

While working on a data science/machine learning project it is important to have a solid project structure to be able to make it reliable and reproducible.

A very good starting point is the project structure shown in the figure below which was provided by Abhishek Thakur in his book Approaching (Almost) Any Machine Learning Problem.

- ✓ **input/**: This folder consists of all the input files and data for your machine learning project. If you are working on NLP projects, you can keep your embeddings here. If you are working on image projects, all images go to a subfolder inside this folder.
- ✓ **src/**: We will keep all the Python scripts associated with the project here. If I talk about a Python script, i.e. any *.py file, it is stored in the src folder.
- ✓ **models/**: This folder keeps all the trained models.
- ✓ **notebooks/**: All jupyter notebooks (i.e. any *.ipynb file) are stored in the notebooks folder.
- ✓ **README.md**: This is a markdown file where you can describe your project and write instructions on how to train the model or to serve this in a production environment.
- ✓ **LICENSE**: This is a simple text file that consists of a license for the project, such as MIT, Apache, etc.

