Supervised ML Models are models that rely on the data labels to train algorithms that will then classify data or predict the outcomes accurately. In this example, we have a dataset of strawberry and apples which when put through a machine learning model, it will predict and classify the data into either apples or strawberries.  
  
The first supervised model we will talk about is logistic regression. We chose this model because this is what is usually used in making categorical predictions, like yes/no, binary or true/false. In our case we are traying to predict which transactions are non-fraudulent (0) or fraudulent (1). When we run the model, we can see that it gives a precision of 82%, a recall of 52% and an f1-score of 64%, which is not the greatest result.

Therefore, we used another prediction classifier, k nearest neighbors. As you can see from the graph, when we separate the data into test (yellow) and training(blue) accuracy, it stablilizes at around 3 neighbors. We can also observe a slight improvement from logistic regression,in the accuracy scores, with a precision of 88%, recall of 74% and f1-score of 81%.

We then tried to do an ROC probability corve of plottin TP rate vs FP rate. We can observe that the highest TP rate is approx. 0.85 and it starts at approx. 0.84, which is not a bad yield.

Because the dataset is imbalanced, we used a prediction classifier that is commonly used in imbalanced datasets, Support Vector Machine (SVM). We used both with and without classweight to measure as accurate as possible. Although the result difference was negligible, the model performed slightly better without class weight, with a precision of 81%, recall of 78% and f1-score of 80%.