

Credit Card Fraud Detection

Shadman Kaif, Abdurrafay Khan, Krutarth Patel, Shanthosh Sivayogalingam

Supervisor: Zeb Tate Administrator: Inci McGreal (Team 2022524, Section #5)



The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

Motivation

There is an increase in e-commerce sales and credit card use which is expected to grow by 7.4% annually. Currently, \$118 billion is lost in legitimate sales due to false positives since the state-of-the-art technology only has 78% accuracy.

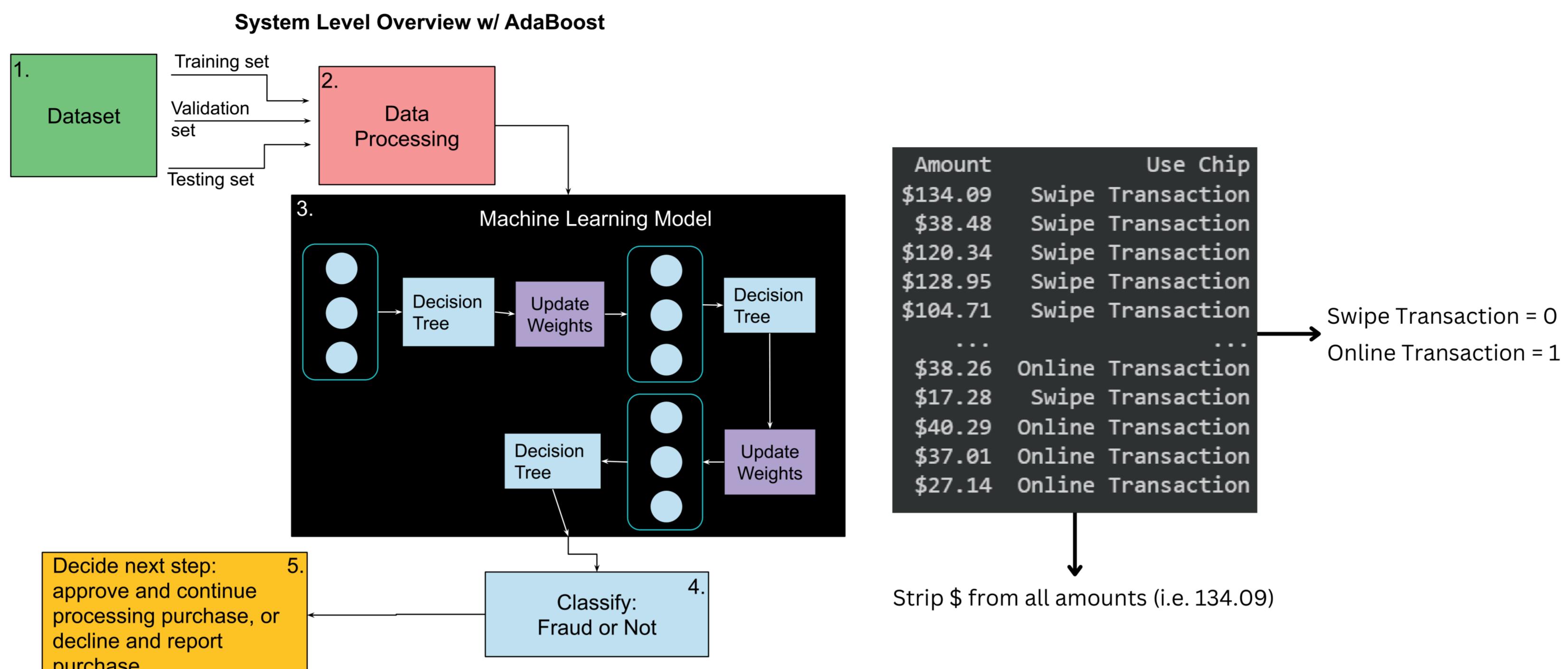


Project Goal & Requirements

Create a linearly classifiable **ensemble** ML algorithm that **minimizes** the number of **false positives** to at least **20-30%** in a credit card dataset

