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 when the user selects a model in order to maximize performance and results. Selecting a model is important as it allows an ideal to be used with the highest performance, prevents overfitting, and increases efficiency. |

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 when you evaluate a model’s performance on data that is independent of the data used to train the model. It is key to choosing the best-performing model because it provides an unbiased assessment of the model’s performance on unseen data. |

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

|  |
| --- |
| Cross-validation is when you use split available data into subsets which is then used as both a training set and a validation set. The purpose of doing cross-validation is that is creates more reliable estimates of a model’s performance compared to a single train-test split. |

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

|  |
| --- |
| Feature engineering is when you create new features or transform existing features to improve the representation of the data. Feature selection is when you select a subset of relevant features to improve the model’s performance and efficiency. The benefits of feature selection are improved model performance, faster training through removing redundant data, easier interpretability, reducing the risk of overfitting, and also focusing on only the most informative features. |

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

|  |
| --- |
| Accuracy measures how correct the overall model’s predictions are by comparing the number of correct predictions to the total number of predictions. Precision focuses on the proportion of correctly predicted possible samples in comparison to the total number of samples predicted to be positive. This means it compares the true and false positives. Recall measures the proportion of correctly predicted positive samples out of all actual positive samples. This means it is comparing the true positives to the true positives and false negatives. The difference between this and precision is the emphasis on the avoidance of false negative errors rather than avoiding false positive errors. |

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