

Collaboration for AI/ML Solutions

Presented by:
Prof. Nandan Sudarsanam
Aakash Sarraf

Statement of Work

	FY 25-26				FY 26-27			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
A) Pension-related SoW								
(i) Risk assessment and management								
(ii) Customer Segmentation								
(iii) Personalized Investment Planning								
B) Insurance-related SoW								
(i) Customer or cohort heat maps								
(ii) Detecting Fraudulent claims								
(iii) Underwriting and Premium pricing								

Detecting Fraudulent Claims

1. Synthetic Data Generation
2. Rule-Based System/Heuristic
3. ML Model
4. RL Framework

Synthetic Data

	Age	Gender	Days in Hospital	Diagnosis	Fraud Type	Amount Billed
0	89	Female	9 days	Diabetes	Wrong Diagnoses	380486.364575
1	51	Female	7 days	Pregnancy	No Fraud	282171.642424
2	7	Male	0 days	Cataract Surgery	No Fraud	126590.549215
3	37	Female	9 days	Pregnancy	No Fraud	97609.907157
4	4	Female	8 days	Routine Check-up	No Fraud	160604.190109

Number of Total Claims = 1564

No Fraud = 1253

Wrong Diagnoses = 99

Phantom Billing = 78

Ghost Patients = 48

Rule Based Fraud Detection

Rule 1: High amount + very short stay (≤ 1 day)

Rule 2: Young (< 30) + high amount

Rule 3: Elderly (> 75) + short stay (≤ 2 days) + high amount

Rule 4: Rare diagnosis categories with high amount

If any of the two conditions are true, flag as a Fraud case.

- ❖ Company will usually investigate a small set of customers which are flagged as suspects.
- ❖ Among them, there are some Fraudulent cases and some Non-Fraudulent case

