



**Altair Knowledge Seeker®**

**Altair Knowledge Studio®**

**Altair Knowledge Studio® for Apache Spark™**

**Version 2020.1 Release Notes**

March 2020

## Table of Contents

1.	Supported Hardware and Software Platforms.....	3
1.1	Overview: Supported Operating Systems .....	3
1.2	System Requirements .....	3
2.	Version Compatibility.....	9
3.	New Features .....	10
3.1	New Features in Version 2020.1 (Released on 31-Mar-2020) Compared to 2020.0.....	10
3.2	New Features in Version 2020.0 (Released on 10-Dec-2019) Compared to 10.9 .....	10
3.3	New Features in Version 10.9 (Released on 16-Aug-2019) Compared to 10.8.4 .....	11
3.4	New Features in Version 10.8.4 (Released on 21-Jun-2019) Compared to 10.8.3 .....	12
3.5	New Features in Version 10.8.3 (Released on 24-May-2019) Compared to 10.8.2 .....	13
3.6	New Features in Version 10.8.2 (Released on 08-Mar-2019) Compared to 10.8.1.....	14
3.7	New Features in Version 10.8.1 (Released on 31-Jan-2019) Compared to 10.8 .....	14
3.8	New Features in Version 10.8 (Released on 11-Dec-2018) Compared to 10.7.3 .....	14
3.9	New Features in Version 10.7.3 (Released on 24-Oct-2018) Compared to 10.7 .....	15
3.10	New Features in Version 10.7 (Released on 27-Mar-2018) Compared to 10.4.....	15
3.11	New Features in Version 10.4 (Released on 01-May-2017) Compared to 9.6 .....	16
3.12	New Features in Version 9.6 (Released on 04-Mar-2016) Compared to 9.5.....	17
3.13	New Features in Version 9.5 (Released on 29-Sep-2015) Compared to 9.4.5.....	18
3.14	New Features in Version 9.4.5 (Released on 17-Jun-2015) Compared to 9.4 .....	20
4.	Known Issues in Version 2020.1 .....	21
5.	Fixed Issues .....	23
5.1	Issues in v2020.0 fixed in v2020.1 .....	23
5.2	Issues in v10.9 fixed in v2020.0 .....	24
5.3	Issues in v10.8.3 and v10.8.4 fixed in v10.9 .....	26
5.4	Issues in v10.8.2 Fixed in v10.8.3 .....	28
5.5	Issues in v10.8.1 Fixed in v10.8.2 .....	28
5.6	Issues in v10.8 Fixed in v10.8.1 .....	28
5.7	Issues in v10.7.3 Fixed in v10.8 .....	29
5.8	Issues in v10.7 Fixed in v10.7.3 .....	31
5.9	Issues in v10.7 Fixed in 10.7 Update of April 26, 2018.....	31
5.10	Issues in v10.4 Fixed in v10.7 and 10.7 Update of April 26, 2018 .....	32
6.	Client Service and Support: Contacting Altair .....	35
6.1	Additional Resources.....	35
6.2	Reporting Issues and Requesting Product Enhancements.....	35

All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means without prior written permission of Altair Engineering, Inc. (Altair). KnowledgeSEEKER, Knowledge Seeker, KnowledgeSTUDIO, Knowledge Studio, KnowledgeREADER, KnowledgeMANAGER, and InsightOPTIMIZER, are service marks, trademarks, or registered trademarks of Altair. All other products, logos, or brand names are either service marks, trademarks, or registered trademarks of their respective owners. Altair data mining technology is protected under Canadian, US and international copyright laws and applications have been filed under Canadian, US and international patent laws.

## 1. Supported Hardware and Software Platforms

### 1.1 Overview: Supported Operating Systems

Altair Knowledge Works products version 2020.1 are available in the standalone and client/server configurations for the following operating systems:

Workstation (Standalone or Client)	Server
Windows 7 Windows 8 and 8.1 Windows 10	<u>Windows Server on x86-64 platform:</u> <ul style="list-style-type: none"> <li>• Windows Server 2012 R2</li> <li>• Windows Server 2016</li> <li>• Windows Server 2019</li> </ul>
Windows Server 2012 R2 * Windows Server 2016 Windows Server 2019	<u>Linux on x86-64 platform:</u> <ul style="list-style-type: none"> <li>• Red Hat Enterprise Linux 6.x and 7.x</li> <li>• CentOS 6.x and 7.x</li> </ul>
* The Standalone configuration for multi-user access on a server OS requires a server-level license	

- All products are supported only on 64-bit systems.
- Tableau integration and Keras modeling features are not supported on Red Hat Linux 6 and CentOS 6.
- Knowledge Studio for Apache Spark requires Apache Spark server. System requirements for this product are described at the end of Section 3.2 below.

