UNIT 8 ASSIGNMENT

Prepare ML Models for the Real World

## Instructions

The questions below relate to the machine learning problem you would like to solve and the data set that you have chosen in the coding assignment. You will explain the type of problem you plan on solving and explain your project plan.

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 assignment page. Information about the grading rubric is available on any of the course assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the assignment.*

Unit 8 Written Portion

# ML Problem Formulation

Answer the questions below about your machine learning problem:

## Questions:

1. List the data set you have chosen.

|  |
| --- |
| A data set on Netflix subscriptions. It offers various of information such as customer information, date of subscription, cost rate, types of movies they watch, etc. |

1. What will you be predicting?

|  |
| --- |
| I’ll be predicting whether or not each customer will keep or discard their Netflix subscriptions. That’ll be our label. |

1. Is this classification or regression problem? If this is a classification problem, is there class imbalance?

|  |
| --- |
| This is a classification problem because we’re trying to find out if the customer will leave their subscription. A yes/no type of prediction. Let’s assume that there’s no class imbalance. |

1. Explain why this is an important problem. In other words, how would a company create value with a model that predicts this label?

|  |
| --- |
| This is an important problem because the company can learn a lot from people who’ve left the subscription by finding the cause. And as a result, knowing the cause allows them to think about what their plans are for the future, their cost projection, movies to purchase to appeal to current and future users. With more subscription, the more success the company will have. |

1. Create a project plan.

* Describe the features that you will choose.
* Choose a model that you will train.
* Explain how you will prepare your data for your model.
* Specify an evaluation metric that is appropriate for your model.
* In your plan, describe your plan to train your model, analyze its performance and then improve the model. That is, describe your model building, validation and selection plan to produce a model that generalizes well to new data.

|  |
| --- |
| * The label I would choose is a column that indicates if the subscription was canceled. The feature I would choose is price. * The model I would use to train would be a logistic regression because we typically use them when the classification problem is binary—which in our case we’re trying to indicate yes/no or true/false. * I would prepare my data by first assigning the features and labels to variables. Then filtering out any NaN values, split them into training and test sets, use the logistic regression model, fit the model, create a prediction variable. I would feed some sample hyperparameter values to see the hyperparameters that allows better performance of the model. Then with the best hyperparameters, I would feed those values into the model, fit it, create prediction variable. * The evaluation metric I would use is accuracy. Since it’s more ideal for classification problems with no class imbalance. * I would then plot the accuracy of the model. Analyze its performance. We can try and improve the model by performing feature selection to see which feature would yield better performance, the number of features selected, etc., see if there’s better hyperparameters. |

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