UNIT 5 ASSIGNMENT

Choosing Your Model

## Instructions

The questions below will prepare you for future interviews as they relate to concepts discussed throughout the week. You’ve practiced these concepts in the coding activities and the exercises   
as well as the coding portion of the assignment. Now let’s formulate your programming into well-reasoned responses.

Except as indicated, use this document to record all your assignment work and responses to any questions. At a minimum, you will need to turn in a digital copy of this document to your facilitator   
as part of your assignment completion. You may also have additional supporting documents that   
you will need to submit. Your facilitator will provide feedback to help you work through your findings.

**Note:** Though your work will only be seen by those grading the course and will not be used or   
shared outside the course, you should take care to obscure any information you feel might be of   
a sensitive or confidential nature.

*Begin your assignment by completing the questions below. Directions to submit your work can be found on the Unit 5 Assignment page online. Information about the grading rubric is available on any of the unit assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the assignment.*

Week 5 Written Portion

# Choosing Your Model

Answer the questions below about selecting the correct models and approaches to solve your machine learning problems.

## Questions:

1. What is model selection and why is performing model selection important?

|  |
| --- |
| Model selection is the process of selecting the appropriate model for a given machine learning problem. Selecting the right model is very important for the best results possible |

1. What is out-of-sample validation and why is this key in helping us choose the best-performing model?

|  |
| --- |
| Out of sample validation is the idea of testing a model on new unseen data. It’s important in helping us choose the best-performing model because the results indicate which model works better with our data |

1. What is cross-validation and what is the benefit of performing cross-validation?

|  |
| --- |
| Cross validation is the process of making sure our model is being trained correctly and in the best way on our training data. It does this by ‘punishing’ the model if a feature hurts the performance of the model. This will prevent the model from overfitting, and it’s like an extra guide for the model to ensure it’s moving in the right direction |

1. What is the difference between feature engineering and feature selection? What are the benefits of feature selection?

|  |
| --- |
| Feature engineering is the process of selecting and transforming raw data into features that can be used for modeling. Feature selection is choosing the best features that would work with our model. Benifits of feature selection: it can help us reduce the complexity of the model, and it helps us select the features that really describe the relationships within our dataset |

1. What are the differences among the classification evaluation metrics accuracy, precision, and recall?

|  |
| --- |
| Accuracy: what percentage of predictions our model got correct  Precision: refers to how well a model can predict key outcomes.  Recall: refers to how many key outcomes were predicted |

*To submit this assignment, please refer to the instructions in the course*.