### 1.2 System Requirements

This section covers minimum and recommended system requirements for all product configurations.

For server-based deployments, system requirements for the server also depend on the maximum number of users expected to be accessing the application concurrently. See the note after the recommended server requirements below.

#### Product Components

The **Standalone configuration** of Knowledge Seeker and Knowledge Studio for Windows Desktop or Windows Server includes the following installation packages:

- KS Workstation
- KS Library for R

The **Client/Server configuration** for Windows Server includes the following installation packages:

Client side:

- KS Workstation

Server side:

- KS Server
- KS Library for R

The Client/Server configuration for Linux includes KS Workstation and KS Server. KS Libraries for R and Python are built in the KS Server package for Linux.

Configurations with Text Analytics Add-on include Lexalytics® Saliency™ Text Analysis Engine.

The Keras modeling features in Knowledge Studio require Docker software as a prerequisite.

**Microsoft .NET Framework 4.6.1 (or higher) and 3.5 Service Pack 1** are prerequisites for KS Workstation. The KS Workstation installation package includes web installers for these components and installs them automatically if they are not found on the system.

KS Workstation includes KS PDF Printer component, which enables the *Copy to PDF* functionality for exporting decision trees, charts and reports to PDF. The printer uses Amyuni PDF Converter by [Amyuni Technologies, Inc.](http://www.amyuni.com)

Knowledge Studio and Knowledge Seeker include R and Python language integration features. R and Python/Anaconda software are prerequisites if these features are required. Python is also required for integration with [Altair Knowledge Hub](http://www.altair.com/knowledge-hub)™ (formerly known as Swarm).

Users who need Tableau integration features require Tableau Desktop or any Tableau workbook viewer on their workstation. Altair does not distribute Tableau software or licenses. If you do not use Tableau but would like to learn more about it, please visit <http://www.tableau.com/> and download an evaluation version.

The following are the minimum and recommended system requirements for all configurations:

## 1. Standalone Configuration

### KS Workstation 2020.1 (Standalone)

Minimum	Recommended
Windows 7, 8.x, or 10 (64-bit) x86-64 based CPU, 2 GHz 4 GB RAM Disk space: <ul style="list-style-type: none"> <li>• 700 MB for program files</li> <li>• 15+ GB for user's project files</li> </ul> Monitor resolution: 1152x864	Windows 7, 8.x, or 10 (64-bit) Quad-core x86-64 CPU, 2 GHz, or better 8 GB RAM or more * Disk space: <ul style="list-style-type: none"> <li>• 700 MB for program files</li> <li>• 20+ GB for user's project files</li> </ul> Monitor resolution: 1680x1050 or better

\* **Note:** The amount of RAM for the desktop version depends on the typical size of analytic datasets and the types of models to be built. If the users will be dealing with very large datasets and need help assessing the required amount of RAM, please email [support@datawatch.com](mailto:support@datawatch.com).

## 2. Client/Server Configuration

### KS Workstation 2020.1 – Client

Minimum	Recommended
Windows 7, 8.x, or 10 (64-bit) x86-64 based CPU, 2 GHz 4 GB RAM Disk space: 700 MB for program files Monitor resolution: 1152x864	Windows 7, 8.x, or 10 (64-bit) x86-64 based CPU, 2 GHz or better 8 GB RAM or more Disk space: 700 MB for program files Monitor resolution: 1680x1050 or better

## KS Server 2020.1

Minimum	Recommended
Windows Server 2012 R2 Red Hat Enterprise Linux 6.x or CentOS 6.x  x86-64 based CPU, 2 GHz 8 GB RAM Disk space: <ul style="list-style-type: none"> <li>• 500 MB for program files</li> <li>• 10+ GB per user for project files</li> </ul>	Windows Server 2012 R2, 2016, or 2019 Red Hat Enterprise Linux 7.x or CentOS 7.x A multi-core x86-64 based CPU(s) with at least 8 cores in total, 2 GHz, or better ** 16 GB RAM or more** Disk space: <ul style="list-style-type: none"> <li>• 500 MB for program files</li> <li>• 15+ GB per user for project files**</li> </ul>

**\*\* Note:** System requirements for the server depend on the maximum number of users expected to be using KS Server concurrently. The KS Server performance also depends on the typical size of analytic datasets, on whether any other applications run on the same server, and other factors. To ensure that your product is compatible with your environment and delivers expected performance, please fill in the Installation Questionnaire. Contact your account representative or email [support@datawatch.com](mailto:support@datawatch.com) to request the questionnaire if necessary.

