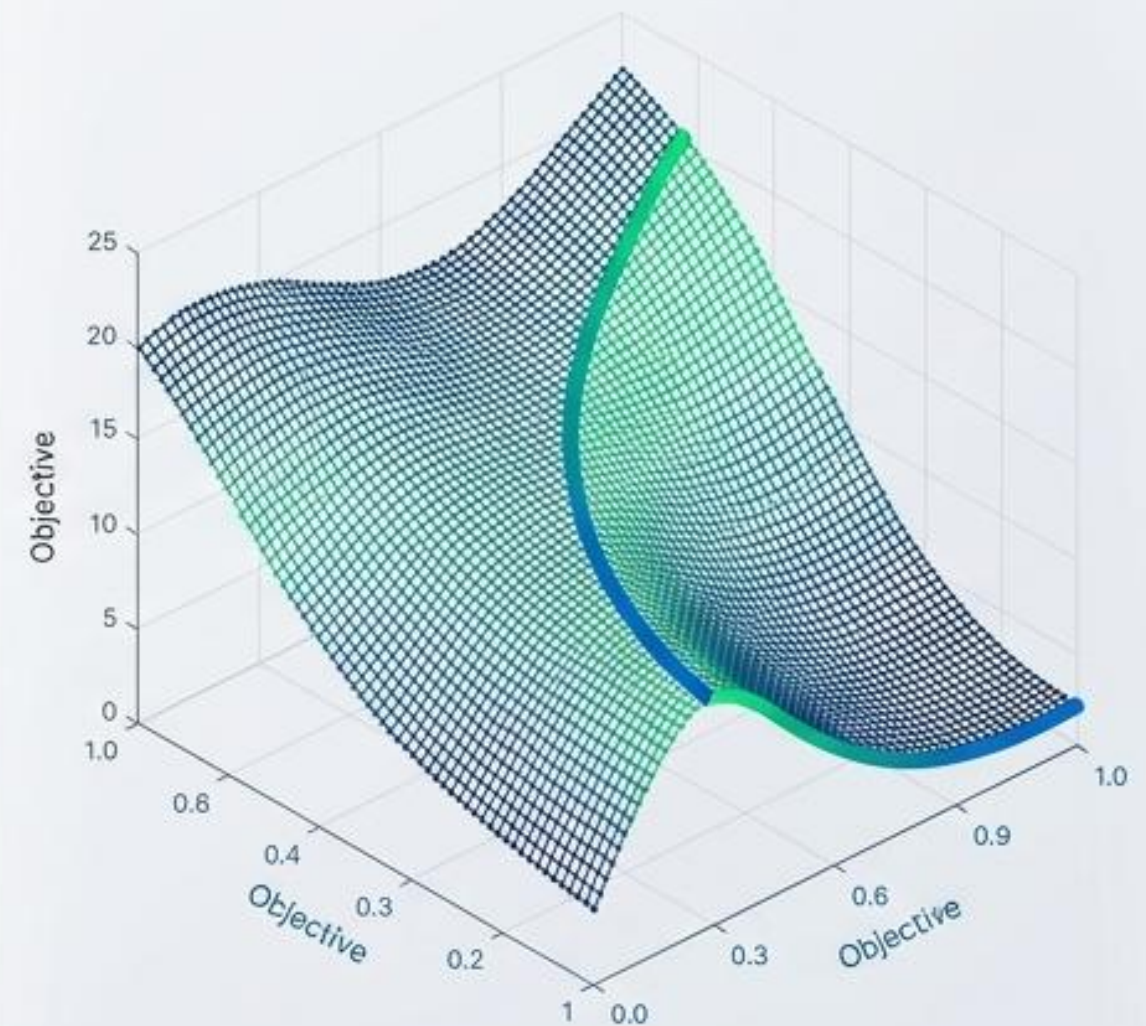




Produced by  PEDI Civil Solutions

Produced by  PEDI Civil Solutions



The Analytics Maturity Framework

Data Analysis

What happened?



Use advanced statistical analyses such as histograms, heatmaps, and descriptive statistics.

Outcome: Understanding baseline variables.

Predictive Analytics

What will happen?

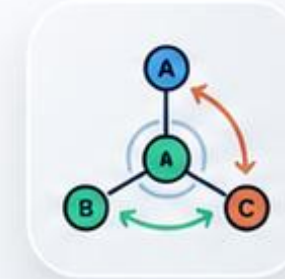


Use machine learning algorithms ranging from simple to complex.

Outcome: Forecasting targets and understanding factor influence.

Multi-Objective Optimization

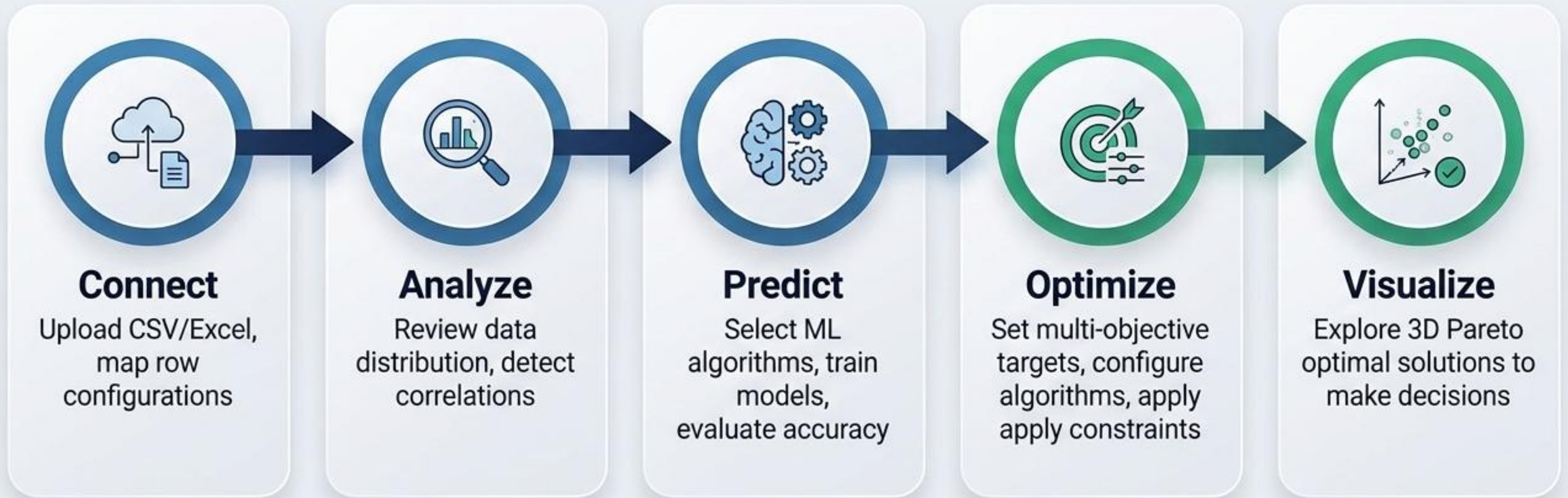
What should we do?



Scientific optimization algorithms.

Outcome: Identifying optimal trade-offs for actionable resource allocation

Overview of the main steps



The DataVision Pipeline: A chronological workflow designed to transform static spreadsheets into dynamic decision engines.

Prepare and Configure the Dataset

Influencing Variables (X)

Optimization Targets (Y)

Required
Metadata Rows

1	Variable Names								
2	Symbols								
3	Units								
4									
5	678	16.74	1009	678	...	611	13	...	1050
6	16.74	16.74	5.5	16.74	...	583	13	...	1050
7	5.5	5.74	1.5	5.5	...	583	13	...	1050
8	0	16.74	0	0	...	1099	13	...	1050
9	0	16.74	0	0	...	1099	13	...	1050
10	0	16.74	0	0	...	583	13	...	1050
11	20	16.74	0	20	...	1090	13	...	1050
12	...	...	...	...	...	...	...	...	...

