FRAUD DETECTION

ASKING THE RIGHT QUESTIONS?

- **What is we making?**
- **Why we made this?**
- **What we propose?**
- **What we deliver?**
- **How we achieved it?**
- **What will you get?**

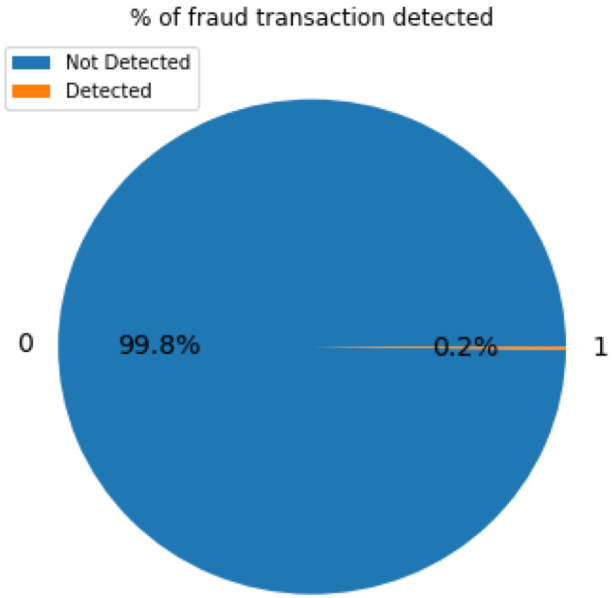
In this demo on **fraud detection**, we are trying to address these questions.

WHAT IS WE MAKING?

- **Fraud detection** is valuable to many industries including some of the mentioned.
  - Banking and financial sectors.
  - Insurance.
  - Law enforcement.
  - Government agencies.
- In recent years we have seen a **huge increase** in Fraud attempts, making fraud detection important as well as challenging.
- **Data mining** and **machine learning** help to foresee and **rapidly distinguish fraud** and make **quick move** to limit costs.

WHY WE MADE THIS?

Note: Data used for the Demo is open source and available on Kaggle.



- With increase in **Fraud attempts**, despite countless efforts and human supervision, **hundreds of millions** are lost due to **fraud**.
- Due to small cases in large population detection of fraud is important as well as challenge.
- The plot above clearly shows the **need** for a system which can be **fast** and **reliable** to mark the transaction which is fraud.
- Since, the current system is letting fraud transaction able to **pass through a system** which is **not labelling them as a fraud**.

WHAT WE PROPOSE?

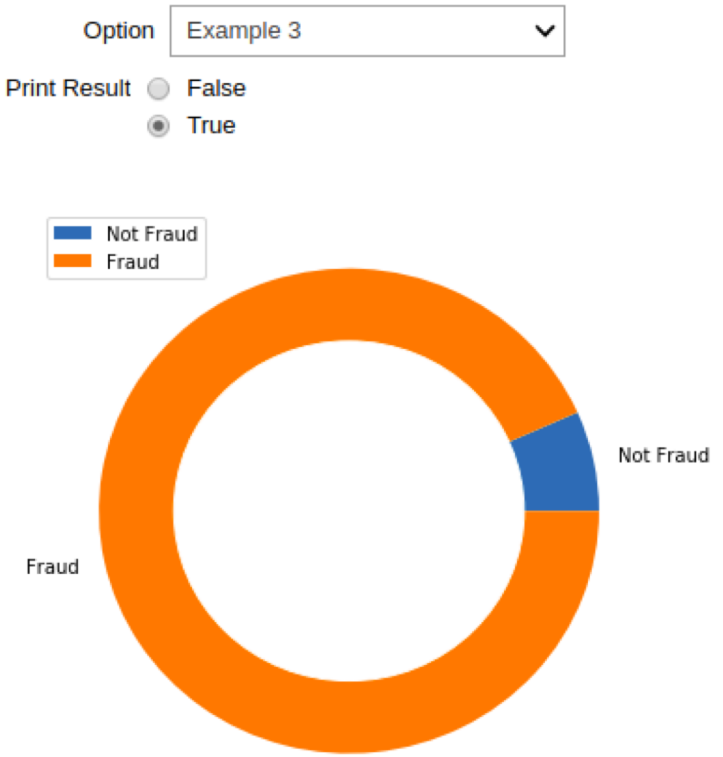
- **Data mining** and **machine learning** help to foresee and **rapidly distinguish fraud** and make **quick move** to limit costs.
- Using data mining tools, a **huge number of transactions** can be looked to **spot pattern** and distinguish fraud transaction.
- Fraud can happen due to –
  - Stolen credit cards.
  - Misleading accounting.
  - Phishing mails.
  - Transaction.
- In this demo we are focusing one **fraud transaction data**.

WHAT WE DELIVER?

- End result is an **interface** to input data and a **trained machine learning** model which is capable of detecting fraud with greater accuracy.

- Machine learning model is scalable(but may require few changes),

  depending on new dataset.
- Model can be converted as per the need of distributed system.
- Model can handle huge data (given that machine is powerful enough to handle the data.)



HOW WE ACHIEVED IT?

- We have performed Exploratory Data Analysis on the sample data which includes:
  - Data Acquisition
  - Data Cleaning
  - Feature Engineering
  - Data Mining
- We have used ensemble technique to make accuracy better, than single learner.
- We have created interface to the system to access machine learning model in user friendly way.
- We have tested different model and technology stack to create better system which is **fast, reliable and rapidly distinguish fraud.**