Victory Analytics

Data Analysis Tool by Machine Learning

Overview

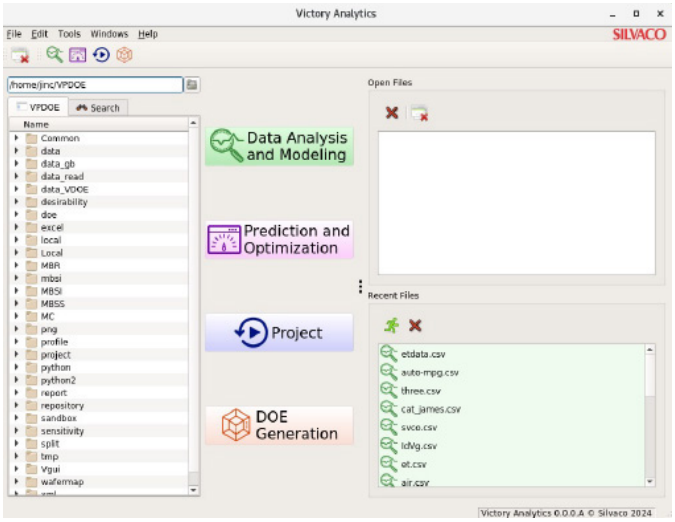
Victory Analytics is a machine learning-based analytics tool created especially for the technology industry's research, development, and manufacturing. Data management and visualization, feature selection and training, modeling, and optimization components make up this system. The adaptability of the software tool enables it to cover both module level in-depth data analysis such as electrical device characterization, process flow optimization, and fab tool recipe optimization and factory level yield improvement such as Fault Detection & Classification (FDC) and Statistical Process Control (SPC).

Key Features

- Multivariate data analysis using a combination of advanced statistics, neural networks, and decision trees algorithm.
- Capabilities for managing input data, such as detecting outliers and missing data imputation
- Support multi-level classification analysis driven by the classical statistics and machine learning algorism
- Advanced Design of Experiments algorithm such as computer-generated D-optimal DOE
- State-of-the-art data filtering and feature selection techniques based on the decision forest algorithm
- Various sensitivity algorithms including Sobol, decision forest, etc.
- Victory Analytics can conduct time series analysis including AR/MA as well as seasonality analysis
- An integrated visualization tool that offers data insight (Chart Designer)
- Provide a python interface that enables complete end-user automation
- Stand-alone product on Windows and Linux OS

Applications for Victory Analytics

- Manufacturing processes: Identify bottlenecks in manufacturing processes, improve quality control, and reduce costs
- Device/Circuit design and optimization: Engineers make better decisions by providing them with insights into their data. Increases the operation margin by the utilization of systematic data training
- Product Yield improvement: Quality control of process specification and/or target metrics. Data-driven decision making

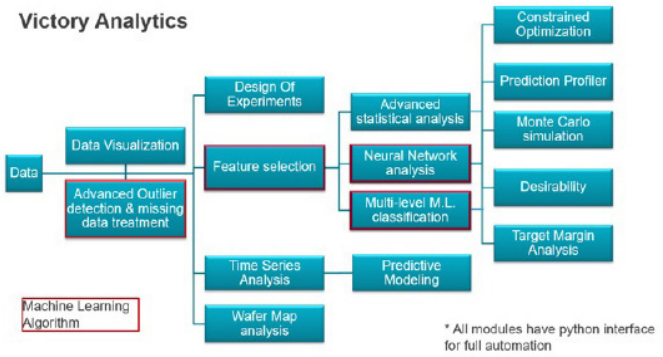


VictoryDoE Main View

Data Analysis and Modeling

Victory Analytics involves collecting, cleaning, and analyzing data to identify trends, patterns, and relationships. It is an iterative process that involves a sequence of steps and requires general understanding of your input and output variables. It is recommended to set up the correct workflow:

- Dataset generation: Generate dataset for modeling including identifying categorical variables, duplicated inputs, etc. Transformation and subset of data if necessary
- Regression term selection: Selection of predictors and regression terms for modeling
- Model build: Build regular or M.L. regression model.
- Model analysis: Model score, sensitivity, and data trend analysis

