

Replication Guide

To replicate the fairness evaluation results, follow these steps:

1. Set Up the Environment

Ensure you have Python installed (recommended: Python 3.8+). Install required dependencies:

```
pip install tensorflow pandas numpy argparse
```

2. Prepare the Model and Dataset

- Place the trained model .h5 file inside the DNN/ folder.
- Place the processed dataset .csv file inside the dataset/ folder.

3. Verify eval2.py Settings

Ensure the script has correct paths and parameters set.

4. Run the Script

Execute the fairness evaluation with the desired feature:

```
python eval2.py gender
```

```
python eval2.py race
```

```
python eval2.py age
```

5. Interpret Results

The script outputs the **IDI ratio** for each tested feature. A higher **IDI ratio** suggests greater model bias for that feature.

6. Why Results Differ from Paper Values?

Since random sampling is involved in data pair generation, the exact IDI ratios will not match the paper's reported values exactly. However, the results should be close and consistent when:

The experiment is repeated multiple times and averaged.

A large enough budget is used (e.g., 100+ pairs).

The dataset and model remain the same as in the paper.