Columns 1 to n

The inputs and factors that affect your business (e.g., number of employees, machine count, market price)

Columns n to m

Must be placed at the end of the dataset. These are the business goals you want to improve (e.g., power consumption, defect rate, production cost)

Step 1: Create a new project

Dashboard View: Initiating Analysis

Total Projects

Manage the entire portfolio of completed and ongoing projects.



Connect data

Upload a new CSV/Excel file or connect to an external source to perform analysis

Connect data



Total Projects

6 +12% vs last period



In Progress

6 +8% vs last period



Completed

0 +16% vs last period



Confirmation

0 +20% vs last period

All

Completed

In Progress

Search projects...

Assignee: All

Status: All

Confirmation: All

Time: All

+ Create Project

Navigate to your portfolio and initiate a new analysis pipeline.

Upload Modal: Preparing Data Entry

Create new project

Project name *

Enter project name (e.g., Q3 demand forecast)

Detailed description

Enter a short project description...

Attached data file (Excel, CSV)



Drag and drop the data file here or click to choose a file

Accepts Excel files (.xlsx, .xls) or CSV files (.csv)

View template sample

Create Project

Accepts formatted Excel (.xlsx, .xls) and CSV files (minimum recommended 200 rows)

Data Rule: Ensure missing values default to 0 and all text symbols are converted to numeric values before uploading.

Dataset Structure Configuration: Defining the Data Schema for Optimization

Map your metadata.

Tell the system exactly which row contains Variable Names, Units, and Symbols.

Dataset structure

ROW CONFIGURATION *1-indexed*

Variable name row

1

Unit row

2

Symbol row

3

Data start row

4

Design objective count

5

Indicate where the actual numeric data begins to prevent parsing errors.

Crucial Step: Input the exact number of Target Variables (Y)
- located at the **end of your sheet**. This tells the engine what needs to be optimized.



Data analysis

Explore, clean, and analyze project data.

Modules > Data analysis

Step 2: Check the data



Import data

Choose a data source



2

Configure analysis

Choose a method



3

Analysis results

Statistics & visuals



Choose another chart type

Continue →

CURRENT PROJECT

DISTRIBUTION PLAN

CHOOSE AN EXISTING DATASET

Bo_du_lieu_demo_val_lieu_san_lap_lan_bien_xay_cang (version 2).xlsx



Dataset structure

Set the rows containing variable names, units, symbols, and the number of design objectives.

TOTAL SAMPLES

151

VARIABLES

22

INPUT X

18

OUTPUT Y

4



Dataset structure

Set the rows containing variable names, units, symbols, and the number of design objectives.

TOTAL SAMPLES

151

VARIABLES

22

INPUT X

18

OUTPUT Y

4



ROW CONFIGURATION (OPTIONAL)

1-indexed

Variable name row

4

Unit row

5

Symbol row

4

Data start row

6

Design objective count

4

Total samples

151

Number of variables (columns)

22

Target variables

Cost, Time, CO₂, Turbidity

Saved models

0

View template sample

View data details

Check data

2

Specify the row numbers for variable names, symbols, and units (if any — enter 0 if not available) for the factors/objectives.

1

Click this area to review the file configuration.

3

Click "Check data" to verify whether the data structure is correct.

4

Review the data and edit or adjust it if needed.

Step 3: Data statistics and analysis

Statistical Options

▶ Run analysis

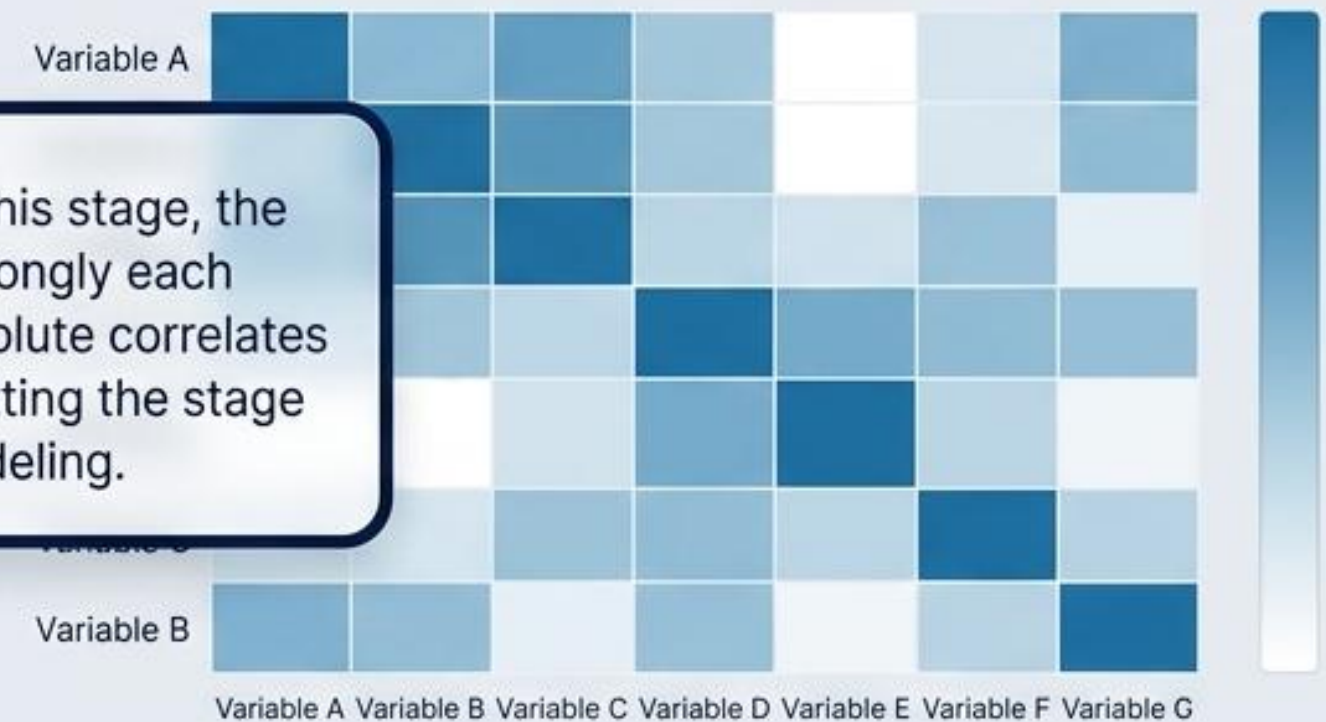


Verify Data Distribution



Influence Analysis: At this stage, the system ranks how strongly each independent variable absolute correlates with the target goals, setting the stage for accurate modeling.

Evaluate Baseline Relationships



Step 4: Review analyses and other visual charts



Import data

Choose a data source



Configure analysis

Choose a method



Analysis results

Statistics & visuals



Choose another chart type

Continue →

Statistical options

Choose the statistics you want to analyze

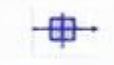
SELECT CHART
ECDF



Histogram



Density chart (KDE)



