UNIT 1 ASSIGNMENT

ML in a Nutshell

## Instructions

Many of the apps and websites you use on a daily basis are examples of applications of machine learning. There are three parts to this assignment where you will analyze an example of your choice.

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

*Complete each assignment part as you progress through the course. Wait to submit the assignment until all parts are complete. Begin your course assignment by completing Part One below. Directions to submit your assignment can be found on the final part of the assignment page at the end of Module 1.3: The ML Lifecycle. 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.*

Part One

# Using ML for Industrial Decision Making

In this part of the assignment, you will identify a real-life company and a product, feature, or application that is driven by a supervised machine learning method. Answer the following questions based on that real-life example.

## Questions:

1. What is your chosen machine learning example?

|  |
| --- |
| Deciding whether to recommend a new movie to a user on Netflix. |

1. State the business objective of the underlying machine learning algorithm.

|  |
| --- |
| Come up with a recommendation system that decides whether to recommend a new movie to a user based on their watch history. |

1. What is the label and what are three features that might be used to predict the label?

|  |
| --- |
| Label: recommend or don’t recommend; three features: previous shows watched, date of shows watched, ratings of previous shows watched |

1. Finally, explain why you think machine learning is the right approach to achieve the underlying objective. (To help your thought process, think about what the alternative, non-ML solution could have been. Note also that sometimes it may be the case that the use of ML by the company is not well motivated.)

|  |
| --- |
| I think machine learning is the right approach here because in a non-ML world, users would not be able to get personalized recommendations. Oftentimes, a user’s watch history is indicative of what they like or don’t like. We can use the watch histories of a certain group of users with similar watch histories to predict the preferences of an individual who falls within that group since they’ll have similar watch histories. |

Part Two

# Recognizing ML Problem Types

In this part of the assignment, you will take your example from the previous part and will further analyze its problem type, classification or regression.

## Questions:

1. What type of problem do you think it represents? Explain why you think your problem is classification or regression given the concepts you explored in this module.

|  |
| --- |
| Binary classification because the recommender system has two choices for the output: recommend or don’t recommend the new movie. |

1. Give another example of a classification or regression problem that you interact with in your daily life, or one that companies or governments might use.

|  |
| --- |
| Predicting the temperature |

Part Three

# The ML Lifecycle

Imagine that you are working for a telecom company. The management of the company is looking to address the problem of customer churn\*. Your task is to predict which customers are likely to churn.   
In your own words, describe the steps that you would take to address this problem. Focus in particular on the following questions:

* Why is it useful to predict the customers that will churn in the future? How can such knowledge serve the business objectives?
* How would you further formalize the problem? Define, in your own words, what inputs would be useful for your model, and how you would define the target quantity or measure that you would try to predict.
* What kind of methods (supervised or unsupervised) would be appropriate to use? Why?
* What kind of data would you ideally use, and what kind of data do you expect to be available?

|  |
| --- |
| 1. Business understanding: It is useful to predict the customers that will churn in the future because we can use this information to see which groups we should target to try to convince them to become our new customers. This will help the company generate more revenue.  2. Data understanding:  a. Input: various customers’ information (how long they’ve been customers of our telecom company, how much/often they text, how much/often they call, how much mobile data do they use per month, how much money they make, etc.) and whether they end up churning  b. Target: whether a particular customer will churn  3. Modeling: supervised machine learning is appropriate here since we need to have training data with the labels in order for our machine learning model to learn from those data points  4. Ideally, I would like to know the start and end dates of each customer’s use of the telecom company’s technology, how much/often they text/call/use mobile data, how much money they make, how many users do they have in their family/group, their gender, their race, their ethnicity, and what companies they choose instead. I’d expect to know all of the above except how much money they make and personal information like gender, race, and ethnicity. |

*\*Customer churn is the loss of customers or clients and happens when customers decide to stop doing business with a company.*

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