Introduction to Data Analysis (DATA 1200)
Assignment #3 – Neural Networks (15% of Final Grade)
Professor: Sam Plati

John Hughes wants to leverage the Leukemia remission dataset **leuanalysisNew.csv** and thinks that maybe a Neural Network model would be a good choice.

The variables are as follows:

Independent Variables:

CELL - cellularity of the marrow clot section

SMEAR - smear differential percentage of blasts

INFIL - percentage of absolute marrow leukemia cell infiltrate

LI - percentage labeling index of the bone marrow leukemia cells

BLAST - absolute number of blasts in the peripheral blood (BLAST)

TEMP - highest temperature prior to start of treatment

Dependent Variable:

REMISS – 0 (no Remission), 1 (Remission)

The previous findings (using Logistical Regression) generated the following results:

```
[[2 2]
 [1 1]]
              precision    recall  f1-score   support

           0       0.67      0.50      0.57         4
           1       0.33      0.50      0.40         2

    accuracy                           0.50         6
   macro avg       0.50      0.50      0.49         6
weighted avg       0.56      0.50      0.51         6
```

Please use the following assumptions in your Neural Network Model:

1. Use 20% for your test size, in other words test_size = 0.2
2. hidden_layer_sizes = (6, 4, 2)
3. max_iter = 10000
4. random_state = 100

NOTE: Please ensure that above constraints are applied to generate the proper analysis

The Ask:

1. Python Code – *2%*
   a) Using Python generate the Neural Network Algorithm. Attach a separate HTML copy of your Python Code with your submission.

   Note: All steps need to be annotated (i.e. Wk6a-NNandDTExample)

2. Create a PowerPoint (PPT) presentation that includes the following:
   a) Cover Page (Title, Name (1ˢᵗ and last) and Student Number)
   b) Present and Explain **four (4) key insights** from the Dataset (i.e. Mean, Std Deviation, etc.) – *4%*
   c) Present and Explain **three (3) key insights** from the classification report (i.e. Precision, Recall, F1) *- 3%*
   d) Compare the classification report from Neural Network Algorithm to the Logistical Regression Algorithm and identify/present **three (3) key insights** with associated explanations. – *3%*
   e) Present and Explain **three (3) ways** to help improve the performance of the Neural Network model. Please justify each of your answers. – *3%*

   Hint: Leverage the Wk6a-NNandDTExample

   Please post your PowerPoint Document (.ppt or .ppx) and HTML Python Code via assignments under Assignment #3 by
   Friday, February 25ⁿᵈ, 2022 @ 11:59 p.m.

Grading Rubric

	Exemplary (14-15)	Proficient (10-13)	Incomplete (7-9)	Needs Improvement (0-6)
Python Code *(2%)*	Python HTML file is complete	Python HTML file is mostly complete. Missing headings or structure.	Python HTML file is incomplete. Incorrect use of heading or code.	Python HTML file is missing or incorrect.
PPT *(13%)*	Cover Page Complete Four (4) Dataset insights presented with explanation/ justification Three (3) Neural Network Model key insights presented and fully evaluated Three (3) comparison insights presented and fully evaluated Three (3) ways to improve the model have been identified with detailed explanations.	Cover Page Complete Four (4) Dataset insights with high-level explanation/ justification Three (3) Neural Network Model key insights presented and with high-level evaluations Three (3) comparison insights presented with high-level evaluations Three (3) ways to improve the model have been identified with only high-level explanations.	Cover Page missing a least one element Less than four (4) insights and/or Missing explanation/ justification Less than three (3) Neural Network key insights presented and/or evaluated Less than three (3) comparison insights presented and/or evaluated Less than Three (3) ways to improve the model have been identified and/or incomplete explanations.	Cover Page Missing Answer is missing or incorrect. Neural network insights missing or incorrect Comparison insights missing or incorrect Missing ways to improve the model and/or incorrect.