Tab 1

ML Final Exam

Total Marks: 100

Submission Instructions

Please submit the following three links. Ensure all links are publicly accessible.

1. **GitHub Repository Link:**
    - Upload your code and requirements to a GitHub repository and make sure to create the repository selecting Public.
2. **Google Colab Link:**
    - Make sure the permission is set to **"Anyone with the link"** with **"Viewer"** or **"Editor"** access.
3. **Hugging Face Deployment Link:**
    - **Crucial:** Test your link in an **Incognito/Private tab** before submitting to ensure it is working for public users.

Dataset Selection

Instructions: Please choose one valid dataset from the options provided below, or select a suitable dataset of your own choice from the internet (e.g., Kaggle, UCI Machine Learning Repository).

- **Provided dataset** : **Final Exam Instructions**

(Note: Ensure the dataset is suitable for a classification or regression task involving standard tabular data.)

Tasks are given in the next page:

Tasks:

1. Data Loading (5 Marks)

- Load the chosen dataset into your environment and display the first few rows along with the shape to verify correctness.

2. Data Preprocessing (10 Marks)

- Perform and document at least 5 distinct preprocessing steps (e.g., handling missing values, encoding, scaling, outlier detection, feature engineering).

3. Pipeline Creation (10 Marks)

- Construct a standard Machine Learning pipeline that integrates preprocessing and the model

4. Primary Model Selection (5 Marks)

- Choose a suitable algorithm and justify why this specific model was selected for the dataset.

5. Model Training (10 Marks)

- Train your selected model using the training portion of your dataset.

6. Cross-Validation (10 Marks)

- Apply Cross-Validation to assess robustness and report the average score with standard deviation.

7. Hyperparameter Tuning (10 Marks)

- Optimize your model using search methods displaying both the parameters tested and the best results found.

8. Best Model Selection (10 Marks)

- Select the final best-performing model based on the hyperparameter tuning results.

9. Model Performance Evaluation (10 Marks)

- Evaluate the model on the test set and print comprehensive metrics suitable for the problem type.

10. Web Interface with Gradio (10 Marks)

- Create a user-friendly Gradio web interface that takes user inputs and displays the prediction from your trained model.

11. Deployment to Hugging Face (10 Marks)

- Deploy the Gradio app to Hugging Face Spaces and ensure it is accessible via a public URL.

Data

1. **Loan Approval Prediction System**
   Link : https://www.kaggle.com/datasets/uciml/loan-prediction-dataset

2. **Stock Price Trend Prediction**
   Link: https://www.kaggle.com/datasets/szrlee/stock-price-history

3. **Medical Insurance Cost Prediction**
   Link: https://www.kaggle.com/datasets/mirichoi0218/insurance

4. **Diabetes Prediction System**
   Link : https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

5. **Employee Attrition Prediction**
   Link:https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

6. **Weather Prediction Using Historical Data**
   Link: https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data

7. **Social Network Ads Click Prediction**
   Link: https://www.kaggle.com/datasets/rakeshrau/social-network-ads

8. **Video Game Sales Prediction**
   Link : https://www.kaggle.com/datasets/gregorut/videogamesales

9. **Mobile Price Classification**
   Link: https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification

10. **Is the Water Drinkable**
    Link: https://www.kaggle.com/datasets/adityakadiwal/water-potability