

Introduction / Background / Motivation

•Elevance Health (EH) is a healthcare company managing insurance, benefits, and patient data for millions.

•Digital fraud causes financial losses and inefficiencies, impacting industries globally.

•In **healthcare**, where sensitive data and transactions are prevalent, fraud detection is critical.

•This project addresses the rising frequency of fraud, which burdens EH and increases costs for consumers.

•*Healthcare fraud accounts for \$455 billion of the \$7.35 trillion spent annually*. Most breaches result from hacking and unauthorized access.

•This project leverages security data to detect fraud in realtime.

•We aim to develop a machine learning-driven fraud detection system to enhance security and trust.

Fraud Detection Methodology Overview

To **detect digital fraud**, we implemented a structured machine learning (ML) process, addressing the challenges outlined in our introduction. This methodology integrates fraud research, data aggregation, preprocessing, model development, and continuous improvement, forming the foundation of our approach.

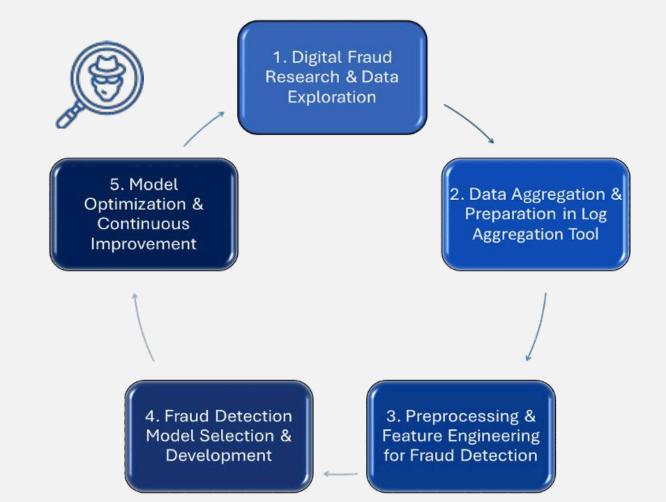


Diagram 1: Fraud Detection Model Development Process

Impact

•Build **trust** with the customers to ensure they stay loyal to EH

• Early detection of suspicious activity in real time to **minimize risk** of potential threat

•Ensure the **integrity and reliability** of EH in the healthcare insurance industry

•Strengthen the overall **security** of **customer data** to safeguard trust with customers

DETECTING DIGITAL FRAUD IN HEALTHCARE

Anna Bajszczak, Landon Berger, Sriya Chakravarthula, Ray Chishty, Sarah O'Farril-Gonzalez (TA), Abha Gupta, Nithil Krishnaraj, Adi Pawar, Božidar Perović, Vishal Ramasubramanian

1. Digital Fraud Research & Data Exploration Exploratory Tactics & EH Log Training Scenarios Exploration Focus •Understand •Learn •Explore log •Narrow indexes & cyber down fraud detection fraud exploration sources needs techniques •Build lookup Investigate functionality •Train on log •Identify impacted •Extract aggregation potential member tool & relevant fraud accounts fraud compliance scenarios indicators MITRE ATT&CK.

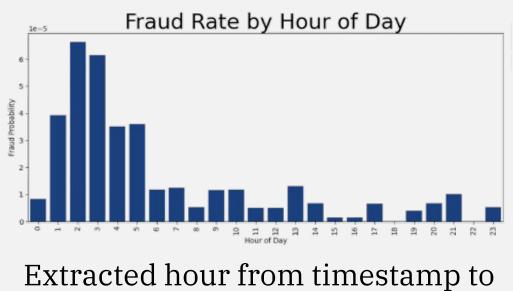
3. Exploratory Data Analysis (EDA) & Feature Engineering for Fraud Detection

EDA 0: Row Data Understanding

Aspect	Summary								
Total Rows	11,897,313								
Total Features	23								
Data Types	Mostly float64, 1 object								
Null Values	No missing values								
Feature Diversity	Mix of low & high cardinality								
Target Variable	is_fraud (binary)								

