LIMKOKWING UNIVERSITY OF CREATIVE TECHNOLOGY
LESOTHO

BIAI 424 • Artificial Intelligence

Class Project

Faculty	Information Communication Technology
Start Date	04/05/2022
Presentation Date	Week 13 (Exact date will be communicated)

INFORMATION FOR CANDIDATE

1. This project is worth **20%** of the overall semester assessment.
2. The total number of marks available for this test is **100** marks.
3. You are allowed to make use of external code up to 20%.
4. Any usage of external code beyond 20% will result in an **automatic** zero (0) awarded for this assessment.
5. Any and all reports are to adhere to the below format:
   Font: Tahoma, size 12, line-spacing 1.5
6. Any references are to follow EIII style.
7. The maximum number of persons per group is three (3).
8. Consult regularly and be mindful of changes in requirements.

You and your groupmates have just been placed on probation by a big bank in Lesotho. The terms of your employment will be based on your performance in the next four (4) weeks. The bank sells different products to its clients and loans are seemingly the most profitable. The bank has however noted an increase in the number of defaulted loans and has tasked you to develop a machine learning based model to assist in the decision making of whether to allocate a loan to an applicant or not.

The bank has granted you access to a database of data that has different fields (predictors); they are however not sure if the parameters provided can be used to predict the probability of defaulting. The data is accessible at in your Google Classroom – you will be provided the link to it.

As you take a look at your data, take care that there are some missing values and outliers which you need to consider and provide reasoning for any decisions made.

Your subtasks are to perform the following **and provide reports** on some of the pipeline steps.

The combined report shall be at least 20 pages.

Task 1	Data gathering and cleaning a. Missing values [5] b. Outliers [5] c. Descriptive analysis [5] Statistical methodologies	[15]
Task 2	Data Prepossessing a. Feature extraction [10] b. Feature tuning [10]	[20]
Task 3	Model Determination [10] a. Provide reasoning for determined model	[10]
Task 4	Model Training [10]	[10]
Task 5	Model Tuning [10]	[10]
Task 6	Prediction dashboard a. Your solution must allow for inserting of predictor values to determine is a loan can be allocated or not	[10]
Task 7	Report format a. You are submitting one (1) printed copy and one (1) softcopy via Google Classroom	[5]

Take care that some of these steps are iterative, you will need to provide reasoning in your report for any and all reasons any decision were altered.

Following the above, you are to prepare a 10-minute presentation of your process; you presentation must include the following:

<table>
<tr><td>Task 8</td><td>Power point presentation
a. Data visualization [4]
At least 4 different figures
b. Steps taken in the model creation [3]
c. Performance matrix and accuracy [3]
d. Presentation [10]</td><td>[20]</td></tr>
<tr><td>Task 8</td><td>You will be provided formatted out of sample data that will be used to test your model; this data will be provided only on the day of presentation.
<table><tr><td>Accuracy</td><td>≤ 45</td><td>≤ 55</td><td>≤ 65</td><td>≤ 75</td><td>≤ 85</td><td>≤ 100</td></tr><tr><td>Mark</td><td>0</td><td>2</td><td>5</td><td>6</td><td>8</td><td>10</td></tr></table></td><td>[10]</td></tr>
</table>

Before completion of you project and presentation, you need to prepare and avail a GitHub link to your complete model.

The solution MUST have the following

1. A readme file to describe the intent of the project [10]
2. A file detailing the model creation. [10]