BOLT Bootcamp 2024

Team 21 - Team Lasantha

Bryan Chang, Khue Do, Steve He

Today's Agenda

01 — 02 — 03 — 04

Proposition

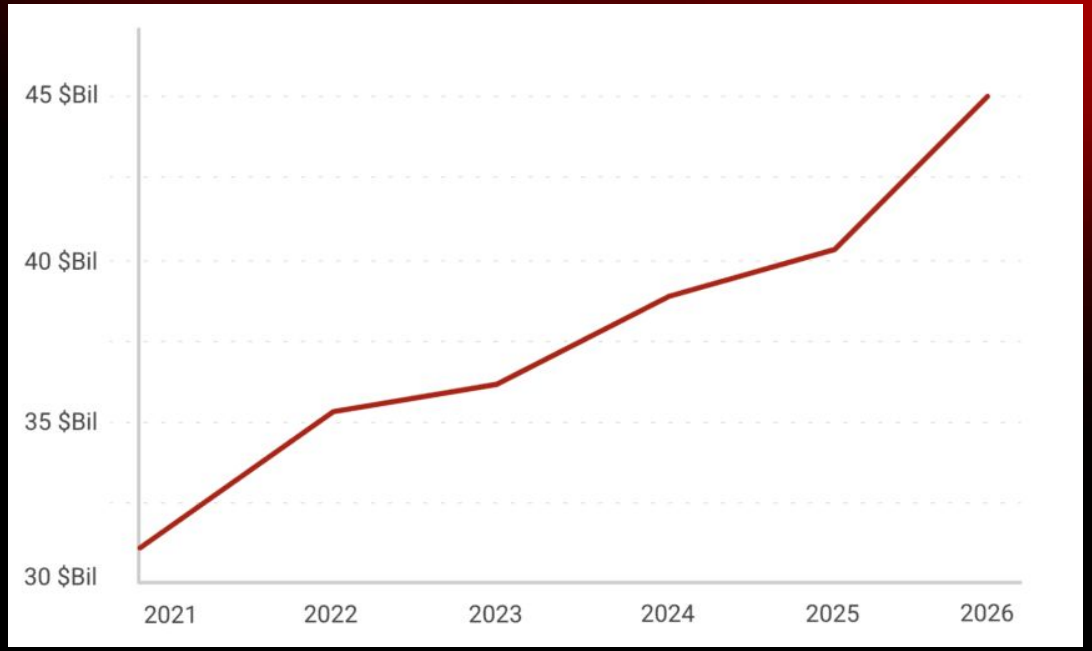
Analytics

Predictive Model

Implementation Plan

Background

CC Fraud Losses



Year

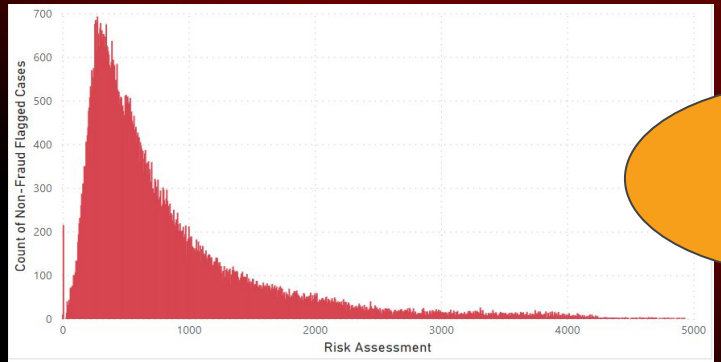
Insight

NullFraud Bank had a 0.265% fraud rate in 2023 - not a stranger to fraud themselves!

