Machine Learning for All

Secure Member Login

EMAIL

email@company.com

PASSWORD   FORGOT?

••••••••

Sign In →

ENTERPRISE AUTH

G Google

Sign up **Register**

Machine Learning for All
INTELLIGENCE, INSIGHTS, PREDICTIONS

+ NEW ML PROJECT

- Tutorial
- **Dashboard**
- ML Workbench
- AI Insights
- Data Repository
- ML Glossary

- Subscription
- Legal

WORKSPACE   DASHBOARD

Dashboard
Track datasets, models, and predictive performance across your workspace.

+ NEW ML PROJECT

ML RUNS	MODELS	DATASETS
83	**83**	**5**
TOTAL TRAINING RUNS LOGGED	ACTIVE MODELS IN YOUR REGISTRY	FILES STORED SECURELY IN THE CLOUD

📈 Latest Training Snapshot   MOST RECENT RUN

MODEL TYPE	ROWS	METRIC
Gradient Boosting	3201	Accuracy 0.861

All Trained Models (83)

Gradient Boosting CLASSIFICATION	loss_incident_prediction_dataset.csv	Features 9 inputs → 1 outputs	Performance Acc: 86.1%	Model ID 1572d34b-c1ad-406b-921c-532371c5048d
Created 15/03/2026				
Gradient Boosting REGRESSION	fried_chicken_feedback_5000.csv	Features 5 inputs → 1 outputs	Performance R²: 0.952	Model ID 4087f36d-0afc-4de5-9259-9c8c50ade42a
Created 14/03/2026				
Gradient Boosting CLASSIFICATION	finance_loan_default_dominant_v3b.csv	Features 7 inputs → 1 outputs	Performance Acc: 85.4%	Model ID 3a9558bb-aab2-4aa0-abc7-b7b83a0caf09
Created 14/03/2026				
Gradient Boosting CLASSIFICATION	finance_loan_default_dominant_v3b.csv	Features 7 inputs → 1 outputs	Performance Acc: 85.4%	Model ID 3aa9b93-406a-4da0-926d-07d84b80a03
Created 14/03/2026				

ML Workbench
Train models from your dataset and deploy predictions to the cloud.

📄 Dataset   0 ROWS

☁️
Drag & drop your file here
or click to browse

.CSV   .XLSX   .XLS

Data Statistics & Preview

ROWS	COLUMNS	NUMERIC	CATEGORICAL
3001	**9**	**8**	**0**

Detailed Column Statistics

vibration_rms_mm_s — float64

COUNT	MEAN	MEDIAN	MODE	STD DEV	VARIANCE
3000	3.46	3.46	0.60	1.21	1.47

MIN	Q1 (25%)	Q3 (75%)	MAX	RANGE	IQR
0.60	2.64	4.28	7.39	6.79	1.64

SKEWNESS	UNIQUE	MISSING
0.03	2187	1 (0.03%)

motor_temp_c — float64

COUNT	MEAN	MEDIAN	MODE	STD DEV	VARIANCE
3000	67.99	68.10	68.80	11.99	143.87

MIN	Q1 (25%)	Q3 (75%)	MAX	RANGE	IQR
35.00	60.10	76.10	112.90	77.90	16.00

SKEWNESS	UNIQUE	MISSING
-0.01	550	1 (0.03%)

hydraulic_pressure_bar — float64

COUNT	MEAN	MEDIAN	MODE	STD DEV	VARIANCE
3000	184.62	184.60	191.40	24.75	612.60

MIN	Q1 (25%)	Q3 (75%)	MAX	RANGE	IQR
98.10	167.90	201.80	260.00	161.90	33.90

Data Labeling

Create Labels from Existing Data
Create a new labeled column based on rules applied to an existing column. Choose manual mode for custom rules or automatic mode for smart labeling.

