Deloitte.

Capstone project
Credit Card Fraud Detection

A presentation by Daniel Pérez Hernández

01 – Content

2 – Why Fraud Detection?

- **Shift towards new technologies.**
- **Most recent pandemics**
- **Lack of risk exposure awareness**

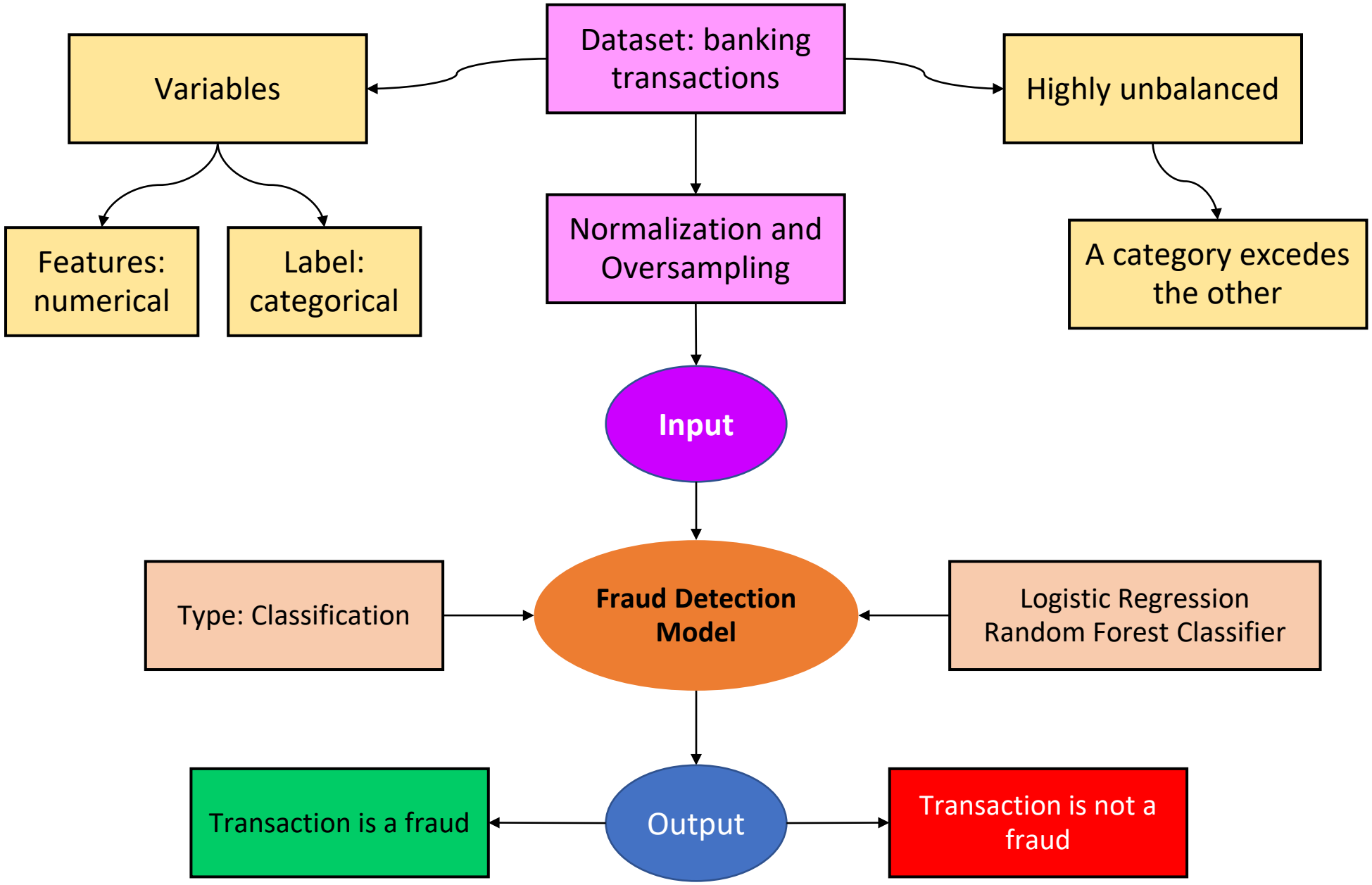
From PwC's Global Economic Crime and Fraud Survey 2022

Artificial Intelligence is vital for financial risk control in cloud environment.
(Wang, et.al; 2021).

