**Title:** CodTech IT Solutions internship – Task Documentation: “Credit Card Fraud Detection” using Machine Learning models.

**Intern Information:**

**Name:** Triveni Dasari

**ID:** ICOD6425

**Introduction:**

Every day, **billions** of credit card transactions are made all over the world. Considering the widespread use of smartphones and the Internet throughout the earth, more and more people are using their credit cards to make purchases online, making payments through apps, etc...  
  
In a scenario such as this one, it is **extremely** important that credit card companies are able to easily recognize when a transaction is a result of a fraud or a genuine purchase, avoiding that customers end up being charged for items they did not acquire.  
  
In this project, I used the **scikit-learn** library to develop a prediction model that is able to learn and detect when a transaction is a fraud or a genuine purchase. I tested four different classification algorithms, **Decision Tree**, **Random Forest**, **Ada Boost Classifier** and **Gradient Boosting** to identify which one of them would achieve the best results with our dataset.

**Implementation:**

In order to implement a predictive model, I used **scikit-learn** library and tested four different classification algorithms to see which one of them would achieve higher accuracy metrics for the dataset.  
  
Pandas, numpy, matplotlib, seaborn and plotly libs were also used to **explore**, **handle** and **visualize** relevant data for this project.  
  
The dataset used for this project was the [Credit Card Fraud Detection](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud) dataset posted on Kaggle, which contains credit card transactions that happened during the month of September, 2013 by european clients for two days.  
  
The dataset has the feature **time**, which shows us the seconds elapsed between each transaction and the first transaction in the dataset. The feature **amount**, containing the transaction amount and the feature **class**, which tells us if that certain transaction is **genuine or a fraud, where 1 = fraud and 0 = genuine**.  
  
Features V1, V2,... V28 are numerical input variables result of a [PCA transformation](https://en.wikipedia.org/wiki/Principal_component_analysis)whose content couldn't be displayed due to their **confidential** nature.

During the development of this project, I've used certain tools from the **sklearn** library that would help me in achieving higher performance metrics for the models, such as the **StandardScaler**, used to alter the scale of **amount** variable and **SMOTE**, from **imblearn** library, used to deal with data imbalance. Both tools were used to avoid creating a model that would be biased towards a determined variable or a determined class.

**Code Explanation:**

* In first step import necessary modules.
* Load the csv file.
* Apply preprocessing and normalize the transaction data, and handle imbalance issues.
* Split the dataset into training and testing sets.
* Train a Classification algorithm by using random forests, to classify transactions as fraudulent or genuine.
* Evaluate performance using metric like precision, recall, and F1 score, and consider techniques like over sampling or under sampling for improving results.

**Usage:**

Credit card fraud detection involves various tools to verify transactions, techniques to authenticate cardholders and technogies, such as statistical analysis and amchine learning, to monitor suspicious transactions and activities.

**Conclusion:**

When we work with a **machine learning model**, we must always **know** for a fact **what it is that we're trying to get from that model**.

In this project, our goal is to **detect fraudulent transactions.** For thatthe random forest classifier has the best performance with an accuracy of 99.94%, precision of 90.9%, recall of 74.07%, and F1 score of 84.6%.