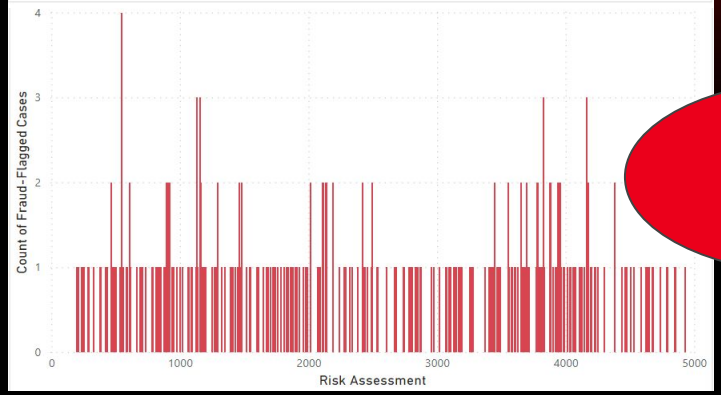
Overview

- Improve fraud detection
- Reduce fraud-related costs
- Protect client transactions
- Enhance customer loyalty
- Adapt to changing consumer habits
- Draw insight from a cleaned dataset

Risk Assessment



NO FRAUD



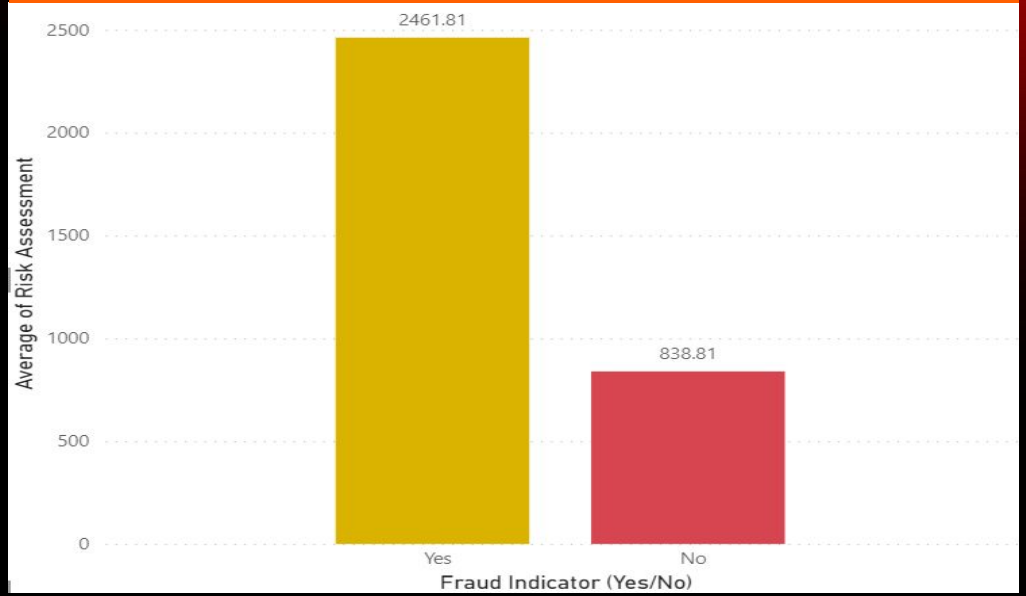
FRAUD

Insight

- Noticeable difference in risk assessment score:
- Mode for non-fraud risk assessment at ~500
- Modes for fraud risk assessment at ~1000 and ~4000

