Fraud Analysis in Canada

Code Enforcers
(Eder, Geraldine & Demilade)

Executive Summary

PROBLEM

Fraud losses **hit CA$647M in 2024** for Canada. Tech is fueling faster, smarter, harder-to-trace schemes.
Fewer victims, bigger losses — fraud is evolving, not disappearing.

DATA

- Open-source data on Canadian Anti-Fraud Centre .
- It consists of reported fraud and attempts by Canadians across the world from **2021 – 2025.**

INSIGHTS

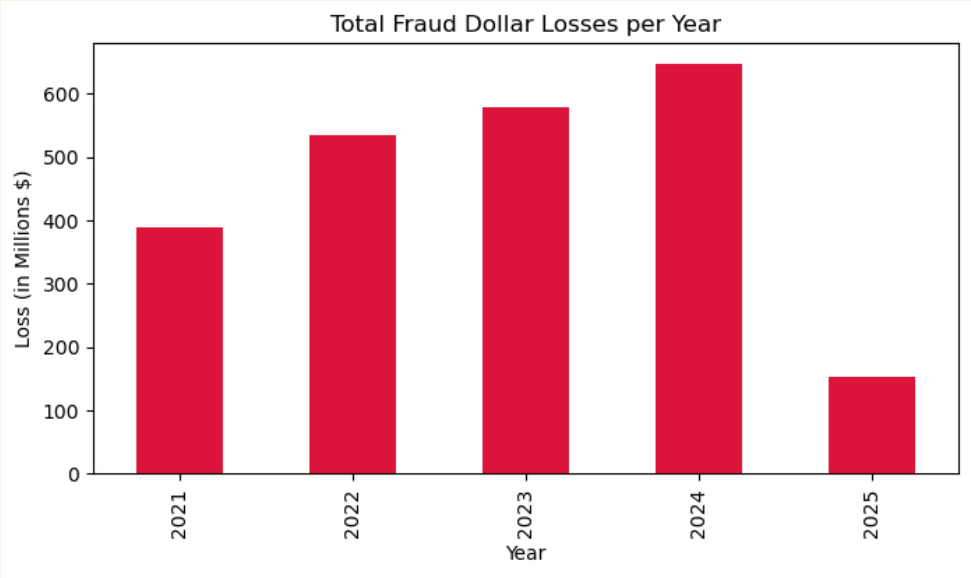
- Victims are those that have been defrauded. **Ontario** has the most reports.
- Victims are mostly **20s – 39**.
- **Investments, Romance and Spear Phishing** are the top 3 fraud categories with the highest financial loss.

RECOMMENDATIONS

- **Trained model using Random Forest** to detect when there is fraud and no fraud. **Accuracy is 88%.**
- Create more awareness among Canadians on popular fraud categories and solicitation methods.

Facts

- Fraud is a growing concern in Canada, with reported losses to the Canadian Anti-Fraud Centre (CAFC) exceeding **CA$2 billion** since 2021 .
- The Canadian Anti-Fraud Centre (CAFC) reported overall losses exceeding **CA$647 million** in 2024, with only 5–10% of incidents formally reported.
- According to Consumer and Government Services Canada, **3.8% of Canadians** experienced mass marketing fraud in the past year – about 975,000 victims annually.