- Utilize ensemble learning
- Achieve 95% overall model accuracy
- Achieve recall score of 50% or higher
- Achieve false negative rate of < 30%
- Processing speed (65,000 transactions/second)
- Adherence to “Protection of Personal Information” policy

System Overview

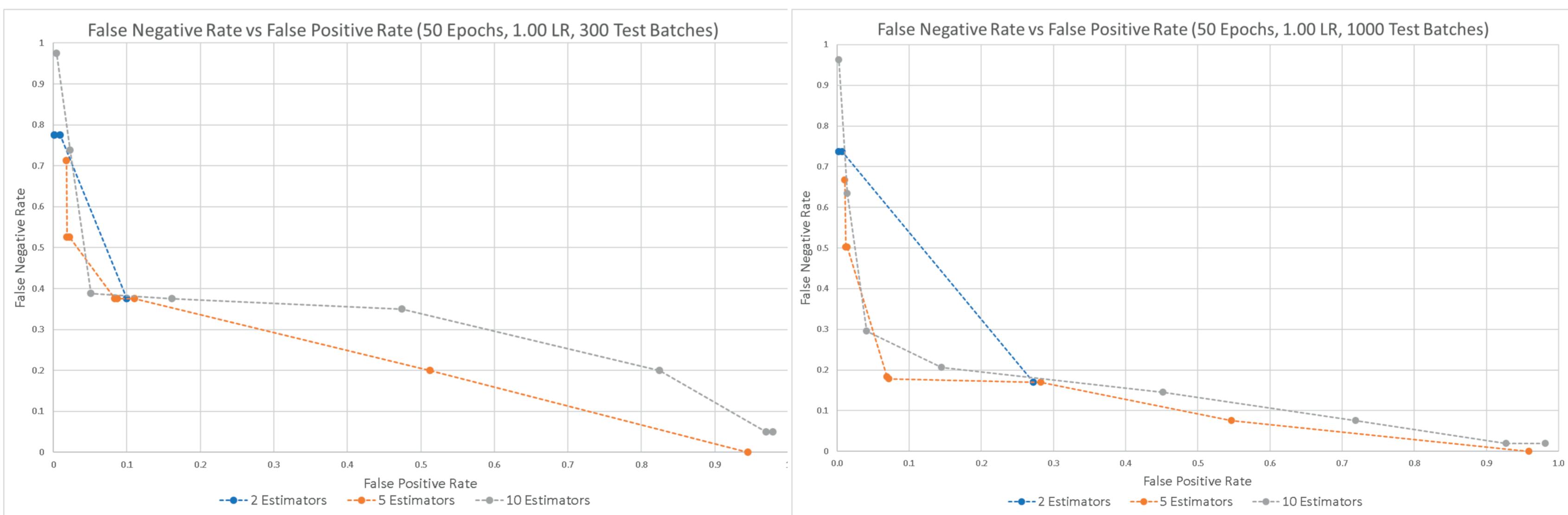


- The IBM dataset has 24 million transactions and is split into training (50%), testing (20%), and validation data (30%)
- Data is processed to convert it from categorical strings into numbers
- Data is then used to train the decision tree AdaBoost model

Testing and Verification

Multiple tests were conducted in order to meet project requirements:

- Tested the model with different hyper-parameters (learning rate, batch size, # of estimators...) to gauge optimal model performance
- Tested multiple iterations to compute model processing speed



