



# Capstone Project 02: Credit Card Fraud Detection

Advanced Certification in  
**Applied Data Science,  
Machine Learning & IoT**  
By E&ICT Academy, IIT Guwahati

Capstone Project	Project 02
Title	Credit Card Fraud Detection
Date of Submission	19/12/2023
Status	Final
Produced by	Avinash Kumar [+91 8010476397]
eMail ID	avi3k16@gmail.com
Batch	Batch-04

## TABLE OF CONTENTS

<b>1</b>	<b>PROJECT OVERVIEW .....</b>	<b>3</b>
1.1	Brief Of Credit Card Fraud Detection Dataset.....	Error! Bookmark not defined.
1.2	Objective .....	3
1.3	About Dataset .....	3
<b>2</b>	<b>BUSINESS REQUIREMENT .....</b>	<b>4</b>
2.1	Overview.....	4
2.2	Challenge .....	4
<b>3</b>	<b>DOCUMENT .....</b>	<b>5</b>
3.1	Document purpose .....	5
3.2	Documents Reference .....	5
3.3	Model Used.....	5
<b>4</b>	<b>SOLUTION .....</b>	<b>6</b>
4.1	Resources Used.....	6
4.2	Project Complition Steps .....	6
4.3	Project Github Link.....	7
5.4	Summary .....	7
<b>6</b>	<b>Acknowledgment .....</b>	<b>8</b>

# 1 Project Overview

## 1.1 Brief Credit Card Fraud Detection Dataset

The datasets contains transactions made by credit cards in September 2013 by european cardholders.

This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

## 1.2 Objective

Identify fraudulent credit card transactions By using Machine Learning Models and Deep Learning Models

## 1.3 About Dataset

**Dataset:** - <https://www.kaggle.com/datasets/arslanali4343/credit-card-cheating-detection-cccd>

API used- `kaggle datasets download -d arslanali4343/credit-card-cheating-detection-cccd`

## 2 BUSINESS REQUIREMENT

### 2.1 Overview

Business requirement to build a robust Model using Machine Learning and Deep Learning to observe the behaviour pattern and attributes of customers to identify or classify which is to be a fraudulent customer precisely

### 2.2 Challenge

Challenge I have faced in this project is that data set was imbalanced and features are highly correlated also data set have outliers , so first I have handled these issues .

I have used various machine learning models like Logistic Regression, KNN , SVC, Gaussian Naïve Baise ,Decision Tree , Random Forest , Ada Boost , Bagging , Gradient Boost ,and ANN Model , to find out which one is giving the best so time consumption for training the ML And DL model was high.

### 3 DOCUMENT

#### 3.1 Document purpose

The purpose of this document is to describe the Machine Learning Model and Deep Learning which can be best fit for business requirement and provide solution in the form of highly accurate model for credit card fraud Detection

#### 3.2 Documents Reference

- 1) Project Proposal for Credit card Fraud Detection
- 2) Credit\_Card\_Fraud\_Detection.ipynb

#### 3.3 Models Used

Machine Learning - Logistic Regression, KNN , SVC, Gaussian Naïve Baise ,Decision Tree , Random Forest , Ada Boost , Bagging , Gradient Boost

Deep Learning – ANN Model

## 4 SOLUTION

It is classification problem so Build ML And Deep Learning models for credit card fraud detection, and do the classification of fraudulent and not fraudulent customers

### 4.1 Resource Used

1. Cloud Platform notebook -Google Colab T4 GPU
2. Python Library – NumPy , Pandas , Matplotlib , Seaborn , Scikit- learn , tensorflow, keras

### 4.2 Project Completion Steps

Predicting heart failure mortality using machine learning classification is a valuable application of AI in healthcare. To create a predictive model, you can follow these general steps:

1. **Data Collection:**
  - Using Kaggle data set API , download data on cloud and make the data frame
2. **Data Pre-processing:**
  - Checked the complete data, like datatype , is there any null values or not
  - Handling of outliers
  - Handle imbalance data set using smote oversampling method
  - Checked the correlated features and handle it , calculate the VIF and drop the very high correlated features
  - Also use the dimension reduction technique to use only required no. of feature and handle correlation .
3. **Data Splitting:**
  - Split your dataset into training and testing sets ,in 70-30 ratio , of complete resampled data
4. **Model Selection:**
  - Choose an appropriate classification algorithm. Common choices include logistic regression, decision trees, random forests, support vector machines, and gradient boosting algorithms.
5. **Apply Machine Learning Models**
  - Train the Machine Learning models like - Logistic Regression, KNN, SVC, Gaussian Naïve Baise, Decision Tree , Random Forest , Ada Boost , Bagging , Gradient Boost to check performance of every model on a data set to pick best ML model
6. **Apply Deep Learning Model**
  - Apply Ann model that gives more than 99 % of accuracy in less training time
7. **Model Evaluation**

- Evaluate the model on the training and testing dataset using appropriate evaluation metrics. common metrics include accuracy, precision, recall, F1-score, ROC-AUC, and classification report and confusion matrix.

#### 4.3 Project GitHub Link

- [https://github.com/avi1712/Data\\_Science\\_Project/tree/master/Capstone\\_Project\\_ML\\_DL](https://github.com/avi1712/Data_Science_Project/tree/master/Capstone_Project_ML_DL)

#### 4.4 Summary

I got good performance of ANN model with --

- Training Loss: 0.0155,
- Training Accuracy: 0.9961,
- Testing Loss: 0.0174,
- Testing Accuracy: 0.9958
- F1\_score- 0.9958513982540814
- Precision - 0.9943291871834963
- Recall - 0.9973782771535581



## 5 ACKNOWLEDGMENT

I would like to appreciate and say thanks to E&ICT Academy, IIT Guwahati and IoT Academy for providing me this opportunity to work on Capstone project which gives the feel of Realtime project. And it definitely value added to enhance my knowledge and skills in Machine Learning And Deep Learning.