- 24% of **the Canadian** have experienced credit card fraud at some point of their life.
- Over **$379 million** were lost to scams and fraud in **2021.**
- In 2021, there were **104,295** fraud reports and **68,076** victims of fraud.
- Those **aged 30 to 39** reported the most cases of credit card theft.

Data Acquisition Process

- The data was collected by Machine Learning Group of ULB(Université Libre de Bruxelles) for research purpose on big data mining and fraud detection.
- These are 284,807 credit card transections made by European during two days in Sep,2013, out of which 492 are fraudulent cases (0.172%)
- Most of the information is transformed with principal component analysis (PCA) except for time and amount.
- Class is the binary response variable, 1 in case of fraud and zero otherwise.

Choice of Classification Method

- Artificial Neural network
- XGBoost
- Logistic regression with MCMC
- Random Forest
- Decision Tree



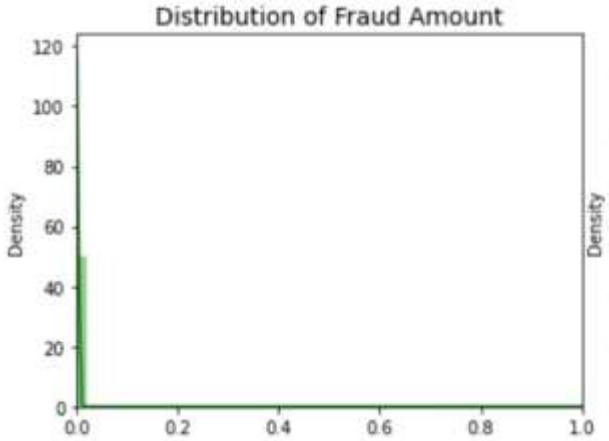
	2019	2023*
Credit Card	34.4%	26.8%
Digital/Mobile Wallet	23.7%	36.6%
Debit Card	17.6%	15.2%
Charge & Deferred Debit Card	10.5%	9.1%
Bank Transfer	5.9%	6.3%
Cash on Delivery	3.2%	1.7%
PostPay	1.0%	0.2%
Other	1.0%	0.3%
Buy Now Pay Later	0.9%	3.0%
Pre-Paid Card	0.9%	0.3%
PrePay	0.9%	0.4%

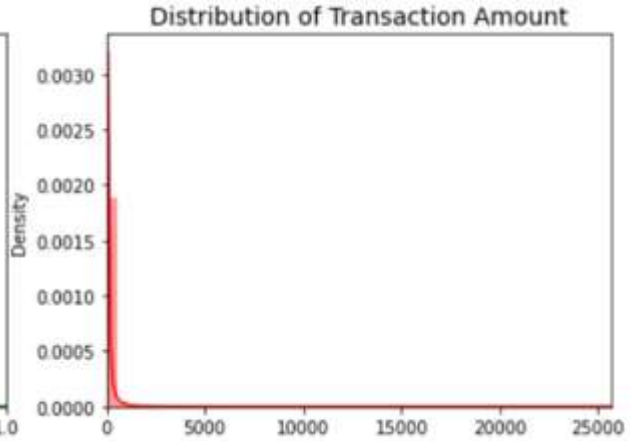
	2019	2023*
Credit Card	40.1%	41.6%
Debit Card	33.5%	31.5%
Cash	14.6%	10.8%
Digital/Mobile Wallet	6.0%	10.1%
Charge Card	3.9%	4.0%
Pre-Paid Card	1.9%	2.0%