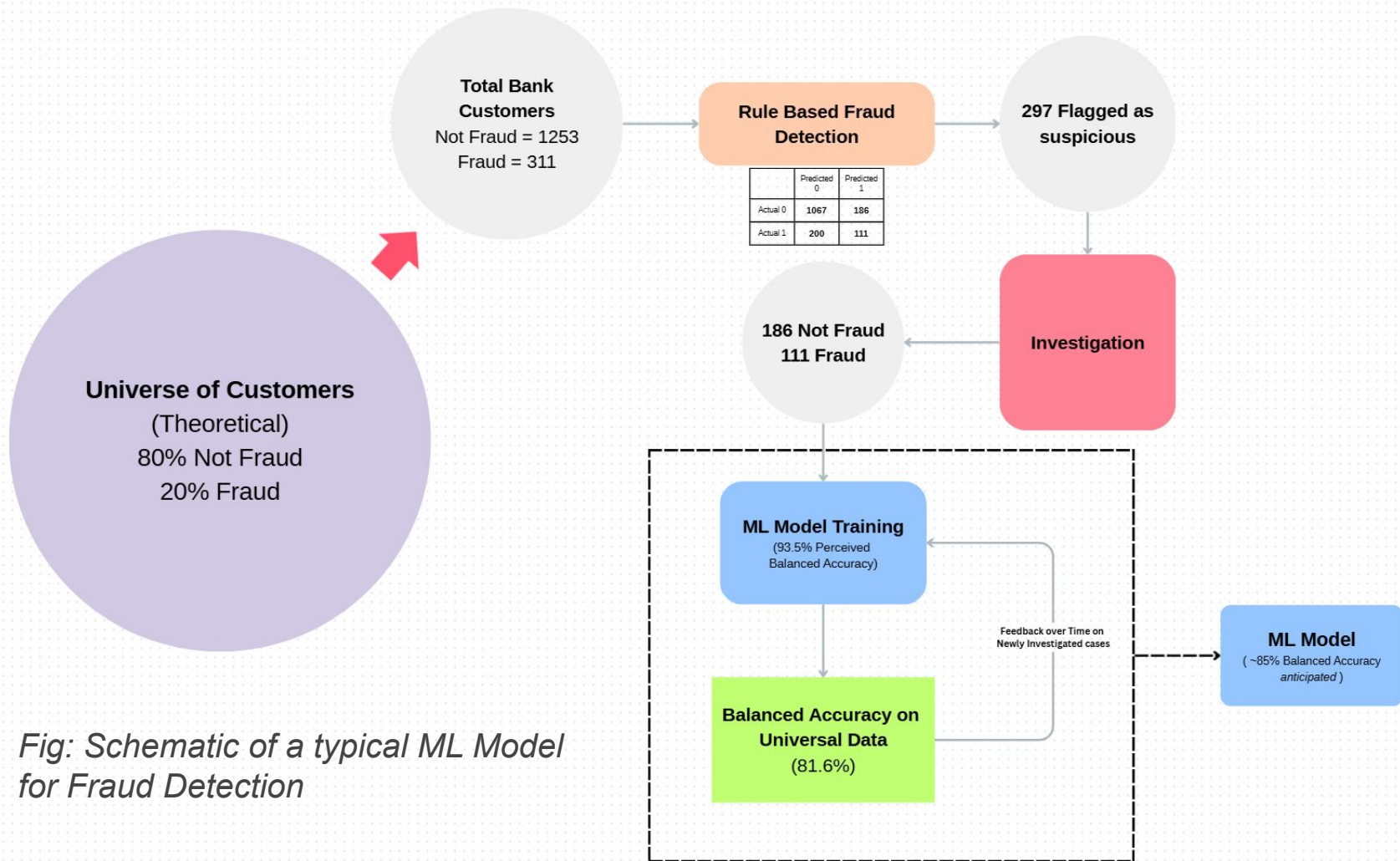
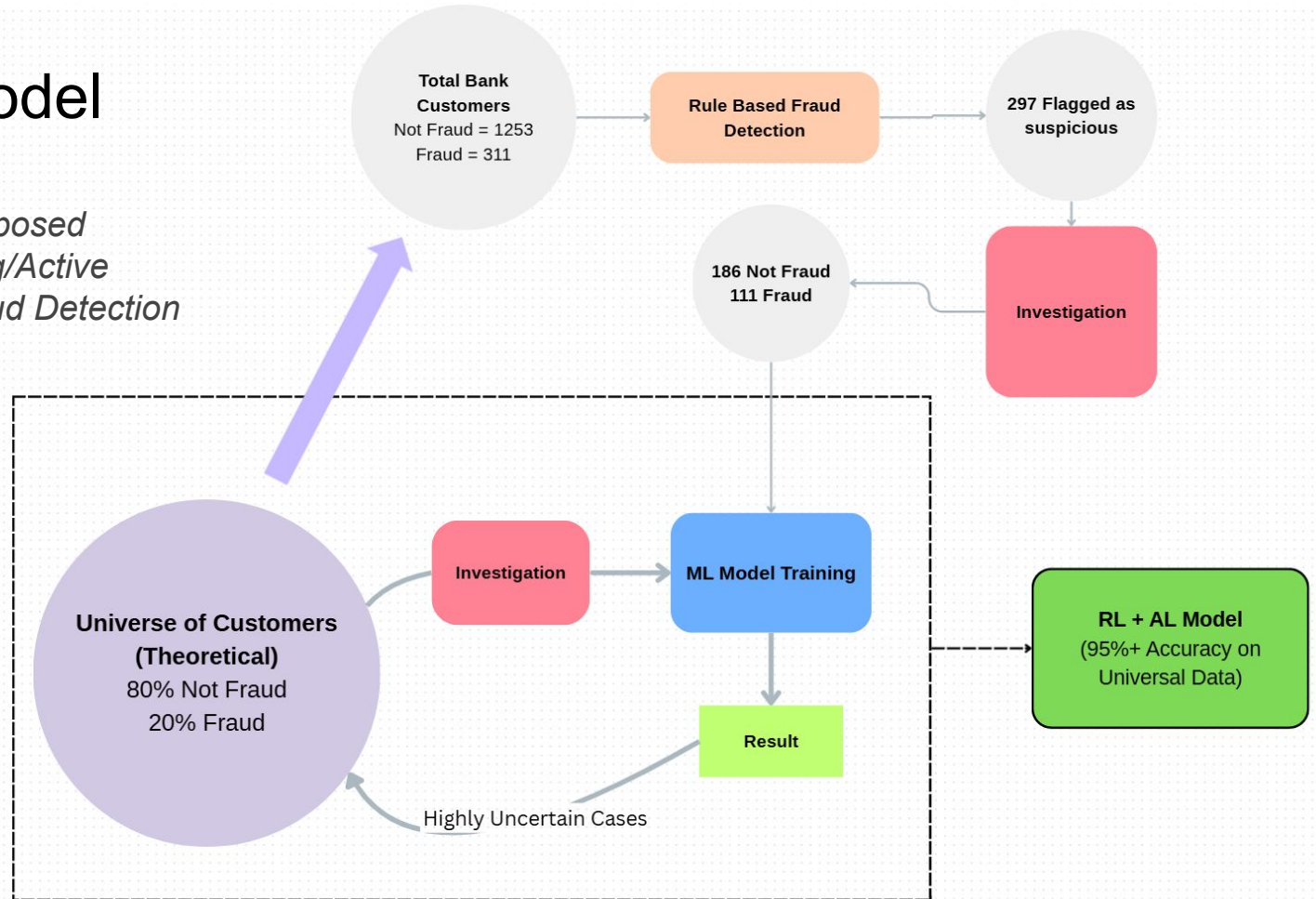


