Group Project (40%)

Machine Learning for Artificial Intelligence

Diploma in Artificial Intelligence and Working into the Future

Release Date: 5th February 2021

Submission Date: 14th March 2021

This is a group-based project (maximum: 3 – 4 students) using PYTHON programming language to analyse a specific problem in the following areas, such as Pharmacy, Library, Holiday Booking System, Medical Practice, Concert hall, Motor mechanic, Sales, Customers behaviour, Primary School, Role-playing game and manufacturing etc. Your group may choose data of any other category based on your interest from Kaggle (www.kaggle.com) or UCI (https://archive.ics.uci.edu/ml/index.php) or any other repository. The dataset should have at least 4000 rows and 10 columns (for example, type of variables may be categorical, continuous and discrete) after cleaning and there is not any maximum limit. You would need to formulate a set of questions in the domain of chosen dataset and the ML project address these questions. Fundamentally, you need to crisply define what the ML modeling objective for Artificial Intelligence is for your project. For example, what are the most important features for predicting X variable? You can start with simple approaches so you can achieve something quickly and then progress to more complicated approaches during this group project.

The project group should consider the following guidelines during the development of Machine Learning (ML) project for Artificial Intelligence.

1. Justification for the selection of machine learning approaches for the chosen problem and dataset.
2. For ML techniques (Classification, Regression, Clustering, Decision Trees, Neural Networks, Reinforcement learning and Text Analytics), you should plan on trying multiple approaches (at least two), with proper parameter-selection techniques and a comparison between the chosen modelling approaches.
3. You should train the ML modelling techniques and subsequently, test the models. Perform a comparison of two or more ML modelling techniques. You may use a statistical approach to argue that one feature is more important than some other feature.
4. Depending on the complexity of the problem, you should use cross-validation approach to justify the authenticity of your ML modelling results.

Your group will present the findings and defend the results in the report (MS Doc/pdf or any other readable format). Your report should capture the following aspects that are relevant to your approach.

i. Brief description of the domain problem along with a chosen data set.

(10 marks)

ii. Motivation for the problem, data set and Challenges faced in ML project

(10 marks)

iii. What is/are the objectives of the problem(s) that are addressed in your project (Classification/ Regression/ Clustering Rules/ Information extraction etc..)

(10 marks)

iv. Characterisation of the data set: source URLS; size; number of attributes; has/does not have missing values; number of examples etc. Discuss the steps to remove the missing values from the dataset.

(5 marks)

v. Train the ML models based on three different splits (Choose one of them, such as 5%, 10%, 15% or 20%, 30%, 40%) and discuss the variation in accuracy/ score obtained from the models in the training as well as testing. Use Cross validation method for the authenticity of your results.

(30 marks)

vi. Interpret the results based on problem specification, address the ML modelling results are not under or over fitted. Justify with arguments.

(10 marks)

vii. Provide the explanation of code that will be used to solve the problem. Comments must be provided along with code.

(10 marks)

viii. Conclusions, Predictions and References (HARVARD style).

(10 marks)

ix. Describe the contribution of each team member in the project clearly and use a bar chart to represent the effort and time spent during this project.

(5 marks)

Maximum Number of Words for the report (1500 words excluding diagrams, code and HARVARD style References).

Note: The names of group members must be uploaded on the link provided on Moodle until 12th February.