ML Coursework strategy outline

Data Preparation

* OneHotKey vs Label Encoding for categorical variables

Decision Tree methods

* Apply SKLEARN decision tree classifier
* Hyperparameter optimization using randomized search vs grid search techniques
* Cross validate & report accuracy

Random Forest methods

* Apply SKLEARN random forest classifier
* Hyperparameter optimization using randomized search vs grid search techniques
* Cross validate & report accuracy and compare with DT analysis

Sources

*LabelEncoder can turn [dog,cat,dog,mouse,cat] into [1,2,1,3,2], but then the imposed ordinality means that the average of dog and mouse is cat. Still there are algorithms like decision trees and random forests that can work with categorical variables just fine and LabelEncoder can be used to store values using less disk space.*

[*https://datascience.stackexchange.com/questions/9443/when-to-use-one-hot-encoding-vs-labelencoder-vs-dictvectorizor*](https://datascience.stackexchange.com/questions/9443/when-to-use-one-hot-encoding-vs-labelencoder-vs-dictvectorizor)