💡 **Labeling Suggestions** `0 required`

These columns should be labeled/encoded before training. ML models need numeric input — text columns **must** be converted.

crane_failure_30d `RECOMMENDED`   boolean · 2 unique · 1 null
Boolean column — encode to 0/1 for better ML compatibility
✦ Best: Label Encoding (true→1, false→0)
[Apply]

vibration_rms_mm_s `OPTIONAL`   float64 · 2187 unique · 1 null
Continuous numeric (2187 unique values) — binning into categories can improve classification
✦ Best: Quartiles (divide into 4 meaningful groups)
[Apply]

LABELING MODE

⚙ **Manual Rules**
Define custom conditions

🪄 ✦ **Automated**
Smart labeling

SOURCE COLUMN   [✓ Select All] [✗ Clear]

☐ vibration_rms_mm_s
☐ motor_temp_c
☐ hydraulic_pressure_bar
☐ load_ratio
☐ wind_speed_m_s
☐ cycles_per_day

NEW COLUMN NAME

[e.g., category, risk_level]

⚙ Change Data Types

Convert columns to different data types before training.

ⓘ Understanding Data Types

INTEGER (int64)

Whole numbers without decimals

✓ Examples:
1, 5, 100, -25, 0, 9999

💡 Use for: Age, Count, Quantity, ID Numbers

DECIMAL (float64)

Numbers with decimal points

✓ Examples:
3.14, 98.6, 0.5, -12.75, 100.0

💡 Use for: Price, Temperature, Weight, Percentage

TEXT (object)

Letters, words, or mixed characters

✓ Examples:
"apple", "New York", "A1B2", "Yes"

💡 Use for: Name, Category, Address, Description

BOOLEAN (true/false)

Only two possible values

✓ Examples:
True, False, Yes, No, 1, 0

💡 Use for: Is Active, Has Feature, Passed/Failed

⚠ **Important:**
Choose the correct data type for better ML model performance. Wrong types can cause errors or poor predictions!

vibration_rms_mm_s	motor_temp_c	hydraulic_pressure_bar
Decimal (float64)	Decimal (float64)	Decimal (float64)

load_ratio	wind_speed_m_s	cycles_per_day
Decimal (float64)	Decimal (float64)	Integer (int64)

maintenance_overdue_days	operator_experience_years	crane_failure_30d
Integer (int64)	Decimal (float64)	Boolean

APPLY TYPE CHANGES   **RESET TO ORIGINAL**

Dataset

3001 ROWS

INPUT FEATURES 8/9

☑ Select All ✗ Clear

☑ vibration_rms_mm_s ☑ motor_temp_c ☑ hydraulic_pressure_bar
☑ load_ratio ☑ wind_speed_m_s ☑ cycles_per_day
☑ maintenance_overdue_days ☑ operator_experience_years ☐ crane_failure_30d

8 of 9 selected

OUTPUT TARGETS 1/9

☑ Select All ✗ Clear

☐ vibration_rms_mm_s ☐ motor_temp_c ☐ hydraulic_pressure_bar
☐ load_ratio ☐ wind_speed_m_s ☐ cycles_per_day
☐ maintenance_overdue_days ☐ operator_experience_years ☑ crane_failure_30d

1 of 9 selected

MODEL SELECTION MODE

Automatic (Best Model)

☑ **STRATIFIED SPLIT (FOR CLASSIFICATION)**
Maintains class distribution in train/test sets

TEST SIZE

0,2

▷ TRAIN MODEL

MODEL
Random Forest
✓ Auto-Selected Best

METRIC
Accuracy 0.899

OUTPUTS
1

All Model Scores (Auto Selection)

Logistic Regression	Random Forest	Gradient Boosting	Decision Tree
0.884	**0.899** ★ BEST	**0.894**	**0.829**

K-Nearest Neighbors	Naive Bayes	SVM	Neural Network (MLP)
0.847	**0.880**	**0.865**	**0.867**

