User Manual

Setup

1. Ensure all dependencies listed in `requirements.pdf` are installed
2. Clone or download the project repository
3. Navigate to the project directory in your terminal or command prompt

Data Format Requirements

The tool expects CSV files with the following format:

- Each row represents a different configuration
- Columns represent configuration options (features)
- The last column must contain the performance metric (target variable)

Example data format:

```
option1,option2,option3,...,performance_value
1,0,1,...,123.45
0,1,0,...,98.76
```

Directory Structure

The tool expects datasets to be organized in the following structure:

```
datasets/
├── system1/
│   ├── dataset1.csv
│   └── dataset2.csv
├── system2/
│   ├── dataset1.csv
│   └── dataset2.csv
...
```

Running the Tool

To run the tool with default settings:

```
python deep.py
```

Jupyter Notebook Alternative

Alternatively, you can use the provided Jupyter notebook:

1. Start Jupyter server:

```
jupyter notebook
```

2. Open the existing notebook in the browser:

```
deep.ipynb
```

3. Execute the notebook:
    - Either run all cells at once using the "Run All" command from the "Cell" menu
    - You can modify parameters directly in the notebook cells before execution

Customizing Parameters

To modify the experiment parameters, open `deep.py` and adjust the following variables at the beginning of the main function. The same code is available as the last cell of the `deep.ipnb` Jupyter Notebook.

```
# Systems to evaluate
systems = ['batlik', 'dconvert', 'h2', 'jump3r', 'kanzi', 'lrzip', 'x264', 'xz', 'z3']

# Number of repeat experiments
num_repeats = 10

# Fraction of data used for training (0.7 = 70%)
train_frac = 0.7

# Random seed for reproducibility
random_seed = 1

# Neural network parameters
epochs = 200
batch_size = 32
patience = 10
```

Output and Results

The tool generates the following outputs:

1. Terminal output with progress updates and summary statistics
2. CSV files in the `results/data/` directory:
    - `all_results.csv` : Summary metrics for all systems and datasets
    - `all_results_detailed.csv` : Detailed metrics for each repeat experiment
    - `training_times.csv` : Comparison of training times for LR and DL models

Visualizations can be obtained by running the `visualize.ipynb` Jupyter Notebook