

# Sezi: Machine Learning Software

## What problem does Sezi solve?

Sezi addresses the general problem of moving from raw data to information, and from information to knowledge (data → information → knowledge).

- **Data quality & heterogeneity** – data acquisition, cleaning, preprocessing to create consistent feature spaces;
- **Predictive uncertainty** – regression and time-series forecasting to estimate future targets and confidence;
- **Population/behavior structure** – classification and clustering to discover segments and rule sets for differentiated policies;
- **High dimensionality** – dimensionality reduction and feature selection to enhance signal-to-noise and runtime;
- **Resource allocation under constraints** – mathematical optimization to recommend feasible, near-optimal plans;
- **Outliers & rare events** – anomaly detection and statistical testing to surface atypical patterns before they degrade outcomes;
- **Algorithmic choice & tuning** – model selection and AutoML to systematically pick and calibrate learners/pipelines;
- **Model staleness in production** – model monitoring and drift detection to trigger retraining and keep performance stable.

## What can Sezi do?

- **ML Model Lifecycle Management** – versioning, tracking, experiment management;
- **Automated ML (AutoML)** – train-test-deploy pipelines with minimal manual intervention;
- **Model Monitoring & Drift Detection** – monitoring deployed models for performance degradation and data distribution shifts;
- **Simulation** – repeatable runs with varying feature vectors and initial parameters to evaluate alternative strategies before execution;
- **Data Acquisition, Cleaning & Preprocessing** – handling missing data, normalization, standardization, integration of heterogeneous data sources;
- **Classification** – supervised learning for categorical outcomes;
- **Regression** – supervised learning for continuous outcomes;
- **Clustering** – unsupervised grouping of data points;
- **Dimensionality Reduction**
- **Optimization** – mathematical and heuristic optimization for parameter tuning and planning tasks;
- **Anomaly Detection** – identifying outliers, fault detection;
- **Statistical Analysis** – descriptive statistics, hypothesis testing, correlations;
- **Model Selection** – comparing algorithms and parameter sets systematically.