



**“Investing in Africa’s Future”**

**COLLEGE OF ENGINEERING AND APPLIED SCIENCES**

**AIN 1113: FUNDAMENTALS OF PROGRAMMING FOR MACHINE LEARNING**

**END OF FIRST SEMESTER EXAMINATIONS**

**NOVEMBER 2025**

**LECTURER: MR. LORENCE DHLAKAMA**

**TIME: 3 HOURS**

## **INSTRUCTIONS**

1. Answer all questions using **Python** and **VS Code** for implementation.
2. Include Comments in your code where necessary.
3. Show all code and outputs clearly.
4. Answer **ALL** questions from **Section A**.
5. Answer any **Two (2)** Questions from **Section B**.
6. Make screenshots of all interfaces and code into a single PDF document with your student number.
7. Create a folder with your student number on your desktop and save all your work in that folder.

**SECTION A [40 MARKS] Answer All questions from this section.**

**QUESTION 1**

- (a) Refactor the following code to follow Python Enhancement Proposal 8 (**PEP 8**) and improve readability: [2]

```
def greet(n):print("Hello "+n)
```

- (b) Using the dataframe below:

```
hospital_staff = [  
    {"hospital": "Parienyatwa", "doctors": 45, "nurses": 120},  
    {"hospital": "Mpilo", "doctors": 30, "nurses": 90},  
    {"hospital": "Chinhoyi", "doctors": 20, "nurses": 60}  
]
```

Use **pandas** to answer all of the following questions:

- (i) Load the above dataframe into pandas [3]
- (ii) Print hospitals with fewer than 25 doctors. [5]
- (iii) Calculate the total number of nurses across all hospitals. [5]
- (iv) Write a program to print the hospital with the highest doctor-to-nurse ratio. [5]

## QUESTION 2

Zimbabwean farmers rely heavily on rainfall to achieve good maize yields. You are given a dataset showing average rainfall (in mm) and maize yield (in tonnes per hectare) across 10 districts. Use linear regression to predict maize yield based on rainfall.

```
data = {
    "district": [
        "Chiredzi", "Gokwe", "Mutoko", "Binga", "Guruve",
        "Zvishavane", "Chipinge", "Karoi", "Beitbridge",
        "Mazowe"
    ],
    "rainfall_mm": [320, 450, 390, 280, 500, 410, 460, 370,
300, 430],
    "maize_yield_tph": [1.6, 2.4, 2.1, 1.3, 2.7, 2.2, 2.5,
2.0, 1.5, 2.3]
}
```

Write a program to:

- (a) Split the data into training and testing sets using `train_test_split`, with a test size of 20% and a `random_state` of 42. [2]
- (b) Instantiate a `LinearRegression` model. [5]
- (c) Train the model using the training data. [5]
- (d) Make predictions on the testing data. [5]
- (e) Print the model's coefficients. [3]

## SECTION B [60 Marks] (Answer Any Two questions from this Section)

### QUESTION 3

Using scikit-learn, write a script to perform a basic linear regression task. You are given a the following dataframe.

```
data = {  
    "rainfall_mm": [300, 350, 400, 450, 500, 550, 600, 650,  
600, 700, 750],  
    "maize_yield_tph": [1.5, 1.7, 2.0, 2.2, 2.4, 2.6, 2.8,  
3.0, 3.2, 3.4]  
}
```

Code a python program to:

- (a) Split the data into training and testing sets using `train_test_split`, with a test size of 20% and a `random_state` of 42. [5]
- (b) Instantiate a `LinearRegression` model. [5]
- (c) Train the model using the training data. [10]
- (d) Make predictions on the testing data. [5]
- (e) Print the model's coefficients. [5]

#### QUESTION 4

Using the dataset below:

Item	Price per kg (ZWL)
Maize	1.20
Tomatoes	2.50
Groundnuts	3.00

- (i) Create dictionary called **market\_prices** to store the above dataset. [5]
- (ii) Load the above dataframe into pandas. [5]
- (iii) Print items priced above ZWL 2.00 per kg. [5]
- (iv) Calculate the average price of all items. [5]
- (v) Write a program to print the cheapest item. [5]
- (vi) Write a program to print the expensive item. [5]

#### QUESTION 5

Given a dataset below create a dictionary called data.

NAME	SALARY	DEPARTMENT
John	4500	Hr
Lina	5200	It
Mark	6100	Finance
Tina	4800	Marketing

- a) Create a DataFrame from the dictionary you created above and display rows where `salary` is above 5000. [5]
- b) Add a column `tax` to the DataFrame, calculated as 10% of the salary. [5]
- c) Plot a horizontal bar chart showing name vs salary. Add title and axis labels. [5]
- d) Using the salary data, train a linear regression model to predict `salary` based on a manually created `experience` column: [2, 4, 6, 3]. [5]
- e) Create a binary column `high_income` (True if salary  $\geq$  5000). Train a logistic regression model to classify high income based on experience. [10]

**END OF EXAMINATION**