

## User Guide: Data Noise Addition Script

### Overview

This script adds noise to specified columns in CSV files located in an input directory. It supports two types of noise distributions: Laplace and Gaussian.

The modified data is saved in a designated output directory.

### Prerequisites

Ensure you have the following installed:

- Python 3.x
- Required Python libraries:

***pip install pandas numpy***

### Input and Output

#### Input

- **CSV Files:** The script processes all .csv files inside the **data/** directory.
- **columns.txt (optional):** A file listing column names (one per line) for which noise will be applied.  
If this file is absent, noise is applied to all numerical columns.

#### Output

- **noisy\_data/:** Contains modified CSV files with noise applied.

### How to Run the Script

1. **Prepare Data:** Ensure CSV files are placed inside the data/ directory.
2. **Optional:** Create a columns.txt file listing specific columns to which noise should be applied.
3. **Run the Script:**

***python script.py***

4. **Choose Noise Distribution:**
  - Enter 1 for Laplace
  - Enter 2 for Gaussian
  - Enter 3 to exit
5. **Enter Epsilon Value:** This controls the amount of noise. Higher values add less noise. Values can range from between 0.05 to 5 for example.

6. **Output Files:** Processed CSV files will be saved in the noisy\_data/ directory.

#### **Notes**

- The script automatically creates the output directory if it doesn't exist.
- If columns.txt is missing, noise is applied to all numeric columns.
- Laplace noise is calculated based on the column median, while Gaussian noise uses the column mean.
- Higher epsilon values reduce noise that is added, while lower values increase it.

#### **Error Handling**

- If an invalid choice is entered for distribution, the script prompts again.
- If no numerical columns are found, the script processes without modification.
- An exception is raised if an unsupported noise distribution is specified.