Smarten
Advanced Data Discovery
Powered by ElegantJ BI

Concept Manual
Assisted Predictive Modeling

Business Intelligence & Advanced Data Discovery

www.Smarten.com

Document Information	
Document ID	Smarten-Concept-Manual- Assisted-Predictive-Modeling
Document Version	Version 1.0
Product Version	Version 5.0
Date	28-August-2018
Recipient	NA
Author	EMTPL

© Copyright Elegant MicroWeb Technologies Pvt. Ltd. 2018. All Rights Reserved.

Statement of Confidentiality, Disclaimer and Copyright

This document contains information that is proprietary and confidential to EMTPL, which shall not be disclosed, transmitted, or duplicated, used in whole or in part for any purpose other than its intended purpose. Any use or disclosure in whole or in part of this information without the express written permission of EMTPL is prohibited.

Any other company and product names mentioned are used for identification purpose only, may be trademarks of their respective owners and are duly acknowledged.

Disclaimer

This document is intended to support administrators, technology managers or developers using and implementing Smarten. The business needs of each organization will vary and this document is expected to provide guidelines and not rules for making any decisions related to Smarten. The overall performance of Smarten depends on many factors, including but not limited to hardware configuration and network throughput.

Contents

1 About this document

This manual explains the concepts required to use Assisted Predictive Modeling features in Smarten Advanced Data Discovery Suite.

1.1 Scope and Organization of Topic Areas

The allocation of memory for the JVM is specified using -X options when starting WildFly (these options may depend upon the JVM that you are using; the examples here are for the Sun JVM).

Chapter 2	Introducing ElegantJ BI - Smarten
Chapter 3	Introducing Assisted Predictive Modeling
Chapter 4	SmartenInsight Process Overview
Chapter 5	Types of Algorithm Techniques
Chapter 6	Sampling
Chapter 7	Example
Chapter 8	Product and Support Information

1.2 Conventions Used

This manual uses typographical conventions in the text to help you distinguish between the names of files, instructions, and other important notes that are relevant during installation. For example:

- Important notes are indicated in a different font color as shown in the example below.

> Note:
> Trend can be applied to a Bar chart, Line chart, Combined chart, and an Area type of Visualization.

- References to documents are highlighted as below:

Reference: **Self-Serve Data Preparation (SSDP) - Concept Manual >Sampling > Sampling-Simple Random Sampling**

2 Introducing ElegantJ BI - Smarten

ElegantJ BI is a full-stack business intelligence tool that employs the "Smarten" approach to Advanced Data Discovery. The solution is composed of a comprehensive set of tools, including Self-Serve Data Preparation, Smart Visualisation, and Plug n' Play Predictive Analytics. These tools are designed to democratize advanced analytics and transform business users into citizen data scientists.

Self-Serve Data Preparation
Self-Serve Data Preparation allows business users to perform data preparation on their own without the assistance of IT staff or data analysts. Users are not restricted by complex tools or forced to wait for IT to deliver crucial data. Guided by smart suggestions and auto-recommendations, business

users can prepare, blend, and transform data and create analysis-ready data quickly and accurately without assistance rather than waiting for central metadata prepared by IT.

Smart Data Visualization

Intuitive Smart Data Visualization tools suggest the best options for visualizing and plotting a particular set or type of data, based on the nature, dimensions, and trend of data so that business users can easily select the appropriate method to clearly and quickly visualize data in a way that is meaningful to the task.

Plug n' Play Predictive Analysis

Assisted Predictive Modeling allows business users to explore predictive algorithms and models without the skill or knowledge of a data scientist or statistician. The solution considers the underlying data and use case, suggests the best-fit algorithm, and then displays output results and data visualization accompanied by an interpretation using simple human language.

3   Introducing Assisted Predictive Modeling

Every organization must plan and forecast results. If the enterprise is to succeed, it must strive for accuracy and identify trends and patterns in the market and industry that will help it predict future results, plan for growth, and capitalize on opportunities.

Predictive analytics for business users leverages machine learning and assisted predictive modeling to help users achieve the best fit and ensure they use the most appropriate algorithm for the data they wish to analyze. With these tools, users can explore patterns in data and receive suggestions to help them gain insight on their own without having to depend on IT or data scientists. The enterprise can provide the tools needed at every level of the organization, with tools and data science for business users that are sophisticated in functionality and easytouse for users at every skill level.

SmartenInsight provides predictive modeling capability and auto-recommendations and auto-suggestions to simplify use and allow business users to leverage predictive algorithms without the expertise and skill of a data scientist. The assisted predictive modeling platform is suitable for business users. These tools allow the organization to apply predictive analytics to any use case using forecasting, regression, clustering, and other techniques to analyze an infinite number of use cases, including customer churn, and planning for and targeting customers for acquisition, identifying cross-sales opportunities, optimizing pricing and promotional targets, and analyzing and predicting customer preferences and buying behaviors.