If Python integration features are used, the prerequisite Anaconda (Python) program files may take up to 5 GB of disk space on the KS Server host.

### **3. Requirements for Knowledge Studio for Apache Spark**

Knowledge Studio for Apache Spark is an application for data mining and predictive analytics on large-scale distributed data structures (HDFS, Amazon S3, Cassandra, etc.) via [Apache Spark](#).

The types of distributed storage supported by Knowledge Studio for Apache Spark are the same as those supported by Apache Spark. The product is supported for any Hadoop distribution satisfying the **Apache Spark server requirements** listed below.

Examples of supported Hadoop distributions:

- Cloudera CDH  
 Supported versions: All CDH versions compatible with Apache Spark 2.1 – 2.4: (CDH 5.7 - 6.x; See “*CDS Powered By Apache Spark Requirements*” [https://www.cloudera.com/documentation/spark2/latest/topics/spark2\\_requirements.html](https://www.cloudera.com/documentation/spark2/latest/topics/spark2_requirements.html))
- Hortonworks Data Platform (HDP)

Supported versions: HDP 2.6.0 or later, with HDFS and YARN. For details, see [https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/bk\\_command-line-installation/content/spark2\\_prerequisites.html](https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.4/bk_command-line-installation/content/spark2_prerequisites.html)

- MapR Converged Data Platform. Includes support for Spark Standalone, Spark on YARN, and Spark on Mesos.
- Databricks Unified Analytics Platform, including Databricks on Azure and AWS. See [Apache Spark on the Databricks cloud](#).

**Note:** Spark cluster configurations with dynamic allocation are not supported for Knowledge Studio for Apache Spark.

System requirements for the Knowledge Studio software components are described at the beginning of Section 3.2 (see above). The following are the requirements for Apache Spark Server.

### Apache Spark server requirements

- Apache Spark version **2.1, 2.2, 2.3, or 2.4**
  - Cluster management system options: Standalone (native Spark cluster) or [Hadoop YARN](#) or [Apache Mesos](#)
- [Python](#) version **3.6 or higher**, 64-bit (Python programming language)
- Required Python packages:
  - Pandas: the [Python Data Analysis Library](#) provides data structures and analysis tools for Python.
  - [NumPy](#): the package for scientific computing with Python
  - [SciPy](#): the fundamental library for scientific computing (version 0.19.1 or higher)
  - [Scikit-learn](#): tools for data mining and data analysis (version 0.19.0 or higher)
  - [PyArrow](#): Python API of Apache Arrow (version 0.9.0 or higher)
- [Jupyter Notebook](#) version **4 or higher**
- [JupyterHub](#) version **1.0 or higher**

The easiest way to get **Python** with all required packages is to install [Anaconda](#) - a free Python distribution that includes many popular packages for data science. The 64-bit Anaconda installer for the latest version of Python can be downloaded from <https://www.anaconda.com>.

For **Jupyter Notebook** prerequisites and installation instructions, see <https://jupyter.readthedocs.io/en/latest/install.html>

For **JupyterHub** prerequisites and installation instructions, see <http://jupyterhub.readthedocs.io/en/latest/>

Note: The use of JupyterHub as a multi-user Jupyter notebook server is strongly recommended. If you prefer not to use JupyterHub, you can use basic authentication

implemented by other mechanisms and configure them according to your security, authentication, and file system access requirements.

Examples of supported types of distributed storage clusters:

- [Amazon Web Services Cloud \(AWS\)](#)  
Note: The Apache Spark environment on Amazon AWS includes Jupyter Notebook and JupyterHub.
- [Microsoft Azure HDInsight](#)  
Note: The Apache Spark environment on Azure HDInsight has Jupyter but does not include JupyterHub by default.
- Custom clusters satisfying the above requirements for Apache Spark.

For your convenience, our Customer Support can provide a script for installing the necessary prerequisites, including the required Python packages. The script is provided upon request. You will need to modify it as you see fit to install the components that you need. Please contact Customer Support at [support@datawatch.com](mailto:support@datawatch.com) to request the script.