Box plot



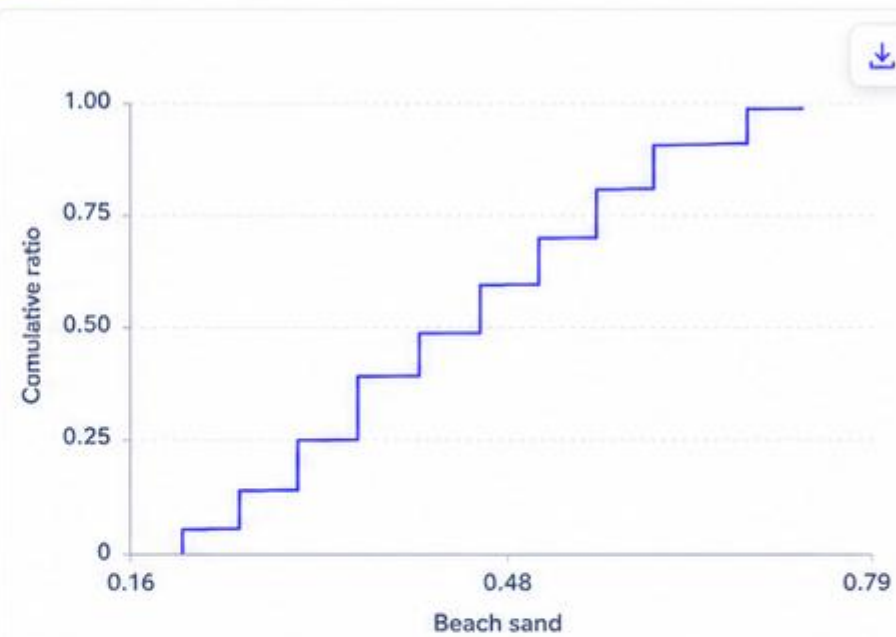
Violin plot

View more chart types

Run analysis

Reset

ECDF - Beach sand

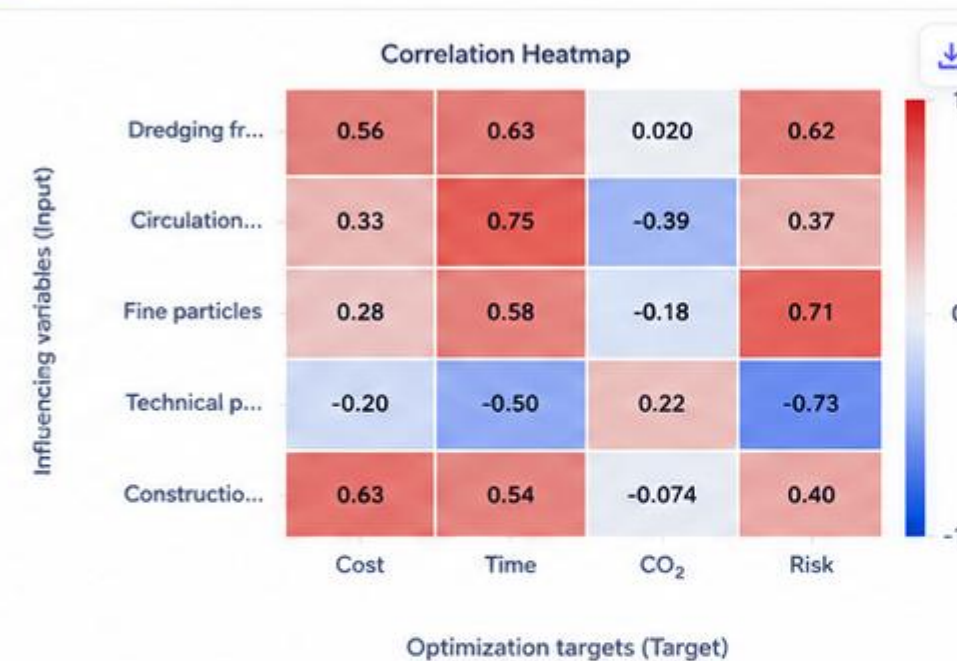


Beach sand



View details

Heatmap - Correlation



Top 5

Top 10

All (18)



Influence analysis

Rank how strongly each variable affects the optimization target by absolute correlation.

Influence analysis

Statistical analysis

Save result

1

Select other factors to explore different statistics (for example operating-time frequency or monthly revenue).

Example: In one year, how many months had profit above \$5,000?

2

Click to review how strongly each factor affects the target and which factor has the greatest influence.

3

Click to view statistical metrics.

4

Click **Save** to store the results in the report.

4

Click **Save** to store the results in the report.

5

Click **Continue** to build the predictive model.



Pipeline

- Data Collection
- Preprocessing
- Model Training
- Evaluation
- Prediction Deployment

Phase 2: Build Predictive Models

Model Selection

XGBoost	0.93	✓
Random Forest	0.89	
LightGBM	0.87	
Linear Regression	0.78	
SVR	0.72	

Model Performance

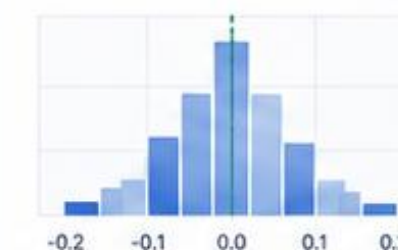


MAE	0.046
RMSE	0.087
MAPE	4.21%

Predicted vs Actual



Error Distribution



Feature Correlation



Data



Preprocessing



Model



Evaluation



Prediction

Step 1.1 View the algorithm guide

View uploaded data

Open guide

TRAINING ALGORITHM

Ridge Regression
Optimized ridge baseline

SINDy
Sparse discovery

SINDy - custom
Custom formula library

XGBoost
Gradient boosted trees

ANN
Neural baseline

View more

START TRAINING

1 Click "Open guide" to view algorithm explanations

Regression model training guide

1. Prepare data: Select the prepared project/dataset from Data analysis, the system automatically detects input X variables and output Y targets.

2. Move to build predictive model: Continue to the prediction model step to train and evaluate the regression model.

3. Choose a prediction method: Choose a suitable training algorithm such as Ridge Regression or SINDy.

4. Train and view results: Select Start training to run the model and move to the prediction results screen.

View template sample

Got it

2 Click the icon to view details

ALGORITHM DOCUMENTS

1. What are the prediction and optimization methods?

1. What are the prediction and optimization methods?

2. What is the SINDy method?

3. What is the difference between SINDy method and ordinary regression?

4. What is ANN?

5. What is the weighted algorithm?

6. What is NSGA II?

7. What is NSGA III?

8. What is SPEA2?

9. What is optimization?

TECHNICAL GUIDE TO PREDICTIVE MODELING, SINDy, AND MULTI-OBJECTIVE OPTIMIZATION

Algorithms, mathematical principles, explainability, evaluation, and DataVision integration

Expanded English edition

DataVision: predictive modeling, equation discovery, and optimization

Data layer
Experiments
Observations
Operational data
External data

Preparation
Quality checks
Scaling
Noise handling
Feature engineering

Modeling
Regression
SINDy discovery
Neural networks
Ensembles

Optimization
LP / NLP / MOP
NSGA-III
SPEA2 / Custom
Constraints

Validation loop: optimized candidates → high-fidelity simulation / field test → new data → model update

Figure 1. End-to-end architecture connecting data, predictive models, equation discovery, optimization, and decision support

Step 1.2 Build the prediction model

1 Data validation

Validate source data

Verify that the inputs and outputs are correctly interpreted by the system.

X (input) 30

Y (output) 5

The system checks the variables of the dataset to ensure they are correctly parsed.

1 Review the dataset: influencing factors – target(s) to optimize

2 Select learning algorithm

Learning algorithm

Ridge Regression

Ridge Regression

SINDy

XGBoost

PINN

Neural Network [PRO]

Gaussian Process [PRO]

Neural ODE [PRO]

Start training

2 Choose the model and continue

3 Algorithm descriptions

Ridge Regression

Well-suited for linear relationships and stable predictions.

SINDy (Sparse Regression)

Discovers governing equations from data with sparse terms.

PINN / Neural Networks

Handles complex systems with high accuracy.