4   SmartenInsight Process Overview

The process of SmartenInsight starts with selecting the Algorithm Technique and data that has to be analyzed. Users can analyze data from Datasets as well as Cubes. The machine learning capabilities of the system automatically selects the best-fit algorithm to create the model for the data and provides the result and data visualization along with the interpretation in a simple language. Users can fine-tune the models, configure visualizations, and personalize them as per their specific requirements. These Assisted Predictive Modeling objects are called SmartenInsight, and these objects can be viewed as stand-alone interactive objects. Users can export SmartenInsight objects in JPG, PDF, and PNG formats.

Typical steps involved in creating a SmartenInsight are shown below:

* Select the Algorithm Technique

- Select the Dataset / Cube
- Select the target and predictor variables
- Apply data filters
- Analyze the model with data visualization and interpretation
- Fine-tune the model
- Customize the visualization
- Apply sampling or change model parameters
- Use SmartenInsight as a stand-alone object andapply the model

5   Types of Algorithm Techniques

Shown below is the list of Algorithm Techniques:

Algorithm Technique	Algorithm
Classification	
	Decision tree
	K-Nearest Neighbor Classification
	Binary Logistic Regression
	Multinomial Logistic Regression
	Support Vector Machine
	Naive Bayes Classification
Clustering	
	K-means Clustering
	Hierarchical Clustering
Association	
	Apriori
Correlation	
	Pearson Correlation
	Spearman Correlation
Forecasting	
	Holt Winters Forecasting
	Auto-Regressive Integrated Moving Average : ARIMA
	Multiple ARIMA
	Auto-Regressive Integrated Moving Average : ARIMAX
	Multiple ARIMAX
Regression	
	Simple Linear Regression
	Multiple Linear Regression
Hypothesis Testing	
	Independent t-test
	Paired t-test
	Chi-squared test
	ANOVA test
Descriptive Statistics	

6 Sampling

Assisted Predictive Modeling provides users with the facility to use sample data instead of the whole data. The size of the sample depends upon the size of data.

SAMPLING

6.1 Sampling—Auto

The auto mode of sampling automatically applies sampling and creates a model with a sample data instead of the whole data. It limits a particular number of records that should be considered for sample data. The size of the sample depends upon the size of data. Users cannot change this limit.

6.2 Sampling—Manual

The manual mode of sampling allows users to change the limit of records that should be considered for a sample. Users can also select the method of sampling.

Users can create samples by using two types of sampling methods:

➢ Simple random sampling
➢ Stratified sampling

Please refer to the following document for more information on the sampling methods.

> Reference: **Self-Serve Data Preparation (SSDP) - Concept Manual >Sampling >Sampling-Simple Random Sampling**
> Reference: **Self-Serve Data Preparation (SSDP) - Concept Manual >Sampling > Sampling-Stratified Sampling**

7 Example

7.1 Regression

The Regression Algorithm helps predict change in one variable based on change in one or more other variables.

Let us consider data containing information about daily flights originating from and landing at various airports in the USA along with weather conditions to predict the delay in arrival of flights by using Regression. The delay in arrival of flights is based on weather conditions at the origin and destination of the flights.

- **Step 1:** Select the target and predictor variables

Target variable	Predictor variables
Arr_Delay	Dep_Delay
	Origin_PrecipitationPreviousHourInches
	Origin_SnowfallInches
	Dest_PrecipitationPreviousHourInches
	Dest_SnowfallInches