- **300 BC, Greece: A shipping merchant kept the "lost" cargo and the claiming from it.**
- **1821: Gregor MacGregor sold non-existing land to investors from Europe.**
- **Recent days: ATM/application/internet banking fraud, amongst others.**

The Association of Certified Fraud Examiners estimates that US organizations lose about 7% of their revenues to fraud.
Nisbet, et.al (2018)

3 – ML Flow Overview


Dataset: banking transactions
Variables
Features: numerical
Label: categorical
Normalization and Oversampling
Highly unbalanced
A category excedes the other
Input
Type: Classification
Fraud Detection Model
Logistic Regression
Random Forest Classifier
Transaction is a fraud
Output
Transaction is not a fraud

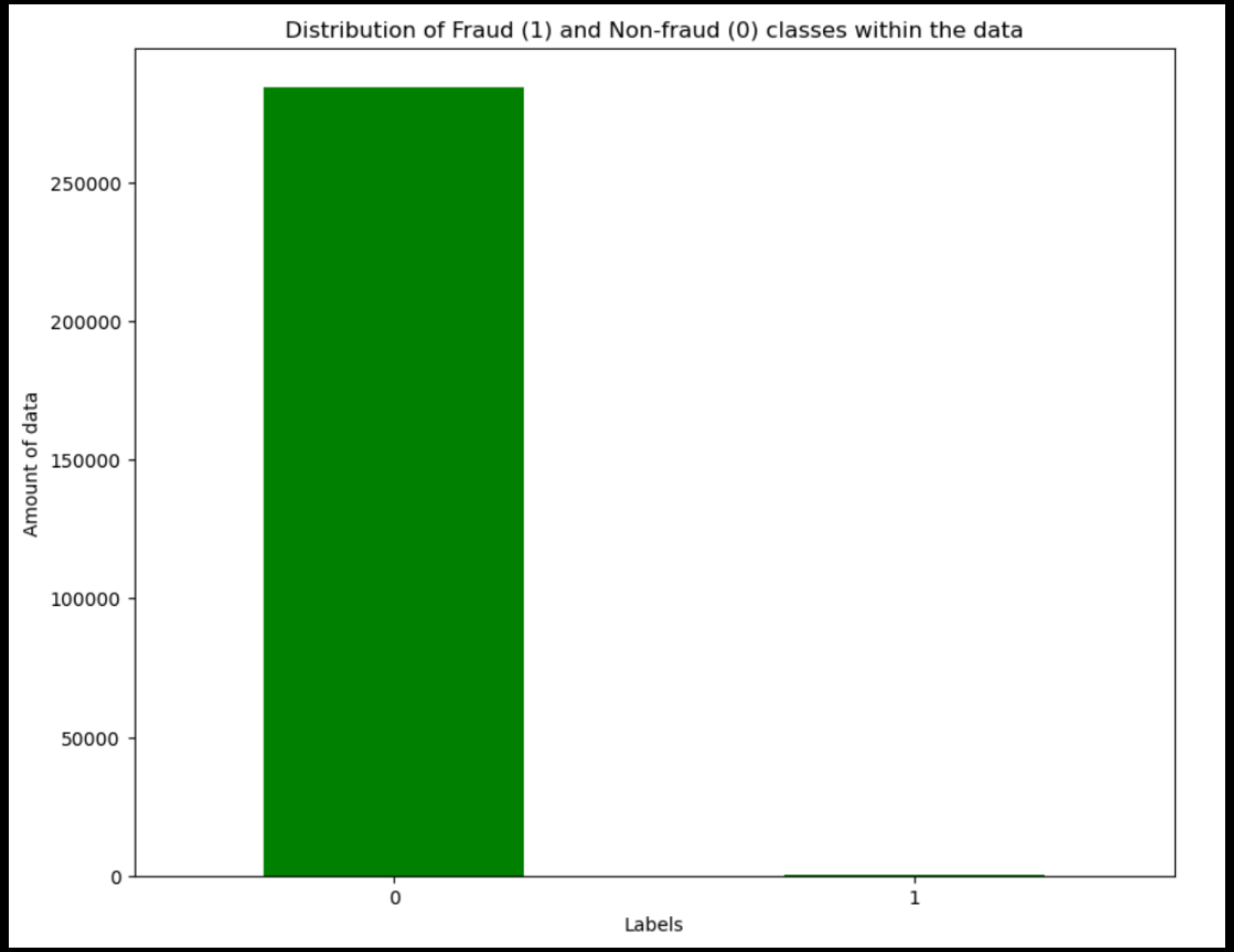
4 – Analyzing the data

Original data

99.83% of the total transactions are labeled as Non-Fraud (284,315)

The rest (0.17%) is labeled as Fraud (495)