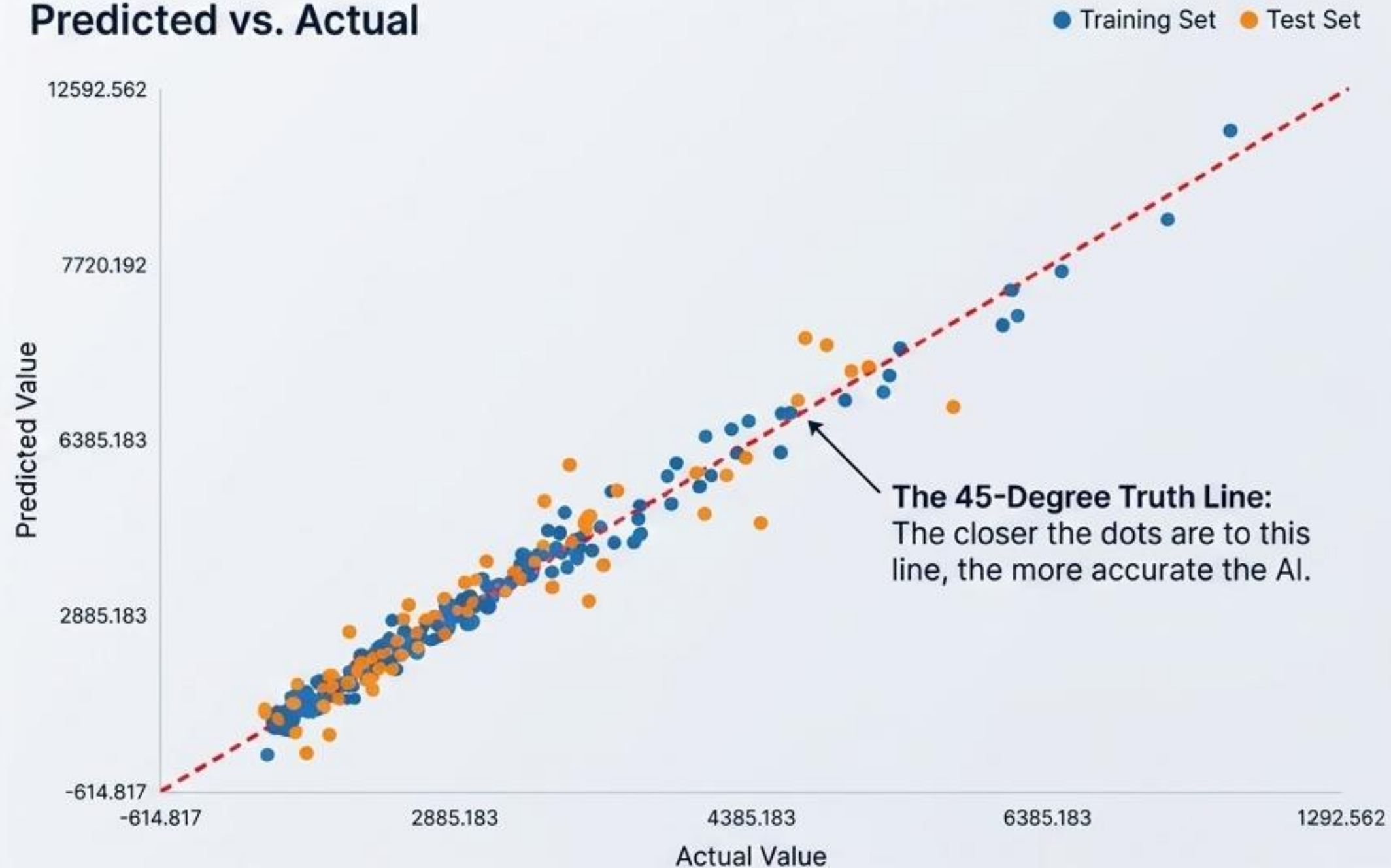
XGBoost

Excels with structured data and robust performance.

3 View details on how each factor affects the target(s) (which factors have the greatest impact on Y?)

Anatomy of a Prediction

Predicted vs. Actual



Metrics Glossary

 **R² (Fit):** **0.9647**

A score from 0 to 1. Higher is better. Measures how much target variance is explained by the model.

 **RMSE:** **380.714**

The average magnitude of your prediction errors. Lower is better.

 **MAE:** **245.125**

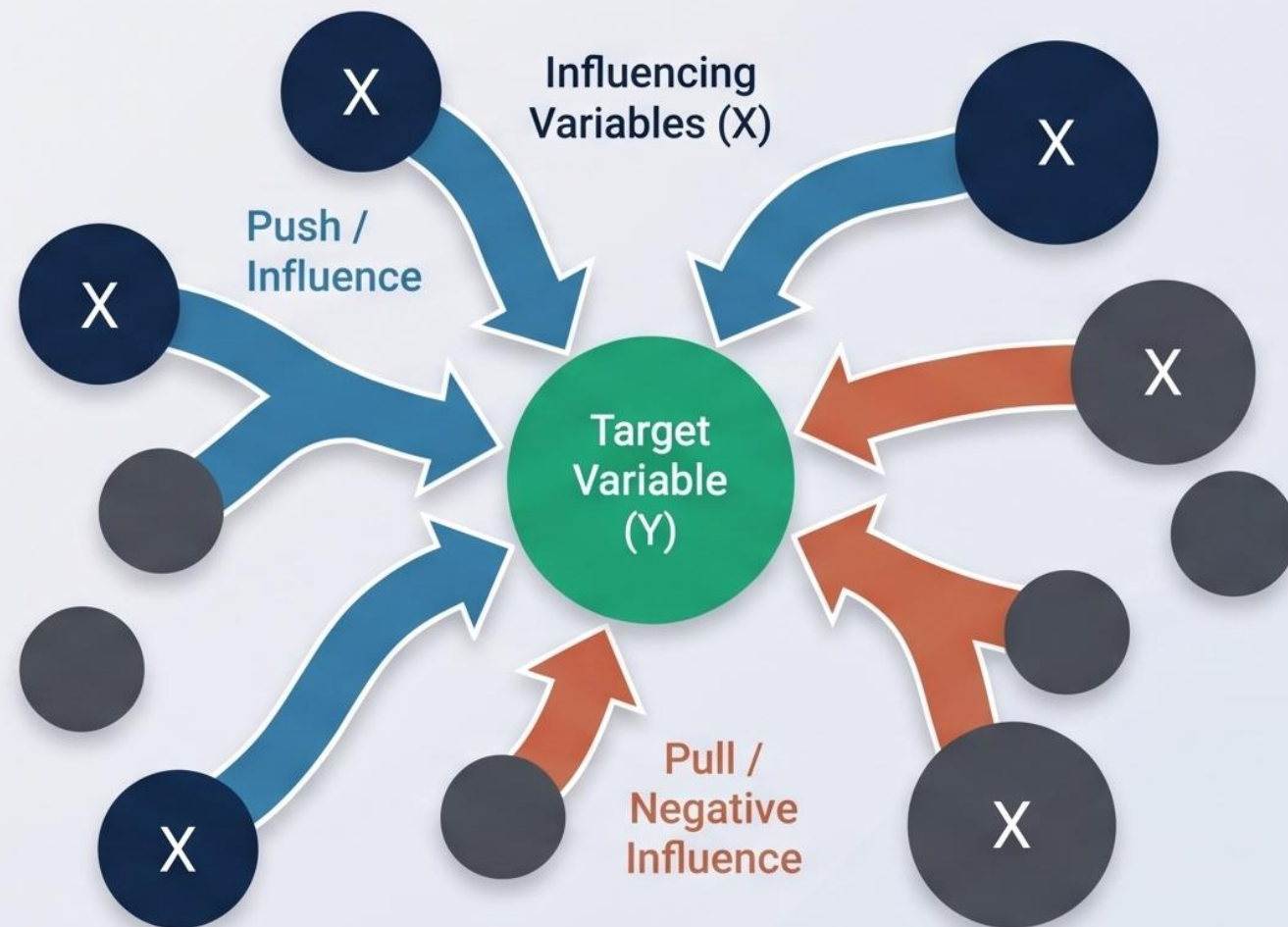
The absolute deviation from the true value.

