Task 6

Task: Dataset Selection and Neural Network Experimentation

Deadline for Presentation: Thursday, April 25th

Team Members: Maximum three allowed

Objective: The objective of this task is to enable trainees to select a classification dataset from either Kaggle or Hugging Face, build and compare several neural network architectures, perform hyperparameter tuning, and report on overfitting. Trainees will present their findings in a concise presentation to their peers within a limited timeframe.

1. **Dataset Selection (15 marks):**
   - Choose a classification dataset from either Kaggle or Hugging Face.
   - Ensure the dataset is approved by the trainer and meets the following criteria:
     - Number of Rows: At least 10,000 rows of data are generally considered suitable
     - Number of Categorical Columns: A minimum of 2 categorical columns is desirable
     - Total Number of Columns: at least 7, including both numerical and categorical features.
     - Offers a clear classification task with labeled data.
     - Provides features that are relevant and appropriate for the classification task.
2. **Neural Network Experimentation (40 marks):**
   - Experiment with building and comparing several neural network architectures on the selected dataset.
   - Explore various aspects of neural network design, including different architectures, activation functions, regularization techniques, and optimization algorithms.
   - Perform hyperparameter tuning on the neural network models to optimize performance.
   - Report on the presence of overfitting and discuss strategies for mitigating it.
3. **Presentation Preparation (15 marks):**
   - Prepare a presentation consisting of a maximum of 10 slides.
   - Ensure the presentation covers the following aspects:
     - Introduction to the selected dataset, including its source, size, and classification task.
     - Overview of the neural network architectures experimented with, including their design and hyperparameters.
     - Comparison of the performance of different neural network architectures and their hyperparameter-tuned versions.
     - Discussion of overfitting in the neural network models and strategies employed to address it.
     - Summary of findings and conclusions drawn from the experimentation process.
4. **Presentation Delivery (30 marks):**
   - Present the findings to peers in the class within a **time limit of 5 minutes**.
   - Deliver the presentation clearly and effectively, addressing all key points outlined in the presentation preparation.

- Engage with peers by encouraging questions and discussion regarding the presented findings.

Total Marks: 100

Note:

- Seek approval from the trainer before finalizing the dataset for experimentation.
- Collaboration, knowledge sharing, and feedback from peers and the trainer are encouraged throughout the experimentation and presentation process.
- The presentation aims to provide a concise summary of the dataset, experimentation process, and findings related to neural network experimentation.