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?

|  |
| --- |
| Predict house price |

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

|  |
| --- |
| Use historical data of housing price to predict prices of house in six months in the chosen area |

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

|  |
| --- |
| Label is house price, three features are location, number of bedrooms, size of the house |

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.)

|  |
| --- |
| Because by using machine learning, it will archive accurate and thoughtful house price predictions. |

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.

|  |
| --- |
| Regression because it will predict the housing prices in six months |

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.

|  |
| --- |
| Spam emails, credit scoring, bank fraud detection |

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?

|  |
| --- |
| * Predicting which customers may leave is helpful because it allows the company to retain those customers proactively. By identifying potential traffic early, the company can implement targeted retention strategies such as personalized coupons, discounts, and improved customer support to drive customer retention. This knowledge can help the company reduce customer churn, increase customer satisfaction, and ultimately increase revenue by retaining valuable customers. * To solve the problem more deeply, relevant input data is collected to train a predictive model. Valuable inputs include: * Customer demographics (age, gender, location). * Usage patterns (call minutes, data consumption). * Billing information (monthly billing, payment history). * Customer support interactions (complaints, inquiries). * Product/Service Functions.   The target number or metric we want to predict is the probability of a customer leaving within a certain period, such as next month. This binary card indicates whether the customer may leave or not.   * In this case, supervised learning methods would be appropriate. Our historical data contains information about customers who churned (positive points) and customers who did not (negative points). Using this labeled data, we can train the machine learning models such as logistic regression, decision trees, and random forests and learn patterns and relationships between input features and target labels. Later, based on feature values, we can predict new customer data we have yet to see before and classify customers as churners or non-churners. * Ideally, we should provide multiple data sources that comprehensively understand customer behavior and interactions between telecom companies. This may include customer records, call logs, billing records, customer support tickets, usage patterns, service/product usage history, customer feedback, and data social network. However, the availability of data in practice may vary. Telecom companies can start with their data, such as basic customer information, billing records, and usage data. As predictive models mature, companies can consider incorporating other data sources to improve the model's accuracy and predictability. |

*\*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*.