For the client/server configuration of Knowledge Studio for Apache Spark, the range of TCP ports 6060 – 6090 on the server must be open for inbound connections from the client machines. In the standalone (desktop) configuration, Knowledge Studio for Apache Spark also uses ports in this range, but this is confined to localhost.

Please refer to the Deployment Guide for installation and configuration guidelines.



## 2. Version Compatibility

- Version 2020.1 can be installed and run side by side with any previous version.
- Uninstalling any version of Altair Knowledge Works product will not remove users' project files or licenses.
- If you keep your current version (2020.0 or older) together with 2020.1 for an evaluation or transition period, it is strongly recommended that you have a separate working directory for each version, since older projects modified in version 2020.1 may no longer be compatible with the original version they were created in. See [Section 5](#) for details about the Working Directory.
- Project compatibility:
  - Projects created in versions 9.3 – 2020.0 can be opened and modified in 2020.1 without any conversion.
  - Projects created in versions 7.0 – 8.7 can be opened in version 2020.1. All analysis objects in the old projects (datasets, trees, models, etc.) can be opened, viewed and edited, **but version 2020.1 will not automatically deduce and reconstruct a workflow from the original project**. Therefore, upon opening an older project, the workflow canvas will initially be blank. Any datasets and models in this project that you intend to use in a workflow need to be manually placed on the workflow canvas from the Project Explorer. This can be easily done with the right-click context menu command “*Put on Workflow Canvas*”, or with the Link command for dataset and model nodes or using the node *Import Model from XML/PMML* from the Model palette.

It is recommended to make a backup copy of your old projects before opening them in version 2020.1.
  - **Forward compatibility is not supported:** models and other objects in projects created or modified in version 2020.1 may **not** be loaded correctly when opened in 2020.0 or older versions.

## 3. New Features

### 3.1 New Features in Version 2020.1 (Released on 31-Mar-2020) Compared to 2020.0

#### Data Import and Export

- Data import from R is now supported for RData files created in R version 3.6.x and higher.

#### Data Preparation and Profiling

- The new **Pivot Table** node in the Manipulate palette for rearranging and summarizing data.
- **Data Preview** feature in the Variable Transformations node. The transformed data can now be previewed before running the transformations.
  - Data Preview in the Expression Editor to preview data while adding a new calculated field.
  - Data Preview in the Dataset Editor to preview top N records of the transformed dataset.
- Field reordering is now supported in the Dataset Editor of the Variable Transformations node.

#### Modeling

- The new **Model Stacking** node is an ensemble model that combines predictive scores produced by multiple models of different types and builds the final model.
- The new **Keras Model** node provides an interface to [Keras](#) - the Python Deep Learning library. The **Keras Scoring** node in the Action palette is provided for deploying Keras models. These features require Docker software as a prerequisite.
  - *Note: Keras modeling features are not supported in the client/server configuration for Red Hat Linux 6 and CentOS 6.*

#### Integration with Tableau®

- Integration with Tableau® software is now supported for Tableau versions 2019.1 and higher.
  - *Note: Tableau integration is not supported in the client/server configuration for Red Hat Linux 6 and CentOS 6.*

### 3.2 New Features in Version 2020.0 (Released on 10-Dec-2019) Compared to 10.9

#### Data Import and Export

- Unified file import: A single **File Import** node for all supported source formats.
- Unified file export: A single **File Export** node for all supported target formats.

- Enhanced data preview on import.
- The new **Folder Import** node imports data from all files in a folder in one take, with options to import data into separate datasets or append to a single dataset.

### **Data Preparation and Profiling**

- The new **Sort, Rank, Lag** node in the Manipulate palette for sorting data, creating ranking fields and lagged variables.
- The new **Multicollinearity Analysis** node in the Profile palette for automatic removal of highly correlated variables.
- The Append node now admits more than two input datasets to append multiple datasets at a time.

### **Modeling**

- The new **Time Series Decomposition** node in the Model palette performs transformation of time series data into component: Series, Trend, Seasonal, and Residuals.
- Attribute Editor in model wizards has a new column to specify options for the infinity value substitution.

### **General usability improvements**

- Unified Code Generation node: A single code generation node is now provided for all supported target code types.

