ML Questions

Intro Questions:

- What is the difference between supervised & unsupervised learning?
- How do you ensure your model is not overfitting?
- How do you ensure your model is not underfitting?
- Why do we split the data into training and testing sets?
- How much data should you be adding to your training set?
- What are the steps to building a model?
- Using the fake dataset called "data" simulate how you would read the data.
- Using the fake dataset called "data" simulate how you would split the test and train data.
- Using the fake dataset called "data" simulate how you would define a decision model.
- Using a fake dataset called "data" simulate how you would define a random forest model.
- Using a fake dataset called "data" simulate how you would fit your model.
- Using a fake dataset called "data" simulate how you would predict your results.
- Using a fake dataset called "data" simulate how you would see the mean absolute error of your model.

Intermediate Questions:

- What are the ways you can handle missing values?
- What are the ways you can handle categorical values?
- What are the benefits of pipelines?
- What is cross-validation and when do you use it?
- What are the main types of data leakage?

- What are the steps for the gradient boosting method?
- What are the steps for the K-Nearest Neighbor algorithm?
- How is the Support Vector Machine Algorithm created?
- Using the fake dataset called "data" show the steps for dropping missing values
- Using the fake dataset called "data" show the steps for using imputation for missing values
- Using the fake dataset called "data" show the steps for dropping categorical values
- Using the fake dataset called "data" show the steps for handling categorical values using label encoding
- Using the fake dataset called "data" show the steps for handling categorical values using one-hot encoding
- Using the fake dataset called "data" show the steps for creating a pipeline
- Using a fake dataset called "data" show the code for cross-validation
- Using a fake dataset called "data" simulate how you would define a gradient boost model
- Using a fake dataset called "data" simulate how to create a K-NN model
- Using a fake dataset called "data" simulate how to create a SVM model