Fig: Schematic of a typical ML Model for Fraud Detection

Proposed Model

Fig: Schematic of a proposed Reinforcement Learning/Active Learning Model for Fraud Detection



Cost Sensitive Losses

		Prediction	
		Not Fraud(N)	Fraud(P)
Actual	Not Fraud(N)	TN Cost = 0	FP Cost = - Investigation(20k)
	Fraud(P)	FN Cost = - Insurance Amt (1.2L)	TP Cost = - Investigation(20k)

Costs include:

1. Insurance Amount
2. Investigation Expenses

Illustrative Example: Loss Due to Insurance Claims

Assuming, Cost of Investigation : Loss due to Fraud = 1x : 6x

- ❖ **Cost of Investigation: x = -₹ 20,000**
- ❖ **Loss due to Fraud: 6x = -₹ 1,20,000**
- ❖ Assuming 100 people are claiming Insurance every quarter

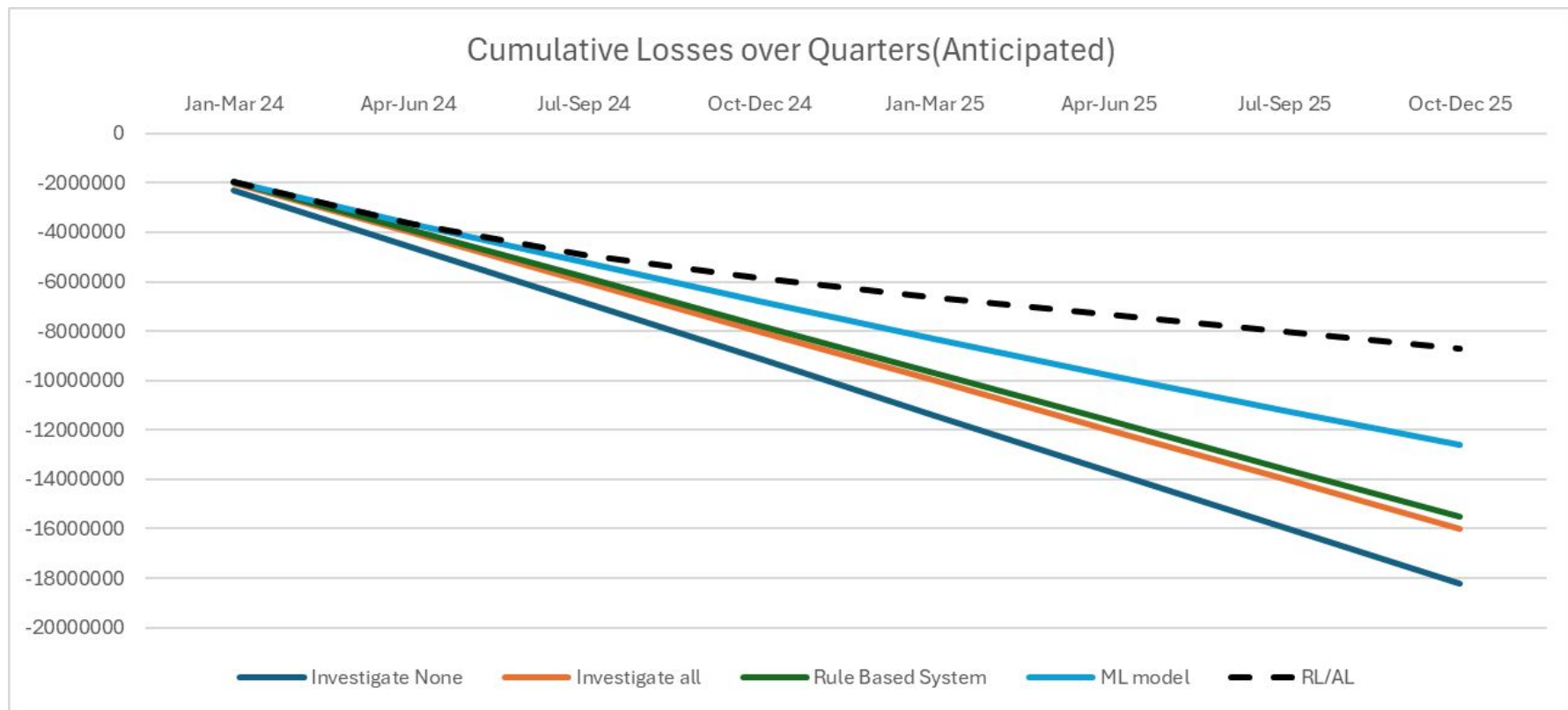
These results are Fabricated but it represents the overall idea and the proposed solution.