EDA 2A: Correlation Analysis

EDA 1: Data Transformation

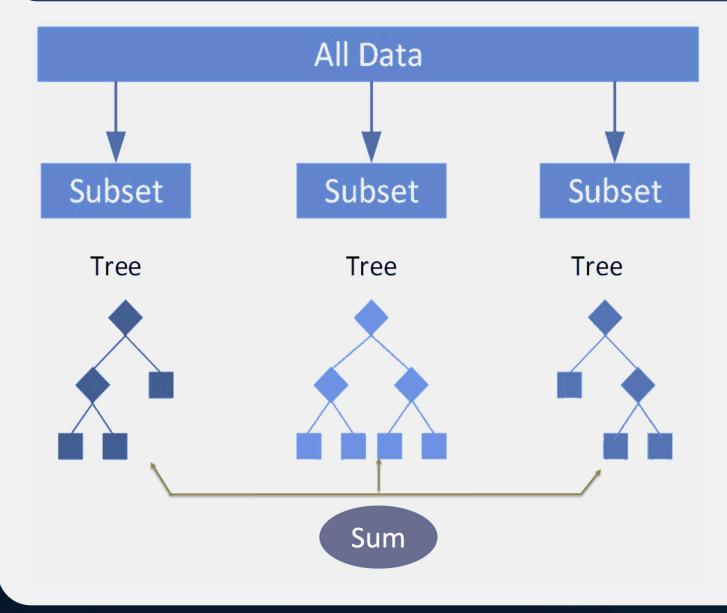


analyze temporal fraud patterns

0 100 0.01 0.02 0.00 0.01 0.00 <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>1</th></t<>																													1
 		0 -	1.00	-0.01	0.01	-0.02	-0.06	-0.00	0.01	0.02	-0.00	-0.00	0.02	0.01	-0.01	0.01	0.01	0.00	0.01	0.00	-0.00	-0.00	-0.01	0.01	0.00	-0.78	0.26	- 1.0	
3 - 0.02 0.01 0.01 0.07 0.01 0.02 0.00 0.02 0.00 0.01 0.01 0.01 0.00 0.00 0.01 0.01 0.01 0.00 0.01		1 -	-0.01	1.00	0.58	0.01	-0.07	0.65	0.45	0.53	0.35	0.00	0.08	0.47	0.16	0.51	0.52	0.04	0.61	-0.04	0.01	0.00	0.04	0.29	-0.00	0.01	-0.03		
3 -0.02 0.01 0.02 0.01 0.02 0.00 0.02 0.00		2 -	0.01	0.58	1.00	0.01	0.01	0.37	0.61	0.64	0.50	-0.00	0.09	0.68	0.09	0.56	0.51	0.05	0.99	0.04	0.03	0.00	0.07	0.28	-0.01	-0.02	-0.04	0.0	
5 -0.00 0.65 0.37 0.01 0.06 0.34 0.32 0.29 0.00 0.06 0.36 0.44 0.36 0.35 0.30 0.01 0.00 0.01 0.02 0.01 0.02 0.01 0.02 0.01 0.02 0.01 0.01 0.02 0.01 0.02 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.01 0.02 0.01 0.01 0.01 0.02 0.01 0		3 -	-0.02	0.01	0.01	1.00	0.07	0.01	-0.02	-0.00	-0.02	-0.00	-0.00	0.00	-0.02	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.04	-0.02	- 0.8	
6 - 0.01 0.45 0.61 0.02 0.04 0.34 1.00 0.01 0.95 0.03 0.02 0.03 0.01 0.00 0.03 0.17 0.01 0.02 0.03 0.01 0.00 0.01 0.02 0.03 0.01 0.00 0.01 0.02 0.00 0.01 0.00 0.01 0.00 0.01 0.01 0.00 0.01 0.00 0.01		4 -	-0.06	-0.07	0.01	0.07	1.00	-0.06	-0.04	0.04	-0.06	-0.00	0.01	0.02	0.03	-0.02	-0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	-0.01	0.11	-0.09		0.
6 0.01 0.45 0.61 0.02 0.01 0.05 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.03 0.01 0.00 0.01 0.00 0.01 0.00 0.00 0.01 0.00 0.00 0.00 0.01 0.00 0.		5 -	-0.00	0.65	0.37	0.01	-0.06	1.00	0.34	0.32	0.29	-0.00	0.06	0.36	0.04	0.36	0.45	0.03	0.39	0.03	0.01	0.00	0.04	0.18	-0.00	0.01	0.02	- 0.6	
8 - 0.00 0.35 0.50 0.02 0.00										and the second																		0.0	9