Total transactions: 284,807

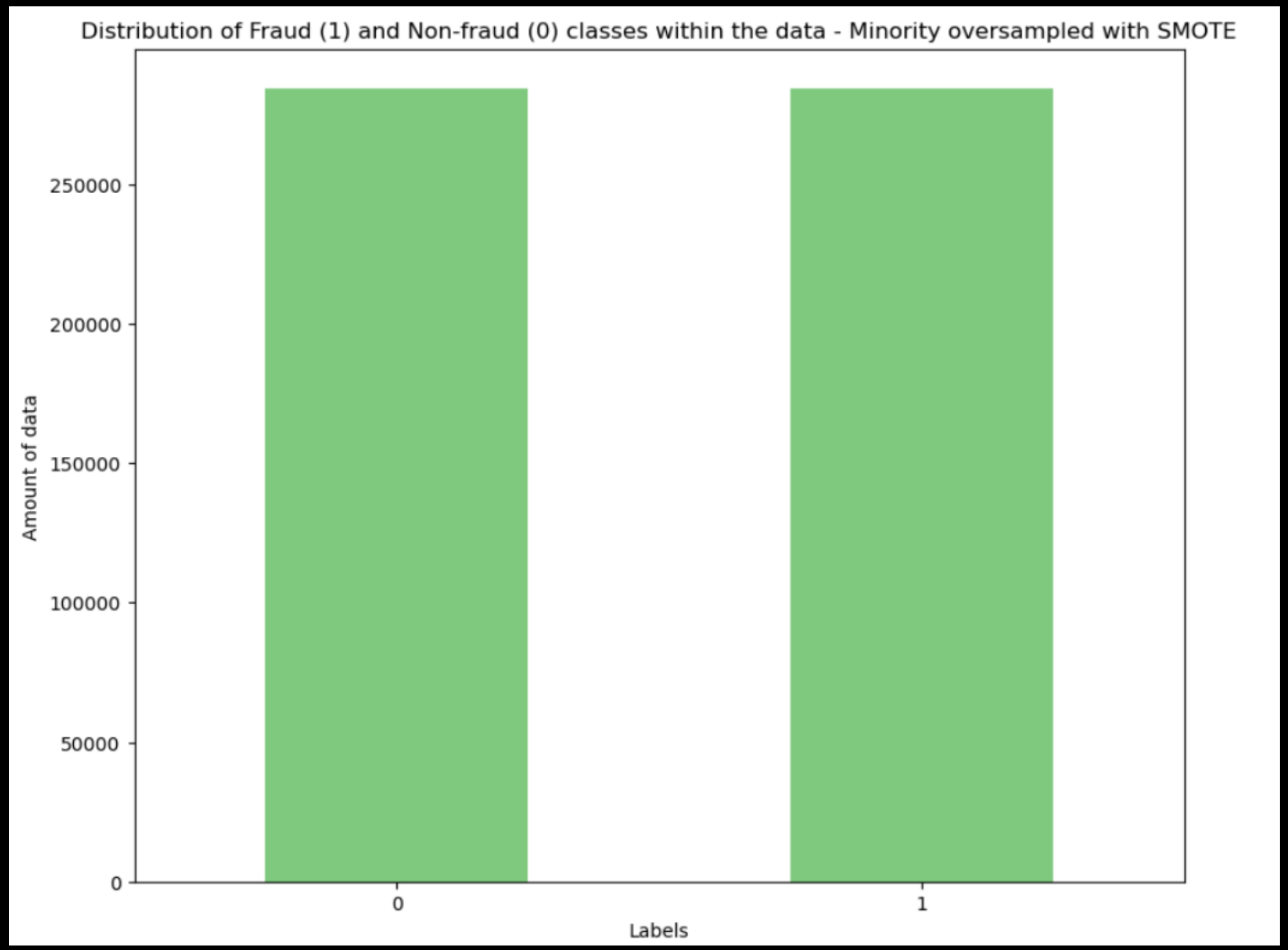


Oversampled data

