**Our Process**

# **Business Understanding**

**What is Credit Card Fraud?**

→ It’s a form of “identity theft.” A person’s card or banking information is used to make a transaction without the person’s knowledge. [1]

→ When the perpetrator uses a physical card, it’s called “card-present fraud”. Else, it’s called “card-not-present fraud” [1]

**Why should we be concerned about this?**

→ In the next decade, credit card fraud may result in $400B in losses globally – Nilson [2]

→ In 2021, credit card fraud rose 70% from last year to $5.8B – FTC [3]

→ From 2.8 million consumers, most frauds are now CNP due to the increase of digital consumerism [3]

**What data from users do we need? And how can we use that data to detect fraud?**

* Purchase history – to determine the purchasing behavior of a consumer and detect anomaly
* Location – can be used to determine if it’s a fraud from a foreign country
* Device ID – can be used if the owner of the card did in fact make those transactions
* IP address – like location, can be used to determine location
* Payment amount – can be used to detect anomalies like a sharp increase or rapid transactions
* Transaction information – can be used to detect fraud by its rapidness and if buy a lot of something in bulk

## Business Question

Let’s say the goal is to:

* Reduce cost and increase revenue
* Develop a model that can predict if a transaction is fraudulent accurately and timely

Other questions to ask:

* Do we have enough data?
* Do we have the right team?
* What are our success metrics?

## Technical Question

**Why should we need an ML solution?**

* There are numerous marketplaces and shops across the globe. Hundreds to thousands to millions of transactions are being made every second. So there are a lot of data to be analyzed
* Using pure manpower and resources would require unmeasurable time and energy
* You would make humans doing repetitive and menial tasks
* With ML and the consultation of experts in the field will make the process of detecting a fraudulent transaction would be quicker, more accurate, and require less resource and time

**Can we build a ML solution?**

* We have a large dataset of transactions labeled as either Fraud or Non-Fraud along with other variables from Kaggle [4]
* Data is stored in the correct format (csv) and is guaranteed to be accessible

**Does the ML solution serve the business requirement?**

* If the project is successful, there will be a reduction cost and increase in revenue
* The ML model will be more efficient, accurate, and faster than humans

## Reference

[1] Inscribe. (2022, August 19). *Credit Card Fraud Detection: Everything You Need To Know*. Fraudulent Document Detection & Automation. Retrieved October 30, 2022, from <https://www.inscribe.ai/fraud-detection/credit-fraud-detection>

[2] Mullen, C. (2021, December 14). *Card industry faces $400B in fraud losses over next decade, Nilson says*. Payments Dive. Retrieved October 30, 2022, from <https://www.paymentsdive.com/news/card-industry-faces-400b-in-fraud-losses-over-next-decade-nilson-says/611521/>

[3] Staff, the P. N. O., & Staff, D. P. I. P. and C. T. O. (2022, February 22). *New data shows FTC received 2.8 million fraud reports from consumers in 2021*. Federal Trade Commission. Retrieved October 30, 2022, from <https://www.ftc.gov/news-events/news/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021-0>

[4] Bannourchaker. (2021, December 9). *Frauddetection\_part1\_eda*. Kaggle. Retrieved October 30, 2022, from <https://www.kaggle.com/code/bannourchaker/frauddetection-part1-eda/notebook>

# **Data Collection**

* Compiled between 1st Jan 2019 - 31st Dec 2020
* 1000 consumers with 800 merchants
* Really unbalanced, fraud is only about 0.172%
* Simulation

## Reference

[5] Shenoy, K. (2020, August 5). Credit Card Transactions Fraud Detection Dataset. Kaggle. Retrieved October 31, 2022, from <https://www.kaggle.com/datasets/kartik2112/fraud-detection>