Risk Assessment

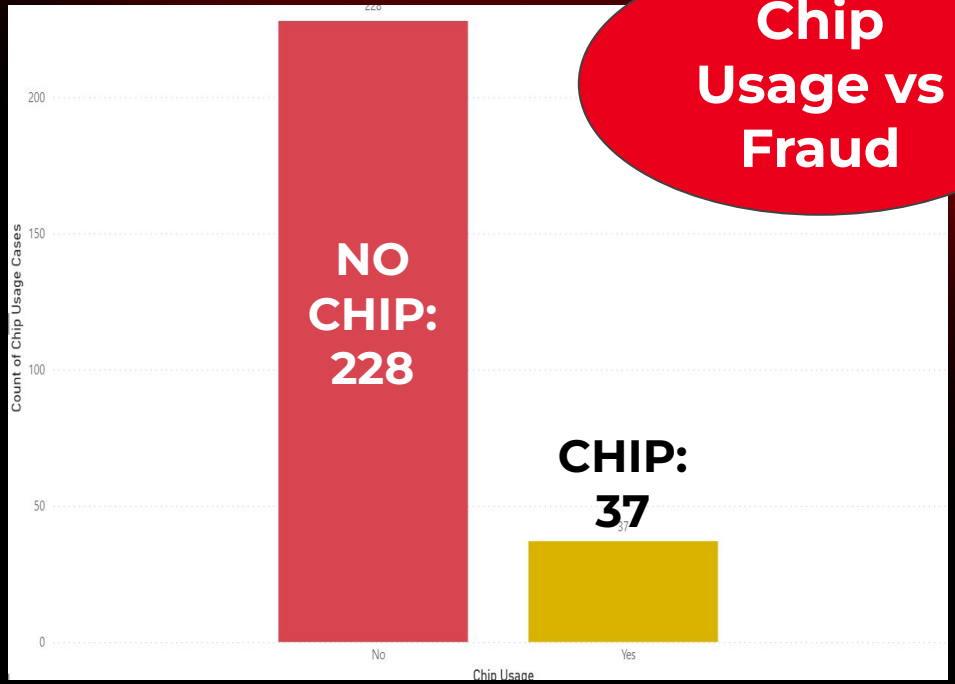
Average Risk Assessment by Fraud Indicator



Insight

That's a ~3x increase in risk assessment value for fraud transactions over non fraud transactions!

Chip Usage

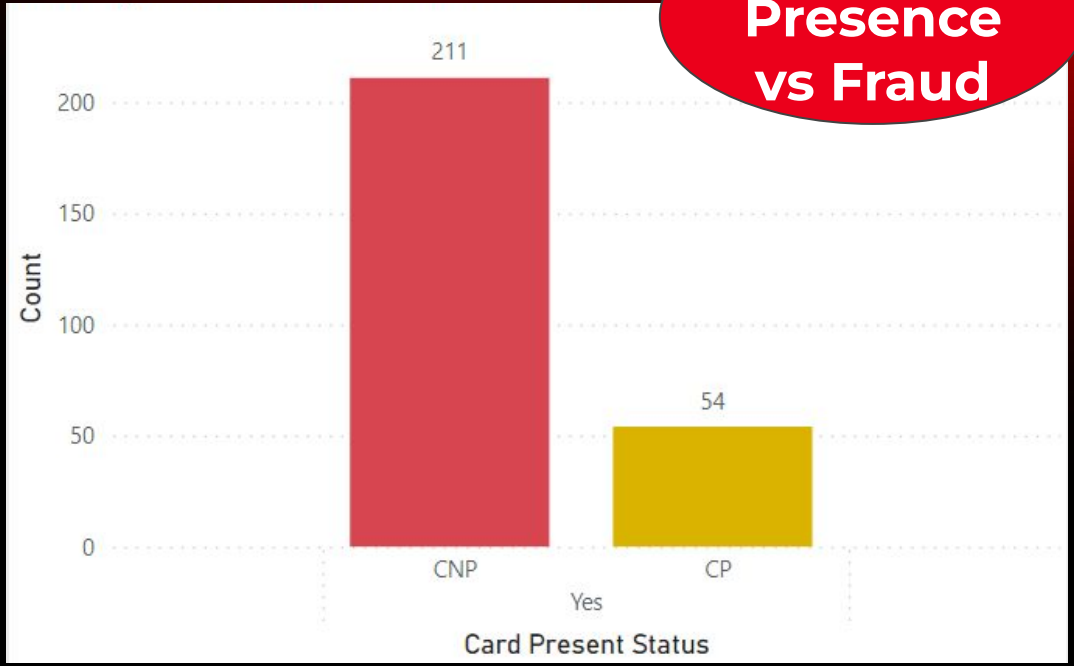


Insight

- Chip usage is nearly 50% of all transactions (54k vs 46k)
- No chip usage carries a fraud rate ~6x higher than that of chip usage