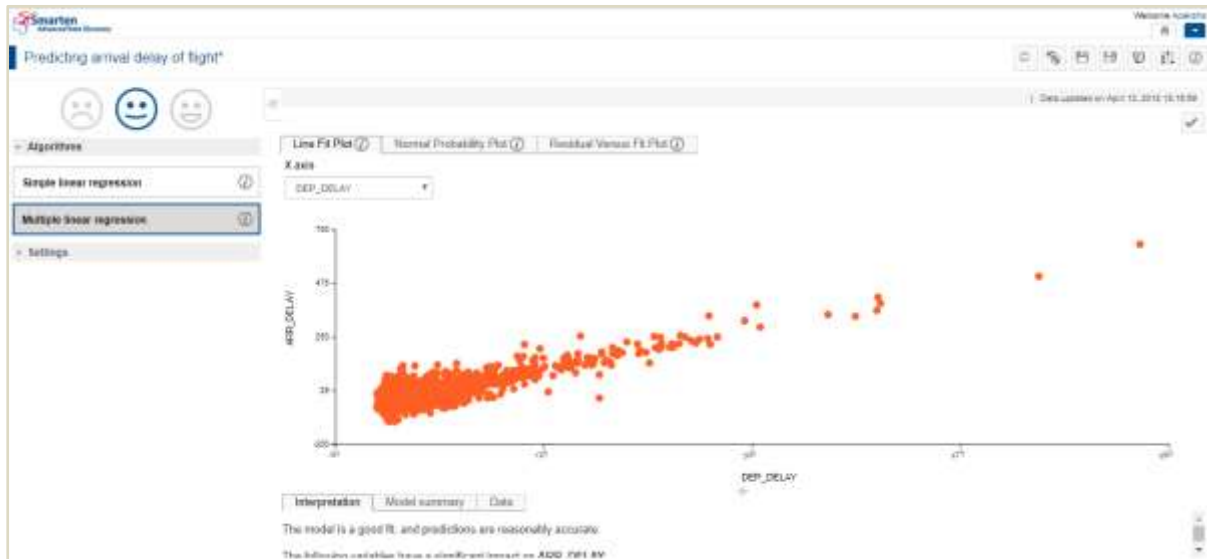Shown below are the target and predictor variables for the Dataset:
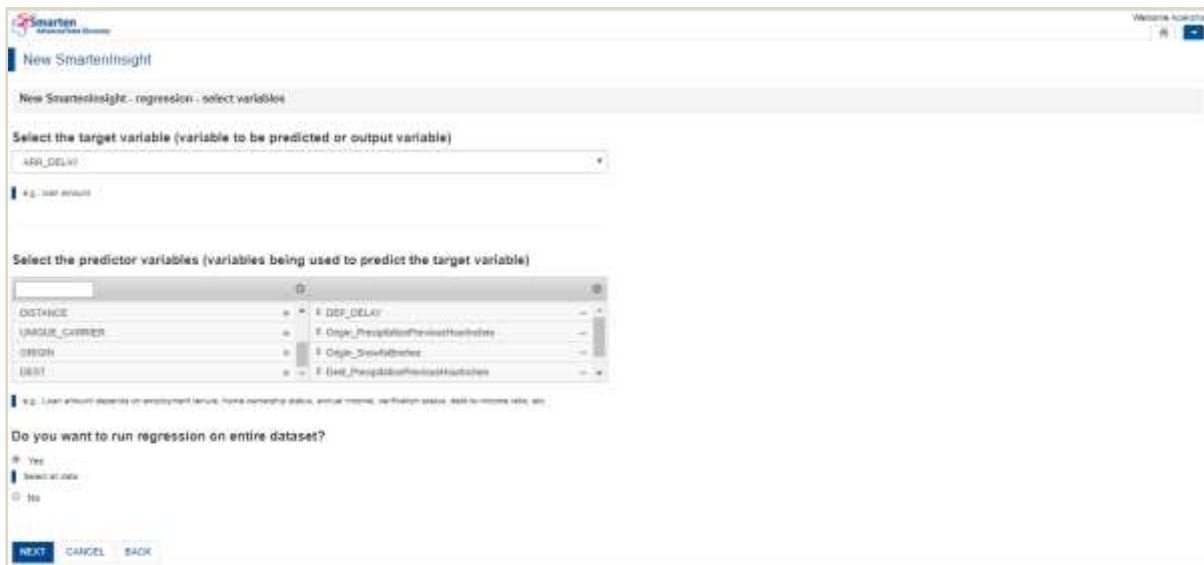
TARGET AND PREDICTOR VARIABLES

- **Step 2:** Analyze the model with data visualization and interpretation

The system selects the "Multiple Linear regression" using machine learning as the best-fit algorithm based on the data and Algorithm Technique and provides the result as shown below:

RESULT—DATA VISUALIZATION

The system also interprets the result in a simple language as shown below:

Multiple linear regression algorithm has been chosen and applied on your data with R Square = 0.665.

The model is a good fit, and predictions are reasonably accurate.

The following variables have a significant impact on **ARR_DELAY**:

- Origin_SnowfallInches
- Dest_PrecipitationPreviousHourInches
- DEP_DELAY

The following variables do not have a significant impact on **ARR_DELAY**:

- Origin_PrecipitationPreviousHourInches

CLOSE

RESULT—INTERPRETATION

From the above result, we can conclude that the delay in arrival of flights depends most on departure delay, precipitation conditions at the destination, and snowfall at the origin of flights.

- **Step 3:** Apply the model

Users can apply a new set of values to the model and view the result as shown below:

APPLY THE MODEL

RESULT

8   Product and Support Information

Find more information about ElegantJ BI-Smarten and its features at www.smarten.com

Support: support@smarten.com

Sales: sales@smarten.com

Feedback & Suggestions: support@smarten.com

Support & Knowledgebase Portal: support.smarten.com