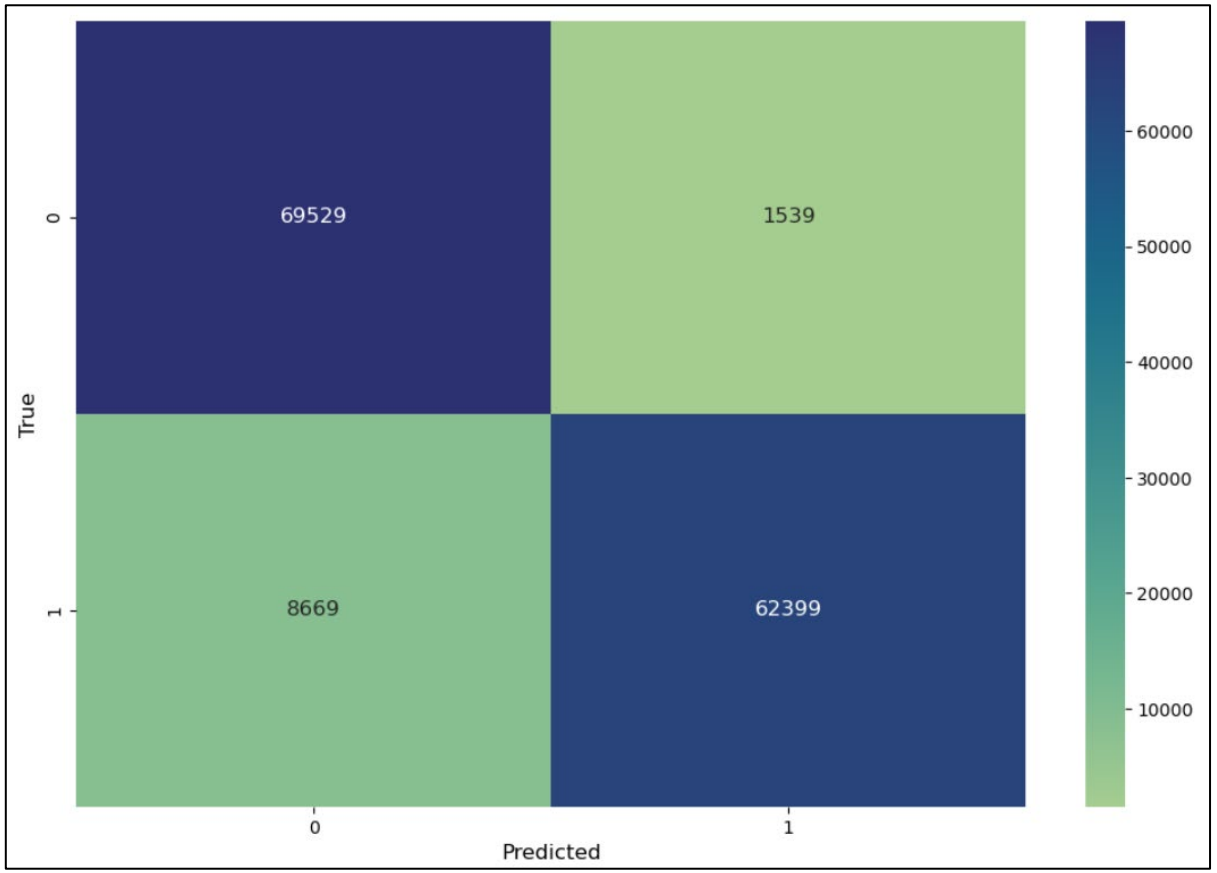
50% of the total transactions are labeled as Non-Fraud (284,315)

