Python Project 1: Funny Money

-by Aniket N Prabhu, ASU ID: 1217704488

1) <u>Data Analysis:</u>

	Variance	Skewness	Curtosis	Entropy	Class
Variance	1.000000	0.264026	-0.380850	0.276817	-0.724843
Skewness	0.264026	1.000000	-0.786895	-0.526321	-0.444688
Curtosis	-0.380850	-0.786895	1.000000	0.318841	0.155883
Entropy	0.276817	-0.526321	0.318841	1.000000	-0.023424
Class	-0.724843	-0.444688	0.155883	-0.023424	1.000000

Correlation tells us how much a variable is dependent on another variable. From the above cross covariance matrix, we can say that Variance is the most correlated to class, Skewness is the most correlated to Curtosis, and Entropy is the most dependent on Skewness. After Variance, Skewness is the most dependent on Class. Hence, according to this analysis, Variance and Skewness are the best predictors of genuine money.

2) <u>Machine Training:</u>

Classification	Accuracy	Combined Accuracy
Perceptron	0.97	0.97
Logistic Regression	0.99	0.99
Support Vector Machine (Linear)	0.99	0.99
Decision Tree Learning	0.99	0.99
Random Forest	0.99	1.00
K-Nearest Neighbor (k=10)	1.00	1.00

From this table, it is clear that K-Nearest Neighbor (KNN) is the best machine learning method out of them for predicting counterfeit bills as it has the highest accuracy as well as the combined accuracy (both 100%). 'Linear' kernel type was chosen for the Support Vector Machine as it gave the best accuracy amongst other kernel types. k=10 was chosen for KNN as values before 10 either gave more than 1 misclassified combined sample (MCS) or would have led to underfitting. The other option for having just 1 MCS were 11 and 12. Above 12 too, KNN gave more than 1 MCS.