7CS030: **Coursework Brief**

(*See Canvas for Submission Details*)

This assessment is designed to test your ability to use various concepts & technologies in AI. You are asked to write Python programs to complete the tasks outlined, you should **also** prepare a document (MS Word or LaTeX) which contains all the information associated with the tasks. Each task is associated with a data set so you should think of yourself as an AI/Data Science specialist being tasked to build a machine learning system for a client who would like to make good decisions based upon the data. Hence, it is not about wowing the client with your programming skills but about providing a clear description of the software developed: including design cycle, analyse of the data set and effectiveness of the model (accuracy scores etc.).

For each of the **tasks 1-3** you will be marked based on the following criteria:

- How the model was constructed:
  - description of the algorithms used and how they were implemented in python,
  - clarity of the program(s), are there appropriate comments, is it easy to use?
  - demonstrate how you can start with simple models and work up to more complex and better performing models,
  - analyse of performance and comparison. **(60%)**
- Any possible visualisation that could be used to help the client understand the model better. (Remember the client is not necessarily an AI/Data Science expert!) **(20%)**
- Analyse of the data, can you make any recommendations to the client? **(20%)**

Make sure you clearly label the tasks and refer to the associated .py file as you explain your work in the document. **Each task should be no more than a page each.** (You have not been commissioned for a year by the client!) You should also make comments within your programs so it is clear how it works to anyone using it. (**Do not go overboard with the comments though!**)

When submitting to canvas make sure you submit all your programs (.py) as well as the document produced (.pdf).

- **Submission Date:** See Canvas Assignment Task.
- **All work which you submit must be your own. Cheating is a serious offense and will be dealt with under the University disciplinary procedures.**

The module team are here to help in anyway we can, so please ask for help if you are having difficulty. Good luck and enjoy as these are really interesting problems!

Task 1: Regression

In this task you are required to apply a machine learning algorithm to the data set houseprice_data.csv which can be downloaded from the assignment task on canvas. This data set contains information about house sales in King County, USA. The data has 18 features, such as: number of bedrooms, bathrooms, floors etc., and a target variable: house price.

Using **linear regression** (simple or multiple), develop a model to predict the price of a house. After developing the model you should also analyse the results and discuss the effectiveness of the model, outlining the improvements when developing the model.

Ideas to consider when completing this task:

- Is there a way of visualising your model? (Possibly just one or two input/feature variable(s).)
- How will you assess the effectiveness of the model?
- Include as many features as you can. Does the model improve?
- How could you make further improvements?
- What can you conclude about your model?

[30 Marks]

Task 2: Clustering

In this task you are required to apply a machine learning algorithm to the data set country_data.csv which can be downloaded from the assignment task on canvas. This data set contains information about a countries child mortality, exports, health spending, etc.

Use **clustering** to investigate this data set. After clustering the data you should analyse the results and discuss what can be concluded by the clusters.

Ideas to consider when completing this task:

- Is there a way of visualising the clusters?
- Can you make any conclusions about the clustering?
- Include as many features as you can. Does the clustering change?
- What advice would you give, in the context of the data, based on the clustering?

[20 Marks]

Task 3: Classification & Neural Networks

In this task you are required to apply a variety of machine learning algorithms to the data set `nba_rookie_data.csv` which can be downloaded from the assignment task on canvas. This data set contains NBA rookie performance with target variable `Target_5Yrs` with 1: if career length >= 5 yrs or 0: if career length < 5 yrs.

The classification problem here is to predict if a player will last 5 years in the NBA. Apply **Logistic Regression**, **Gaussian Naive Bayes** and construct **Neural Networks**. After developing the various models you should also analyse the results and discuss the effectiveness of the models, outlining the improvements when developing the models and compare the approaches/algorithms used (strengths and weaknesses).

Ideas to consider when completing this task:

- Apply various algorithms to the problem. *Caution*: Use a small number rather than many, analyse in depth rather than being superficial and repetitive.
- Is there a way of visualising the model(s)?
- How will you assess the effectiveness of the model(s)?
- Include as many features as you can. Does the model improve?
- Compare the models produced.
- How could you make further improvements?
- What can you conclude about your model?
- How strong is the relationship between the predictor and target variables?

[40 Marks]

Task 4: Ethics of AI

In this task you are required to write a short essay (< 500 words) to discuss the following ethical dilemma in AI:

The Trolley Problem is a well known problem in ethics, discuss the trolley problem in the context of autonomous vehicles.

Marks will be awarded based on how well you meet the three criteria:

- Understanding - In-depth, authoritative, full understanding of key issues with evidence of originality. (4 marks)
- Depth of knowledge - Key issues analysed, selective source(s) used to support argument/discussion. (3 marks)
- Structure - Coherent and compelling work logically presented. (3 marks)

[10 Marks]