Name: Sai Vamshi Atukuri
Question/Problem: Detecting credit card transaction frauds.

Problem Statement:

Let's assume a hypothetical bank. The bank identifies any transaction greater than 2000 as a possible fraud and take a confirmation from customer before approving it. Due to this strategy, the bank is facing customer service issues and high incurred losses. So, the bank wants to build a new strategy based out of machine learning now.

Data:

The dataset contains transactions made by European customers with fraud transactions flagged. ~285K transactions are spanned across two days where only ~500 are fraud transactions. Most of the variables are masked using PCA transformations due to confidentiality issues. The dataset is found at https://www.kaggle.com/mlg-ulb/creditcardfraud.

Solution:

The most popular method to detect the frauds is to use supervised learning to classify them. Another issue to solve is the imbalanced dataset (only 0.17% event rate). Hence different classification algorithms and sampling techniques are explored and compared. Different metrics like precision, recall, accuracy, F1 score, AUC, losses saved, and wrong calls sent are calculated.

Model with a recall of precision and recall greater than 90% are built which shows that all the frauds and non-frauds are correctly captured solving the issue posed by the bank.