🎓

# Capstone 3 Project Proposal

| | |
|---|---|
| ◴ Created | @November 9, 2023 12:35 PM |
| ⊙ Class | Capstone Three |
| ⊙ Type | Project |
| ☑ Reviewed | □ |

## The Problem:

According to the Federal Trade Commission, credit card fraud is the single most common form of identity theft.

According to an Infosys report, credit card fraud rose by 21% in 2021 and is on pace to cause over $43 billion in losses by 2026. To add salt to the wound, it’s estimated that over 80% of credit cards currently in use in the United States have been compromised.

The importance of a bank to safeguard its customers' credit card information is paramount, and even the most vigilant of cardholders can fall victim to the ever-increasing schemes of those who would wish to steal their information.

I remember sitting in a Starbucks with my wallet on the table only to have a man run by my table, swipe my wallet, and run out before I even had a chance to react. These types of physical intrusions are common, yes, but in today’s digital world, it’s far more common to have your information stolen electronically without you even knowing it.

The ability of a bank or credit card institution like American Express to quickly classify credit card charges as fraudulent or not enables customers to feel safe, secure, and confident that their information is under careful watch. It also prevents massive financial losses to these banks and institutions.

Fraud detection ML models are one tool banks use in the fight against credit card theft, and they can be a more effective way of identifying fraud than conventional methods.

💡 Some benefits include:

**Faster Detection Time**: ML models can help quickly identify shifts from regular spending patterns in real-time. The farther the transaction in question is from the standard pattern, the more likely it’s fraud.

**Higher Accuracy of Identified Charges**: Sometimes conventional methods will block genuine transactions from going through; I think we’ve all experienced that at some point if we’ve gone on vacation outside the country. With the right data, an ML model can better identify the real factors that encompass a fraudulent transaction and more accurately predict what’s really fraud and what isn’t.

Once those real factors and patterns are identified, we can use the model on larger and broader datasets to further refine and paint the best picture we can of what’s fraud and what isn’t.

💡 **The end goal of any model is this**: Is what we’re trying to predict useful and valuable to predict?

If the answer is no or you’re unsure, then maybe a model isn’t right for the given situation.

For the case of credit card fraud, the case is clear, and the value it brings to banks and financial institutions is well known.

My goal is to use data on 550,000 credit card transactions to build a baseline model to predict fraud. I want to understand which features most impact a transaction being classified as fraud, and from there drive further analysis.

My main error metric I’ll examine will be recall because I’d say a false negative does more harm to a consumer and to a financial institution than a false positive does. Both are important, but I want to minimize false negatives.

The dataset will also be highly imbalanced, so accuracy will need to be examined in the broader context. If 99.2% of all transactions are not fraud, then having an accuracy of 99.2% could just mean I predicted all charges as not fraud and completely missed the point here.

I’m excited to take on this challenge and think it will be a great send-off for the program.