 **MAPE:** **4.32%**

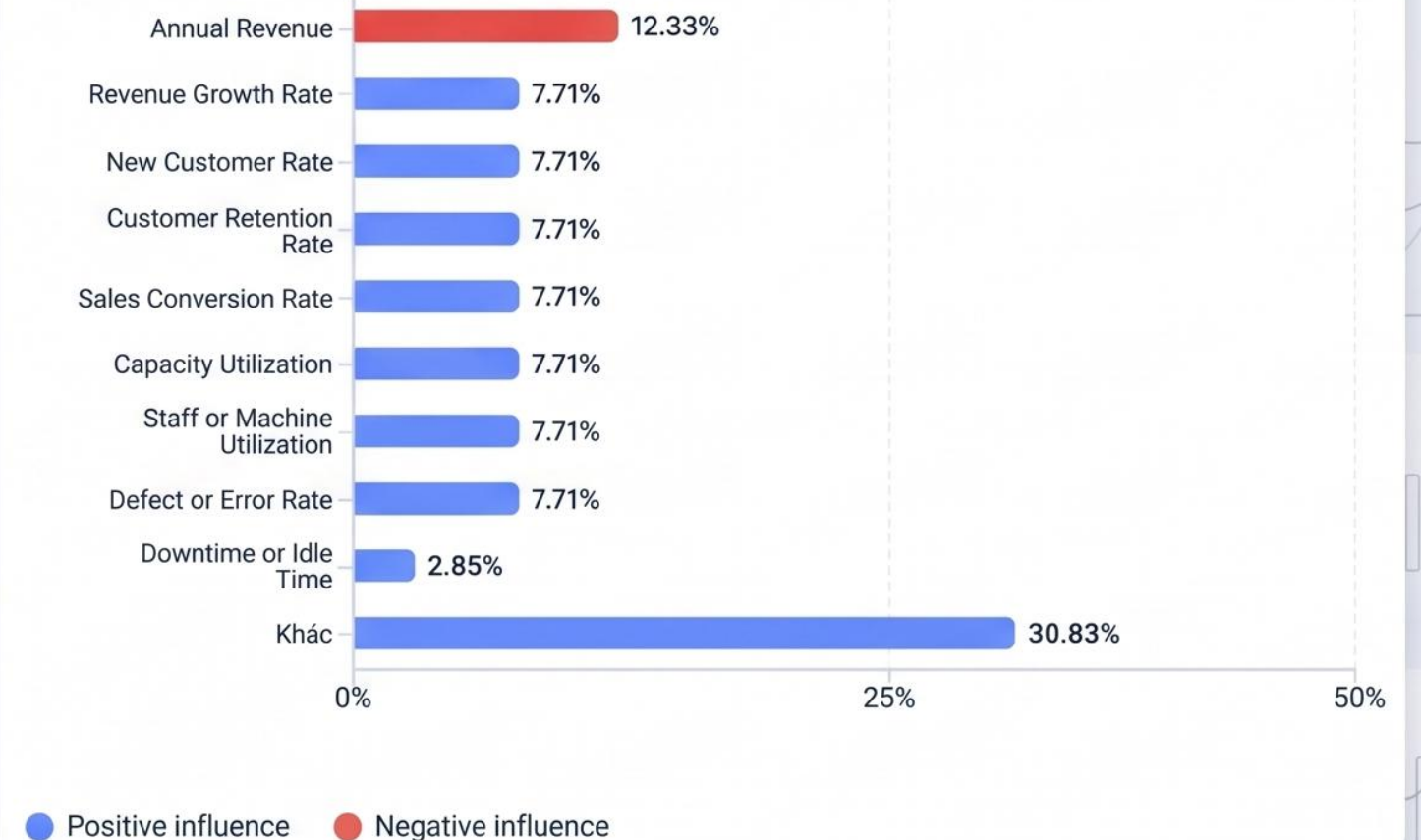
The error expressed as a clear percentage.

Step 2: Charts analyzing the influence and importance of the variables.

Influence Map



Parameter Plot



Positive Influence (Marine Blue): As the input increases, the target increases.

Negative Influence (Terracotta): As the input increases, the target decreases.

DataVision automatically maps these physics.





If the selected optimization algorithm does not produce the desired results, return to choose a different optimization algorithm.

Regression Analysis

Applied to the full dataset

← Back to previous section

↓ Save result

i Guide

↺ Change input variables (Go back)



Prepare data



Configure model



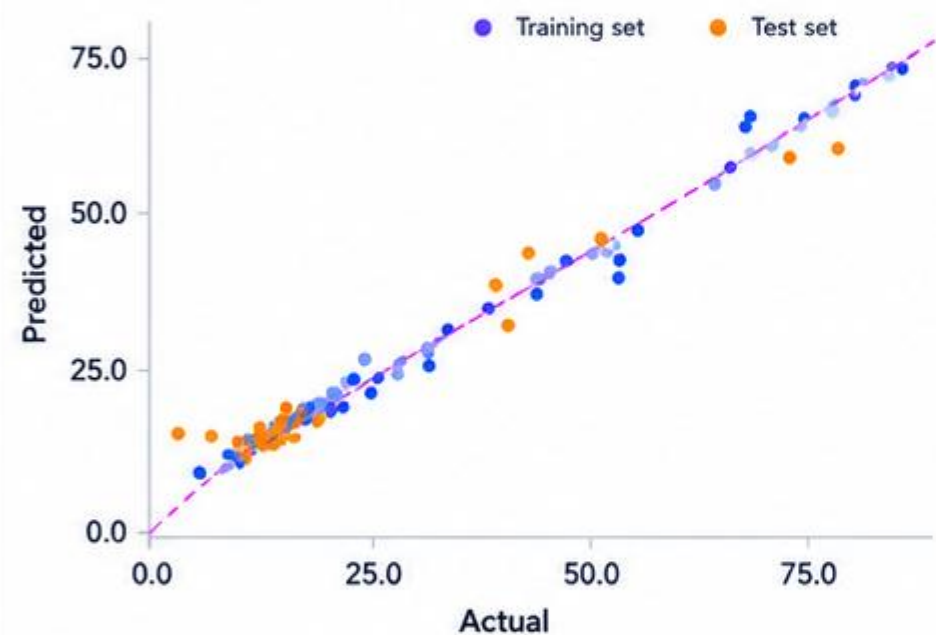
Train and evaluate



4 Interpret results

R4 - CS MPa (Predicted vs. Actual)

R^2 (Training): 0.960 R^2 (Test): 0.960 R^2 (Overall): 0.960 Adjusted R^2 : 0.958

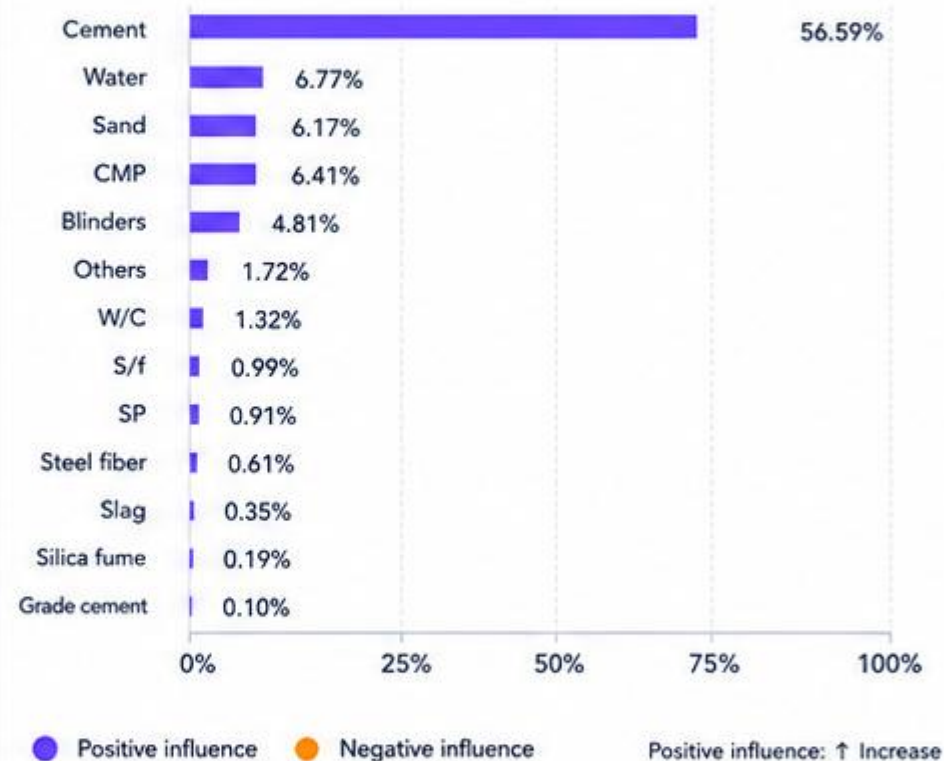


● Training dataset

Status: Operating normally



Factor Influence



This chart shows prediction accuracy. The higher the R^2 correlation coefficient, the stronger the model's predictive performance.