## **3.3 New Features in Version 10.9 (Released on 16-Aug-2019) Compared to 10.8.4**

### **Data Preparation and Profiling**

- Data filtering capabilities
  - **Filter** node in the Manipulate palette for selecting or deleting records satisfying user-defined conditions.
  - An **interactive filter panel** in dataset views for filtering data without going back to the workflow and creating a new dataset. Toggle between filtered and unfiltered views with a single click.
- Variable Transformations
  - Four new helpers in the Variable Transformations node: *Text Join*, *Text Split*, *Date & Time Join*, *Date & Time Split*.
    - *Text Join* and *Date & Time Join*: concatenate text or date & time values and constants from several fields.
    - *Text Split* and *Date & Time Split*: split a text, date, time, or timestamp field into several fields representing text or date/time components, respectively.
  - Percentage of missing values can be previewed in the Dataset Editor of Variable Transformations node wizard.

- Data types can be easily changed in the Dataset Editor grid of the Variable Transformations node wizard.
- Expressions for derived variables can be instantiated using a column in the Dataset Editor grid of the Variable Transformations node wizard.
- **Make Time Series** node for converting datasets to evenly-spaced time series via temporal aggregation (e.g., creating daily time series by producing daily averages).
- **Time Series tab** and **Autocorrelation tab** in datasets produced by the Make Time Series node visualize evenly-spaced time series and auto-correlation plots, respectively.
- Specifying group-by fields in the Aggregation node wizard is now optional.

### Decision Trees and Strategy Trees

- Two Bonferroni adjustment types - Biggs and Kass - can now be set in the user interface of the Decision Tree wizard and via Options & Preferences.
- Tree growth constraints: Minimum allowable node size for non-terminal and terminal nodes can now be specified as a percentage.
- User-defined labels for branches in decision tree and strategy tree splits.

### New model types

- **Factor Analysis** node in Model palette.

### General usability improvements

- Format options for node comments on the workflow canvas include background color, font color, and font size.

### R Integration

- R integration requires the [Rserve](#) package (Binary R server). It must be installed as a prerequisite.

## 3.4 New Features in Version 10.8.4 (Released on 21-Jun-2019) Compared to 10.8.3

### Optimization Improvements

- In optimization problems with complex aggregate functions in constraints, the optimizer uses approximations to the Hessian matrix instead of computing the exact matrix.
- Aggregate functions with much greater complexity can now be included in constraint expressions. For example, constraints of the following form are now allowed:
  - $2 * \text{SUM}([\text{Channel}] = \text{'Email'}) \leq 1000$
  - $\text{SUM}([\text{Var1}] * [\text{Decision\_Var}]) / \text{SUM}([\text{Var2}] * [\text{Decision\_Var}]) \leq \text{SUM}([\text{Var3}] * [\text{Decision\_Var}])$

### 3.5 New Features in Version 10.8.3 (Released on 24-May-2019) Compared to 10.8.2

#### Improved Integration with Altair Knowledge Hub™

Five more types of models can now be published to [Knowledge Hub](#):

- Strategy Trees
- Scorecards (both Scorecard and Scorecard Editor nodes)
- Deep Learning models (Neural Networks)
- Constrained Regression
- Regularization

This means the publishing of all types of predictive models is now supported.

#### Outlier Detector Improvements

- The **Univariate Distance** method replaces the Standard Deviation method, which is now a special case of the former. Aside from selecting the variables to be tested for outliers, users also have a choice of 3 location measures and 3 scale measures. Users can define grouping variables to calculate the measures and detect outliers within groups rather than the entire dataset.

#### New Variable Transformation Features

- **Standardization helper** in the Variable Transformations node allows you to standardize (rescale) numeric variables - one or multiple at a time. The user has a choice of 6 location measures and 5 scale measures, which can be combined as desired.

- **Safe division**

The SAFE\_DIVISION() function is provided as an alternative to ordinary slash division (/) to better handle integer arguments and extreme cases such as division by zero and the infinity value.

- Safe division of integers always returns a real number (a value of type Number, not rounded or truncated).
- Safe division of a positive or negative number by zero produces positive or negative infinity, respectively.
- Safe division can be configured to produce a specific user-defined value for the cases of zero divided by zero. This value can be specified as the third argument of the SAFE\_DIVISION() function.

- **Infinity**

Division and other operations can return positive or negative infinity. The infinity symbol  $\infty$  is now used in all views where this value occurs. In the SQL code in Variable Transformations or Strategy Tree calculations, the notation +/- {INF} can now be used to designate positive or negative infinity. Algebraic operations involving infinity are defined according to the usual mathematical conventions for the extended real number line.

