# Project 4

Stephanie. Nebiat, Bryan, Adebola

## Final Grade: /100 Well done!

**Data Model Implementation (25 points)**

* A Python script initializes, trains, and evaluates a model (10 points)

10/10 It does.

* The data is cleaned, normalized, and standardized prior to modeling (5 points)

5/5 It is.

* The model utilizes data retrieved from SQL or Spark (5 points)

5/5 It does.

* The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared. (5 points)

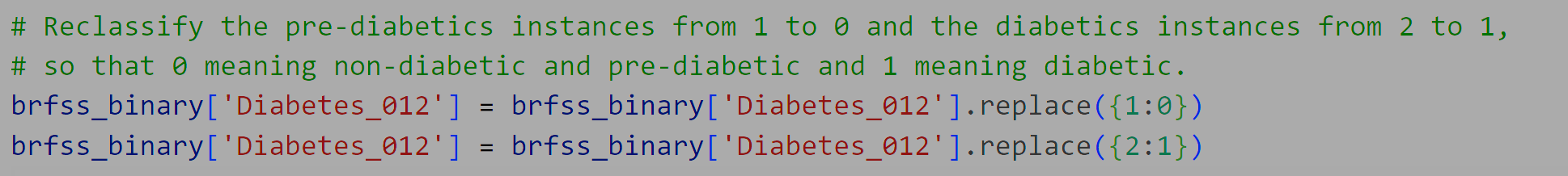
5/5 It does… barely. 😊

**Data Model Optimization (25 points)**

* The model optimization and evaluation process showing iterative changes made to the model and the resulting changes in model performance is documented in either a CSV/Excel table or in the Python script itself (15 points)

8/15 How did you determine which columns to keep. This was a great opportunity to use PCA. It should have been possible to run the code locally or in Colab so that you could systematically identify which columns to keep.

In this step:



Did you factor in the diabetes type? If so, Type 1 diabetes would be so different from Type 2 that you could just drop them. Also, women with gestational diabetes are different too and should probably be dropped.

What explains the drop from 433K rows to 76K rows. It seems like you have excluded a lot of rows. Does this invalidate your analysis? It might. You need to explain this better.

Where is the optimization process shown? You fiddled with the model a lot to get to 75%. How did you determine which columns to keep, how many epochs to run, how many layers in your neural network to use, and so on. Where’s the elbow curve? Or anything that shows how you got to 75%?

* Overall model performance is printed or displayed at the end of the script (10 points)

10/10 It is.

**GitHub Documentation (25 points)**

* GitHub repository is free of unnecessary files and folders and has an appropriate .gitignore in use (10 points)

10/10 It is and does.

* The README is customized as a polished presentation of the content of the project (15 points)

15/15 Excellent job.

**Group Presentation (25 points)**

* All group members speak during the presentation. (5 points)

5/5 Yes.

* Content, transitions, and conclusions flow smoothly within any time restrictions. (5 points)

4/5 I love the title slide. Very clean and excellently designed. The one point was deducted because in business presentations, you need to be succinct in the written and spoken words. I felt like there was a lot of data on the slide and there was a lot of spoken commentary as well. Too much, in fact. Don’t be afraid to leave a lot of data on the slide that the audience can delve into later if they will get the deck to work with after the presentation. That way, you just need to hit the highlights of the content, rather than speaking all the content.

* The content is relevant to the project. (10 points)

9/10 It was! I loved your deck. Very good. The only thing wrong is that you neglected to explain how you iterated through the options to come to the final 75% model.

* The presentation maintains audience interest. (5 points)

4/5 I was engaged for most of it. Unfortunately, there was too much speaking and you lost me at times. Also, please practice speaking in public. Or in private in a mirror. Practice until you are comfortable presenting, or until you can fake it well. It’s ok to pretend to be comfortable presenting. Still great job!