These charts analyze how variables influence changes in the target outcome. Expert knowledge helps interpret how each factor affects the target. For example, temperature and working time may affect employee productivity.



This chart helps evaluate the degree of influence that different factors have on the target. For example, which factors most strongly affect construction quality.

GUIDE: REGRESSION ANALYSIS & PREDICTION

Predict the project's target outcomes based on the regression model.

Regression analysis

Fit on the full dataset

Target prediction

Save result

Guide

Change input variables (Go back)

Prepare data

Configure model

Train and evaluate

4 Interpret results

Influencing values

Invest rate (%)

3.25

Land area (%)

2.1

New store count (%)

4.5

Recycle rate (%)

6.2

Discount rate (Code)

102.4

Average distance (km)

23450

Competitor rate (%)

78.5

Store capacity (m²/year)

320.5

Weather forecast 1 (cm/m)

295.3

Weather forecast 2

512.6

Run prediction

Forecast results based on trained model

Reset values

Restore default input values

Prediction results

Profit (million JPY)

125.45

million JPY

Time (days)

8.73

days

CO₂ emissions (tons)

45.62

tons

Traffic (NTU)

112.35

NTU

OPTIMIZATION PROPOSALS



Optimize profit

Adjust X1, X2, X8 to increase forecasted profit

View details



Optimize sales

Increase X5, X10 to improve sales volume

View details



Minimize risk

Tune X2, X4, X8 to reduce market exposure risk

View details

Feature importance

See more (10 variables) >

24.5%



Recycle rate (%)

18.5%



Competitor rate (%)

12.5%



New store count (%)

8.9%



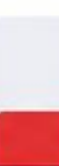
Store capacity (m²/year)

8.1%



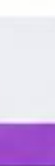
Land area (%)

6.6%



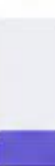
Discount rate (Code)

5.9%



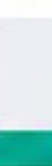
Weather forecast 1 (cm/m)

4.1%



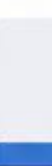
Average distance (km)

4.0%



Invest rate (%)

1.8%



Weather forecast 2

2

Enter the project input information.



1



Click here to proceed to target prediction — forecast the project outcomes in advance.

3



Press Run to make the prediction; the results will appear below.

4



The factors that strongly influence the project targets.



GUIDE: PERFORM MULTI-OBJECTIVE OPTIMIZATION

Follow these 3 simple steps to choose models and start multi-objective optimization.



EVALUATION METRICS (Cost)

R² All

Fit on the full dataset

0.8574

RMSE All

Error on the full dataset

4.136

MAE (Mean absolute error)

Mean absolute deviation

3.140

MAPE (Mean absolute percentage error)

Mean percentage deviation

1.242%



Multi-objective optimization settings

1



Click to
continue with
optimization

2



Choose the most
appropriate model
for each objective

3



Click Continue to
proceed with
multi-objective
optimization



Multi-objective optimization settings

Choose the trained model/formula to pass to multi-objective optimization.

+ Add

Undo



Cost

Target variable (Y)

CHOOSE FORMULA/MODEL

MODEL FORMULA

Linear (🏆 Best) (R² Test = 0.801) ▾

Cost = -62.7019 - 6.85573 × Pitch (%) + ...

Time (days)

Target variable (Y)

CHOOSE FORMULA/MODEL

MODEL FORMULA

Cubic (🏆 Best) (R² Test = 0.823) ▾

Time (days) = -47.3697 - 1.44665 × Pitch (%) + 36...

CO₂ Emissions (tons)

Target variable (Y)

CHOOSE FORMULA/MODEL

MODEL FORMULA

ExpCombined (🏆 Best) (R² Test = 0.893) ▾

CO₂ Emissions (tons) = exp(3.83501 + 0.00959443 × Pitch (%))...

Turbidity (NTU)

Target variable (Y)

CHOOSE FORMULA/MODEL

MODEL FORMULA

Linear (🏆 Best) (R² Test = 0.893) ▾

Turbidity (NTU) = 448.465 + 2.87389 × Pitch (%) + ...

Cancel

Accept and proceed to optimization 🚀



Optimization Workflow

- 01 Objectives
- 02 Constraints
- 03 Algorithm Setup
- 04 Optimization
- 05 Evaluation

Optimization Performance

Pareto Solutions	128
Objectives	4
Constraints	12
Run Time	00:14:32

Optimization Algorithms

NSGA-II	0.94	✓
SPEA2	0.88	
PSO	0.85	
MOEA/D	0.82	

Criterion: Hypervolume (↑)

Phase 3: Multi-objective Optimization



Objectives



Constraints



Algorithm

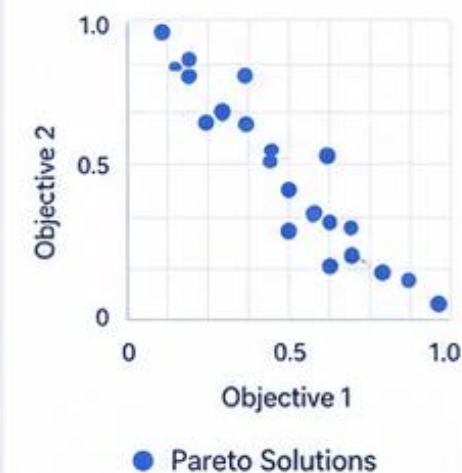


Pareto

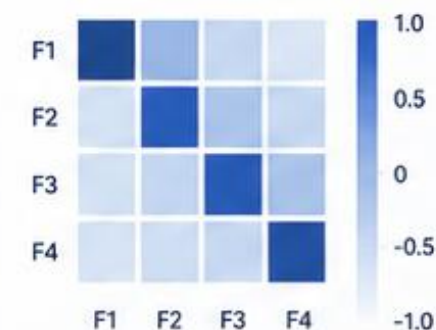


Decision

Pareto Front



Objective Trade-offs



Step 5: Multi-Objective Optimization

Step 1: Define the Direction

Select **Min** or **Max** for each objective.

SET DESIGN OBJECTIVES

[Select default](#)

#	OBJECTIVE NAME	VARIABLE (Y)	OPTIMIZATION DIRECTION	WEIGHT	THRESHOLD / TARGET VALUE
1	Profit (Y1)	Y_time	Max	0.25	Min - Max
2	Net Profit Margin (Y2)	Errors) Y_doc_error	Max	0.25	
3	Cost-to-Revenue Ratio (Y3)	Y_rating	Max	0.25	
4	Operational Efficiency Score (Y4)	Y_debt	Max	0.25	Min - Max

Assign weights if one business goal is more vital than the others.

Step 2: Size the Population

Define **Population Size** and **Generations**.

ALGORITHM AND POPULATION SETTINGS

OPTIMIZATION ALGORITHM

NSGA-II (Multi-objective genetic algorithm)

POPULATION SIZE

100

EVALUATION LIMIT

350000

Total Solutions Explored = Population Size × Generations.
(Default: 100 pop × 10 gen).

Algorithm Note: DataVision utilizes NSGA-II, an industry-leading genetic algorithm designed specifically to evaluate thousands of multi-objective scenarios simultaneously.

Step 6: Set Up Advanced Constraints Before Optimization (Optional)

Variable Range Limits

E.g., You only have between 10 and 50 machines available.



Discrete Values

E.g., Shipping routes can only be option 1, 2, or 3, not fractions.



Sum Constraints

E.g., Total budget for Variables A + B + C cannot exceed \$100,000.