- **PASTE() function**

The PASTE() function of three or more string arguments concatenates them starting from the second one. The first argument is used as a separator between the terms in the resulting string.

### 3.6 New Features in Version 10.8.2 (Released on 08-Mar-2019) Compared to 10.8.1

#### Angoss Analytics Software Suite is now a part of Altair Knowledge Works

The products of Angoss Analytics Software Suite are now a part of *Altair® Knowledge Works™*.

- *Angoss KnowledgeSTUDIO* is now called *Altair Knowledge Studio®*
- *Angoss KnowledgeSTUDIO for Apache Spark* is now called *Altair Knowledge Studio® for Apache Spark™*
- *Angoss KnowledgeSEEKER* is now called *Altair Knowledge Seeker®*
- *Angoss KnowledgeREADER* is now available as *Altair Knowledge Seeker®* with *Text Analytics Add-on*.

#### Altair HyperWorks Units Licensing

Altair HyperWorks Units is a value-based license management system enabling metered usage of an entire suite of products.

All *Altair® Knowledge Works™* products are enabled for a single pool of recyclable HyperWorks Units. The Text Analytics Add-on is only available with registration-key licensing.

#### New Platform Support

- Support for Windows Server 2019 has been added.

### 3.7 New Features in Version 10.8.1 (Released on 31-Jan-2019) Compared to 10.8

#### New Split Search Method in Decision Trees

- The new split search method **Forward Exhaustive** is now available in the Decision Tree training parameter settings in addition to two existing methods (*Cluster* and *Exhaustive*).
  - The new method is optimized for ordered variables. It involves a combination of optimization techniques to increase the chances of arriving at the global optimal solution in a fast and efficient way. A user-defined parameter allows to tune the method to improve its accuracy at the expense of performance.
  - The method is also available in Ensemble Trees (Random Forest, Bagging, and Boosting) and in the Outlier Detector node.

### 3.8 New Features in Version 10.8 (Released on 11-Dec-2018) Compared to 10.7.3

### Integration with Swarm

- Angoss analytics software applications now integrate with Swarm.
  - Angoss datasets can be exported to Swarm.
  - Swarm files can be imported into Angoss datasets
  - Angoss models can be published to Swarm (in PMML format). Supported model types for model export: Decision Trees, Logistic Regression, Linear Regression, Regularization, and Constrained Regression.
- Note: Swarm integration features require Python version 3.7 or higher.

## 3.9 New Features in Version 10.7.3 (Released on 24-Oct-2018) Compared to 10.7

### Decision Trees and Ensemble Tree Models

- Two new measures added in Decision Trees: *Unadjusted log p-value* and *Bonferroni-adjusted log p-value*.

### R and Python Language Integration

- R integration is now supported for R versions 3.5.x.

### Platform Support

- Angoss software is no longer supported on 32-bit operating systems.

## 3.10 New Features in Version 10.7 (Released on 27-Mar-2018) Compared to 10.4

### Workflow and Project Features, User Interface Improvements

- Project clean-up command removes datasets and other objects no longer used anywhere in the project workflows.
- Context menu command "*Find Reference in Workflow*" in Project Explorer finds workflow nodes that use the selected dataset, model, etc.
- Nodes in palettes are now arranged in the alphabetic order
- Angoss Variable Selection node has been moved to the **Profile** palette
- Multi-Level Neural Network model node renamed to *Deep Learning*
- HTML tables in reports can be sorted by columns

### Data Manipulation and Profiling

- Filtering Join node (semi-join and anti-join)

### Model Deployment / Code Generation

- R and Python code generation for Decision Trees, Strategy Trees, and Ensemble Trees
- Generation of code in the Language of SAS for Market Basket Analysis and Ensemble Trees

### Integration with Apache Spark (in KnowledgeSTUDIO for Apache Spark)

- Advanced login to Spark allowing users to specify SparkContext parameters
- Cross Tabs in the Dataset view
- Variable Selection node
- Variable list now available for Spark functions
- Reshape node
- Filtering Join node (semi-join and anti-join)
- Substitute Missing Values node

#### Platform Support and Database Access

- Angoss Server 10.7 is supported on Windows Server 2016
- Angoss Server for Linux no longer includes DataDirect ODBC drivers. It uses the [unixODBC](#) driver manager. We recommend that native ODBC drivers be used for the required databases. Please contact Angoss Support if you need assistance in configuring database connections on Linux.

