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?

|  |
| --- |
| Determining if the credit card transaction is flagged as fraud or nonfraudulent. |

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

|  |
| --- |
| To ensure that customers can safely use their credit cards and that there’s no identity theft and to ensure that customers’ accounts will remain safe. |

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

|  |
| --- |
| Label: if it’s fraud or not.  Features: transaction amount, transaction distance, monthly total number of transactions |

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.)

|  |
| --- |
| Machine learning is the right approach because we will be analyzing large amount of historical data and to learn from them. Utilizing this method will help determine if future data (transaction) is fraud or not. You cannot hard code to determine if something is fraud. |

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.

|  |
| --- |
| I believe this type of problem is a classification problem. Specifically binary classification because the problem itself requires a yes/no type of prediction. |

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.

|  |
| --- |
| Another example of a classification problem in which I’ve interacted with in my daily life is whether or not an email is considered spam or not. I’ve seen several of my emails filtered into the spam portion of the Gmail, and I barely see any spam in my inbox. |

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?

|  |
| --- |
| It’s useful to predict the customers that will churn in the future because the company’s objective is generate revenue and this is interconnected with the customers. Their goal is to gain more customers. A customer churn is exactly the opposite of what they need and it’s detrimental to their company. If we can analyze data and predict the similarities that will make customers churn, the company can then target those similarities and improve. This type of knowledge will be helpful when formulating our business objectives because it’ll allow us to understand what does the company needs, their goals, and overall to assess the situation.  The features I would collect data on is if the customers are satisfied with customer service, and different specific components of what the telecom company does. The examples will be in the form of yes/no. Then there will be a label that indicates if they churn or not, and it’ll be in the form of yes/no.  A supervised learning would be appropriate to use because the data I’m collecting requires labeled input and output data. But specifically, I’m collecting a classification learning algorithm because the prediction is in the form of a yes/no—if they’ll churn or not.  The data I’ll ideally use is different specific services the telecom company provides to every customer and if they’re satisfied with the cost rates as well. To collect these opinions from customers we can have them fill out surveys. |

*\*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*.