Project Architecture - *22/01/2026*

Exploration Notebooks

Exploration.ipynb

This notebook is used for data exploration and provides a detailed implementation of the required transformations and preprocessing steps.

Explo&PCA.ipynb

This notebook is used for data exploration analysis and PCA computation.

OptimizeParameters4.ipynb

This notebook is used for hyperparameter optimization for the model implemented in `LSTMn°4.ipynb`

OptimizeParameters5.ipynb

This notebook is used for hyperparameter optimization for the model implemented in `LSTMn°5.ipynb`

OptimizeParameters6.ipynb

This notebook is used for hyperparameter optimization for the model implemented in `LSTMn°6.ipynb`

PredictionQD6.ipynb

This notebook is used to run the entire pipeline in order to compute prediction for station QD6.

PredictionW14.ipynb

This notebook is used to run the entire pipeline in order to compute prediction for station W14.

PredictionRecents.ipynb

This notebook is used to run the entire pipeline in order to compute prediction for the other recent stations.

Modules

utils.py

This module provides essential functions for the end-to-end data pipeline, from raw data cleaning to final submission formatting. It also includes various functions for visualizing predictions and evaluating their results.

NAR_models.py

This module contains the Non Autoregressive (NAR) deep learning architectures used for predicting passenger affluence in the Transilien train network.

modelsQD6.py

This module contains the deep learning architectures used for predicting passenger affluence for station QD6.

modelsW14.py

This module contains the deep learning architectures used for predicting passenger affluence for station W14.

AR_models.py

This module contains the 1st autoregressive (AR) deep learning architecture used for predicting passenger affluence in the Transilien train network.

Prediction Attempts

LSTMn°4.ipynb

This notebook contains the prediction pipeline for the 1st model (NAR LSTM)

LSTMn°5.ipynb

This notebook contains the prediction pipeline for the 2nd model (AR LSTM)

LSTMn°6.ipynb

This notebook contains the prediction pipeline for the 3rd model (Enriched NAR LSTM)

CSV Files

QD6predictions

This file contains the frozen predictions for station QD6 from `PredictionQD6.ipynb`.

W14predictions

This file contains the frozen predictions for station W14 from `PredictionW14.ipynb`.

RECENTpredictions

This file contains the frozen predictions for the other recent stations from `PredictionRecents.ipynb`.

train_f_x.csv	-	*y_train_sncf.csv*	-	*x_test.csv*
x_train		y_train		x_test

x_train_new.csv

Features Enriched x_train with `CreateFeatures.ipynb`.

y_train_new.csv

Features Enriched y_train with `CreateFeatures.ipynb`.

y_test_example.csv

This file is an example of submission provided by the platform.

y_predicted_LSTM4

This file contains the predictions for general stations with the first model from `LSTM4.ipynb`. The file does not already contain the predictions for the recent station and the index order is not already the good one for submission.

y_predicted_LSTM5

This file contains the predictions for general stations with the second model from `LSTM5.ipynb`. The file does not already contain the predictions for the recent station and the index order is not already the good one for submission.