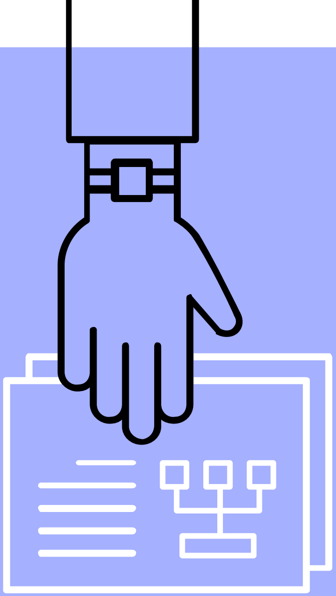
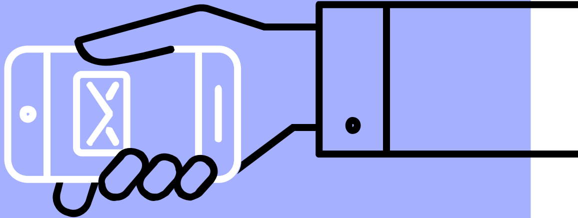
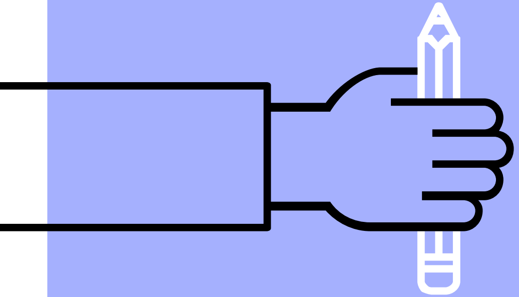


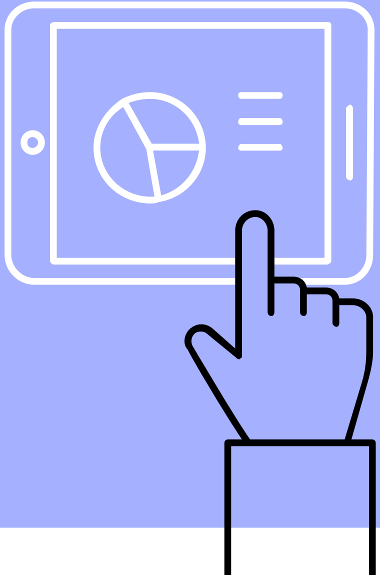


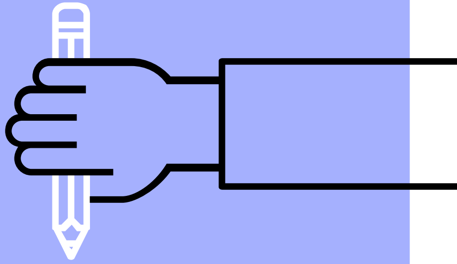
PROJECT PROPOSAL

Credit Card Transaction Fraud Detection

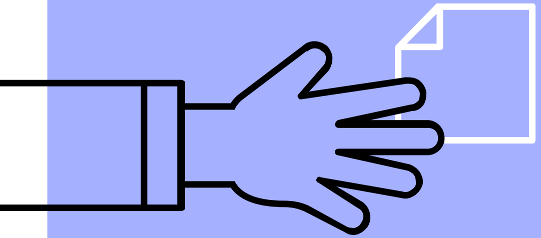
Group 6 - Scarlett Dong, Deep Prajapati, Mitchel Smith, Sylvia Yang







Why this topic?



$9,470,000,000

per year

Whoa! That’s a lot of money, don’t you agree?

Motivation

With the rapid growth of credit card use in e-commerce and the booming of touchless payment activities during Covid-19 pandemic, credit card fraud has been a growing problem worldwide.

As the top one in cases of fraud transaction, US has reported losses of $9.47 billion per year* in 2018 and is generally frustrating. With so many transaction, an average of 1 billion per day worldwide**, it’s impossible for businesses to review every transaction by hand. In addition, the attempted fraud transactions rose 35% in dollar amount in April 2020 during global pandemic***.

An automated detection program must be developed in order to decrease the number of auto.