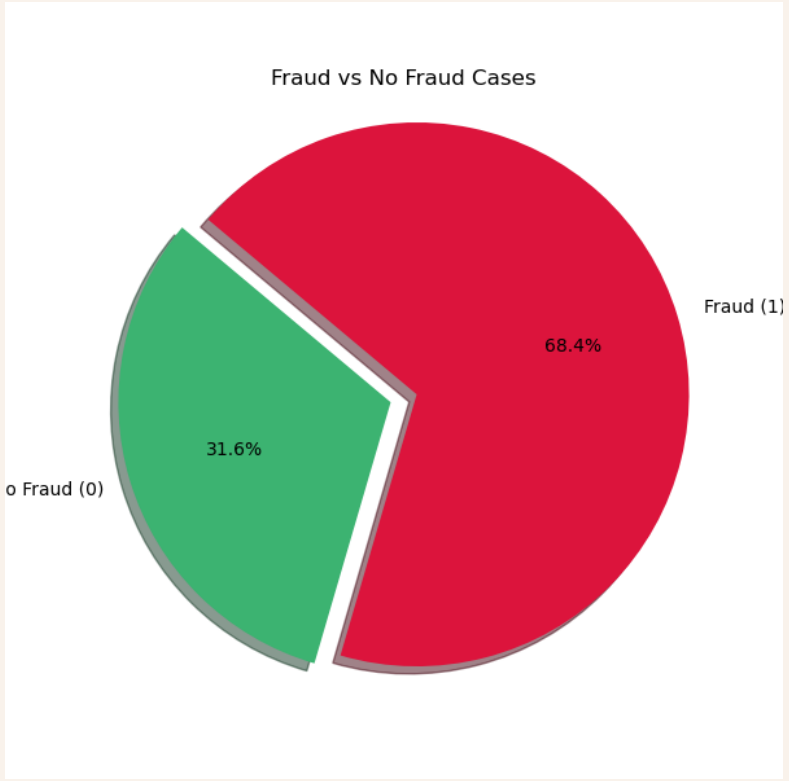
Data Approach

CLEANING
- Removed missing values
- Removed unknown data for certain columns
- Dropped unnecessary columns
- Renamed columns

ASSUMPTIONS
- In the provided data - victims (1) and attempt & others (0)- Fraud vs No Fraud.
- For model and insights, we filtered for reports made in Canada-only.

LIMITATIONS
- Underreporting from Victims
- Delays between incident and report date
- No fraud includes multiple non-harmful categories

ETHICS
- No personal identifiers
- Data used solely for public interest, research, and education
- Respect for data privacy

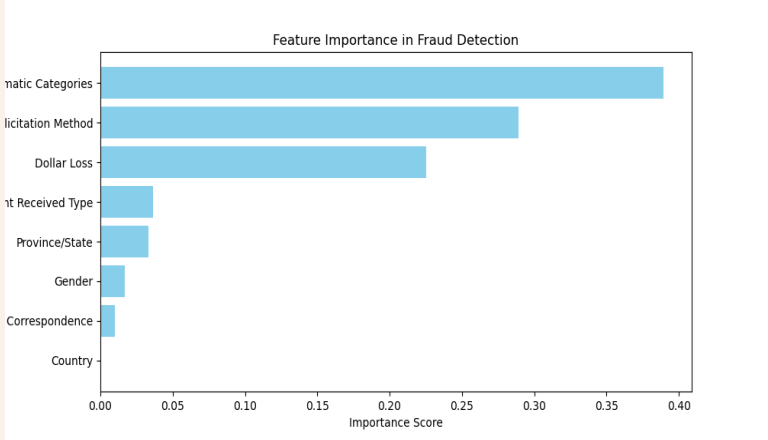


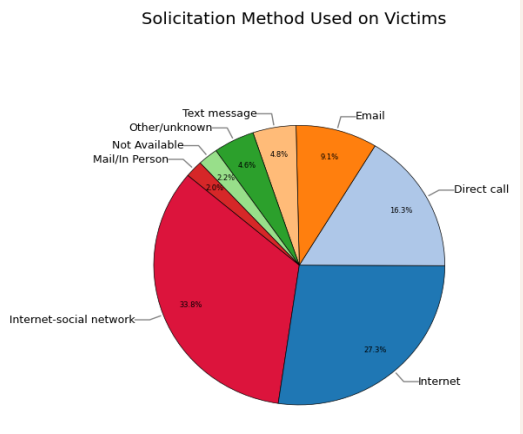
Features of Importance

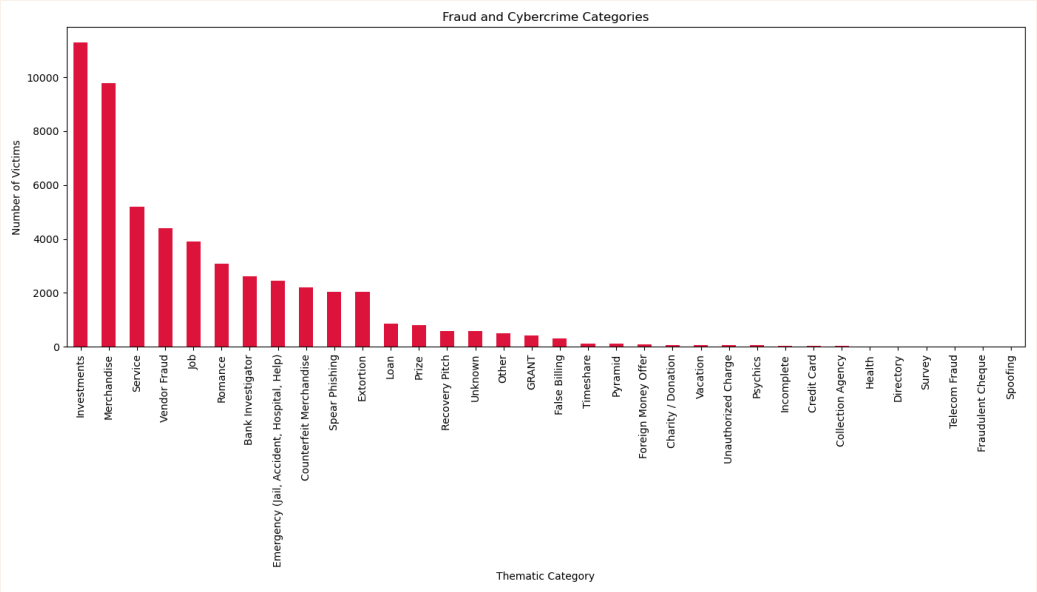
- Based on the horizontal bar chart, **fraud & cybercrime categories, solicitation methods and dollar loss** are the features that are most important.