Feature Importance

8 features

vibration_rms_mm_s
motor_temp_c
load_ratio
maintenance_overdue_days
hydraulic_pressure_bar
cycles_per_day
wind_speed_m_s
operator_experience_years
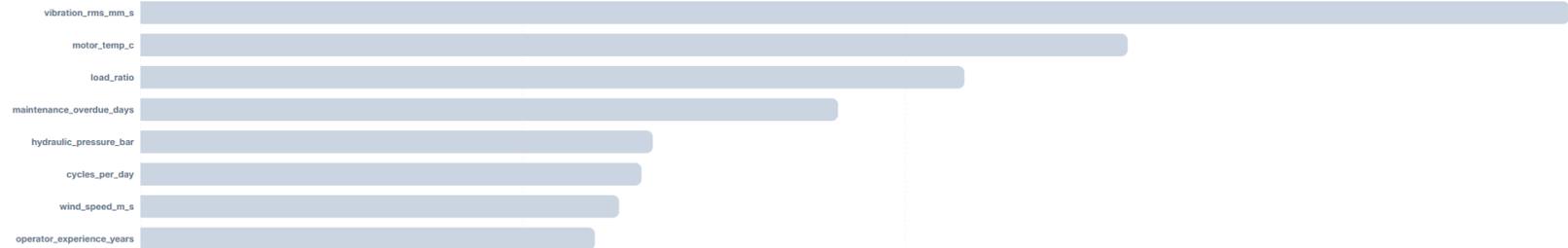
Manual Prediction

VIBRATION_RMS_MM_S

5

MOTOR_TEMP_C

50

HYDRAULIC_PRESSURE_BAR

150

LOAD_RATIO

0,5

WIND_SPEED_M_S

5

CYCLES_PER_DAY

500

MAINTENANCE_OVERDUE_DAYS

50

OPERATOR_EXPERIENCE_YEARS

5

PREDICT

⊘ **Prediction Results**

CRANE_FAILURE_30D

`CLASSIFICATION`

0.00

ⓘ **MODEL INFORMATION**

crane_failure_30d: `classification`

🎯 **Prediction Optimizer**

Fix one input feature and find the top 3 combinations of other features that produce the best predictions for your target output.

`999/999 RUNS REMAINING TODAY`   `⏱ COOLDOWN: 8S`

📊 ~14.962.140.849.768.138.000.000 possible combinations (smart-sampled to 100.000) — Large dataset, may take longer

⚡ CLASSIFICATION — FINDS HIGHEST PROBABILITY FOR TARGET CLASS

FIXED INPUT FEATURE

vibration_rms_mm_s

FIXED VALUE

5

TARGET OUTPUT

crane_failure_30d

TARGET CLASS (CODE)

1

`⏱ WAIT 8S`

`✓ 50.000 combinations evaluated (of 6.838.272.691.850.154.000 total)`   `⚡ Smart sampling used`   `Fixed: vibration_rms_mm_s = 5`   `Class: 1`

#1   **95%**
                                                                                                               PROBABILITY

vibration_rms_mm_s: 5.0 | motor_temp_c: 89.2 | hydraulic_pressure_bar: 192.6 | load_ratio: 0.954 | wind_speed_m_s: 3.53 | cycles_per_day: 591.0 | maintenance_overdue_days: 79.0 | operator_experience_years: 2.7

#2   **95%**
                                                                                                               PROBABILITY

vibration_rms_mm_s: 5.0 | motor_temp_c: 90.8 | hydraulic_pressure_bar: 162.2 | load_ratio: 0.61 | wind_speed_m_s: 5.56 | cycles_per_day: 404.0 | maintenance_overdue_days: 107.0 | operator_experience_years: 5.9

#3   **95%**
                                                                                                               PROBABILITY

vibration_rms_mm_s: 5.0 | motor_temp_c: 75.4 | hydraulic_pressure_bar: 161.0 | load_ratio: 0.833 | wind_speed_m_s: 6.88 | cycles_per_day: 392.0 | maintenance_overdue_days: 75.0 | operator_experience_years: 2.5

🏆 Global Optimizer

Find the top 10 input combinations across ALL features that produce the best predictions for your target output.