### **3.11 New Features in Version 10.4 (Released on 01-May-2017) Compared to 9.6**

#### Workflow and Project Features

- Redesigned icons and toolbars
- Redesigned and reorganized functional palette
- Redesigned nodes and connectors on the workflow canvas.
- Two connector style options

#### Decision Tree and Strategy Tree Enhancements

- Refreshed and improved design of tree nodes - new color palette, node border style, fonts, and split line style.

#### In-memory Analytics on Large-Scale Distributed Data via Apache Spark (available in KnowledgeSTUDIO for Apache Spark)

- Pre-built nodes for data preparation, modeling, model evaluation and deployment with Spark.
- Analytics on all types of distributed data structures supported by Spark (Hadoop Distributed File System (HDFS), Amazon S3, etc.).
- Support for Object Stores, Distributed File Systems, Network Shares and other Enterprise data repositories: Angoss can directly load data from several new data sources: Hdfs:// (Hadoop); har:// (Hadoop Archive); s3:// , s3a:// (Amazon S3); ftp://; File:// (Network Shares); viewfs:// (Hadoop NS Vols)
- Analytics supported for Hive tables, Text (CSV), Parquet, ORC, Avro - the most common data formats and frameworks in distributed storage systems. This includes both load and export functions.
- All interface modes supported: 1) visual workflow-based graphical user interface; 2) code-based interface for programming in languages for distributed computing.



- Data preparation & manipulation functions for dataframes in distributed storage via Apache Spark: append, deduplicate, join, aggregate, filter, sample, create new variables.
- Predictive modeling and cluster analysis on distributed data via Apache Spark: Decision Trees, Random Forest, Boosting, Strategy Trees, Regression, Clustering, and Naive Bayes model.
- Models developed on distributed data via Apache Spark can be deployed in the same Big Data environment. Their performance can be evaluated and compared.
- Capability to write custom programs in Python and embed these at any stage of the Angoss workflow. Enhanced capability to use Jupyter notebook for interactive editing and running code via a web browser.
- Python code log: View and copy the Python code for any analytic operation performed in Angoss software with Spark.
- Scheduling and Automation: Automated scoring of data on a scheduled basis with Spark; execute entire workflows within Spark

#### Optimization Improvements (available with the InsightOPTIMIZER add-on license)

- Add multiple optimization scenarios within any workflow.
- Automate/Schedule optimization scenarios to run in the background.
- Improved reporting: Comparisons between scenario outputs; sensitivity analysis

### **3.12 New Features in Version 9.6 (Released on 04-Mar-2016) Compared to 9.5**

#### **I. All Products**

##### Data Preparation Features

- Outlier Detector: Automatically detect outliers in a dataset and flag or change them for improved modeling
- Variable Importance: Order variables by their importance to aid in variable selection for modeling
- Identification of special missing values
- Transforming multiple columns at a time: Define and apply the same transformation to multiple fields
- Variable Selection automation: Select model variables according to their predictive power

##### Decision Tree and Strategy Tree Enhancements

- Code generation for tagged nodes, sub-branches, or current node
- Treatments for non-terminal nodes

##### Workflow Management Enhancements

- Undo delete node
- Automatic documentation of modeling nodes
- Self-documentation: Automatic logging of operations in the Microsoft Word format

### Integration with Qlik

- Integration with Qlik Sense - business intelligence software for interactive data visualization - provides the ability to create visualizations and dashboards with Qlik Sense directly within Angoss Workflows.

## **II. KnowledgeSTUDIO**

### Predictive Models and Scorecards

- Partial Least Squares Regression (PLS Regression)
- Constrained Regression
- Regularization
- Cross Validation
- Schema node (a template of input variables to simplify model development)
- Survival Analysis (requires R Language Integration)
- Scorecard Editor: Manually change score points in an existing scorecard
- Stability and Characteristic Report for scorecards and models with binary dependent variable

## **III. InsightOPTIMIZER**

- General Dataset-Level Optimization (linear and nonlinear)
- General Record-Level Optimization (linear and nonlinear)

## **3.13 New Features in Version 9.5 (Released on 29-Sep-2015) Compared to 9.4.5**

### **I. All Products**

#### Ensemble Trees

Three ensemble models based on Decision Trees (3 new nodes in the Modeling palette):

- Random Forests
- Bagging (Bootstrap Aggregating)
- Boosting

#### Python Language Integration

- The [Python language](#) Integration feature allows you to write Python language code within the Python Code process node in the Angoss workflow and use its outputs in subsequent operations in Angoss.

