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?

|  |
| --- |
| Bank of America fraud detectors |

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

|  |
| --- |
| To detect fraudulent business related to one’s bank account and alert the bank and account owner |

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

|  |
| --- |
| A label is what you can infer from a data point (like if it’s fraud). 3 features that can be used to predict this label are: credit card charge, debit balance, and location of transaction. |

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.)

|  |
| --- |
| ML is the right approach to detect fraud in banking because a pre-coded algorithm is not likely to flag fraud that happens to align with past data. ML is how we achieve generalization which allows for answers to situations when we encounter new, never-before-seen data. |

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.

|  |
| --- |
| This problem is a binary classification as there is one label (is this fraud?) in which we should be able to answer “yes” or “no”. |

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.

|  |
| --- |
| An example of a regression problem are the weather forecasts. There are infinite possible label values which are associated with real numbers such as 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?

|  |
| --- |
| It is useful to predict the customers that will churn in the future because such a trend can majorly affect company growth and the very existence of a company. Companies rely on customers and the number of customers needs to grow.  Useful inputs for the model would be to have a profile of customers, with information about their income, age, and level of satisfaction about the telecom company. Using these inputs, I want to predict if a customer is likely to churn.  An unsupervised method would be appropriate because we are not sure what is causing customer churn so we need to identify a pattern in the data. Knowing a pattern that is associated with customer churn helps us address the problem.  Ideally, we want data in the form of customer profiles in order to find a pattern that may explain why customer churn in happening at the company. Realistically, the company may not have this condensed information on every one of their customers so we would have tons of data on each customer to create our own profiles and find patterns. |

*\*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*.