📊 ~14.962.140.649.768.138.000.000 possible combinations (smart-sampled to 100.000) — Large dataset, may take longer

🏷️ CLASSIFICATION — FINDS HIGHEST PROBABILITY FOR TARGET CLASS

TARGET OUTPUT

crane_failure_30d ⌄

NUMBER OF RESULTS

Top 10 ⌄

TARGET CLASS (CODE)

1

⏳ WAIT 11S

✓ 100.000 combinations evaluated (of 14.962.140.649.768.138.000.000 total) ⚡ Smart sampling used Class: 1

RANK	VIBRATION_RMS_MM_S	MOTOR_TEMP_C	HYDRAULIC_PRESSURE_BAR	LOAD_RATIO	WIND_SPEED_M_S	CYCLES_PER_DAY	MAINTENANCE_OVERDUE_DAYS	OPERATOR_EXPERIENCE_YEARS	PROBABILITY
🥇	6.36	91.70	167.10	0.50	7.72	521.00	45.00	0.20	**95%**
🥈	5.40	97.00	160.30	0.75	2.38	495.00	146.00	7.30	**95%**
🥉	5.85	93.70	151.70	0.80	5.64	594.00	82.00	8.30	**94%**
#4	5.26	91.30	152.50	0.63	5.89	445.00	66.00	7.70	**94%**
#5	5.85	84.70	112.30	0.64	2.81	549.00	131.00	6.80	**94%**
#6	5.06	89.70	144.10	0.68	7.28	404.00	79.00	2.90	**94%**
#7	5.93	89.00	174.40	0.63	9.13	422.00	116.00	5.00	**93%**
#8	5.20	74.00	157.40	0.85	7.52	494.00	122.00	2.00	**93%**
#9	5.32	88.50	196.30	0.82	8.49	395.00	76.00	1.70	**93%**
#10	5.83	92.60	174.80	0.89	3.78	380.00	87.00	2.30	**93%**

💡 **Optimization Summary**

The #1 input combination has a 95% probability for class 1. Evaluated 100.000 combinations (sampled from 14.962.140.649.768.138.000.000 possible).

AI-Powered Insights
GET EXPERT ANALYSIS OF YOUR ML MODELS

ⓘ What is this?
Let AI analyze your machine learning models and provide easy-to-understand insights, recommendations, and business value explanations. Perfect for non-technical users who want to understand what their models are telling them.

🗒 Select Trained Model to Analyze

YOUR TRAINED MODELS (85)

Random Forest - Classification - 📄 crane_failure_prediction_dataset.csv (Acc 89.7%) - 15/03/2026

TYPE	TASK	FEATURES	PERFORMANCE
Random Forest	Classification	8 inputs	89.7%

Select AI Provider

OpenAI GPT Models	Claude Anthropic Models	✨ Gemini Google Models	LM Studio Local Server

⚙ API Key Configuration

SELECT MODEL

⭐ Gemini 2.5 Flash (Recommended) — FREE / $0.30/$2.50 · 1M ctx

---

📊 ML Data Insights Report
Sunday, March 15, 2026 · 09:25 PM

MODEL OVERVIEW

MODEL TYPE		TASK	
Random Forest		Classification	

MODEL ID		GENERATED BY	
33fa86ac-7f6c-4e8c-abd2-ea3b0d9f3c5a		GEMINI	

PERFORMANCE METRICS

METRIC	VALUE
f1	0.8786
best_model	Random Forest
accuracy	0.8968
recall	0.8968
all_model_scores	[object Object]
precision	0.8866

FEATURE IMPORTANCE

RANK	FEATURE	IMPORTANCE
#1	vibration_rms_mm_s	0.0%
#2	motor_temp_c	0.0%
#3	load_ratio	0.0%
#4	maintenance_overdue_days	0.0%
#5	hydraulic_pressure_bar	0.0%
#6	wind_speed_m_s	0.0%
#7	cycles_per_day	0.0%
#8	operator_experience_years	0.0%

AI ANALYSIS & RECOMMENDATIONS

Here's a detailed analysis of your machine learning model, translated into simple, non-technical language for a business user:

Here's a detailed analysis of your machine learning model, translated into simple, non-technical language for a business user:

Machine Learning Model Analysis: Predicting (Likely Machine Faults/Performance Issues)

Your Goal: To use data to predict a specific outcome, likely classifying whether a machine is about to fail, has a fault, or is performing optimally versus sub-optimally.

Our Tool: We've used a sophisticated prediction system called a "Random Forest" model. Think of it as a committee of many smaller decision-makers (like mini-experts) that each vote on an outcome, and then the final decision is based on the majority vote. This approach makes it very robust and accurate.