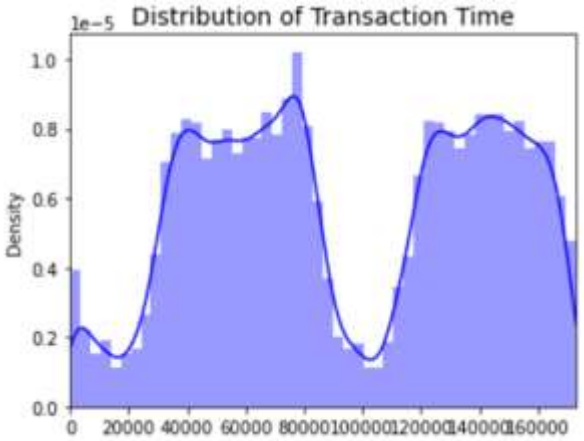
Data

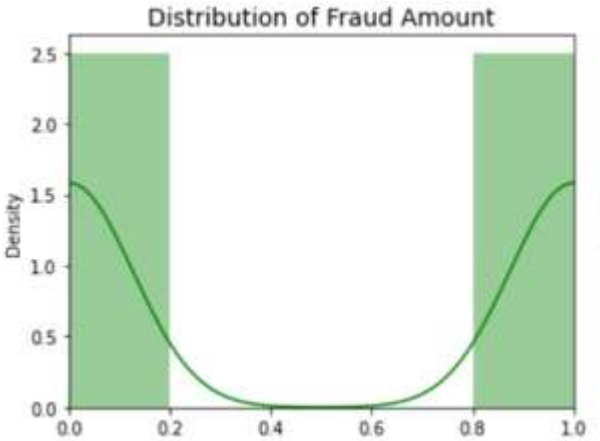
Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V28	Amount	Class
0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.021053	149.62	0
0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	0.014724	2.69	0
1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	-0.059752	378.66	0
1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	0.061458	123.50	0
2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	0.215153	69.99	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.918215	7.305334	1.914428	...	0.823731	0.77	0
172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.024330	0.294869	0.584800	...	-0.053527	24.79	0
172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.296827	0.708417	0.432454	...	-0.026561	67.88	0
172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.686180	0.679145	0.392087	...	0.104533	10.00	0
172792.0	-0.533413	-0.189733	0.703337	-0.506271	-0.012546	-0.649617	1.577006	-0.414650	0.486180	...	0.013649	217.00	0