- Scenario 1-3: Models are not updated over time.
- Scenario 4: Model is based on the Historical Investigated cases, and it's updated on new investigated cases. This is Biased, because of the Initial Rule and Data.
- Scenario 5: Model has evolved from Rule Based to static ML and then to Feedback based ML model

Scenario Analysis

Model	Jan-Mar 24	Apr-Jun 24	Jul-Sep 24	Oct-Dec 24	Jan-Mar 25	Apr-Jun 25	Jul-Sep 25	Oct-Dec 25	Total Loss
1 Investigate None	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 22,80,000	-₹ 1,82,40,000
2 Investigate all	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 20,00,000	-₹ 1,60,00,000
3 Rule Based System	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 19,40,000	-₹ 1,55,20,000
4 ML model	-₹ 19,40,000	-₹ 16,80,000	-₹ 16,10,000	-₹ 15,60,000	-₹ 15,20,000	-₹ 14,70,000	-₹ 14,30,000	-₹ 14,10,000	-₹ 1,26,20,000
5 RL/AL	-₹ 19,40,000	-₹ 16,80,000	-₹ 12,80,000	-₹ 9,40,000	-₹ 7,80,000	-₹ 7,20,000	-₹ 7,00,000	-₹ 6,80,000	-₹ 87,20,000

Illustrative Results: Loss plotted by quarter

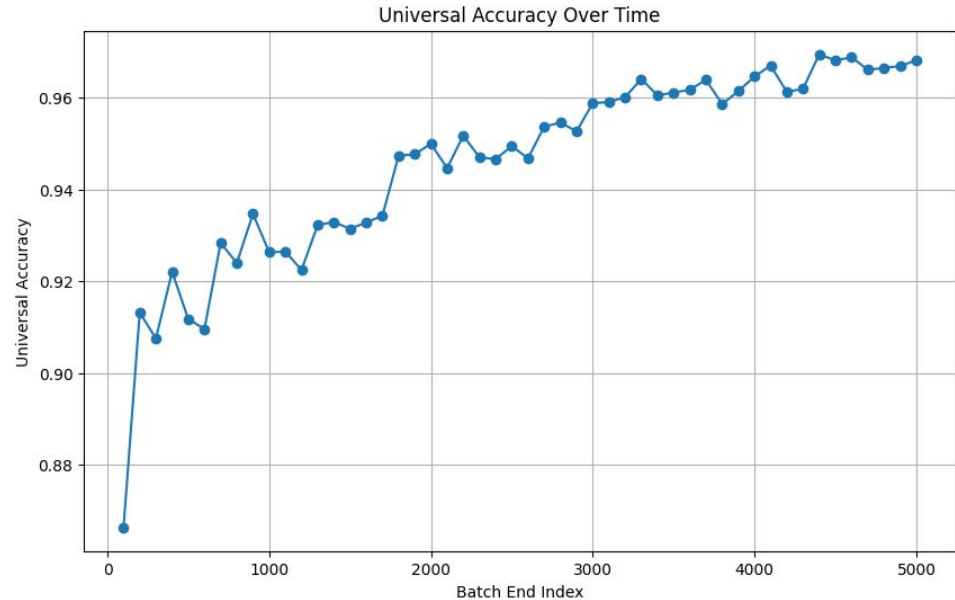


Results on Simulated Data: Improvement of Universal Accuracy when trained using RL/AL approach