9 - 0.00 0.00																			a second s						-0.04	-0.03	-0.06		<u> </u>
b 10 0.02 0.08 0.09 0.00 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 <th< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>the second s</td><td>and the second second second</td><td>A DESCRIPTION OF THE OWNER OF THE</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>- 0.4</td><td>ť</td></th<>									the second s	and the second second second	A DESCRIPTION OF THE OWNER OF THE																	- 0.4	ť
11 -0.01 0.47 0.68 0.00 0.02 0.36 0.38 0.51 0.28 0.00 0.26 1.00 0.08 0.47 0.15 0.10 0.00 0.01																													8 0.
12 -0.01 0.16 0.09 -0.02 0.03 0.04 0.02 0.08 0.00 0.05 0.13 0.01 0.02 0.01 0.01																													Ξ
 13 - 0.01 13 - 0.01 14 - 0.01 15 - 0.00 10 - 0.01 0.52 0.51 0.00 0.01 0.01 0.52 0.51 0.00 0.01 0.01 0.52 0.51 0.00 0.01 0.01 0.52 0.51 0.00 0.01 0.01 0.01 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.01 0.02 0.01 <td></td> <td></td> <td>- 0.2</td> <td>-</td>																												- 0.2	-
A - 0.01 0.52 0.51 0.00 0.04 0.5 0.0 0.04 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.0 0.5 0.5 0.0 0.5																													ě
1 0.01 0.02 0.00 <t< td=""><td>m</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>a second s</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></t<>	m																		a second s										1
16 - 0.01 0.61 0.99 0.00 0.01 0.39 0.62 0.66 0.50 0.11 0.75 0.12 0.63 0.50 0.07 1.00 0.05 0.33 0.01 0.02 -0.04 -0.02 -0.04	Fe																		Construction of the local division of the									- 0.0	Č 0.
17 - 0.00 -0.04 0.04 -0.00 0.01 0.03 0.01 0.04 -0.00 -0.01 0.01 0.03 0.01 0.01 0.02 0.01 0.02 0.01 0.02 0.01 0.01 0.01 0.00<		15 -	0.00																										5
$ \begin{bmatrix} 17 - 0.00 & -0.04 & 0.04 & -0.00 & 0.01 & 0.03 & 0.03 & 0.01 & 0.03 & 0.01 & 0.04 & -0.00 & -0.01 & 0.13 & 0.00 & 0.03 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & -0.00 & -0.01 & -0.02 \\ 18 - 0.00 & 0$														A REAL PROPERTY AND INCOME.					and the second										ž
