Lab 4 - From Data to Model

🧹 Step 1: Data Processing

Explore and run the notebook : `notebooks/00_data_engineering.ipynb` using VSCode or JupyterHub Notebooks to understand the data engineering tasks.

Ensure you are in the root of the source code repo

```
cd house-price-predictor
```

Clean and preprocess the raw housing dataset:

```
python src/data/run_processing.py   --input data/raw/house_data.csv   --output
data/processed/cleaned_house_data.csv
```

validate that `data/processed` path you see the cleaned data as

```
README.md              cleaned_house_data.csv
```

Also Explore and run the notebook : `notebooks/01_exploratory_data_analysis.ipynb` to understand how Experimental Data Analysis looks like.

🧠 Step 2: Feature Engineering

Explore and run the notebook : `notebooks/02_feature_engineering.ipynb` to understand how feature engineering is done.

Apply transformations and generate features:

```
python src/features/engineer.py   --input data/processed/
```

```
cleaned_house_data.csv   --output data/processed/featured_house_data.csv   --
preprocessor models/trained/preprocessor.pkl
```

validate that it generates the `preprocessor.pkl`

```
ls models/trained/
```

[sample output]

```
README.md        preprocessor.pkl
```

📈 Step 3: Model Experimentation & Trainging

Explore and run the notebook : `notebooks/03_experimentation.ipynb` totun the model experiments.

This will generate `configs/model_config.yaml` .

If you have not run this notebook, download the sample config from model_config and add it to `configs/model_config.yaml`.

Train your model and log everything to MLflow:

```
python src/models/train_model.py   --config configs/model_config.yaml   --data
data/processed/featured_house_data.csv   --models-dir models   --mlflow-
tracking-uri http://localhost:5555
```

You could track Model Experiments/Results using http://localhost:5555/

validate the model file is been generated

```
ls models/trained/
```

[sample output]

```
README.md        house_price_model.pkl      preprocessor.pkl
```

Now the model is trained and ready to be packaged and setup for inference.

#courses/mlops