MTU
Ollscoil Teicneolaíochta na Mumhan
Munster Technological University

# MACHINE LEARNING.

Lab 01: Python Environment

## BACKGROUND.

The goal of this lab exercise is to install the Anaconda Python 3.8 environment (https://www.anaconda.com/products/individual) and to implement some simple functions to make sure you are setup for the lab sessions to come.

### Task 1.

The Fibonacci numbers are the numbers in the following integer sequence:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...

By definition, the first two numbers in the Fibonacci sequence are 0 and 1, and each subsequent number is the sum of the previous two, i.e. $f_i = f_{i-1} + f_{i-2}$.

Create a program that creates a list and will populate it with the first 40 Fibonacci numbers.

The program should then ask the user to enter an integer value between 1 and 40 to indicate which number in the Fibonacci series they would like to see and the application should display that number. For example, if the user enters 13, the 13th number is 144.

### Task 2.

Write a program that reads the contents of the Bram Stoker's Dracula. The eBook can be downloaded at http://www.gutenberg.org/cache/epub/345/pg345.txt.

The objective of this program is to read all the data from this file and output words that occur at a specific frequency within the text.

Your program should read all words from the file. It should record all words that have a character length of at least minWordLength along with their frequency of occurrence in the novel.

Your program should then print each word along with the frequency of the word, which occurs more often in the novel than minWordOccurence.

The result of minWordLength=5 and minWordOccurence=300 should be:

- “which” -> 636
- “could” -> 458
- “would” -> 408
- “there” -> 508
- “shall” -> 410

**Task 3.**

In this question you will perform a basic analysis of a bike sharing dataset available at

https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset.

Find the CSV file “day.csv” and create a program that goes through each line and reads the weather condition indicator and number of bike rentals recorded. Note, that the first line does not contain a valid data record.

It should then output the number of days in the dataset with clear weather conditions and the average number of bike rentals on clear days.

(The result should be 463 days with an average of 4877 bike rentals per day)