

ANALYTICS & MACHINE LEARNING WITH BIOVIA PIPELINE PILOT

VALIDATED DATA SCIENCE TOOLS FOR EVERY STEP
IN THE SCIENTIFIC DATA ANALYSIS CYCLE
DATASHEET

Analytics is the key to unlocking meaningful patterns in data. In science based industries it requires awareness of complex data types to process large chemical, genomic, proteomic, textual and image data streams. BIOVIA Pipeline Pilot's capabilities harness multiple disciplines to help organizations understand complex systems and decipher everything from molecular behavior to optimal production processes.



Simplify your data science workflow. The Analytics and Machine Learning Collection for Pipeline Pilot gives you the tools for everything from data ingestion, cleaning and exploration, to model building, validation, deployment, optimization, and design of future experiments – all in a single environment.

BIOVIA PIPELINE PILOT ANALYTICS AND MACHINE LEARNING BENEFITS

It's Fast and Easy

- Build on a robust framework for model building, validation and deployment
- Utilize a comprehensive set of pre-implemented prediction algorithms or those implemented by in-house experts
- Data pipelining architecture scales to enormous data sets
- Use code-optional drag and drop, or write your own scripts in R or Python

Purpose-built for Science

- Our algorithms are designed to work with various numerical, chemical, biological, textual, and image data types
- Employ custom, domain-specific analyses with a library of curated scientific component collections

Validated Models

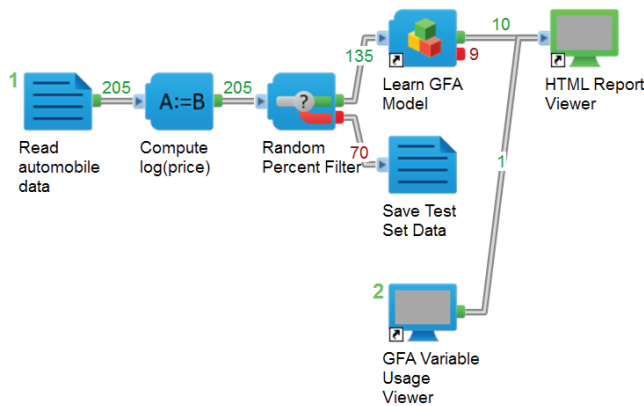
- Build better models with our built in validation techniques
- Easily interpret model results with interactive visualizations
- Increase your odds of success with tools that ensure models are applied properly when making predictions

Multiple Model Deployment Options

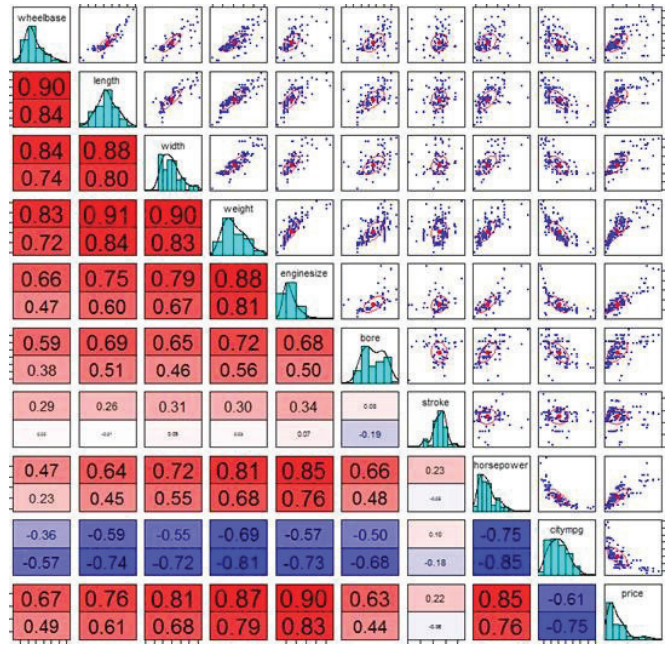
- All models are automatically saved in Pipeline Pilot for easy re-use, sharing or deployment
- Deploy models across BIOVIA and 3rd party applications or as web services to achieve faster, more efficient processes to drive innovation

Key Capabilities

- Apply any of 15+ machine learning (ML) methods to your scientific and engineering data
- Merge, join, characterize, and clean your data sets
- Perform exploratory analysis, including PCA, clustering, and multi-dimensional data visualization
- Build fast, scalable Bayesian classification models
- Use the GFA method's genetic algorithm for variable selection and building regression ensemble models
- Build accurate, easy-to-use RP Forest regression and classification models
- Use R-based ML methods such as support vector machines, neural networks, and XGBoost without writing R scripts
- Employ the ML framework for cross-validation, hyperparameter tuning, and variable importance assessment for any type of model
- Use regression and classification model evaluation viewers to assess and compare model test set performance
- Use built-in applicability domain measures and error models to assess sample-specific prediction confidence
- Apply Pareto optimization to multi-objective optimization problems



An example machine learning protocol



Even simple characterization analysis can yield valuable insight. Our R Pairs Plot displays pairwise relationships within multivariate data and helps you visualize strongly or weakly correlated variables

Our 3DEXPERIENCE Platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the 3DEXPERIENCE Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 170,000 customers of all sizes in all industries in more than 140 countries. For more information, visit www.3ds.com.



3DEXPERIENCE®