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?

|  |
| --- |
| Calculate estimated insurance premium in auto insurance company. |

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

|  |
| --- |
| Try to reduce the risk assessment and provide personalized pricing |

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

|  |
| --- |
| Features: gender, age, accident history. Label: insurance cost |

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.)

|  |
| --- |
| It leverages large datasets to accurately predict individual risk profiles, leading to personalized and fair insurance premiums. Traditional methods rely on broad categories and heuristics, lacking the precision. |

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.

|  |
| --- |
| It is a regression problem because we need to calculate the exact amount of premiums. |

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.

|  |
| --- |
| Companies might use binary classification to decide promoting a candidate or not. They might use performance rate and colleagues’ evaluation as features. |

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?

|  |
| --- |
| The company can analyze the reason why their customers are leaving and improve their product. Input could be the plan type, price, loyalty, data usage, etc. Target quantity would be binary variable indicating whether a customer would churn or not. Supervised learning is appropriate because we can define a clear target variable and have historical data to train the model. Ideal data is customer demographic profile, full history from previous telecom company, usage metrics, and detailed billings. Expected available data would be basic customer information, usage summaries, service interaction logs, and billing records. |

*\*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*.