Data mining assignment 1

You’re encouraged to work in a group. However, please make sure that if you work in a group, it is no more than two people. There are only 17 students in this class that let’s try not to cluster together. Please also indicate your teammate name on the first line of your solution as a comment. A group only needs to submit once. If it is a coding question(s), the solution needs to be saved in a real Python executable file, a file whose name ends with “.py”. Please do not submit a PDF file for a coding question unless noted otherwise.

You solution will be graded on a curve if you work alone. This is a very simple assignment. Please try to do it by yourself without resorting to all the AI that’s available on the internet. You are not hired because you can use ChatGPT. An account or a football coach can do that just as well. The due date is 06/18 (Sunday).

This assignment has two parts. The first part consists of two very simple exercises. The second part is more of analysis question type. Please submit a short report to detail your thought process for the second part. There is a piece of code named ‘assigment1\_decision\_tree.ipynb’ attached to this file. You need to read through it to finish the second part. You need to write a few lines of code. But since this is a report type of questions, you may simply include the code in the report.

Good luck!

Part I (60 points)

Q1. (30 points)

The following sequence is formed using words and numbers:

(1) The first number is 1

(2) In the first number, there was one 1 so the second number in the sequence is

11

(3) In the second number, there were two 1s so the third number in the sequence

is 21

(4) In the previous number, there was one 2 and there was one q, so the fourth

number is 1211

(5) In the previous number, there was one 1, one 2, and two 1s, so the fifth

number is 111221...

This sequence can continue infinitely.

**Write a function (a static method if you will) that, given a staring number m and an integer n, returns the next n numbers in order.**

So, for example, given a starting number of 1 and an integer 4, the sequence

returned should be [11, 21, 1211, 111221]

Or, given a starting number of 11 and an integer 3, the sequence returned should

be [21, 1211, 111221].

Q2. (30 points)

**Write a function that does this**:

Given any two lists A and B, determine if:

List A is equal to list B; or

List A contains list B (A is a superlist of B); or

List A is contained by list B (A is a sublist of B); or

None of the above is true, thus lists A and B are unequal

Specifically, list A is equal to list B if both lists have the same values in the same

order. List A is a superlist of B if A contains a sub-sequence of values equal to B.

List A is a sublist of B if B contains a sub-sequence of values equal to A.

Examples:

If A = [] and B = [] (both lists are empty), then A and B are equal

If A = [1, 2, 3] and B = [], then A is a superlist of B

If A = [] and B = [1, 2, 3], then A is a sublist of B

If A = [1, 2, 3] and B = [1, 2, 3, 4, 5], then A is a sublist of B

If A = [3, 4, 5] and B = [1, 2, 3, 4, 5], then A is a sublist of B

If A = [3, 4] and B = [1, 2, 3, 4, 5], then A is a sublist of B

If A = [1, 2, 3] and B = [1, 2, 3], then A and B are equal

If A = [1, 2, 3, 4, 5] and B = [2, 3, 4], then A is a superlist of B

If A = [1, 2, 4] and B = [1, 2, 3, 4, 5], then A and B are unequal

If A = [1, 2, 3] and B = [1, 3, 2], then A and B are unequal

Part II (40 points)

Please read the code inside the file named ‘assignment1\_decision\_tree.ipynb’. To spare you from the trouble of configuring some packages dependencies, **you are advised to use google colab instead of a copy of Python on your computer**.

Please answer all of the four questions which are asked in the form of ‘comments’ inside the file and include all the answers in a report such as a PDF file.