MATLAB for Data Science

Explore data; build machine learning models; do predictive analytics

Joko Triloka
joko.triloka@darmajaya.ac.id

Overview

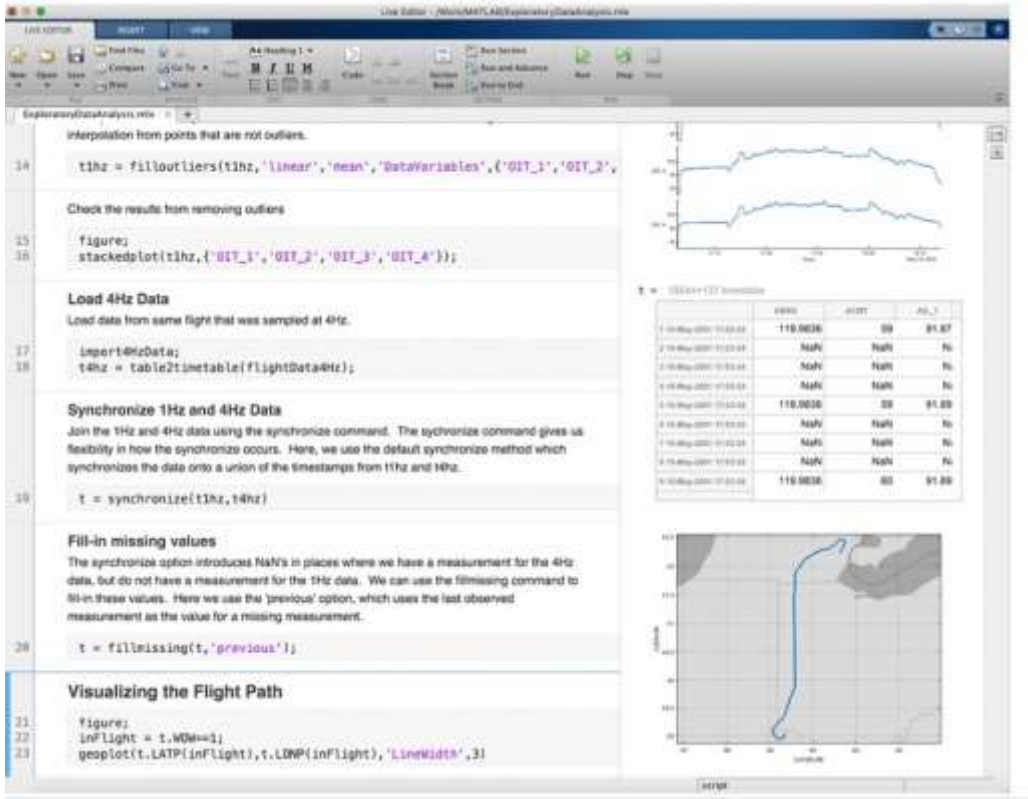
MATLAB® makes data science easy with tools to access and preprocess data, build machine learning and predictive models, and deploy models to enterprise IT systems.

- Access data stored in **flat files, databases, data historians, and cloud storage**, or connect to live sources such as **data acquisition hardware** and **financial data feeds.**
- Manage and clean data using **datatypes and preprocessing capabilities** for programmatic and interactive data preparation, including **apps for ground-truth labeling.**
- Document data analysis with **MATLAB graphics** and the **Live Editor** notebook environment.
- Apply **domain-specific feature engineering** techniques for sensor, text, image, video, and other types of data
- Explore a wide variety of modeling approaches using **machine learning and deep learning apps**.
- Fine-tune machine learning and deep learning models with **automated feature selection, model selection, and hyperparameter tuning** algorithms.
- Deploy machine learning models to **production IT systems**, without recoding into another language.
- Automatically convert machine learning models to **standalone C/C++ code**

Why Use MATLAB for Data Science?

Exploratory Data Analysis

Spend less time preprocessing data. From time-series sensor data to images to text, MATLAB datatypes significantly reduce the time required to preprocess data. High-level functions make it easy to synchronize disparate time series, replace outliers with interpolated values, filter noisy signals, split raw text into words, and much more. Quickly visualize your data to understand trends and identify data quality issues with plots and the Live Editor.



Exploratory Data Analysis

>>> First Step: Preprocessing data

Data cleaning, smoothing, grouping

Data can require preprocessing techniques to ensure accurate, efficient, or meaningful analysis. Data cleaning refers to methods for finding, removing, and replacing bad or missing data. Detecting local extrema and abrupt changes can help to identify significant data trends. Smoothing and detrending are processes for removing noise and polynomial trends from data, while scaling changes the bounds of the data. Grouping and binning methods identify data characteristics by groups.

Live Editor Tasks

Clean Missing Data	Find, fill, or remove missing data in the Live Editor
Clean Outlier Data	Find, fill, or remove outliers in the Live Editor
Find Change Points	Find abrupt changes in data in the Live Editor
Find Local Extrema	Find local maxima and minima in the Live Editor
Smooth Data	Smooth noisy data in the Live Editor
Remove Trends	Remove polynomial trend from data in the Live Editor
Join Tables	Combine two tables using key variables in the Live Editor

End of Preprocessing Data

Next….Data Visualization……