Our Final Design Solution

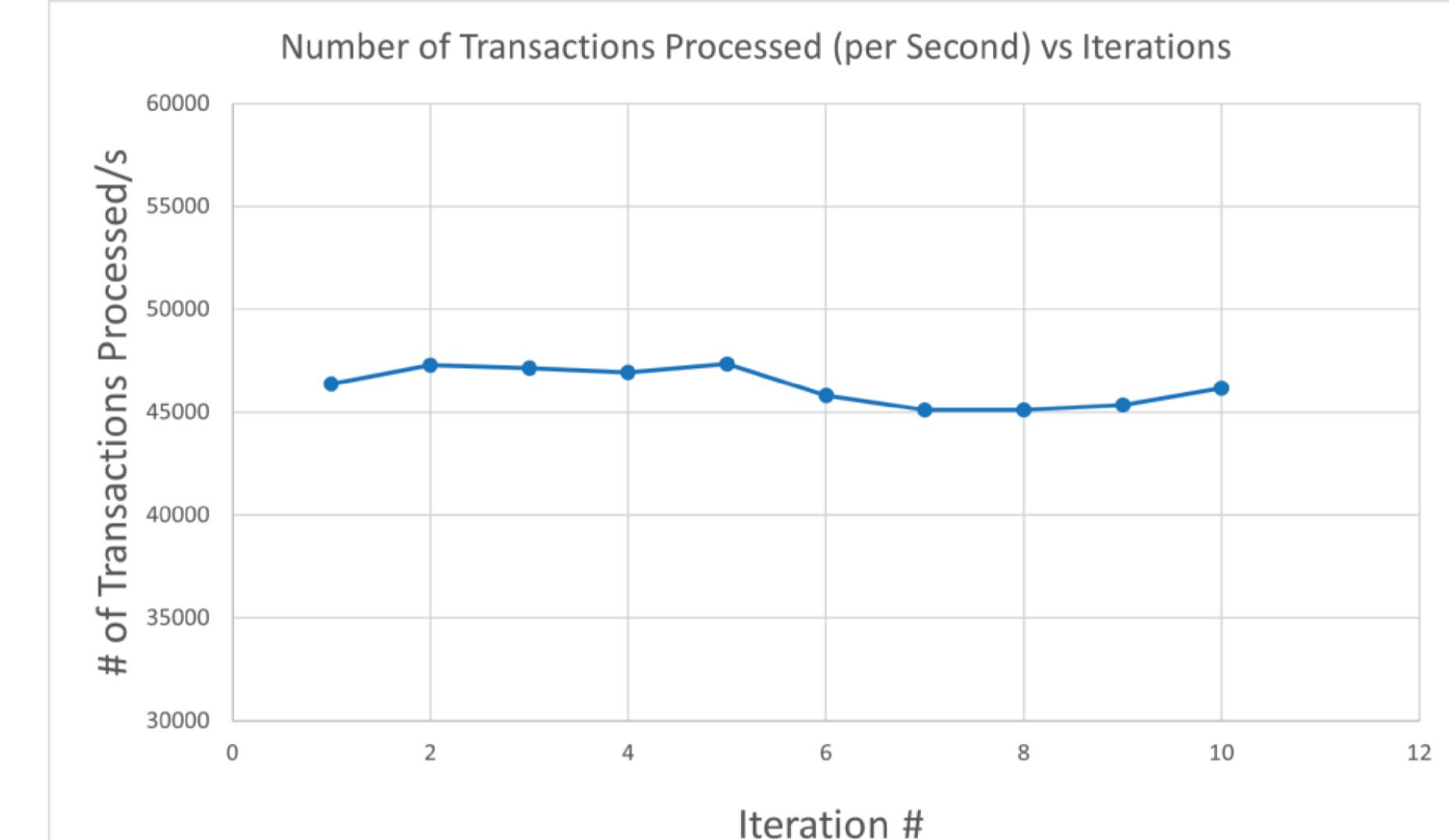
An Adaboost model trained with the following parameters:

- **Learning Rate = 1.00, Batch Size = 16**
- **# of Estimators = 10, Prediction Threshold = 50%**

Results

- Accuracy = 95.9%
- False Positive Rate = 4.10%
- False Negative Rate = 29.6%
- Recall = 70.4%
- Speed = 46,257 transactions/sec

Actual	
Predicted	Positive
Positive	150 True Positive
Negative	4575 False Positive
Positive	63 False Negative
Negative	107212 True Negative



Conclusion

- Final model achieves all objectives except transaction speed
- Better hardware can improve transaction speed
- Model can be implemented in a bank system to combat fraud
- Further testing can improve the model
 - Explore different types of weak classifiers for Adaboost