- The fraud categories with the most victims were **Investments and Merchandise.**

- The solicitation method used on most victims are internet – **social network** (33.8%), **internet** (27.3%) and **Direct call** (16.3%).





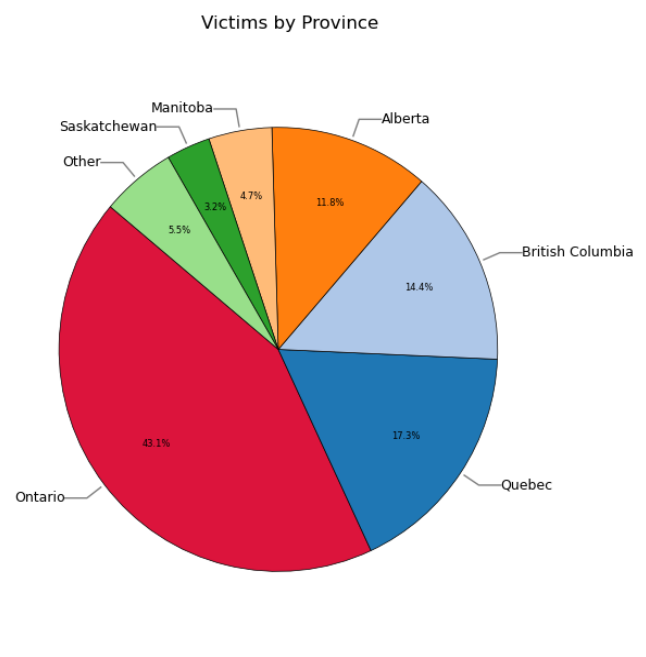


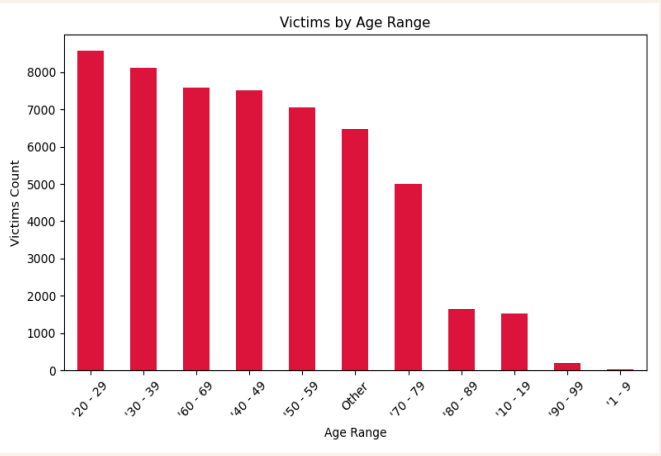
Data Insights

- **Ontario** is the most reported province of Fraud by victims. This is expected as Ontario is the most populated province in Canada, accounting for 38–40% of Canada's total population.

- The age range **of 20s – 39** have the most victims. This is expected as the top solicitation methods are social media, internet and direct call, which are most common within the age range.

- **Investments, Romance and Spear Phishing** are the top 3 fraud categories with the highest financial loss.





