By: Hewan Sirak

Dataset Link: https://www.kaggle.com/datasets/kartik2112/fraud-detection/data

Github Link: https://github.com/hewansirak/iCog-Trainings/tree/main/HyperParameter_Tuning

The goal of this project was to train multiple models, tune hyperparameters for one, and evaluate the performance of the models to show how tuning makes a huge difference and decreases computational cost. The dataset chosen was bulky and I used the fraud_Train csv to make it easier.

I used the models: Logistic Regression and Random Forest Regression and evaluated them using Root Mean Squared Error (RMSE) and R-squared (R2).

Logistic Regression:

- Provides a linear approach to classification being relatively fast to train and interpret.
- However, it may struggle with complex, non-linear relationships within the data.

Random Forest Regression:

- This ensemble method builds multiple decision trees and aggregates their predictions.
- It is capable of capturing complex, non-linear relationships and more robust to outliers and noise.
- However, it is computationally more expensive and took too much time that I had to use hyperparameter tuning
- I employed RandomizedSearchCV to optimize the model's hyperparameters (n_estimators, max_depth, min_samples_split) due to the large dataset and computational constraints I couldn't let it finish training without any tuning it took longer times

These are the results I got at the end

Model	RMSE	R2	Training Time (approx.)
Logistic Regression	0.07	0.04	2 minute
Random Forest Regression	0.03	0.80	45 minutes