Predicting Fraud in Credit Card Transactions

ITCS 5156-051 Project
Spring 2024 Semester
Chris Piacesi

Credit Card Fraud

According to a 2022 Press Release from Nilson Report:

- Payment card fraud losses worldwide exceeded $32 billion in 2021.

- Nearly $12 Billion in the US.

- Over the next 10 years, the industry is projected to lose an accumulated $397 billion worldwide.

- Online purchases leave merchants more vulnerable to fraud.

Nilson Report [www.nilsonreport.com]. (2022b, December 22). *Payment card fraud losses reach $32.34 billion* [Press release]. Retrieved April 13, 2024, from https://www.globenewswire.com/news-release/2022/12/22/2578877/ 0/en/Payment-Card-Fraud-Losses-Reach-32-34-Billion.html

Existing Approaches

★ Improved Strategy for High Recall Using KNN, LDA, and Linear Regression[1]

• Decision Tree, Logistic Classification, and Random Forests [2]

• Support Vector Machines [3]

[1] Chung, Jiwon, and Kyung-Ho Lee. "Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression." *Sensors*, vol. 23, no. 18, Sept. 2023, p. 7788. https://doi.org/10.3390/s23187788.

[2] Afriyie, Jonathan Kwaku, et al. "A Supervised Machine Learning Algorithm for Detecting and Predicting Fraud in Credit Card Transactions." *Decision Analytics Journal*, vol. 6, Mar. 2023, p. 100163. https://doi.org/10.1016/j.dajour.2023.100163.
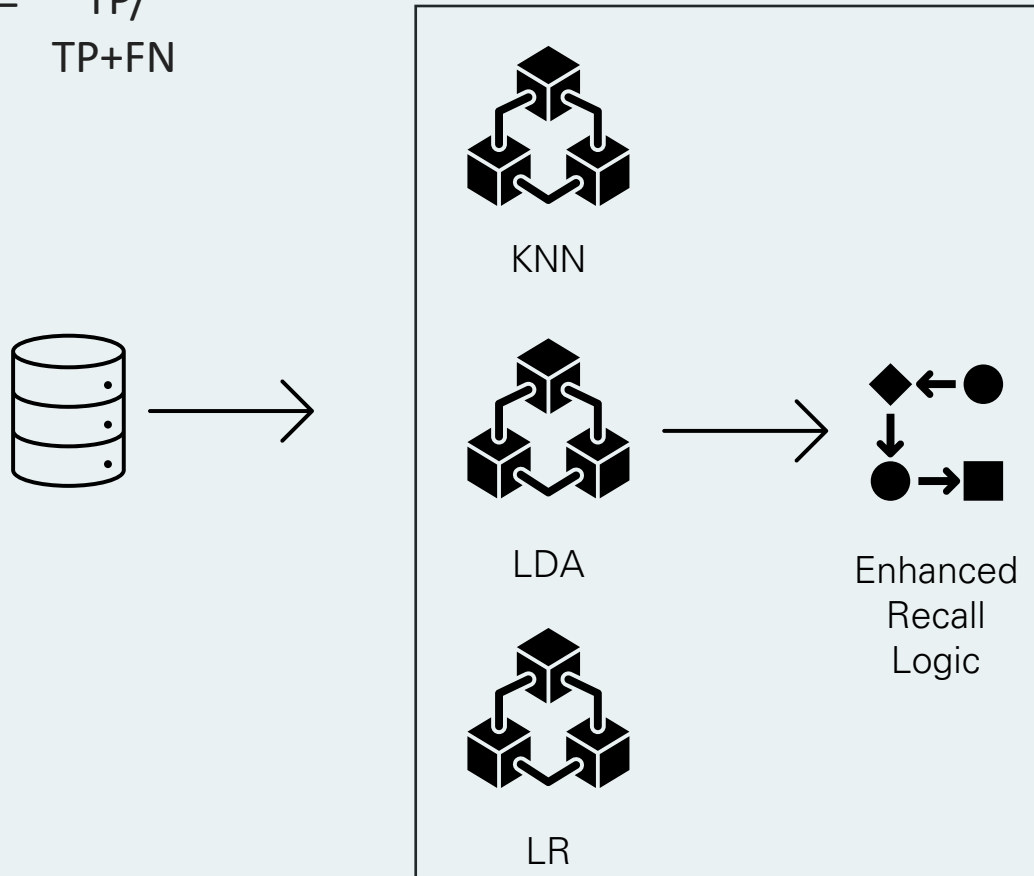
[3] Xia, Jianglin. "Credit Card Fraud Detection Based on Support Vector Machine." *Highlights in Science Engineering and Technology*, vol. 23, Dec. 2022, pp. 93–97. https://doi.org/10.54097/hset.v23i.3202.