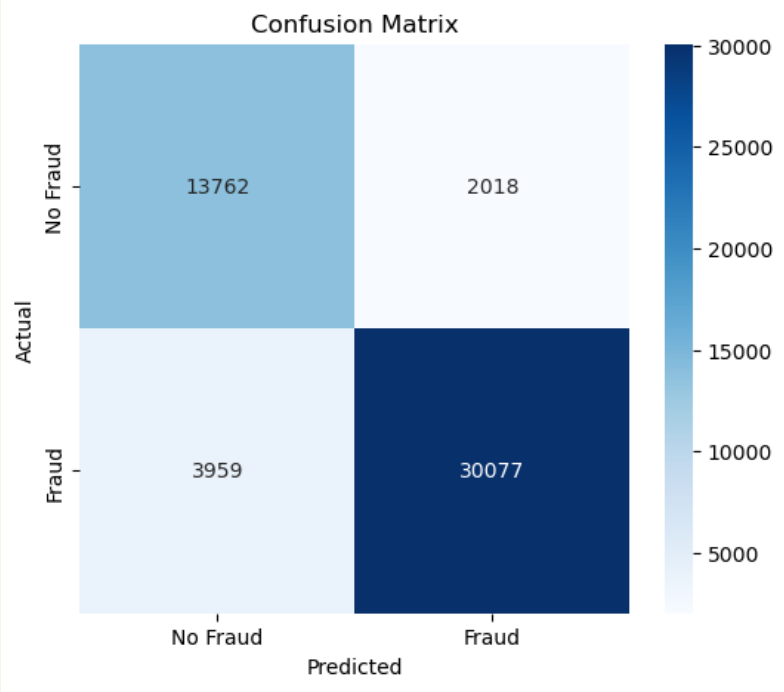
Fraud Category	Loss (in Million $)
Investments	1184.42
Romance	255.6
Spear Phishing	246.29
Job	119.28
Service	80.66
Extortion	74.16
Unknown	56.25
Bank Investigator	45.61
Merchandise	44.07
Other	41.68
Emergency (Jail, Accident, Hospital, Help)	27.4
Recovery Pitch	25.27
Vendor Fraud	24.56
Prize	18.65
Loan	16.76
Foreign Money Offer	16.0
Timeshare	8.35
GRANT	6.42
Counterfeit Merchandise	3.19
False Billing	1.68
Psychics	1.17
Pyramid	0.89
Vacation	0.31
Charity / Donation	0.31
Incomplete	0.16
Collection Agency	0.13
Credit Card	0.11
Unauthorized Charge	0.07
Directory	0.05
Fraudulent Cheque	0.01

Random Forest Model – Machine Learning

- For RF Model - 2018 false positives (false Fraud) and 3959 false negatives (false No-Fraud). **The accuracy score is 88%.**

- When compared to KNN optimized, which can be seen on the next slide, the model has few false frauds, which makes the recall for 'No-fraud' high **at 87%.** Slight decrease in false 'No-fraud', hence the recall for Fraud is close at **88%** - this means it catches 88% of actual frauds.

- The F1 score is a balanced score between precision and recall, which are both over 80%



	precision	recall	f1-score	support
0	0.78	0.87	0.82	15780
1	0.94	0.88	0.91	34036
accuracy			0.88	49816
macro avg	0.86	0.88	0.87	49816
weighted avg	0.89	0.88	0.88	49816

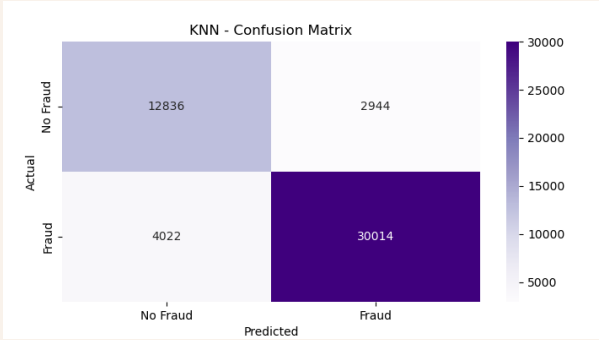
K-nearest neighbors (KNN) -Optimization

- For **KNN=5**, 2944 false positives (false Fraud) and 4022 false negatives (false No-Fraud). **The accuracy score is 86.01%.**

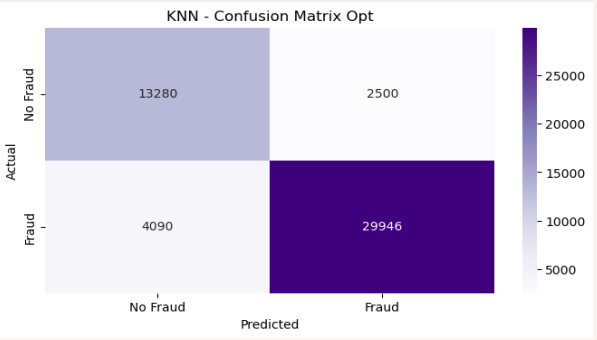
- For **KNN=15,** 2500 false positives (false Fraud) and 4090 false negatives (false No-Fraud). **The accuracy score is 86.77%.**

- At **optimized KNN of 15**, the model has fewer false frauds, which makes the recall for 'No-fraud' higher **at 84%.** Due to a slight increase in false 'No-fraud', the recall for Fraud is close at **88%** - this means it catches 88% of actual frauds.