CONSTRAINT CONFIGURATION

e.g., $A1 + A2 \leq A3$ Add

☐ Activate

POL_code POD_code Origin_to_transit_days
Transit_time Delay_days Route_type Volume_m3
Volume_m3 Container_type
Freight_rate_usd_m3 Trucker Transport_cost
Inland_km Inland_days Labor Handling_days

+ - * / < > = () Clear draft

Constraints ensure the AI doesn't just find a mathematically perfect solution, but a practically executable one.

GUIDE: REVIEW PARETO OPTIMAL SOLUTIONS

Explore the list of Pareto optimal solutions and inspect each one to view the required project inputs / settings needed to achieve target outcomes.

PARETO OPTIMAL SOLUTIONS

✓ Total Pareto solutions: 126

 Solution Proposals

 Export results

 Save result

 Continue



STT	D_INDEX	Revenue (BILLION VND)	Profit (BILLION VND)	CO ₂ Reduction (t)	Production (t)	NTU (NTU)	Water (m ³)	Energy (t oil eq.)	Waste (t)	Landfill (ha)
1	0.41	241.50	145.95	13233.03	52.55	0.64	0.08	0.16	0.16	0.10
2	0.45	242.26	148.31	12958.27	53.20	0.64	0.06	0.16	0.16	0.10
3	0.47	242.72	149.79	13436.78	50.10	0.66	0.17	0.13	0.13	0.09
4	0.47	243.29	147.33	13379.24	50.48	0.73	0.07	0.20	0.19	0.10
5	0.48	241.58	149.47	13038.25	54.03	0.67	0.17	0.18	0.17	0.10
6	0.48	241.49	151.09	13765.40	47.88	0.73	0.07	0.20	0.19	0.10
7	0.49	242.71	152.57	13473.18	48.16	0.67	0.23	0.12	0.12	0.08
8	0.54	241.46	148.16	13396.56	54.90	0.47	0.09	0.12	0.12	0.07

1



Click here
to review each
optimal solution.

SOLUTION PROPOSAL

Solution 1 (D_Index: 0.41)

< Previous

Next >

Total: 126 Pareto solutions



EXPECTED RESULTS

Achieved results of optimization objectives

Revenue
Unit: Billion VND **241.50**

Profit
Unit: Billion VND **145.95**

CO₂ Reduction
Unit: t **13233.03**

Production
Unit: t **52.55**

Target values are calculated based on the regression model.



SOLUTIONS TO ACHIEVE OBJECTIVES

Project input values/settings required to achieve the desired outcomes

NTU
Unit: NTU **0.64**

Water
Unit: m³ **0.08**

Energy
Unit: t oil eq. **10.18**

Waste
Unit: t **0.08**

Labor (Empl.)
Unit: People **10.28**

Landfill
Unit: ha **0.08**

Recycle Rate
Unit: % **0.33**

Materials
Unit: ton **22.01**

Total Investment
Unit: k\$ **12780.16**

IRR
Unit: % **81.77**

Select another solution from the dropdown or navigate buttons to compare input factors.

Close

2



These are
the target results /
project outcomes
that can be achieved.

3



These are
the required
project inputs /
settings
to achieve the
desired results.

RUN LOG - OPTIMIZATION RUN DETAILS



✓ Optimization complete!

Run time: 00:00:40 | Estimated time: 00:06:20 | Completed at: 16:00:52 7/8/2026

1. Conditions and constraints

- Independent variables (X): Input 1, Input 2, Input 3, Input 1, Input 2, Input 3, Safety Impact Factor, Safety Impact Factor, Environmental Impact Factor
Input 1, Input 2, Choke, Quality Factor, Time Factor, Environment

- Target variable (Y): Safety Impact Factor (Safety Impact Factor)

- Target limits (Y Range / Design Goals):

Safety Impact Factor Unbounded

- Variable limits (Individual XI Range):

Input 1 € [929000, 44974000]
Input 2 € [83, 1901]
Input 3 € [1, 5]
Input 1 € [929000, 44974000]
Input 2 € [83, 1901]
Input 3 € [1, 5]
Safety Impact Factor € [5, 25]
Safety Impact Factor € [1, 5]
Environmental Impact Factor € [929000, 44974000]
Input 1 € [83, 1901]
Input 2 € [1, 5]
Input 3 € [929000, 44974000]
Choke € [1, 5]
Quality Factor € [5, 81.4]
Time Factor € [5, 71.5]
Environment € [5, 71.5]

- Generations: 10 generations
- Population size: 100 individuals

PARETO OPTIMAL SOLUTIONS

Total Pareto solutions: 1

Export results

Save result

Continue



No.	Dom. Index	Objective (Y): Safety Impact Factor	Input 1	Input 2	Input 3	Input 1	Input 2	Input 3	Cost	Quality	Time	Environment
1	0.00	39.52	44540314.65	620.13	1.33	25455228.98	1192.82	2.51	23.96	85.41	18.73	23.96



After reviewing the Pareto results, click "Continue" to move to the next step.