Objective

High Recall to prevent the misclassification of genuine fraud cases as non-fraud instances.

Recall = $\frac{TP}{TP+FN}$

KNN

LDA

Enhanced Recall Logic

LR

- If pKNN[i] is 0 or pLDA[i] is 0, and pLR[i] is less than mvLR, then pOR[i] is set to 0.
- Conversely, if pKNN[i] is 1 or pLDA[i] is 1, and pLR[i] is greater than mvLR, then set pOR[i] to 1.
- If neither of the conditions is met in a particular row, pOR[i] simply takes on the value of pKNN[i].

Datasets

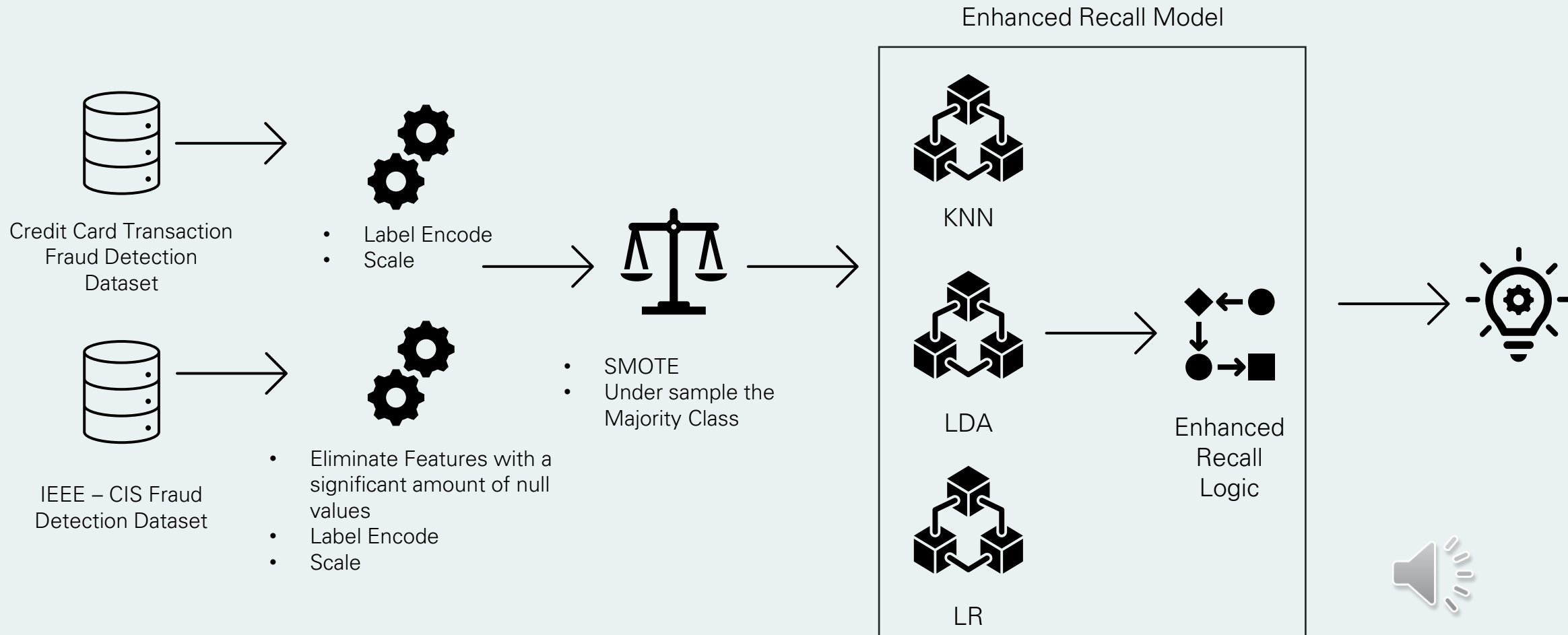[IEEE-CIS Fraud Detection | Kaggle](#)

The data comes from Vesta Corporation's real-world e-commerce transactions and contains a wide range of features from device type to product features. Vesta Corporation is the forerunner in guaranteed e-commerce payment solutions and provides fraud detection and prevention platform.