---

1. Model Performance Summary (How Good Is It?)

Overall, this model is a **strong performer** and demonstrates excellent capability in predicting the desired outcome.

* **Top Choice:** Among all the different types of prediction models we tested (including K-Nearest Neighbors, Logistic Regression, Neural Networks, etc.), the **Random Forest model was the clear winner**, achieving the highest scores. This means we've picked the best tool for the job based on the data we have.
* **Highly Accurate:** The model is **correct nearly 90% of the time** (specifically, 89.7%). This is a very high level of accuracy for real-world predictions.
* **Reliable in Identifying Problems:** When it says there's a problem or a specific category, it's generally right, and it also does a great job of catching most of the actual problems that occur.

---

2. What the Metrics Mean for Your Use Case

Let's imagine the model is trying to predict "Machine Fault" (the "positive" outcome) versus "Normal Operation" (the "negative" outcome).

* **Accuracy (0.897 or 89.7%):**
* **Meaning:** This is the overall percentage of times the model makes a correct prediction, whether it's correctly identifying a fault or correctly identifying normal operation.
* **In Practice:** Out of 100 predictions, it will be right almost 90 times. This is a solid indicator of overall performance.

* **Recall (0.897 or 89.7%):**
* **Meaning:** This tells us how good the model is at catching **all the actual problems**. If there were 10 actual machine faults, how many of them did the model successfully flag?
* **In Practice:** If 10 machines truly have a fault, our model will detect nearly 9 of them (8.97, to be precise). This is crucial if missing a problem (a "false negative") is costly or dangerous. It means we're doing a good job of *not missing* critical events.

* **Precision (0.887 or 88.7%):**
* **Meaning:** This tells us how often the model is **right when it predicts a problem**. If the model says there's a fault, what's the chance it's truly a fault and not a false alarm?
* **In Practice:** If the model flags 10 machines as having a fault, almost 9 of those will genuinely have a fault. This is important for avoiding unnecessary interventions or wasted resources from "false alarms."

* **F1 Score (0.879):**
* **Meaning:** This is a balanced score that considers both Recall and Precision. It's especially useful when both catching problems and avoiding false alarms are important.
* **In Practice:** An F1 score close to 0.9 means the model achieves a strong balance – it's both good at finding most problems *and* reliable in its predictions, making it a generally trustworthy system.

ML Glossary
TERMS, METRICS & DEFINITIONS

Search terms, metrics, concepts..

ALL | REGRESSION METRICS | CLASSIFICATION METRICS | MODEL TYPES | DATA PROCESSING | MODEL CONCEPTS

22 DEFINITIONS FOUND

R² (R-Squared)
REGRESSION METRICS

Coefficient of determination that measures how well the model explains the variance in the target variable.

FORMULA
```
R² = 1 - (SS_res / SS_tot)
```

📘 INTERPRETATION
Range: 0 to 1 (can be negative for poor models). Higher is better. $R^2=1$ means perfect predictions, $R^2=0$ means the model is no better than predicting the mean.

💡 EXAMPLE
$R^2=0.85$ means the model explains 85% of the variance in wine quality scores.

MSE (Mean Squared Error)
REGRESSION METRICS

Average of squared differences between predicted and actual values. Penalizes large errors more heavily.

FORMULA
```
MSE = (1/n) × Σ(y_actual - y_predicted)²
```

📘 INTERPRETATION
Range: 0 to ∞. Lower is better. MSE=0 means perfect predictions. Units are squared, so harder to interpret directly.

💡 EXAMPLE
MSE=0.5 for wine quality means average squared error is 0.5 (quality scale²).

---

7 DEFINITIONS FOUND

Random Forest
MODEL TYPES

Ensemble method that builds multiple decision trees and combines their predictions. Reduces overfitting and improves accuracy.

📘 INTERPRETATION
Works well for both classification and regression. Handles non-linear relationships. Good default choice for many problems.

💡 EXAMPLE
Used for predicting wine quality based on chemical properties.