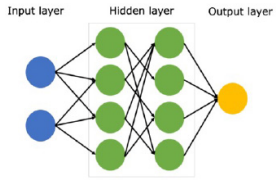


Advanced Filtering and Feature Selection

- The workspace is to manage simulation projects, which are organized like directory structures
- VictoryDoE supports multiple workspaces
- At each project, users can assign a flag and attach comments
- All essential file management tools are provided: copy (clean), paste, archive, clean project, lock, etc.
- Project view allows users to do search and sort based on the file name, time, flag, file size, etc.
- Shared workspace is useful in the collaborative workflow in which multiple users share projects
- Simulation examples are available through default example workspace

Neural Network Parameters

Victory Analytics Neural Network is based on the supervised learning configuration and Feed Forward Architecture. Having a non-linear activation function in neurons does makes a N.N. non-linear, which is the biggest advantage of NN regression over conventional polynomial regression method. The figure below is the schematic diagram of N.N. configuration in Victory Analytics.



Neural network configuration:

Architecture:
- Feed Forward Architecture

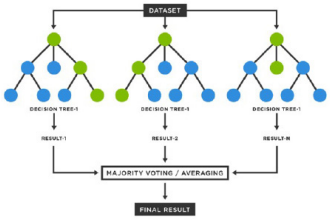
Hyper parameter:
- Number of hidden layers
- number of neurons per hidden layer
- Number of Epochs

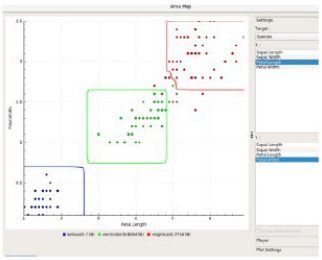
Activation functions:
- Linear, Sigmoid, HyperbolicTangent, Soft Plus, Soft Sign, Exponential Linear, etc

Control:
- MinMax, Standardization
- Loss function control
- Adaptive Moment Estimation optimization
- L1/L2 regularization
- Train/Validation/Test groups

Multi-level Classification by Decision Forest Algorithm

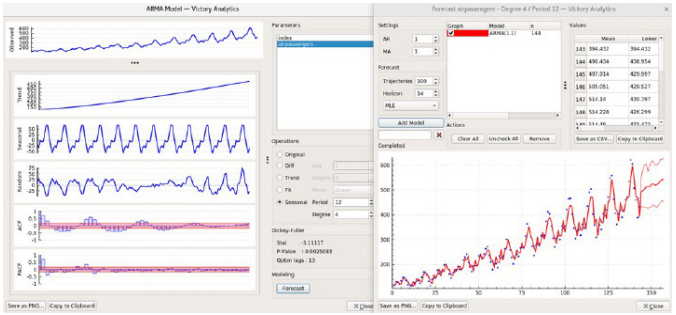
Victory Analytics provides multi-level classification in which the target variable has more than two categories. The engine behind is the decision forest algorithm. It produces multiple decision trees, each one trained on a slightly different subset of the data then take a majority vote for classification. The figure below shows the schematic diagram of decision forest algorism and the final results showing 3-level classification.





Time Series Analysis

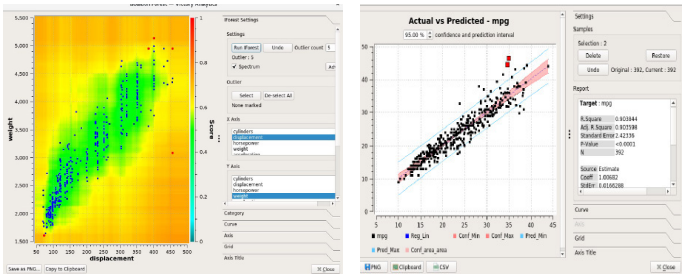
Time series analysis is a method for studying data collected at specific points in time and uncover patterns and trends within that data, as well as forecasting future values. Victory Analytics is based on ARMA (Autoregressive Moving Average), a statistical method used for time series analysis and forecasting. It combines two simpler models, AR (Autoregressive) and MA (Moving Average), to capture the influence of past values and random errors on a future value in the time series.