KNN = 5



KNN = 15



KNN = 15

	precision	recall	f1-score	support
0	0.76	0.84	0.80	15780
1	0.92	0.88	0.90	34036
accuracy			0.87	49816
macro avg	0.84	0.86	0.85	49816
weighted avg	0.87	0.87	0.87	49816

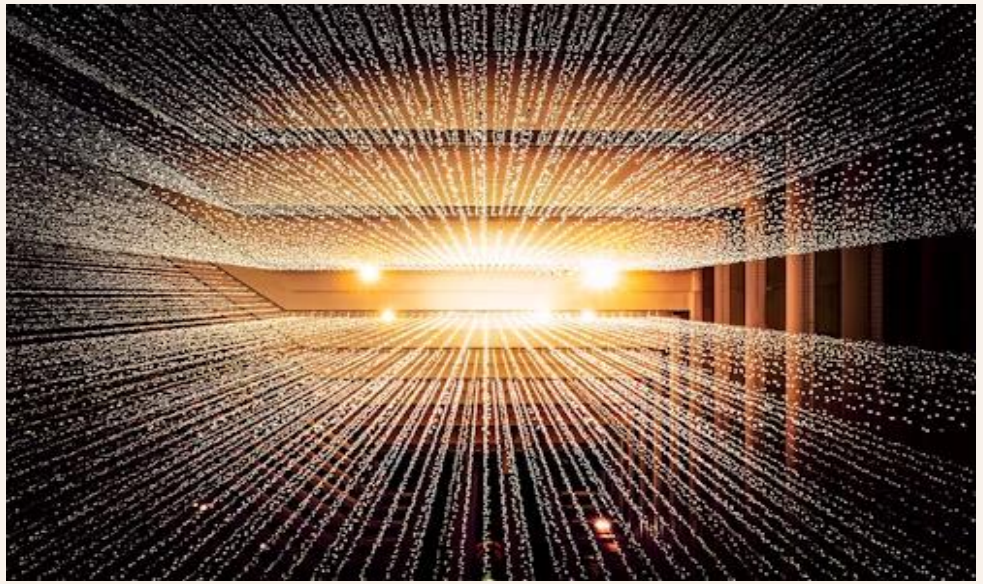
Comparison of Models

Summary

Accuracy of 2 models and Optimization

	Random Forest Classifier	K Neighbors Classifier (K=5)	Optimized K Neighbors Classifier (K=15)
Accuracy Score	0.88002	0.86017	0.86771

	precision	recall	f1-score	support
0	0.78	0.87	0.82	15780
1	0.94	0.88	0.91	34036
accuracy			0.88	49816
macro avg	0.86	0.88	0.87	49816
weighted avg	0.89	0.88	0.88	49816



Random Forest is better as it has a higher accuracy score above.
Also, it;

- handles high-dimensional & mixed data well.
- is better at capturing complex patterns.
- is robust to noise & outliers. + many more

Recommendation

Fewer victims, bigger losses – fraud is evolving, not disappearing. Trained model using Random Forest is recommended as it has an accuracy score of 88%. To **improve model robustness and accuracy**, consider gradient **boosting models like XGBoost**, which has high performance on tabular data and often outperforms RF/KNN.

Stakeholder Impact:

- **Law Enforcement & Government –** Identify high-risk fraud provinces- Ontario, Quebec & B.C. and optimize resource allocation. Pre-empt large value scams from Investments or Romance.
- **Financial Institutions –** Integrate fraud probability scores into transaction monitoring systems.
- **Public & Media –** Raise awareness of fraud patterns in Investments and Merchandise that had the most victims and support public education initiatives.

💡 A Fraud predictive interactive app is in the works. This is to enable the General Public to predict if a fraud will result in financial loss based on certain questions.

Questions?

Thank you for your Attention!

"We can't prevent every scam—but we can use data to detect and disrupt fraud faster." – The Code Enforcers