**A screenshot of a computer program

AI-generated content may be incorrect.Documentation of github project organization**

Figure: Directory Structure

**Directory Details**

**1. data/**

* **Purpose**: To store datasets used in the project.
* **Subfolders**:
  + **raw/**: Original, untouched data files (e.g., .csv, .json, images).
  + **processed/**: Cleaned, transformed, or preprocessed datasets.
* **Best Practice**: Never modify files in raw/. Always save intermediate results in processed/.

**2. notebooks/**

* **Purpose**: For Jupyter notebooks used in experimentation and exploration.
* **Examples**:
  + 01\_data\_analysis.ipynb: Exploratory Data Analysis (EDA).
  + 02\_model\_trials.ipynb: Quick prototyping of models and ideas.

**3. src/**

* **Purpose**: Contains reusable, production-ready Python scripts.
* **Examples**:
  + data\_processing.py: Scripts for data cleaning, splitting, and transformations.
  + model.py: Definitions of machine learning models.
  + train.py: Scripts for training models.
  + evaluate.py: Scripts to compute performance metrics.

**4. experiments/**

* **Purpose**: To store logs, results, and outputs of different experiments.
* **Examples**:
  + Hyperparameter tuning results.
  + Performance metrics comparison files.
* **Best Practice**: Use tools like MLflow or W&B to track experiments here.

**5. models/**

* **Purpose**: To store trained models, checkpoints, and weights.
* **Examples**:
  + random\_forest.pkl: Serialized model file (e.g., using joblib or pickle).
  + model\_v1.h5: TensorFlow or PyTorch model files.

**6. utils/**

* **Purpose**: Contains utility scripts for common tasks.
* **Examples**:
  + plotting.py: Functions for visualizing data and results.
  + metrics.py: Functions for computing evaluation metrics.
  + logger.py: Custom logging configurations.

**7. config/**

* **Purpose**: To store configuration files for parameters and settings.
* **Examples**:
  + config.yaml: Training hyperparameters, file paths, etc.
  + logging\_config.json: Logging levels and formats.

**8. tests/**

* **Purpose**: Contains unit and integration tests to ensure code reliability.
* **Examples**:
  + test\_data\_processing.py: Tests for data transformation functions.
  + test\_model.py: Tests for model creation and inference.

**9. docker/**

* **Purpose**: Contains Docker-related files for containerizing the project.
* **Examples**:
  + Dockerfile: Instructions to build the project’s Docker image.
  + docker-compose.yml: Setup for multi-container configurations.

**10. Root Files**

* **requirements.txt**: List of Python dependencies (e.g., pandas, scikit-learn).
* **README.md**: Project overview, setup instructions, and usage guide.
* **.gitignore**: Specifies files and folders Git should ignore (e.g., data/processed/, \*.pkl).
* **LICENSE**: Specifies licensing terms for the project.

**Difference between src and notebook folder:**

**src Folder (Source Code)**

* **Purpose**: This folder contains the core implementation of your project. It holds reusable scripts and modules that perform specific tasks.
* **Contents**:
  + Scripts for each part of the ML pipeline, such as data preprocessing, model training, evaluation, etc.
  + These scripts can be imported and used in other parts of the project.
* **Why It’s Important**:
  + Promotes modularity: You can easily reuse functions and classes.
  + Clean separation of code: Keeps core logic away from experimentation (done in notebooks).

A screen shot of a computer

AI-generated content may be incorrect.

**notebooks Folder (Jupyter Notebooks)**

* **Purpose**: This folder is used for exploratory data analysis (EDA), initial prototyping, and experimentation.
* **Contents**:
  + Jupyter notebooks focused on:
    - Visualizing and understanding the data (EDA).
    - Trying out quick experiments.
    - Comparing models and hyperparameter settings.
  + Often less structured than the code in the src folder.
* **Why It’s Important**:
  + Keeps experimentation separate from production-ready code.
  + Serves as a visual and interactive space to explore ideas.

A black background with white text

AI-generated content may be incorrect.