Advanced Filtering: iForest and Actual vs Prediction Method

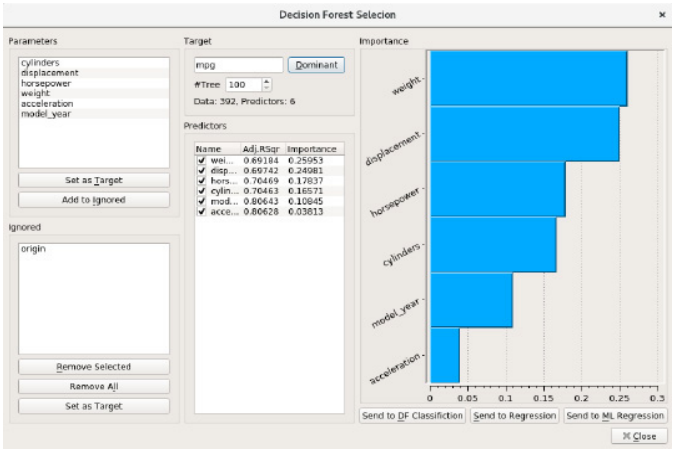
Isolation Forest is a state-of-the-art anomaly/outlier detection algorithm based on decision tree technique in Machine Learning. It was built because anomalies are the data points that are “few and different”. Victory Analytics provides iForest algorithm to effectively detect the anomality.

In addition, there is a bottom-up approach in which the model is completed and weed out the outlier from the Actual-Predict graph. In this approach, users can select the outliers from the trend distribution and enhance the data model.



Feature Selection by Decision Forest Algorithm

When building a random forest, each decision tree splits the data based on features that predict the target variable. By tracking how often a feature is used for splitting across all the trees in the forest, Victory Analytics reports how important that feature is for the overall model.



Optimization and Desirability

Victory Analysis provides various optimization algorithms. The choice of optimization options is attributed by various conditions such as non-linearity of the model and number of inputs and outputs. The optimization algorithms are:

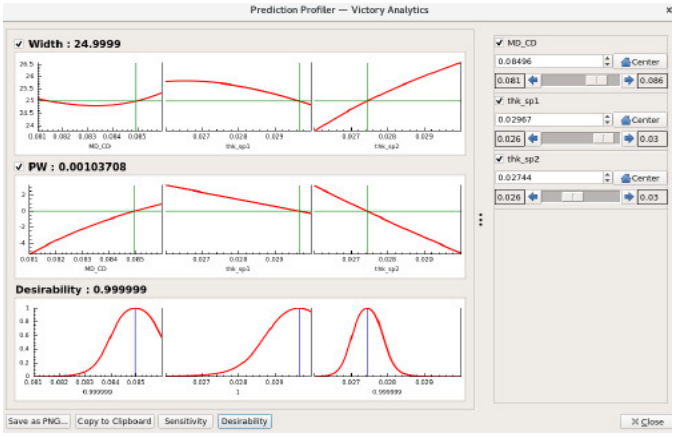
Local Optimization Algorithms

- Levenberg-Marquardt
- Hooke-Jeeves

Global Optimization Algorithms

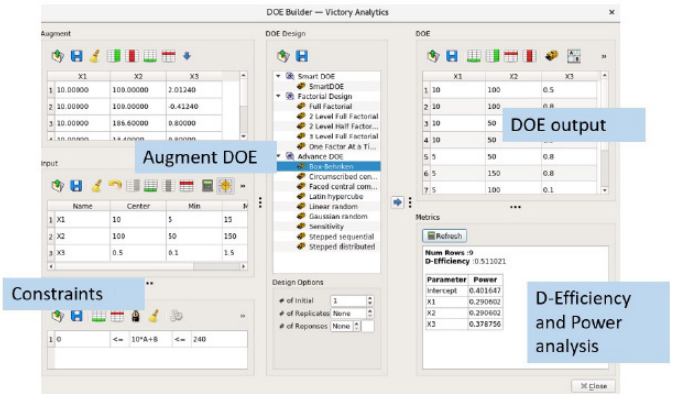
- Simulated Annealing
- Parallel Tempering
- Genetic Algorithm
- Differential Evolution

In conjunction with optimization algorithm, Victory Analytics provides Desirability analysis. It is a data analysis technique that quantifies the multiple output target into a single index. It is predicated on the assumption that there is an ideal set of values for the data and that any divergence from this ideal set of values is bad.



