Replication

Data and Model Preparation

1. Datasets should be placed in the dataset/ directory. The following datasets are used:
    - processed_adult.csv
    - processed_communities_crime.csv
    - processed_compas.csv
    - processed_credit.csv
    - processed_dutch.csv
    - processed_german.csv
    - processed_kdd.csv
    - processed_law_school.csv
2. Pre-trained DNN models should be in the DNN/ directory with naming convention:
    - model_processed_adult.h5
    - ...

To replicate the complete results with 30 trials per dataset (as presented in the report):

```
python main.py
```

For a quicker verification with fewer trials, use command-line arguments:

```
python main.py --trials 3 --population 20 --generations 10
```

To test only one specific dataset eg: processed_adult:

```
python main.py --dataset processed_adult --trials 5
```

Expected Outputs

After running the experiments, you should see:

1. Terminal output showing progress for each dataset, including:
    - Generation-by-generation updates (for GA)
    - IDI ratio for each approach
    - Statistical test results
2. The results/ directory should contain:
    - summary_comparison.csv: Table with metrics for all datasets
    - simplified_comparison.png: Visualization comparing GA vs Random Search
    - Multiple comprehensive_comparison_*.png files (one per dataset)

Reproduction Notes

- Results may vary slightly due to the stochastic nature of both algorithms
- The statistical significance and general trends should remain consistent
- Differences in hardware (particularly CPU/GPU performance) may affect execution time but not the results