*US Reported losses for 2018

**Global Transactions per day

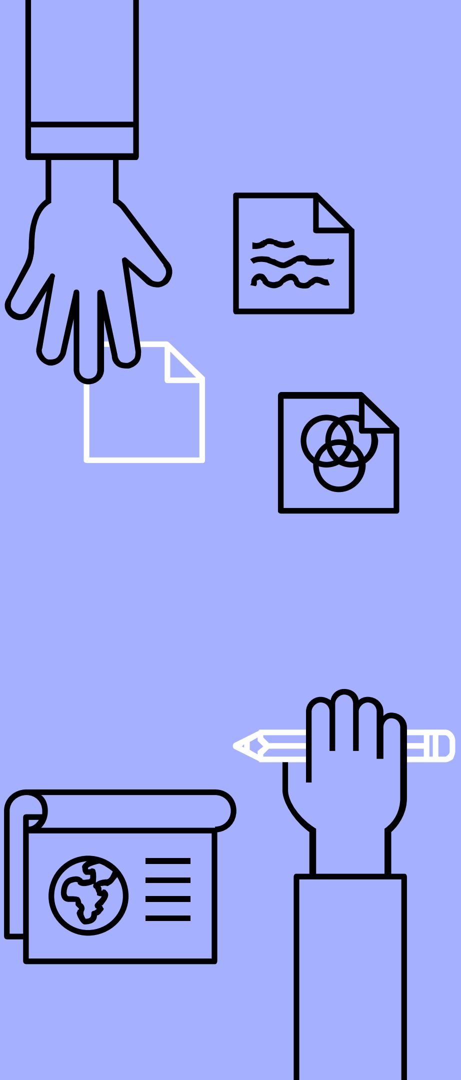
***fraud transactions rise during pandemic

Problem Statement

There’s too much data for a human to review every transaction and it nearly impossible to do so. We must identify a subset of the transactions for further review by human.

Ideally this isn’t just a random sample but is an educated guess where the subset is only likely fraudulent charges.

The problem, then, is how do we identify likely fraudulent charges?



Catalina, Jimena. (2020). Salerio Powerpoint template. Retrieved from https://www.slidescarnival.com/salerio-free-presentation-template/1875

Literature Survey

- Credit Card Fraud Detection with a Neural-Network
- Credit Card Fraud Detection Using Hidden Markov Model
- Credit card fraud detection using machine learning techniques: A comparative analysis
- Credit Card Fraud Detection - Machine Learning methods
- Cost sensitive modeling of credit card fraud using neural network strategy

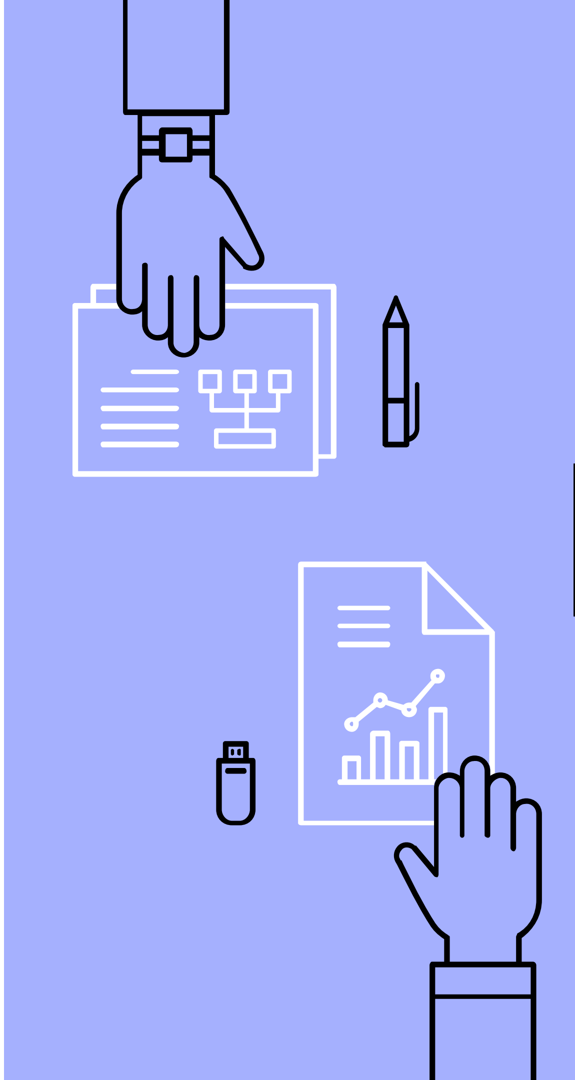
Credit Card Fraud Detection with a Neural-Network

By Sushmito Ghosh and Douglas L. Reilly, 1994

Trained Neural-Network on 650 thousand bank accounts each with 50 data fields.

