Week 8: Daily Morning Challenge

Day 1: Tuesday 3rd March 2020

Question 1: What is a decision tree and briefly explain how it works

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. A decision tree can be used to visually and explicitly represent decisions and decision making. A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data). In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

Question 2: Outline two algorithms that can be used for decision tree classification

C4.5 → (successor of ID3)
CART → (Classification And Regression Tree)

Question 3: What is the role of API development as part of a data scientist skillset

API is an acronym for Application Programming Interface that software uses to access data, server software or other applications and have been around for quite some time. An API in data science works just like any standard API by creating an abstraction layer for data scientist to integrate machine learning into the day-today applications that they develop.