Logistic Regression
MODEL TYPES

Statistical model for binary/multi-class classification. Predicts probability of class membership.

📘 INTERPRETATION
Simple, interpretable, fast. Works well for linearly separable data. Good baseline model.

💡 EXAMPLE
Classifying emails as spam or not spam.

Linear Regression
MODEL TYPES

Finds best-fit straight line relationship between input features and continuous target.

FORMULA
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon$$

📘 INTERPRETATION

📖 Complete Guide

Learn how to predict anything with Machine Learning — no coding required!

This tutorial walks you through every feature step by step. Follow along and you'll train your first AI model in minutes.

📋 What You'll Learn

1 What is Machine Learning?	**2** Prepare Your Data (CSV/Excel)
3 Upload Your Dataset	**4** Clean Your Data
5 Label & Encode Data	**6** Check Data Types & Convert
7 Select Input & Output Features	**8** Choose a Model & Configure
9 Train Your AI Model	**10** Make Predictions
11 Save, Load & Manage Models	**12** Practice Datasets
★ Pro Tips & FAQ	

💡 1. What is Machine Learning?

Machine Learning (ML) is like teaching a computer to make predictions based on examples you give it. You don't write rules — the computer *figures out the rules by itself* from your data.

🍗 Simple Example: Fried Chicken

Imagine you want to predict how delicious fried chicken will taste. You collect data about past batches:

🌡 Frying Temperature + ⏱ Cook Time + 🥫 Sauce Amount + 🌶 Spice Level + 🍗 Crispiness

→ ⭐ **Predicted: Deliciousness Score**

The AI learns: "Higher temp + longer cook time + more spice = more delicious!" and can predict future batches.

---

12 Practice Datasets

Ready to practise? Download these **industry-themed synthetic datasets** and train your own models. Each dataset is designed to teach you different ML concepts — from classification to regression.

💊 Pharma — Trial Response
`Classification`

Predict whether a patient will respond to a clinical trial based on dosage, adherence, biomarkers, and demographics.

`dose_mg` `adherence_rate` `baseline_severity_score` `biomarker_A` `genotype_variant` `patient_age` `sex` `→ trial_response`

2,500 rows · 8 columns ⬇ **Download**

🏥 Health — Length of Stay
`Regression`

Predict how many days a patient stays in hospital based on triage severity, infection status, comorbidities, and BMI.

`triage_acuity_level` `infection_status` `comorbidity_index` `bed_pressure` `patient_age` `bmi` `sex` `→ length_of_stay_days`

2,500 rows · 8 columns ⬇ **Download**

💰 Finance — Loan Default
`Classification`

Predict whether a borrower will default on a loan using credit score, debt-to-income ratio, loan amount, and term.

`credit_score` `debt_to_income` `loan_to_income` `loan_amount_usd` `annual_income_usd` `term_months` `borrower_age` `→ loan_default`

2,500 rows · 8 columns ⬇ **Download**

🏗 Maintenance — Crane Failure
`Classification`

Predict crane failure within 30 days using vibration, motor temperature, hydraulic pressure, and maintenance history.

`vibration_rms_mm_s` `motor_temp_c` `hydraulic_pressure_bar` `load_ratio` `wind_speed_m_s` `cycles_per_day` `maintenance_overdue_days` `operator_experience_years` `→ crane_failure_30d`

3,000 rows · 9 columns ⬇ **Download**

🦺 HSE — Incident Prediction
`Classification`

Predict workplace safety incidents within 30 days using hazard scores, PPE compliance, overtime, training levels, and shift conditions.

`hazard_score` `ppe_compliance` `overtime_hours_week` `training_compliance` `equipment_condition_score` `shift_type` `weather_severity` `contractor_ratio` `→ hse_incident_30d`

3,200 rows · 9 columns ⬇ **Download**

🎯 How to Use These Datasets

1. Download any dataset above
2. Go to **ML Workbench** and upload the CSV
3. Click **Automatic Clean** to prepare the data
4. Encode text columns (e.g., sex, shift_type, genotype_variant)
5. Select the target column as **Output** and the rest as **Inputs**
6. Train with **Auto mode** first — then experiment with different models!

And many other functions ....