arti ANALYTICS

Artificial Intelligence Engineer Assignment

BACKGROUND

Purpose of this assignment is to evaluate candidate's AI engineering skills in deep learning, problem-solving abilities, and adherence to coding standard in Python. Completed assignment should be submitted no later than 17.11.2024. 12:00 CET.

ASSIGNMENT OVERVIEW

1. Develop a simple feed forward neural network with three layers with 50 neurons in each layer.
2. Use the MNIST dataset to prepare data for training: https://netorgft13262551-my.sharepoint.com/:x:/g/personal/tarik_hubana_artianalytics_com/Eck9DIe00x9Jib_ricZhanIBUx0g_0X8pBImZEtMhiafzw?e=4FYNKK
3. Split the data in 10 datasets where each dataset consists only one class labels (numbers from 0 to 9)
4. Create 3 groups of sub-datasets for each dataset from step 3 by:
   a. Augmenting each dataset with another 5% rows compared to initial size, where additional 5% rows are randomly selected from other 9 classes of datasets from step 3 (10 new datasets)
   b. Augmenting each dataset with another 10% rows compared to initial size, where additional 10% rows are randomly selected from other 9 classes of datasets from step 3 (10 new datasets)
   c. Augmenting each dataset with another 15% rows compared to initial size, where additional 15% rows are randomly selected from other 9 classes of datasets from step 3 (10 new datasets)
5. Train all models on datasets from step 4 (30 in total) and use early stopping to determine the number of iterations (and epochs) for each neural network configuration training with each dataset.
6. Develop three ensemble models based on step 4 datasets (4.a, 4.b and 4c model) to test how mixture of additional data impacts performance (5%, 10% and 15%) and implement voting or other ensemble method to obtain final accuracy.
7. Measure the total learning time for each of three models (training of 10 sub-models), and calculate the total number of iterations per three models (sum of iterations of sub-models)
8. Test three ensemble models with test data.
9. Plot the dependence of accuracy (x-axis) and training time (y-axis) on one scatter plot, and dependence of accuracy (x-axis) and number of iterations (y-axis) on second scatter plot.
10. Comment and discuss the final results in analysis posted in GitHuB readme.

SUBMISSION INSTRUCTIONS

Code should be submitted via GitHuB, and analysis via GitHuB readme or word document. Please include assumptions made and how to run the application in readme section. Provide the link to the repository via email to tarik.hubana@artianalytics.com, migdat.hodzic@artianalytics.com and anesa.jakubovic@artianalytics.com