Doctor Anywhere - Take Home Assignment

For AI Engineer

Stroke Prediction

You are given a task by the Leadership team (CEO/CPO) to develop a robust machine learning model to predict stroke. You are responsible for the end-to-end and success of this AI project. For simplicity, in this assignment the scope is simplified and the dataset is already prepared for you so you can focus more on the ML modeling.

Task

1. Develop ML models to predict stroke given the dataset. You must only use the provided dataset to train your model.
2. Use and compare two variant of models:
   a. Gradient boosting (recommended libraries: scikit, XGBoost)
   b. You are free to choose other variant of models e.g. random forest, neural network
3. In your solution perform followings (you may add more if you want/feel necessary to)
   a. EDA and share insights can be told from the data
   b. Data processing/cleaning if applicable
   c. Model selection & optimisation: Cross Validation, hyperparameter optimisation, evaluation (with your selected metric(s) for the task)
   d. Train your final model
4. If you're shortlisted for next round, you'll present & explain your solution (using jupyter notebook) to interviewer

During the interview, you will present & explain the approach & result in your solution.

Evaluation Criteria for The Assignment

1. Processes on approaching the problem and ML development. Your solution should demonstrate your skills on understanding of the data, ML development, models, model selection, hyperparameters optimisation, and evaluations. (50%)
2. Model performance in provided development dataset and a test dataset (you are not given). (30%)
3. Presentation & readability of the solution and code (20%)

Submission

1. The deadline for this assignment is 4 working days from the time this assignment landed in your Inbox.
2. Code solution in jupyter notebook file,
3. Final optimized trained model files,
4. Your code should include a function:
    a. to take input of your trained models and a test data that has the same format like the train dataset,
    b. Make prediction
    c. Return prediction and your chosen metric(s),
       The function looks something like this:
       def run_prediction(model, test_data, ground_truth_label):
           …
           return prediction_output, report_metrics
5. Submit these files in single zip format and name it "mltest_{yourname}.zip".
6. Please submit by 'Replying All' to the email thread which you received this assignment from. Your submission should be in the form of a URL link. (Eg; A Google Drive link where we are able to access and download your deliverable zip)