This distinguished between types of fraud i.e. Stolen Card vs Cloned Card.

Notably required access to huge amounts of personal data. Access to the full account and 50 data fields is a lot. For comparison, the Kaggle dataset we are using only has access to transactions and the 30 data fields we are using have been pre-processed so that we don’t know what they represent.



Credit Card Fraud Detection Using Hidden Markov Model

By Srivastava et al 2008

Uses the Hidden Markov Model for learning.

This method focuses on identifying fraud committed through theft of credit card information, either the physical card or verifying information. One thing worth noting is that it doesn't know the type of item being purchased, which could be helpful for detecting fraud.

Another important note is that this approach ignores a major aspect of credit card fraud which is the opening of credit cards using fake or stolen identities. Since this method focuses on identifying patterns and discrepancies for an individual account, it would work poorly for new cards.

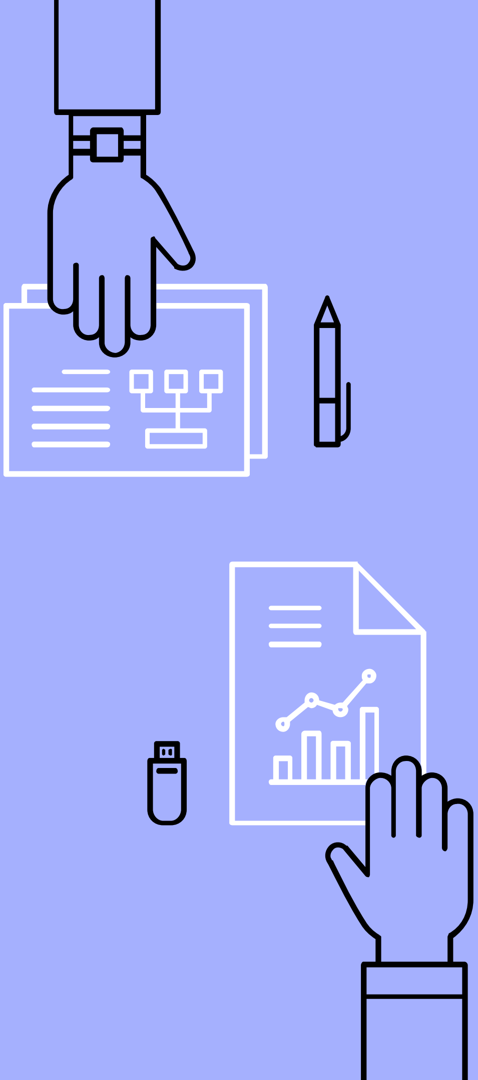
Credit card fraud detection using machine learning techniques: A comparative analysis

By John O. Awoyemi; Adebayo O. Adetunmbi; Samuel A. Oluwadare etc. at 2017

Many techniques have been applied to credit card fraud detection, artificial neural network, genetic algorithm , support vector machine, frequent itemset mining, decision tree, migrating birds optimization algorithm , naïve bayes.

A comparative analysis of logistic regression and naive bayes is carried out in.

The paper reports the experimental results and discussion about the comparative analysis.



Credit Card Fraud Detection - Machine Learning methods

By Dejan Varmedja; Mirjana Karanovic; Srdjan Sladojevic; etc. at 2019

There are two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, cvv code, type of card and other.

The purpose of this paper is to analyze various machine-learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection.

The dataset can be downloaded from Kaggle and it contains 284,807 transactions where 492 transactions were frauds.

Cost sensitive modeling of credit card fraud using neural network strategy

By Fahimeh Ghobadi; Mohsen Rohani at 2016

With the rapid growth of credit card use in e-commerce, credit card fraud has become a rising issue in banking transactions. In order to reduce the risk of losing reputation of credit card issuers, fraud detection has become an important subject.

This paper, using the real transaction data from a big Brazilian credit card issuer, develops a model with the name of Cost Sensitive Neutral Network (CSNN). This model is based on Artificial Neural Networks and Meta Cost procedure in order to deal with imbalance data. The advantages of this model is to save costs and increases detection rate.

This paper is going to help our project with dealing with issue of imbalanced data with Meta Cost procedure.

Our

IDEAS

- Using the Kaggle Dataset and Isolation Forest detect fraud.
- Using the same dataset and Linear Regression detect fraud for comparison.

