Assignment

Faculty Name:	Information Technology
Module Code:	Machine Learning Algorithms
Module Name:	ITMLA2-B44 Block 4
Module Coordinator:	N/A
Internal Moderation:	Community of Practice
Copy Editor:	Mr Kyle Keens
Total Marks:	100
Submission Date:	25/11/2022

This module is presented on NQF level 6.

Mark deduction of 5% per day will be applied to late submission, up to a maximum of three days. Assignments submitted later than three days after the deadline or not submitted will get 0%. [1]

This is an individual assignment.

This assignment contributes 40% towards the final mark.

[1] Under no circumstances will assignments be accepted for marking after the assignments of other students have been marked and returned to the students.

Instructions to Student

1. Remember to keep a copy of all submitted assignments.
2. All work must be typed.
3. Please note that you will be evaluated on your writing skills in all your assignments.
4. All work must be submitted through Turnitin[2] and the full Originality Report should be attached to the final assignment.
5. Each assignment must include a cover page, table of contents and full bibliography, based on the referencing method applicable to your faculty as applied at Eduvos.
6. Use the cover sheet template for the assignment; this is available from myLMS.
7. Students are not allowed to offer their work for sale or to purchase the work of other students. This includes the use of professional assignment writers and websites, such as Essay Box. If this should happen, Eduvos reserves the right not to accept future submissions from a student.

Assignment Format

Students must follow the requirements when writing and submitting assignments as follows:

- Use Arial, font size 12.
- Include page numbers.
- Include a title page.
- Print submissions on both sides of the page.
- Write no more than the maximum word limit.
- Ensure any diagrams, screenshots and PowerPoint presentations fit correctly on the page and are referenced.
- Include a table of contents.
- Use the accurate referencing method throughout the assignment.
- Include a bibliography based on the applicable referencing method at the end of the assignment.
- Include the completed Assessment/Project Coversheet (available on myLMS).
- Check spelling, grammar and punctuation.
- Run the assignment through Turnitin software.

[2] Refer to the Eduvos Policy for Intellectual Property, Copyright and Plagiarism Infringement, which is available on *my*LMS.

Essential Embedded Knowledge and Skills Required of Students

- Report-writing skills
- Ability to analyse scenarios/case studies
- Understanding of subject field concepts and definitions
- Ability to apply theoretical knowledge to propose solutions to real-world problems
- Referencing skills

Resource Requirements

- A device with Internet access for research
- A desktop or personal computer for typing assignments
- Access to a library or resource centre
- Prescribed reading resources

Delivery Requirements (evidence to be presented by students)

- A typed assignment[3]
- A Turnitin Originality Report

Minimum Reference Requirements

At least five references for first year, ten references for second year and fifteen references for third year.

Additional reading is required to complete this assignment successfully. You need to include the following additional information sources:

- Printed textbooks/e-books
- Printed/online journal articles
- Academic journals in electronic format accessed via ProQuest or other databases
- Periodical articles e.g., business magazine articles
- Information or articles from relevant websites
- Other information sources e.g., geographic information (maps), census reports, interviews

Note
• It is crucial that students reference all consulted information sources, by means of in-text referencing and a bibliography, according to the applicable referencing method.

[3] Refer to the Conditions of Enrolment for more guidance (available on myLMS).

Section A

Learning Objective

Test the student's understanding of Machine Learning and Algorithm

Assignment Topic

Problems with Machine Learning and Algorithm

Scope

Week 1-5

Question 1 — 40 Marks

Study the scenario and complete the question(s) that follow:

Credit Worth Assessment: Technical Consultant

Poer Bank is a prolific South African bank. Due to the bank competitive loan interest rate, the company has faced a surge of many clients borrowing money from the bank, however not always paying back their loans on time. Thus, the company would like to minimise the risk of potential customers and have more informed decision on if it should grant a loan or not to a potential debtor. Poer Bank has therefore contacted your business intelligence expert team to assist them with this endeavour. Aware of the fact that Banks keep records of clients' transactions, you have requested Poer Bank to give access to customer credits information for the past 10 years to undergo the investigation. The Bank has granted you access to some of its data in its relational databases which they have exported in a voluminous SQL file under your disposition ("**poer_credit_data.sql**").

Source: Yves Matanga, 2022

1.1 Using the SQL file provided, create a local database (all tables included) on your system. Using Python and relevant SQL queries within Python, create a dataframe or list of customer records with the following attributes: **customer age**, **gender**, **marital status**, **education level**, **amount loaned** as well as **payback percentage** at end of term.