Card Presence

Card Presence vs Fraud



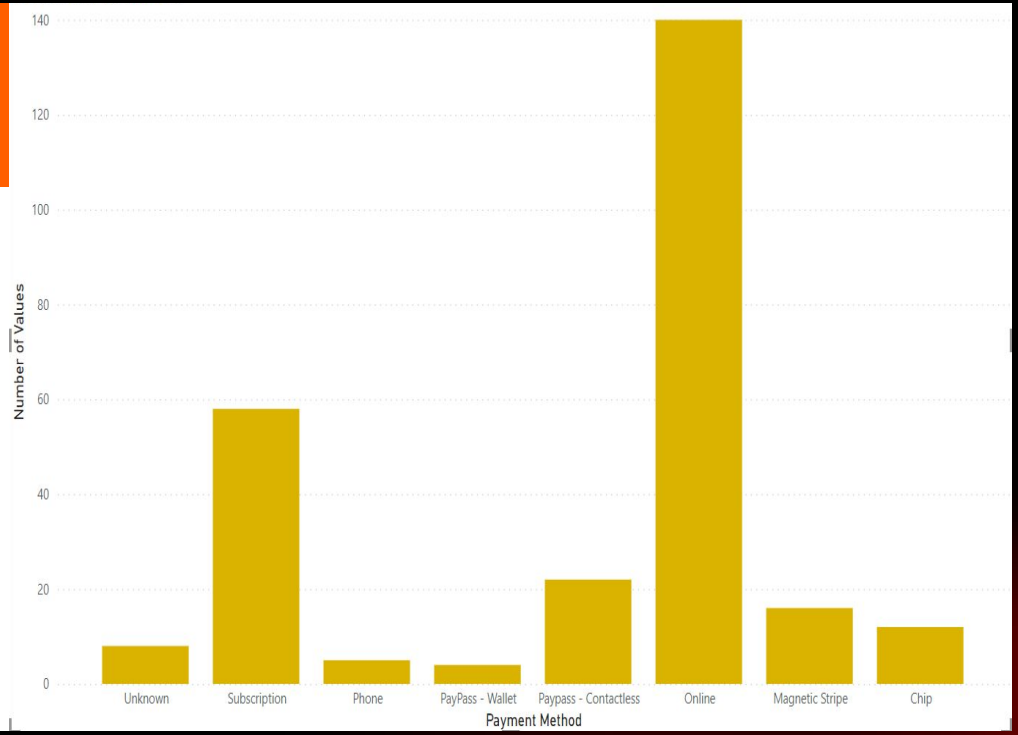
Insight

- **Card Presence is ~50% of all transactions**
- **No card presence carries a fraud rate ~4x higher than that of card presence**

Payment Method

Shift in Payment Methods

- Some payment methods had higher fraud rates than 0.265%: online, phone, subscription, unknown
- Less usage of paypass payment methods
- Change in user habits: digital 21st century



Future Variables

Variable of Interest	Reasoning
Cross-border Transactions	Not convincingly higher, but seems intuitive
Repeated Card Fraud	43/183 cards responsible for multiple instances of fraud, small sample size
Month & Time of Month	Spikes in fraud during spring and fall, needs more time

Logistic Regression Model

Objectives

- 70% training set - 30% testing set
- GOALS : enhance efficiency, reduce overfitting, and improve model performance

Choice of Model

- LOGISTIC REGRESSION used to make our prediction
- Stepwise variable selection method, namely forward selection

Variables Selected

- Risk Assessment
- Transaction Value
- Cross-Border Transaction
- Payment Method : subscription, postal, phone, unknown
- Card Present Status
- Chip Usage