19 -0.00 0.00																												0.2	
20 - 0.01 0.04 0.07 0.00 0.04 0.03 0.04 0.02 -0.00 0.01 0.06 0.01 0.03 0.05 0.00 0.00 0.00 1.00 0.02 -0.04 0.02 0.04 21 - 0.01 0.29 0.28 0.00 0.01 0.18 0.17 0.27 0.11 0.00 0.02 0.21 0.21 0.22 0.33 0.13 0.01 0.00 0.00 -0.02 -0.04 0.02 -0.04 22 - 0.00 -0.00 -0.01 0.00 -0.01 -0.04 0.01 0.00 -0.01 -0.00 0.00 -0.00 -0.02 -0.04 -0.02 -0.04 23 - 0.78 0.01 -0.02 0.04 -0.01 -0.02 -0.01 -0.02 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.02 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.00 -0.01 -0.01 -0.01 -0.01																													
21 - 0.01 0.29 0.28 0.00 0.01 0.18 0.17 0.27 0.11 0.00 0.09 0.51 0.02 0.21 0.21 0.22 0.33 0.13 0.01 0.00 0.02 1.00 0.00 -0.02 -0.04 22 - 0.00 -0.00 -0.01 0.00 -0.01 -0.00 -0.01 -0.04 0.01 0.00 -0.01 -0.00 -0.01 -0.00 0.00																												0.4	0.
22 - 0.00 -0.00 -0.01 0.00 -0.01 -0.00 -0.01 -0.04 0.01 0.00 -0.01 -0.00 -0.01 -0.00 -0.00 0.00		20 -	-0.01	0.04	0.07	0.00	0.00																					0.1	
230.78 0.01 -0.02 0.04 0.11 0.01 -0.02 -0.03 -0.01 0.00 -0.02 -0.01 0.01 -0.02 -0.02 -0.00 -0.02 -0.01 -0.00 0.00 0.02 -0.02 -0.01 1.00 -0.21 24 - 0.26 -0.03 -0.04 -0.02 -0.09 0.02 -0.03 -0.06 -0.01 0.00 -0.01 -0.05 -0.02 -0.01 -0.03 -0.01 -0.04 -0.02 -0.00 -0.00 0.04 -0.04 -0.01 -0.21 1.00 0.00 0.02																													
24 - 0.26 -0.03 -0.04 -0.02 -0.09 0.02 -0.03 -0.06 -0.01 0.00 -0.01 -0.05 -0.02 -0.01 -0.03 -0.01 -0.04 -0.02 -0.00 -0.00 0.04 -0.04 -0.01 -0.21 1.00			_																									0.6	
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24		24 -	0.26	-0.03	-0.04	-0.02	-0.09	0.02	-0.03	-0.06	-0.01	0.00	-0.01	-0.05	-0.02	-0.01	-0.03	-0.01	-0.04	-0.02	-0.00	-0.00	0.04	-0.04	-0.01	-0.21	1.00		0.
			ò	i	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		