Design of Experiments

- Generate DOE table from various DOE algorithms
- Option to add Augmented and Constraints DOE features
- Provides state-of-the-art DOE algorithms such as computer-generated DOE from D-optimization algorithm
- Visualization of DOE table
- DOE builder supports Fraction of Design Space (FDS) plot
- Power analysis of DOE is conducted in the DOE builder



Monte-Carlo Simulation for Cp/Cpk Analysis

Monte-Carlo simulation is performed based on the mean and sigma value of the input variables. The output distribution after more than 100k simulation represents the potential distribution in the real world. From MC simulation user can achieve

- Cp/Cpk estimation
- Fail rate against USL/LSL
- Various distribution functions other than normal distribution are provided

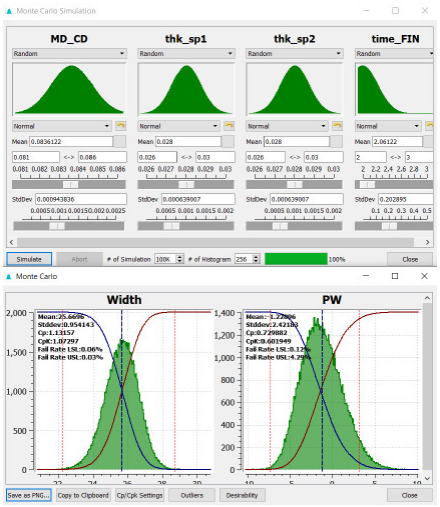
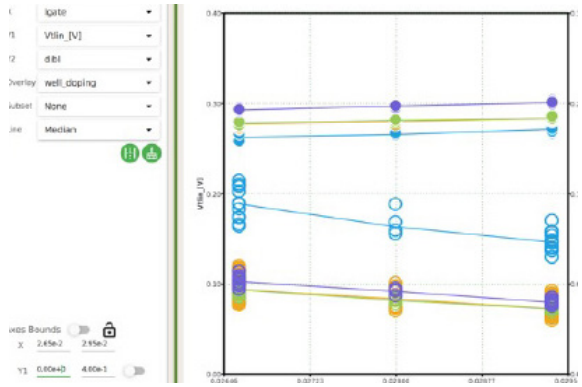


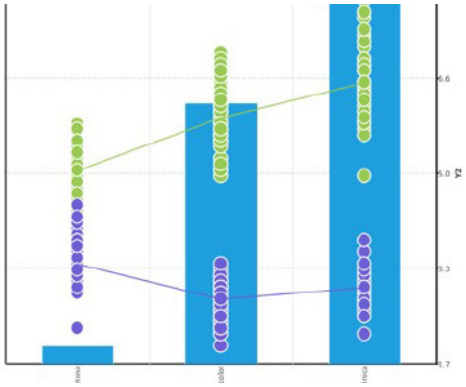
Chart Designer

Chart Designer is a data visualization tool that offers insight into the dataset.

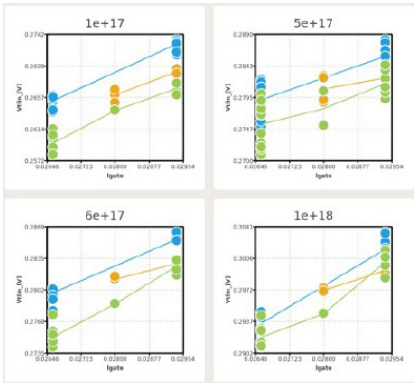
- X-Y plot with with X-axis as a string value
- Bar chart
- Scatter plot
- Box plot



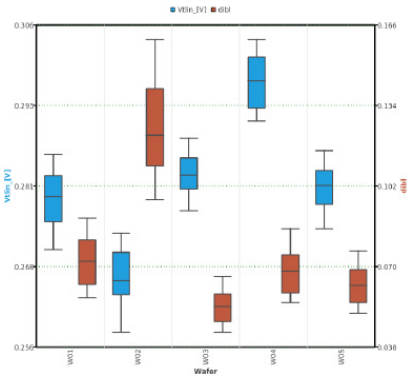
(a)



(b)



(c)



(d)

(a) X-Y plot with displaying Y1 and Y2

(b) Overlay of scatter and bar chart

(c) Multiple X-Y plot based on the subset of the data

(d) Box plot