Logistic Regression Model

Model Result

Model accuracy : 85.13%

Precision : 70%

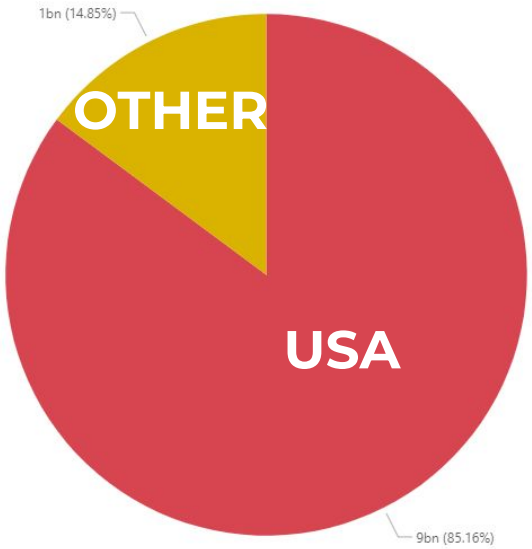
Predicted	Actual	n
<chr>	<chr>	<int>
0	0	24720
1	0	4307
0	1	21
1	1	49

Insight

- **GOAL : Find a sweet balance between high accuracy and high precision**
- **PROCESS : Classifier is adjusted to improve the precision rate (true positive probability) -> results in more false positives lowers the overall accuracy**

Further Improvements

Transaction Value



Fraud by Country

	merchant_country	fraud_incident_rate
	<chr>	<dbl>
1	ARM	0.0208
2	BRA	0.0220
3	CAN	0.00311
4	DEU	0.0202
5	DOM	0.0303
6	ECU	0.05
7	ESP	0.000982
8	FIN	0.0909
9	FRA	0.00702
10	GBR	0.00588
11	GIB	0.5
12	IND	0.0690
13	IRL	0.005
14	ISL	0.0312
15	ITA	0.00853
16	JPN	0.005
17	KEN	0.125
18	LBN	0.389
19	MEX	0.00142
20	NLD	0.00387
21	PER	0.0476
22	SGP	0.0739
23	USA	0.00204