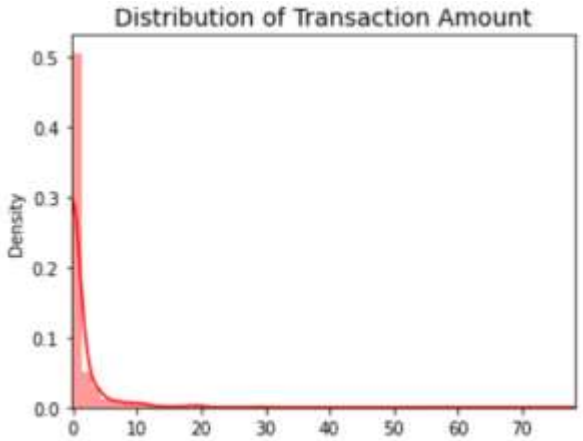
Transforming Unbalanced Data

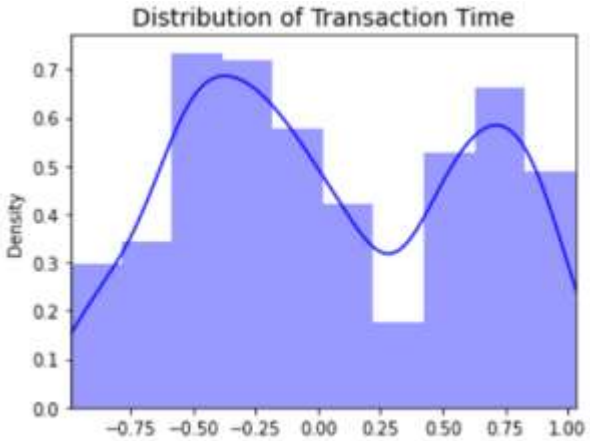




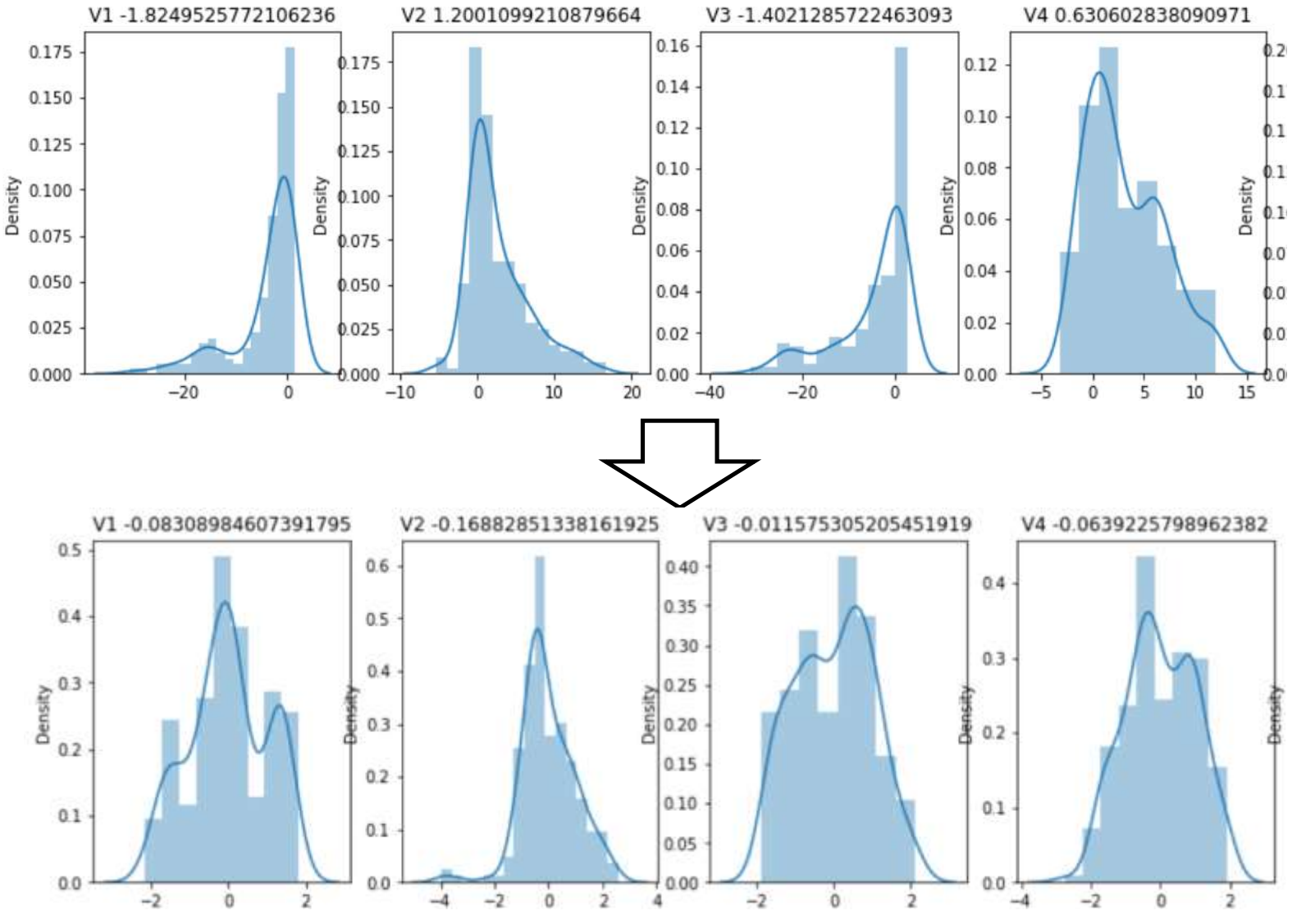




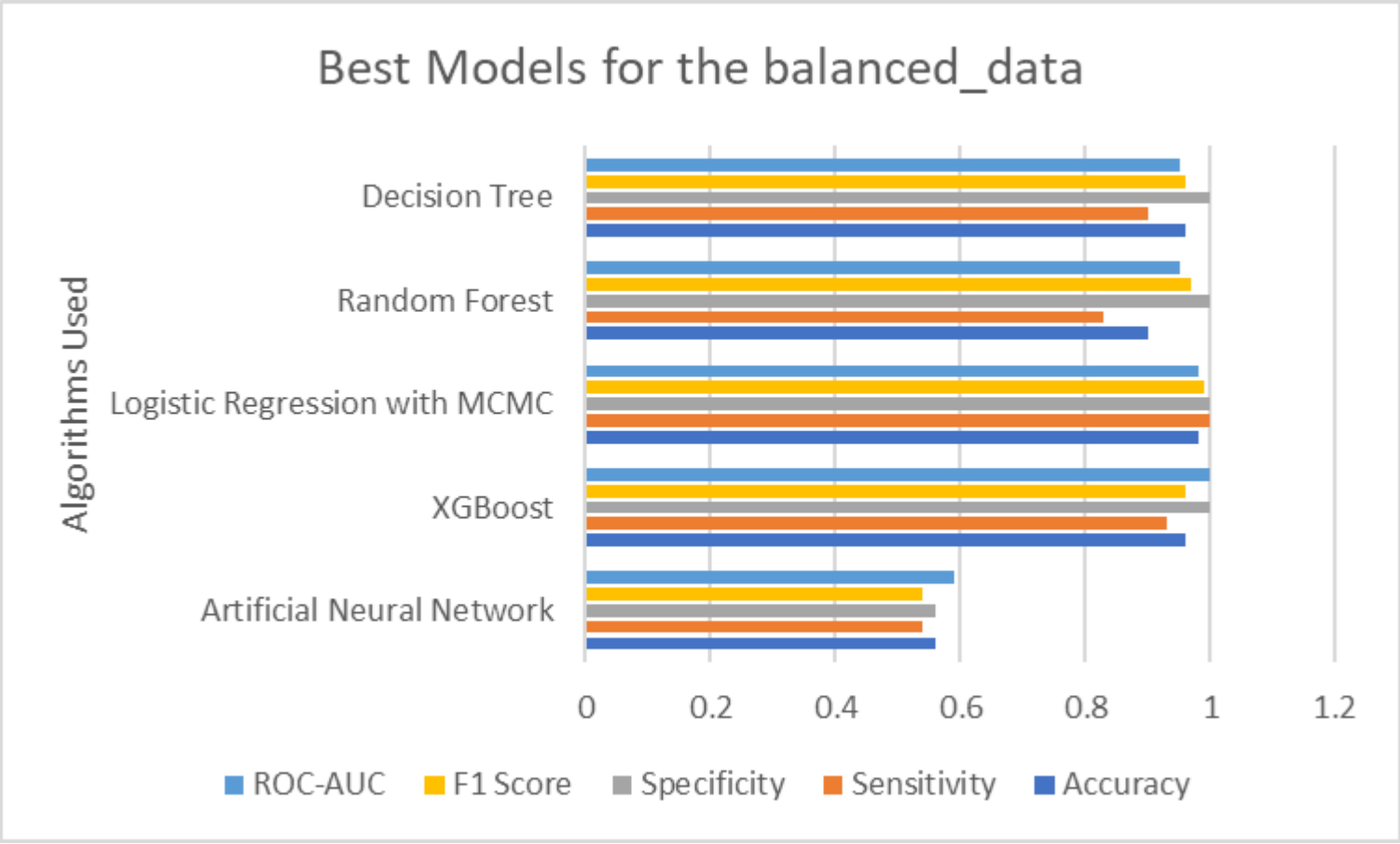




Transforming Skewed Data



Best Model for our Balanced Data



Future Work

- This data was used with the PCA transformed data as given V1 to V28 as the bank data are confidential. We would like to use a more comprehensive data in the future for our analysis.
- Model building using different deep learning methods such as CNN, RNN and see the accuracy on the unbalanced as well as balanced data.
- Model building using different classification techniques such as KNN, Naive Bayes, Support Vector Machine etc.
- MCMC approach was quite affective while applying it to the logistic regression method, we would like to apply it to different algorithms and see the accuracy.

Questions?