**University of Wolverhampton**

**Faculty of Science and Engineering**

**Department of Mathematics and Computer Science**

**Module Assessment**

|  |  |
| --- | --- |
| **Module** | 5CS037 – Concepts and Technologies of AI |
| **Module Leader** | Hiran Patel |
| **Semester** | 1 |
| **Year** | 2020/21 |
|  |  |
| **Assessment** | Portfolio |
| **% of module mark** | 100% |
| **Due Date** | See below |
| **Hand-in – what?** | **Portfolio as specified in this document** |
| **Hand-in- where?** | Canvas |
|  |  |
| **Pass mark** | 40% |
| **Method of retrieval** | Submit the resit assessment (will be distributed at the end of the module) by end of resit week (July) |
| **Feedback** | Individual feedback via Canvas, in addition verbal feedback is available in class. |
| **Collection of marked work** | N/A |

**Assessment overview**

This portfolio is split up into 3 separate tasks which will test your knowledge of applying machine learning algorithms to real datasets and evaluating the current ethical implications of AI. Each task should be zipped up into a single zip folder containing all source code, resource files (datasets) and reports for the submission on Canvas.

Throughout this module, you will be taught many machine learning algorithms (regression and classification) such as Linear Regression, Multivariate Regression, Neural Networks (NN), Random Forest, Decision Trees, Support Vector Machines (SVM), Naïve Bayes and Deep Learning (building on Neural Networks). You will be given an explanation on how they work along with some examples. Your task is to apply this knowledge to different datasets and explain difference in accuracy (if any).

Throughout this assessment, you will be using 2 (or 3) datasets, if you are unsure of the suitability of the ones you have found, please confirm them with myself (Hiran) before you begin these tasks.

1. Regression Task (predicting specific numerical values) (40% - 100 marks)

For this task, you will find a dataset and predict a specific value using 3 different types of Machine learning algorithms (one must be using neural networks) and write a small report. Your first sub-task is to find a dataset suitable for regression (table data with mainly numerical values). For regression models, unless you are using a neural network (even then, it can vary), you will not need to normalise the data unless your research says otherwise. From your dataset, print out a scatter matrix of all numerical data. You should also aim to visualise any string data too. You will need to analyse the string data, group them so you can generate numerical data from it, then create some graphs. Even though you may have string data in the dataset, it may not be useful for regression so you will not be penalised for this task if you do not use that feature. You then need to pick 3 algorithms; these can be any of your choice, but you will need to explain how they work in your report. Once you have fit the model and made some appropriate predictions, state the accuracy of the model after using the test data, and evaluate its usefulness for your dataset. In the report, you must explain in detail as to why the model has performed in such way and suggest ways of improving the model.

**Create a regression model using algorithm 1 – state which algorithm and submit working source code (20 marks) – Model can be either Support Vector Regressor, Decision Tree Regressor, Random Forest Regressor or Multiple Linear Regression.**

**Create a regression model using algorithm 2 – state which algorithm and submit working source code (20 marks) – Model can be either Support Vector Regressor, Decision Tree Regressor, Random Forest Regressor or Multiple Linear Regression.**

**Create a model using Neural Networks – submit working source code (20 marks)**

**Report (40 marks)**

1. **Showing accuracies of each model.**
2. **Explanation of how the three models work (include references)**
3. **Suggest on different methods to improve your models. E.g. Remove/add data? If so, which features would you build on? Think about visualising the data first and seeing which features do not help with the prediction. For NN models, change the number of neurons, layers etc?**
4. Classification Task (predicting a class for a new input) (40% - 100 marks)

For this task, you will find a dataset and classify a set of records to a specific target using 2 different types of Machine learning algorithms and write a small report. Your first sub-task is to find a dataset suitable for machine learning. For many classification models, you will need to normalise the data so refer to methods on how to normalise numerical and string data. If there is a feature which has groups of strings, use the one-hot encoder, this will add extra columns in the dataset but will become much easier to train. For this task, you will be asked to use one CSV file type dataset (can use from task 1 if there is a classification feature). Once you have fit the model and made some appropriate predictions, state the accuracy of the model after using the test data, and evaluate its usefulness for your dataset. In the report, you must explain in detail as to why the model has performed in such way and suggest ways of improving the model.

**Create a classification model using algorithm 1 – state which algorithm and submit working source code (20 marks) –** Model can be either **Support Vector Machines, Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbours or Neural Networks** (you are not limited to these but these will be taught on the module).