4. Model Selection

Feature Index

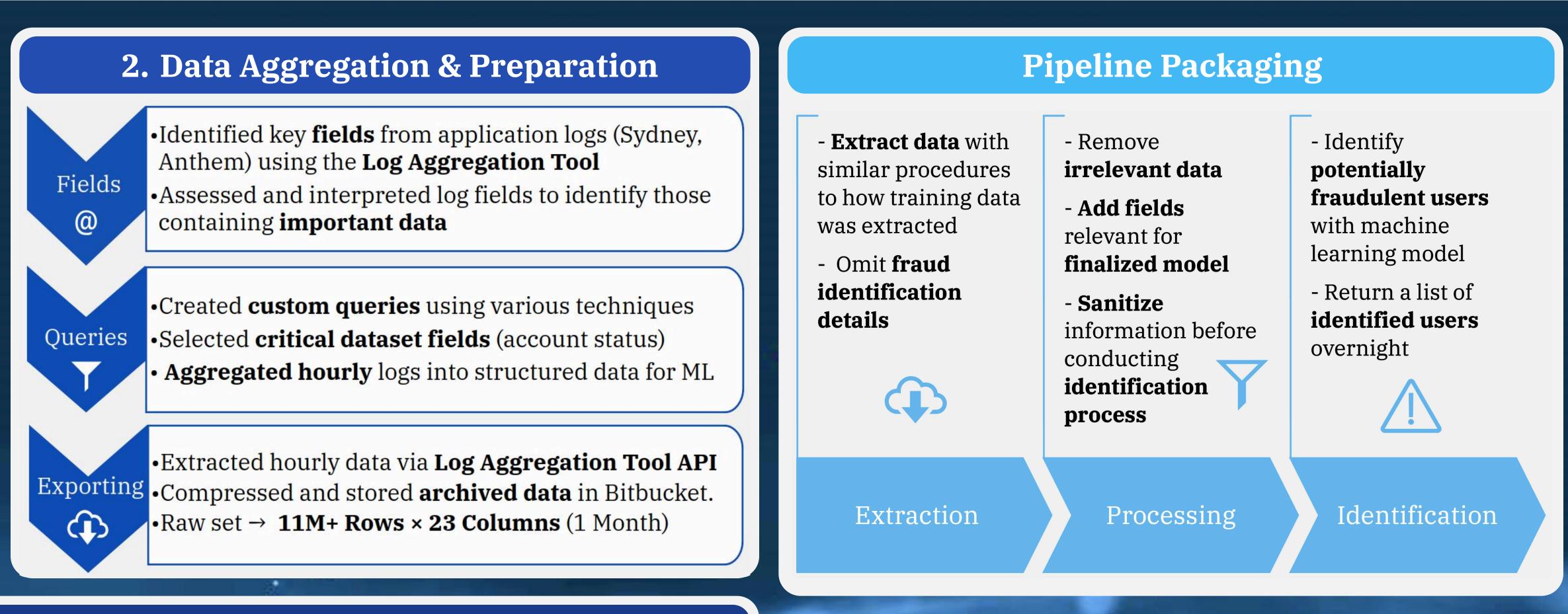


- **Split data** into subsets
- Train model with split data
- Assess performance (accuracy, precision)

The XGBoost model performed best

The Data Mine Corporate Partners Symposium 2025

This poster was created as part of an academic course and is intended solely for educational purposes. The content is based on synthetic data and hypothetical scenarios. It does not represent actual work, product, policies, or practices of Elevance Health, nor should it be interpreted as reflecting the views or official positions of the company.



EDA 2B: Adaptive Synthetic Sampling (ADASYN) for Fraud Cases

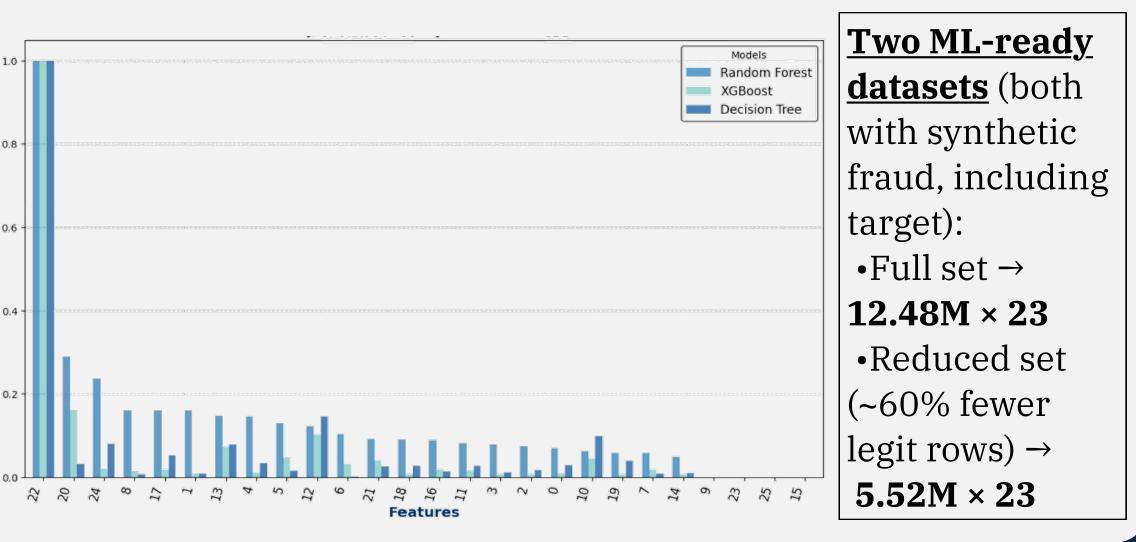
Before Resampling • ~0.001% Fraud (96 cases)