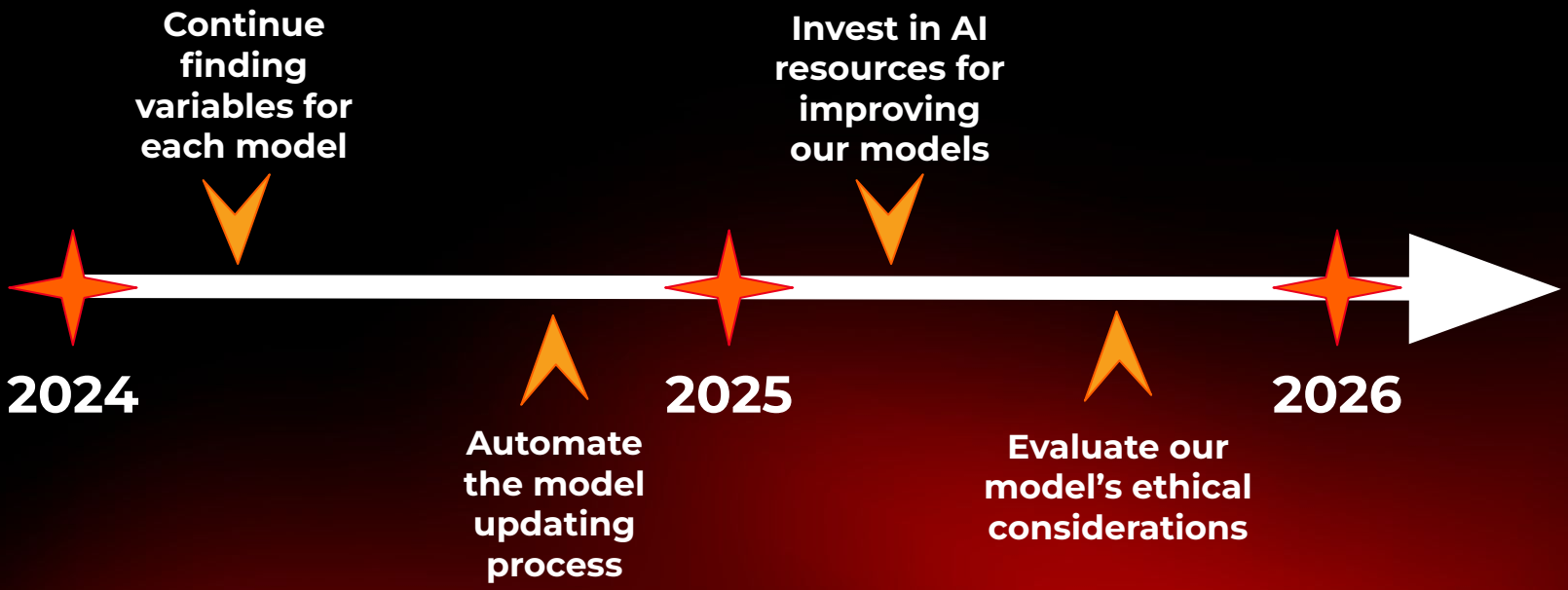
Adaptive Analytics

- Model by **region** rather than on a global scale
- Capable of updating models in **constant time**

KPIs

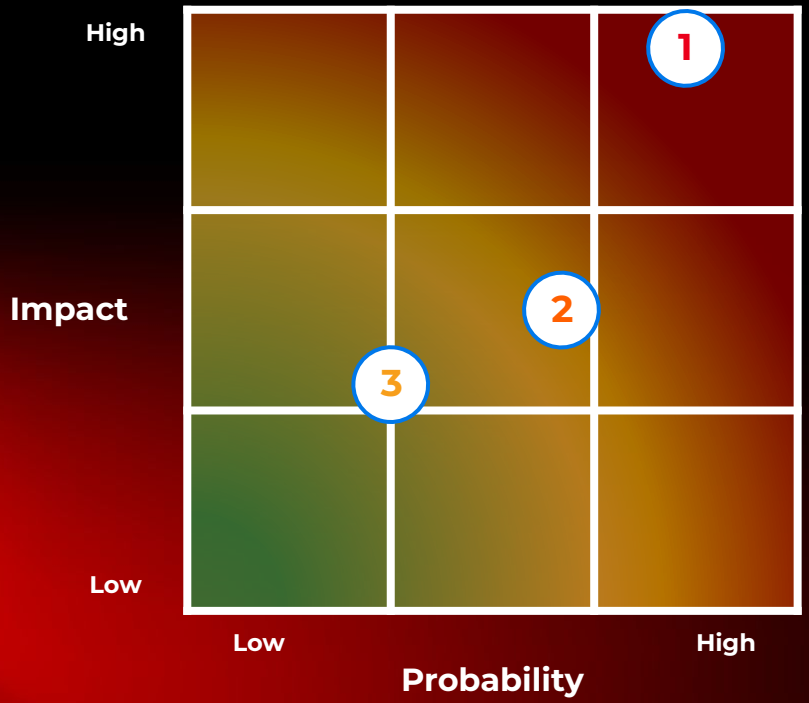
Key Performance Indicator	Current Performance	Target Performance
F1 Score: $\frac{2 * (\text{precision} * \text{sensitivity})}{\text{precision} + \text{sensitivity}}$	70%	85% - 90%
True Positive	70%	85%
False Positive & False Negative	30% & 15%	15% & 10%
Cost of Addressing Errors	N/A	Decrease!

Timeline



Proposition | Data Analysis | Predictive Modelling | **Implementation**

Risks and Mitigation



1 Cybersecurity

Mitigation: Regular penetration testing

2 Regulatory Compliance

Mitigation: Ensure transparency of processes

3 Model Performance Degradation

Mitigation: Model monitoring and validation

Proposition > Data Analysis > Predictive Modelling > Implementation

Thank you!

Team 21 - Team Lasantha

https://github.com/khuedo22/BOLT-Datathon-2024