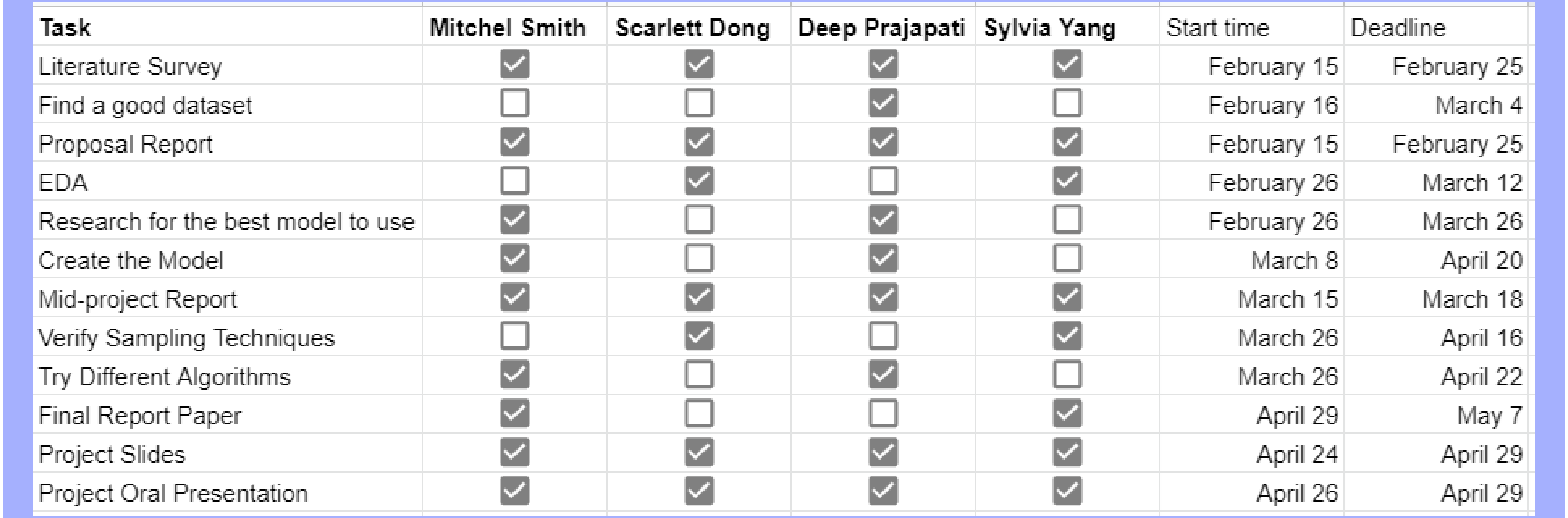
Our Questions

- How does an isolation tree work?
- How good can the results be from a dataset that is highly unbalanced?
- Do certain methods work especially well on unbalanced data?

Our Expectations

- Answer the questions on the previous slide.
- Being able to use similar methods/solutions in future scenarios (such as at a job or other research)

We Value Our Teamwork!



Task	Mitchel Smith	Scarlett Dong	Deep Prajapati	Sylvia Yang	Start time	Deadline
Literature Survey	☑	☑	☑	☑	February 15	February 25
Find a good dataset	☐	☐	☑	☐	February 16	March 4
Proposal Report	☑	☑	☑	☑	February 15	February 25
EDA	☐	☑	☐	☑	February 26	March 12
Research for the best model to use	☑	☐	☑	☐	February 26	March 26
Create the Model	☑	☐	☑	☐	March 8	April 20
Mid-project Report	☑	☑	☑	☑	March 15	March 18
Verify Sampling Techniques	☐	☑	☐	☑	March 26	April 16
Try Different Algorithms	☑	☐	☑	☐	March 26	April 22
Final Report Paper	☑	☐	☐	☑	April 29	May 7
Project Slides	☑	☑	☑	☑	April 24	April 29
Project Oral Presentation	☑	☑	☑	☑	April 26	April 29

THANKS!

Any questions?

Sylvia Yang <syang51@uncc.edu>

Mitchel Smith < msmit590@uncc.edu>

Scarlett Dong <sdong5@uncc.edu>

Deep Prajapati <dprajap1@uncc.edu>

PPT template: Catalina, Jimena. (2020). Salerio Powerpoint template. Retrieved from https://www.slidescarnival.com/salerio-free-presentation-template/1875