The other 50% is labeled as Fraud (284,315)

The new data has doubled the size of the original

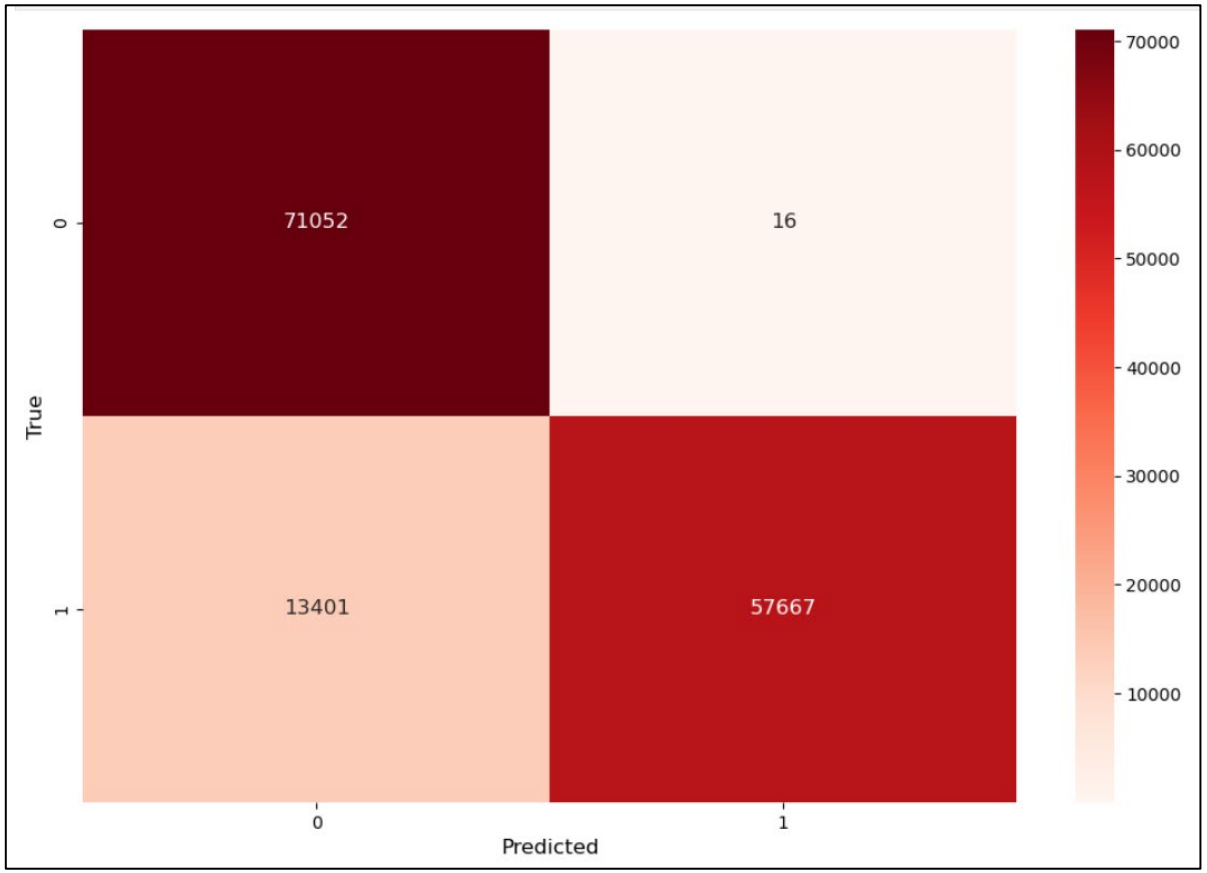


5 – Model predictions



Logistic Regression

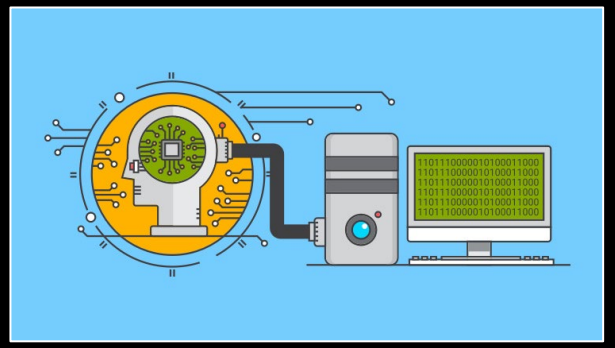
- Correctly predicts Fraud transactions by 98.23% from both classes
- Of all the Fraud predicted transactions, 97.59% were truly Fraud
- Of all the true Fraud predicted transactions, 87.80% were correctly predicted
- F1-Score of 92.43%



Random Forest Classifier

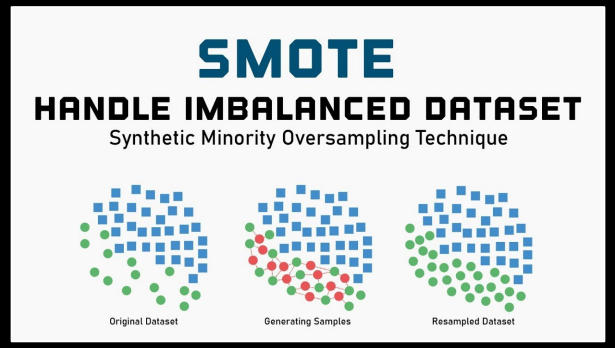
- Correctly predicts Fraud transactions by 97.87% from both classes
- Of all the Fraud predicted transactions, 99.97% were truly Fraud
- Of all the true Fraud predicted transactions, 81.14% were correctly predicted