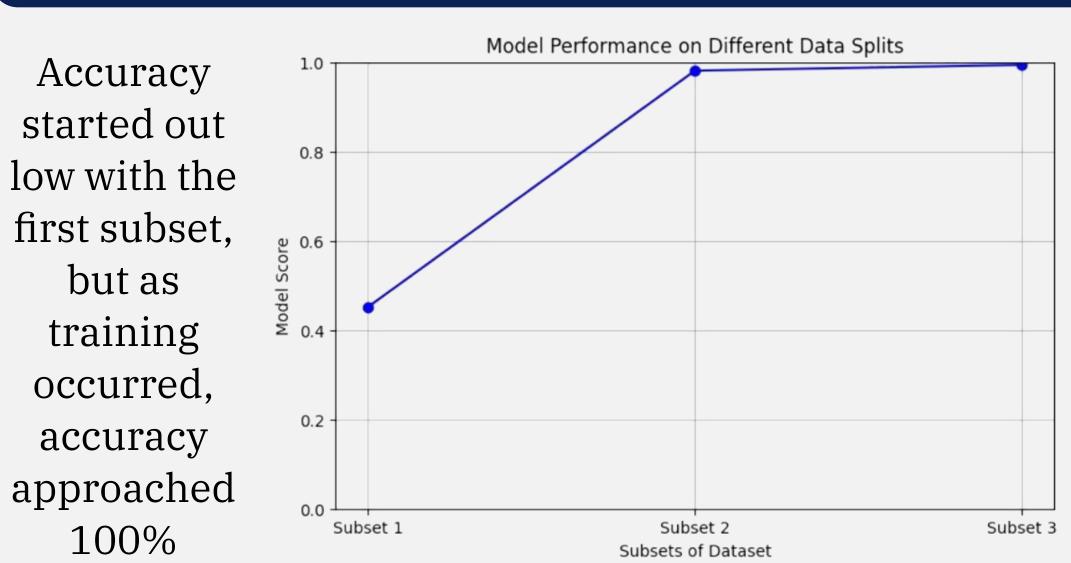
After Resampling9.1% Fraud (~1.1M cases)

EDA 2C: Feature Importance from Models

•Feature 22 ranked highest across all models •22 features + target, after removing 4 low-importance features



5. Model Deployment





Conclusions

- •EH handles sensitive patient data, emphasizing the need for a **secure fraud detection system**.
- •Our team analyzed EH data using the Log Aggregation Tool API, cleaned datasets, and built an ML model for fraud detection.
- •We analyzed threat actor patterns and established a **resilient detection protocol** powered by ML

Future Directions

- •Utilize **GPU power** for data collection and ML training
- •Consult mentors for **virtual environment resources**.
- •Explore log data more deeply to **develop advanced queries** for enhanced fraud detection insights.
- •Improve collaboration for higher productivity.
- Test models for fraud detection accuracy.

Acknowledgements & Resources

- Thank you to The Data Mine, Corporate Partner mentors, and faculty mentor for their support.
- Project Guidance: EH Cyber Defense and Security Analytics Dept. & Mustafa Abdallah Program
- **Support**: Maggie Betz, Bryce Castle

Resources:

¹ Crossing the Global Quality Chasm: Improving Health Care Worldwide: National Academies Press (US); 2018 ² HIPAA Journal. Editorial: Lessons from 2024 Healthcare Data Breaches.