(7 marks)

1.2 To estimate the solvency of potential customers borrowing at Poer Bank, you have decided to build a multilayer perceptron (MLP) neural network regression model that evaluate the relationship between the customer age, marital status, education level against the payback percentage. Use 75% of your data to train the model and 15% for testing the model. You may use discrete numbers (single = -1, divorced = 0, married = 1) to condition the categorical variable: "**marital_status**"

a. To better estimate the data inner relationships, normalize the model features (age, marital status, education level). (5 marks)

b. Using Python, train the MLP regressor and plot two scatter plots to estimate the goodness of fitness for the trained data (training data vs model prediction) and the tested data (test data vs model prediction). Calculate the coefficient of determination R^2 for both sets (see appendix section). Retrain the model until you obtain a satisfactory goodness of fit for both trained set and tested set.
(You may the early stopping technique while training your model) (20 marks)

[Sub Total 25 Marks]

1.3 Having obtained the model estimate the extent to which potential debtors having the following profile will pay their debt to the end of term:

age	gender	Marital status	NQF	Loan request	Pb
28	f	married	7	90000	
28	m	married	7	90000	
40	m	single	8	1000000	
40	m	divorced	8	1000000	
50	f	single	7	120000	
50	f	single	7	800000	
50	f	married	5	800000	
28	f	single	7	90000	
65	f	single	7	90000	
25	m	single	5	380000	

(8 marks)

All graphs must have titles, x-axis, and y-axis labels as well as grids. Print your python code below each graph and results in your assignment documentation.

End of Question 1

Question 2 — 35 Marks

Study the scenario and complete the question(s) that follow:

DSS: Automating HR employee selection

Prek Hi Tech is an IT company in Sandton that design top notch software. The company would like to up its competitiveness in the market by selecting quality workforce. Over the years, the company has noticed that not all implied hired perform at the highest standard. Hence, the company has profiled all its employees for the past five years with a quality score capturing the following attributes: age, NQF level, previous experience prior to employment including their appreciation of their work quality.

Prek Hi Tech uses an online recruitment process to stock potential employees' details. Based on the profiling information stocked ("**Prek_HR_data.csv**"), the company would like to design an automatic selection criterion that pre-select the potential quality workers which will could through an actual interview process. As a machine learning guru in the company, you have been tasked to design the algorithm.

Source: Matanga, NY (2022)

2.1 From the data in Prek HR data ("**Prek_HR_data.csv**"), using Python generate a dataset that has as attributes: the **NQF level**, **age** and **experience level** of each profile employee against **quality score** as target. Split your dataset into a training set and validation set (15 % test data, 85% training data) and submit your python code for the effect. (6 marks)

2.2 . Design a decision tree classifier from the training data and compute the classification accuracy of both the training set and test set. Plot the decision tree (tree rules). You may use "**graphviz**" Python library to plot the decision tree rules. (19 marks)

2.3 Evaluate if the following candidate profiles will perform well or not at Prek Hi Tech based on the train model.

Age	NQF	Experience Level (num years)	Quality score
23	7	1	?
30	9	3	?
60	8	3	?
45	10	8	?
36	6	10	?

(5 marks)

2.4 Based on the result in 3 and the decision tree, Determine the profile of an eligible candidate for Prek Hi Tech. (5 marks)

All graphs must have titles, x-axis, and y-axis labels as well as grids. Print your python code below each graph and results in your project documentation.

End of Question 2

Question 3 — 25 Marks

Study the scenario and complete the question(s) that follow:

Optimising crops production: Merka Agri

Merka Agri is a wholesaler producer of corns. The company makes use of fertilizer to accelerate the growth of its crops added for 30 days period from planting prior to the harvesting. The amount of fertilizer obviously as an impact of the growth of the corns and the company would like to have an adequate estimate of the relationship between the amount of fertilizer added overall against the weight of the final corn. You have conducted an experiment for 100 crops whereby a fixed amount of mg per pot is added every day and at the end of 30 days the weights of the corns have been recorded. You are currently in the analysis of the data ("**merka_argi_corn_experiment.csv**").

Source : http://www.stat.cmu.edu/~hseltman/309/Book/chapter9.pdf

3.1. Generate a Scatter plot that shows the relationship between the amount of fertilizer per day (mg) and the weight of the corn. Label your x and y-axis appropriately as well as provide a title for your plot (6 marks)

3.2. Using Python, estimate by linear regression the relationship between the corn final weight and the fertilizer addition (mg) from Merka Agri data. (Use all your data for training the linear regressor). Write down the linear model of the corn growth:

weight = a***fertilizer_addition** + **b**

(10 marks)

3.3. Generate a new plot of your linear model, superimpose your scatterplot in question 1. (6 marks)

3.4. What is the expected weigh of a corn planted in Merka Agri farms if the 50mg of fertilizer is regularly added for 30 days prior to harvesting? (3 marks)

End of Question 3

Section B

Plagiarism and Referencing

Eduvos places high importance on honesty in academic work submitted by students and adopts a zero-tolerance policy on cheating and plagiarism. In academic writing, any source material, e.g., journal articles, books, magazines, newspapers, reference material (dictionaries), or online resources (websites, electronic journals, or online newspaper articles), must be properly acknowledged. Failure to acknowledge such material is considered plagiarism; this is deemed an attempt to mislead and deceive the reader and is unacceptable.

Eduvos adopts a zero-tolerance policy on plagiarism; therefore, any submitted assessment that has been plagiarised will be subject to severe penalties. Students who are found guilty of plagiarism may be subject to disciplinary procedures, and outcomes may include suspension from the institution or even expulsion. Therefore, students are strongly encouraged to familiarise themselves with referencing techniques for academic work. Students can access the Eduvos referencing guides on myLMS.