**Create a classification model using algorithm 2 – state which algorithm and submit working source code (20 marks) -** Model can be either **Support Vector Machines, Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbours or Neural Networks** (you are not limited to these but these will be taught on the module).

**Report (60 marks)**

1. **Showing and comparing accuracies of each model.**
2. **Explanation of how your two models work (include references)**
3. **Suggest on different methods to improve your models. E.g. Remove/add data? If so, which features would you build on? Think about visualising the data first and seeing which features do not help with the prediction. For NN models, change the number of neurons, layers etc?**
4. Report on ethical issues surrounding Artificial Intelligence (20% - 100 marks)

For this task, you will write a report on the ethical issues surrounding AI. AI has been the forefront of many new strategies to improve aspects of industry such as sales, productivity (robotics), design etc. As much as there are many positive outcomes using AI, there are many ethical issues to consider when using AI. For example, is it OK to remove employees (human) and replace them with robots? Or would you trust a machine learning algorithm to generate a medical treatment plan for you? Your job is to write a report which outlines the key ethical issues, here are some you may want to consider:

* **Unemployment caused by AI**
* **Safety of AI regarding data security**
* **AI becoming Biased**
* **AI making mistakes – what are the consequences?**
* **Should AI be used to monitor and control people?**

The above points are suggestions, through your research you may find more interesting points. You should have no more than 5 headings and each section should contain three references at minimum. There is no set word count but if you would like a guide, between 1500 and 2000 words would be adequate.

**Important Message**

You may be asked to clarify your assessment after moderation has taken place. This is to ensure the work has been completed by yourself.

You must achieve 40 percent overall to pass this module. There will be a resit opportunity during resit week (July) to achieve a pass.

**Submission of work**

Your completed work for assignments must be handed in on or before the due date. ***You must keep a copy or backup of any assessed work that you submit.  Failure to do so may result in your having to repeat that piece of work.***

**Penalties for late submission of coursework**

Standard Faculty of Science and Technology arrangements apply.

**ANY late submission (without valid cause) will result in 0 marks being allocated** **to the coursework**.

**Procedure for requesting extensions**

If you have a valid reason for requiring an extension you must request an extension using e:vision. **Requests for extension to assignment deadlines should normally be submitted at least one week before the submission deadline and may be granted for a maximum of seven days (one calendar week).**

**Retrieval of Failure**

A pass of 40% or above must be obtained overall for the module (but not necessarily in each assessment task).

**Where a student fails a module they have the right to attempt the failed assessment(s) once, at the next resit opportunity (normally July resit period).  If a student fails assessment for a second time they have a right to repeat (i.e. RETAKE) the module.**

**NOTE: STUDENTS WHO DO NOT TAKE THEIR RESIT AT THE NEXT AVAILABLE RESIT OPPORTUNITY WILL BE REQUIRED TO REPEAT THE MODULE.**

**Mitigating Circumstances (also called Extenuating Circumstances).**

If you are unable to meet a deadline or attend an examination, and you have a valid reason, then you will need to request via e:vision **Extenuating Circumstances.**

**Feedback of assignments**

You will be given feedback when you demonstrate your work.

You normally have **two working weeks** from the date you receive your grade and feedback to contact and discuss the matter with your lecturer. See the Student’s Union advice page <http://www.wolvesunion.org/adviceandsupport/> for more details.

**Registration**

Please ensure that you are registered on the module. You can check your module registrations via e:Vision You should see your personal tutor or the Student Support Officer if you are unsure about your programme of study. The fact that you are attending module classes does not mean that you are necessarily registered. A grade may not be given if you are not registered.

**Cheating**Cheating is any attempt to gain unfair advantage by dishonest means and includes **plagiarism** and **collusion.** Cheating is a serious offence. You are advised to check the nature of each assessment. You must work individually unless it is a group assessment.

**Cheating** is defined as any attempt by a candidate to gain unfair advantage in an assessment by dishonest means, and includes e.g. all breaches of examination room rules, impersonating another candidate, falsifying data, and obtaining an examination paper in advance of its authorised release.

**Plagiarism** is defined as incorporating a significant amount of un-attributed direct quotation from, or un-attributed substantial paraphrasing of, the work of another.

**Collusion** occurs when two or more students collaborate to produce a piece of work to be submitted (in whole or part) for assessment and the work is presented as the work of one student alone.