- F1-Score of 89.58%

6 – Conclusions



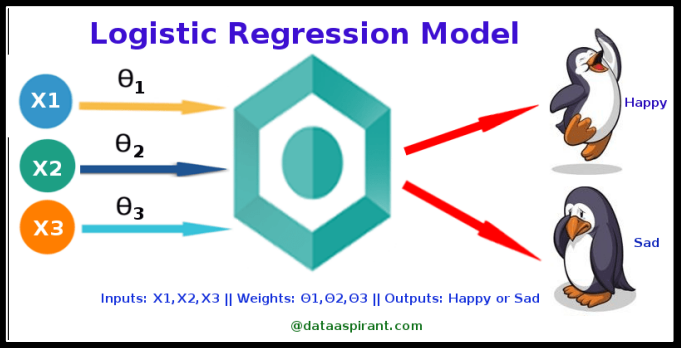
Machine learning algorithms can aid in detecting fraudulent activity in financial transactions at a low price and high efficiency.

For that, historical transactions data must be used for training the model.



Imbalanced datasets can be treated through statistical techniques for a better analysis and prediction.

However, there are some disadvantages that are worth exploring (e.g. it oversamples uniformly)



For this capstone project, the most efficient algoritm was the Logistic Regression Algorithm, although alternatives as Random Forest Classifier can be an alternative path if needed.