For the ML Model:

Initial Cross Validation Balanced Accuracy: **92.5%** (Perceived Accuracy)

Initial Universal Accuracy: **85.1%**



Results on Simulated Data

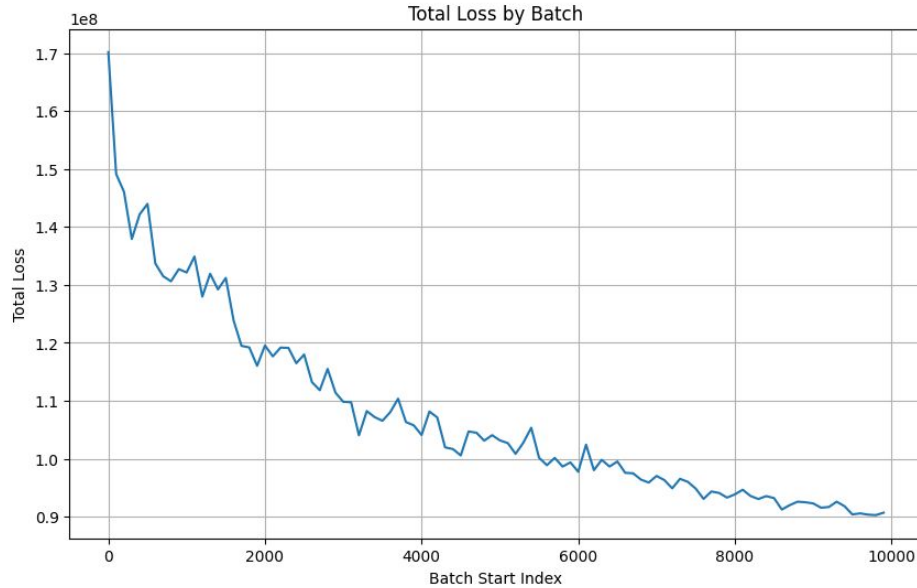


Fig 1: Drop in Business Loss over time because of Model Improvement

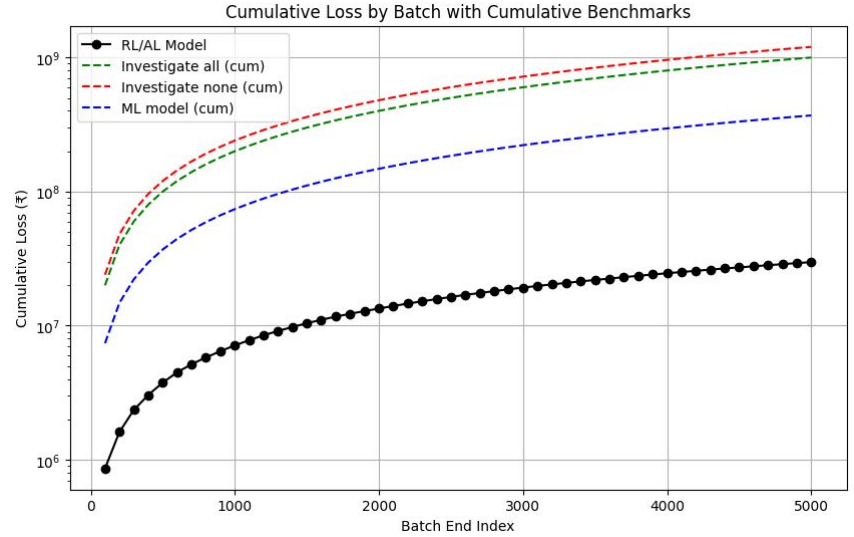


Fig 2: Comparison of cumulative losses for 4 different scenarios

Future Work

1. Building the RL + AL framework on the Actual Data
2. Cost Sensitive Framework (*treating mistakes as losses weighted by their real financial impact.*)
3. Binary class to multi class extension
 - a. No Fraud
 - b. Wrong Diagnoses
 - c. Phantom Billing
 - d. Ghost Patients

Thank You