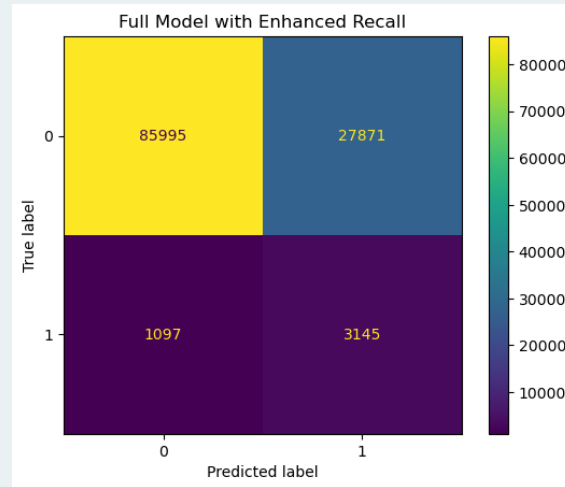[Credit Card Transactions Fraud Detection Dataset (kaggle.com)](#)

Simulated data contains 1000 customers doing transactions with a pool of 800 merchants from January 1st, 2019 through December 31st, 2020.

My Approach – Improved Strategy for High Recall Using KNN, LDA, and Linear Regression

Enhanced Recall Model

Credit Card Transaction Fraud Detection Dataset

- Label Encode
- Scale

- SMOTE
- Under sample the Majority Class

IEEE – CIS Fraud Detection Dataset

- Eliminate Features with a significant amount of null values
- Label Encode
- Scale

KNN

LDA

LR

Enhanced Recall Logic

6

Results

Full Model with Enhanced Recall

	precision	recall	f1-score	support
0	0.99	0.76	0.86	113866
1	0.10	0.74	0.18	4242
accuracy			0.75	118108
macro avg	0.54	0.75	0.52	118108
weighted avg	0.96	0.75	0.83	118108

KNN
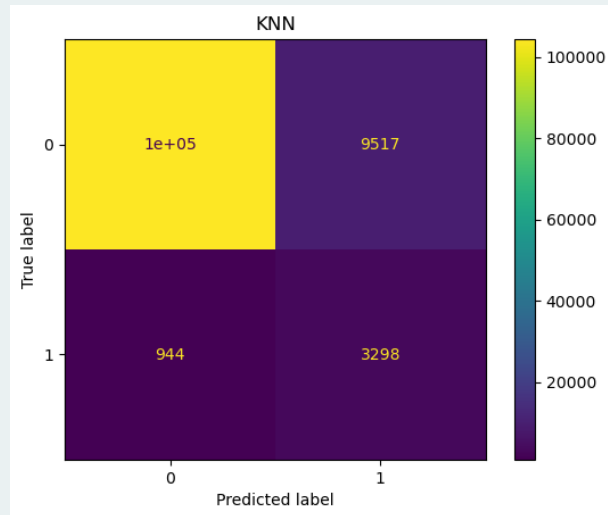
	precision	recall	f1-score	support
0	0.99	0.92	0.95	113866
1	0.26	0.78	0.39	4242
accuracy			0.91	118108
macro avg	0.62	0.85	0.67	118108
weighted avg	0.96	0.91	0.93	118108

LDA

	precision	recall	f1-score	support
0	0.99	0.75	0.86	113866
1	0.10	0.74	0.18	4242
accuracy			0.75	118108
macro avg	0.54	0.75	0.52	118108
weighted avg	0.96	0.75	0.83	118108

LR
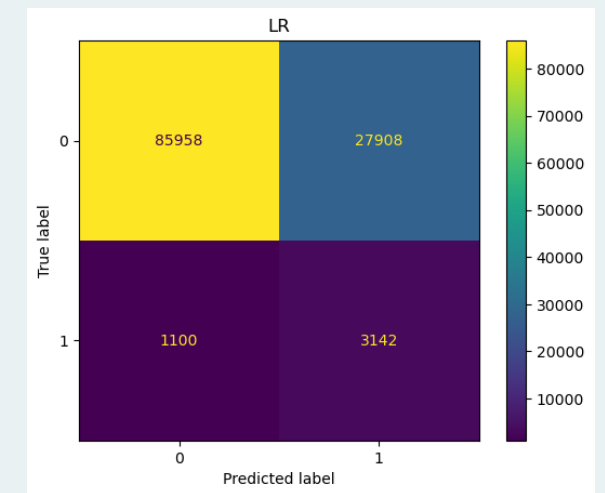
	precision	recall	f1-score	support
0	0.99	0.75	0.86	113866
1	0.10	0.74	0.18	4242
accuracy			0.75	118108
macro avg	0.54	0.75	0.52	118108
weighted avg	0.96	0.75	0.83	118108

Conclusions and Future Work

- Poor performance of the LDA and LR models is bringing down the performance of the Ensemble Learning approach.

- Much more sophisticated approaches are required to get accurate results with high recall.

- Vesta Corporation promotes its use of not only the application of ML, but also AI, and Human Intelligence for its fraud detection and prevention platform.

- Additional Feature Engineering and Model Tuning to improve performance.

- Time Series Analysis

Chris Piacesi

cpiacesi@uncc.edu

https://www.linkedin.com/in/chris-piacesi

Thank you