FOUR MATLAB® LIVE SCRIPT NOTEBOOKS · POWERED BY THE CARIMO ML KIT

# MACHINE LEARNING NOTEBOOKS

**4** MATLAB NOTEBOOKS

The complete machine-learning pipeline in MATLAB, taught on real hardware data: from raw sensor signals to deployed, uncertainty-aware models.

IN ONE BREATH

Four MATLAB Live Script notebooks deliver a full, rigorous machine-learning curriculum: 68 guided experiments spanning data exploration, classification, regression and clustering. Every experiment runs on real, labelled sensor data generated by the CARIMO ML Kit (sold separately), so students learn on genuine, physically-grounded signals, with an honest supervised-vs-unsupervised gap rather than clean synthetic toy sets.

**68**
guided experiments

**75+**
ML models & algorithms benchmarked

**1.9M**
rows of real sensor data

## Inside the four notebooks

· the full machine-learning pipeline, hands-on in MATLAB

### Basic Experiments — 11 experiments

Exploratory data analysis and feature engineering

The on-ramp: data audit → outlier filtering → scatter exploration → feature ranking → PCA → first classifiers → lab report.

**Techniques:** ANOVA F-score · Mutual Information · mRMR feature ranking; Decision Tree, Random Forest, SVM, k-NN; PCA & t-SNE; cross-validation, bootstrap, silhouette, ARI.

**Signature finding:** a three-way ensemble feature ranking surfaces the "Two Systems" structure: electrical features predict the load class, while vibration features carry no signal.

### Classification — 19 experiments

Full supervised pipeline: single-model deep-dive → 38-model benchmark → feature selection → PCA → hyperparameter tuning → cost-sensitive learning → interpretability → held-out validation → deployment.

**Techniques:** decision trees, random forest, gradient boosting, SVM (linear/RBF), k-NN, logistic regression, neural networks, naïve Bayes; stratified & block-level 10-fold CV; **Bayesian hyperparameter optimisation** (Gaussian-process surrogate, expected-improvement); misclassification-cost matrices; **SHAP · LIME · partial-dependence**.

**Rigour that impresses:** block-level cross-validation eliminates temporal leakage; ~99% accuracy across a 38-model benchmark.

### Regression — 21 experiments

Predict vibration intensity from accelerometer features: OLS baseline → 25-model benchmark → OOB permutation importance → Ridge sweep → Bayesian AutoML → residual diagnostics → conformal intervals → deployment.

**Techniques:** OLS, Ridge, Lasso, ElasticNet, SVR, k-NN, Random Forest, Gradient Boosting, Bagged Trees, **Gaussian Processes**, Quantile Regression; **conformal prediction intervals**; moving-window feature engineering on 500 Hz streams; block temporal split; `fitrauto` AutoML.

**Graduate-grade:** Gaussian-process regression with calibrated uncertainty, plus distribution-free conformal intervals with guaranteed coverage.

### Clustering — 17 experiments

Unsupervised discovery: K-Means deep-dive → 12-algorithm benchmark → feature selection → PCA → optimal-K discovery → density methods → hierarchical + Birch → GMM (BIC) → stability → 2-D explorer.

**Techniques:** K-Means++, K-Medoids, Agglomerative (Ward), Birch, **DBSCAN, OPTICS, HDBSCAN**, Mean-Shift, Affinity Propagation, Spectral, Gaussian Mixtures; PCA · t-SNE · UMAP; silhouette, Gap statistic, BIC, ARI; bootstrap cluster stability.

**The lesson that sticks:** a 12-algorithm leaderboard where the best unsupervised score (ARI ≈ 0.04) exposes why labels matter: on the very same data, classification labels at ~99%.

CARIMO ML NOTEBOOKS

# Where the data comes from, and what you need to run it

## The data source: CARIMO ML Kit (sold separately)

The notebooks run on real, labelled sensor data generated by the **CARIMO ML Kit**, a compact, sensor-rich benchtop platform built around a real PMDC motor. **Hot-swappable magnetic load disks** and a vibration knob inject genuine physical disturbances, streamed live across 12 signal channels. Students use the datasets that ship with the notebooks, or **capture and label their own** on the kit. It is a genuine hardware data source that gives the labs their physically-grounded character.

- 12 V PMDC motor + quadrature encoder (1336 CPR)
- Coin-type vibration motor + front-panel 0–100% intensity knob
- ARM Cortex-M4F microcontroller (with FPU) · 12-bit ADC
- Hot-swap magnetic load disks of increasing magnet count, loading the motor into distinct classes
- 3-axis MEMS accelerometer (±2 g, 1 mg) + current sensor (±3 A, ±2.5%)
- USB 2.0 · timestamped CSV export · live-plot GUI (12 signal channels)

**Lab 1: Basic, Classification & Clustering**
Nine-class load dataset · ~1.1 M rows · 11 features · 50 Hz

**Lab 2: Regression**
Vibration dataset · ~780 K rows · 500 Hz high-rate accelerometry

**Guided · Standard · Research**
depth switcher for first-years to researchers

**10 + 10 + 10**
quizzes · exercises · variations, every experiment

**Interactive & reportable**
in-browser demos + lab-report templates

## Methodological rigour for research-grade teaching

**Leakage-aware validation**
Block-level and held-out temporal splits, not naïve k-fold.

**Calibrated uncertainty**
Gaussian-process posteriors + distribution-free conformal intervals.

**Interpretable & reproducible**
SHAP / LIME / PDP on versioned, documented datasets.

**Honest benchmarks**
75+ models head-to-head; the supervised-vs-unsupervised gap shown, not hidden.

## What you need to run it

### MATLAB: your own licence

A recent MATLAB release with the relevant toolboxes (e.g. **Statistics & Machine Learning Toolbox**; **Deep Learning Toolbox** for the neural-network labs). **Licences are the user's own and are not included.**

### The interactive lab manuals

Each notebook is paired with a companion HTML lab manual that opens in any browser (no server, no install), with the depth switcher, interactive demos, quizzes and exercises.

### CARIMO ML Kit (sold separately)

The hardware kit generates the datasets that ship with the notebooks, and lets students capture and label their own real sensor data.

Ideal for machine-learning, data-science and AI teaching laboratories, from first-year undergraduate to graduate research.

## A complete, rigorous machine-learning curriculum in MATLAB, on real hardware data.

Bring hands-on machine learning to your laboratory. Request a demo or a quote today.

**Dr. Sravya Paluri** · Chief Operating Officer

contact@carimo.tech | +91 90823 97027 | www.carimotech.com

CARIMO Technologies Pvt. Ltd., recognised by the Government of India as an innovative AI-based startup (2024), incubated at SINE, IIT Bombay.
MATLAB® is a registered trademark of The MathWorks, Inc. • © 2026 CARIMO Technologies Pvt. Ltd. • Specifications subject to change.