Show all 1 results

Use horizontal/vertical scroll bars in the table to view all data

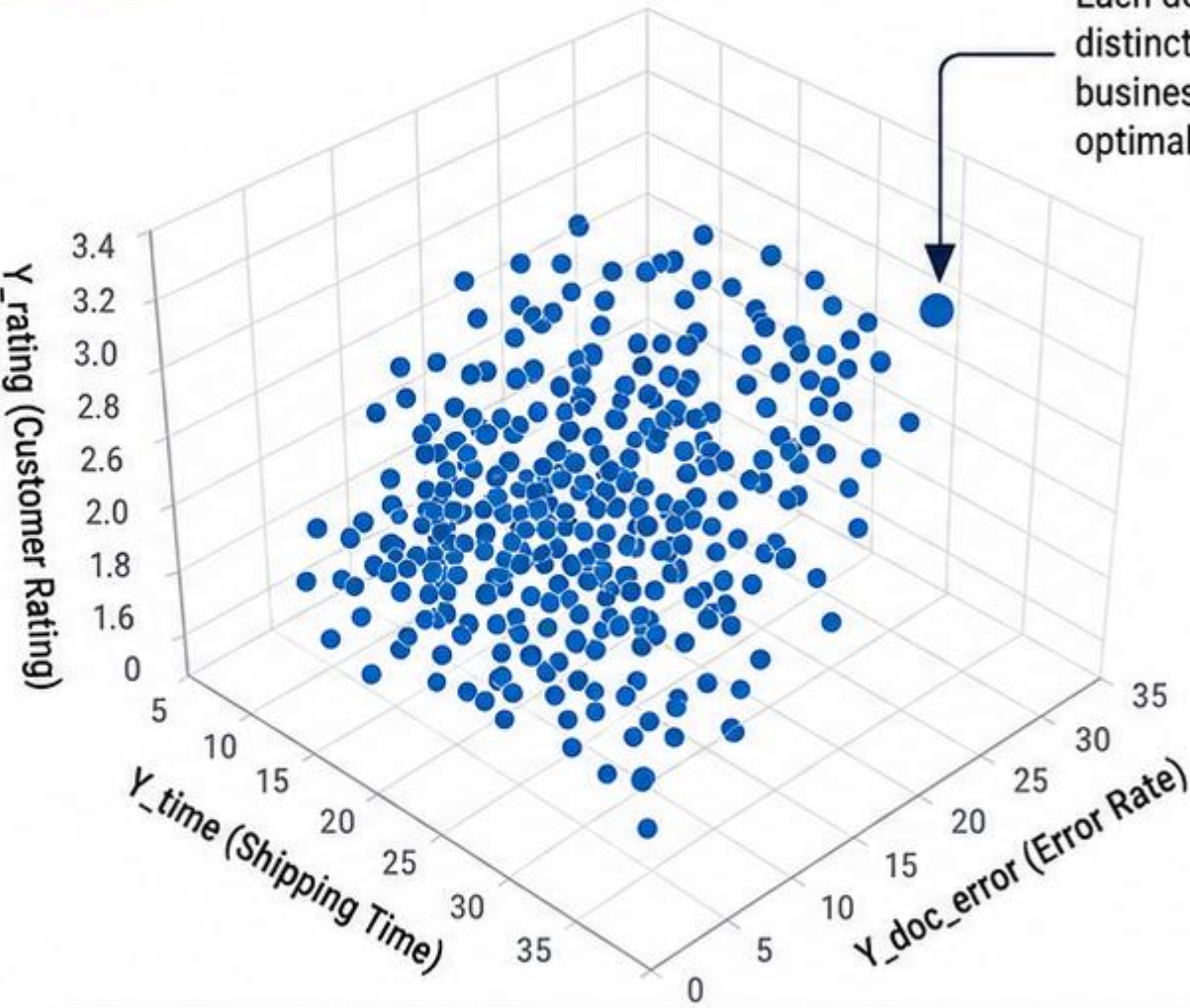


Step 7: Export Optimization Results and Continue to Explore Visual Analytics

Premium technical slide

3D Scatter Chart

3D SCATTER CHART



Pan Zoom Fit Camera Reset Auto rotate ☒

◆ PARETO OPTIMAL SOLUTIONS

Total Pareto solutions: 200

No.	D_INDEX	Y_TIME SHIPPING TIME (DAYS)	Y_DOC_ERROR DOCUMENT ERROR RATE (ERRORS)	Y_RATING CUSTOMER RATING (STARS)
1	1	0.81	24.35	1.41
2	2	0.81	24.35	1.41
...	...	...	...	...
4	13	0.89	25.77	1.39
5	14	0.87	25.38	1.39
6	16	0.88	24.30	1.41
7	17	0.89	24.32	1.40
8	17	0.88	24.60	1.39
9	13	0.89	25.77	1.39

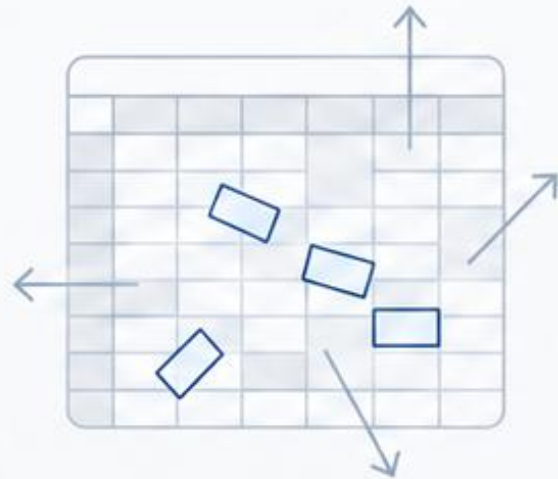
 **Actionable Instruction:** Use the visualizer to pan, zoom, and isolate clusters. Select a specific point on the 3D frontier to reveal the exact input variable configuration required to achieve that outcome.

Value Realization Journey



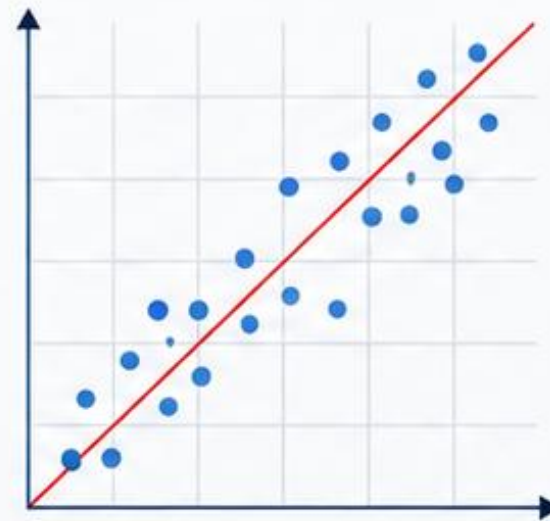
Explore result visualizations and switch to chart views like the examples shown.

Hindsight (The Data)



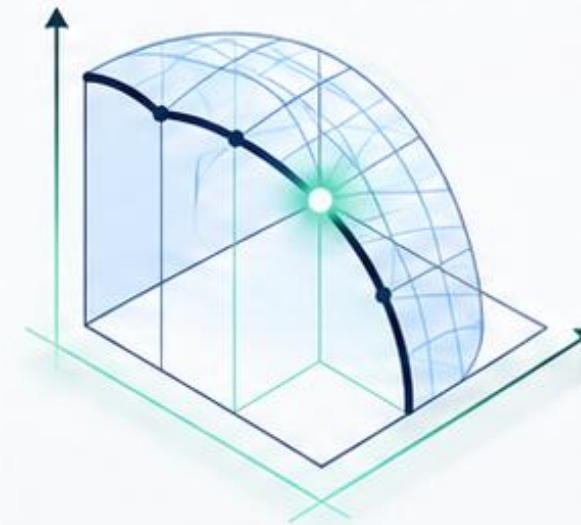
You provided fragmented historical records.

Foresight (The Model)



DataVision mapped the governing physical and business laws.

Insight (The Action)

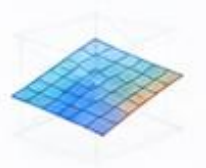


You possess a prescriptive blueprint for resource allocation, risk mitigation, and target achievement.

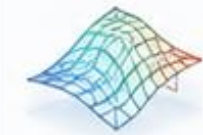
Chart views



Scatter 3D



Surface 3D



Surface Mesh 3D



Scatter 2D



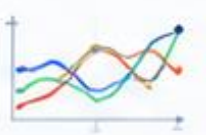
Line chart



Bar chart



Radar Chart



Compare lines

Stop analyzing what happened. Start optimizing what's next.

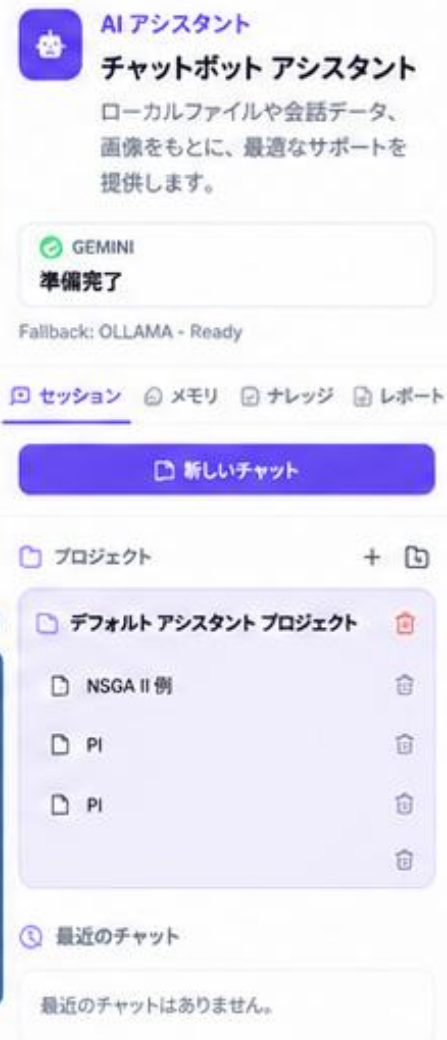
サポート部門へのお問い合わせ



ご不明な点があれば
AIに質問するか、
/admin と入力して
サポートチームへ
連絡できます。
このプロセスには
時間がかかる場合が
ありますので、
しばらくお待ちください。

AI チャットボット アシスタント インターフェース

データの分析、プロジェクトの管理、インサイトの取得をインテリジェントにサポートします。



コンテキストと履歴:
過去のセッション、
最近のチャット、
プロジェクト情報を表示し、
ユーザーがいつでも
会話を継続できるように
します。

データの会話
fast でチェックのようにファーストを使用します。

こんにちは！私は DataVision アシスタントです。画像や
ファイルをアップロードするか、チャットを開始してください。
プロジェクトについて何を知りたいか教えてください。
画像、ファイル、テキストチャットの確認に対応しています。

チャットボットのご利用案内:
歓迎メッセージを書いてください。
ユーザーがサポートを受けるために
チャットボットを使用し、会話の
履歴を保存し、AI 生成エラーを
アップグレード 改善するための
提案を行います。

データについて入力してください...