#### Decision Tree Enhancements

- Force Split: Optimal Binning option

- AutoGrow option to allow or prevent split on the same variable on two consecutive levels
- Tag Node feature enhancements
- Node Comment feature improvements

#### Workflow Management Enhancements

- Node alignment
- Show/hide grid in the workflow canvas
- Snap node to grid
- Copy workflow layout to the same or other workflows
- Cut and paste nodes and parts of the workflow
- Copy datasets from other projects
- Auto-arrange layout; Undo auto-arrange

## II. KnowledgeSTUDIO

#### New stopping criterion for Stepwise Regression: Entropy difference

- A separate node for WOE transformations in the Data Transformations palette

#### Scorecard Enhancements

##### **Improved and redesigned Weight of Evidence transformations**

- A separate node for WOE transformations in the Data Transformations palette
- More charts and statistics available in the Weight of Evidence Editor (characteristic curves, Gini index)
- WOE Transformations Report
- Three advanced optimization types for Weight of Evidence can now be performed separately or all together
- Export of WOE transformations to a WOE Template.

##### **Dynamic WOE binning optimization in regression models**

- New advanced option for regression models: The boundaries of Weight of Evidence bins can be automatically adjusted while training the model. A dataset with adjusted bins for WOE variables is an additional optional output of the wizard.

## III. KnowledgeREADER and Products with Text Analytics Add-on

#### Text Analysis supported in Workflow

- The Text Analysis palette includes the Text Analysis, Association Map, Text Analysis Comparison, and Trend Analysis nodes

#### Lexalytics' Saliency Text Analysis Engine upgraded to Version 6

- New language packs: Korean, Italian, Dutch, Japanese, Malay and Singlish languages are now supported

#### IV. InsightOPTIMIZER

**Linear and nonlinear optimization** for marketing/CRM, credit risk management, and other areas. Example: channel optimization in marketing campaign management based on response models.

- For linear optimization, Angoss uses a linear programming solver based on the revised Simplex method.
- Nonlinear optimization uses the Boundary and linear equality/inequality constrained optimization (BLEIC).

### 3.14 New Features in Version 9.4.5 (Released on 17-Jun-2015) Compared to 9.4

#### Tableau Integration:

- Starting from version 9.4.5, Angoss includes integration with [Tableau®](#) - business intelligence software for interactive data visualization. Integration with Tableau is supported in the desktop mode and in the client/server mode for Windows and Linux.

Integration with Tableau Server and Tableau Desktop allows users to create and publish workbooks for Tableau dashboards and share them throughout the organization from within the Angoss application framework. Workbooks from Tableau Server dashboards can be downloaded to Angoss applications and incorporated into Angoss workflows as templates for creating new workbooks.

#### Improvements in Workflow, Scorecards, and Decision and Strategy Trees:

- Workflow enhancements:
  - Easy switching from object views back to the workflow using a toolbar button.
  - The capability to put desktop documents on the Workflow canvas. MS Office documents and any other types of files can now be added to the Workflow canvas for project documentation purposes.
- Scorecard enhancements: A new option in Scorecard Scaling allows users to incorporate Base Score in Score Points.
- Decision Trees and Strategy Trees: Improved Node Comment feature: more user-friendly and intuitive user interface for inserting, editing, and formatting node comments.

**A detailed history of new features in all Angoss releases since version 4 is available upon request.**

## 4. Known Issues in Version 2020.1

### Knowledge Studio for Apache Spark

- Spark cluster configurations with dynamic allocation are not supported at this time.
- An exception occurs when validating a model with Boolean dependent variable and mapping it to an integer field in the input dataset.

### Integration with [Altair Knowledge Hub](#)™

- Limitations:
  - Knowledge Studio and Knowledge Seeker Models with discrete dependent variables with missing values (nulls) or empty strings are not supported in Knowledge Hub. The workaround is to use the Variable Transformation node to substitute missing values or empty strings before creating the model.
  - In Knowledge Hub, when scoring an exported Knowledge Studio model that has independent variables of Date or Time type, the Date and Time variables need to be explicitly mapped to Timestamp fields in the input dataset in Knowledge Hub. Alternatively, convert Date and Time variables to Timestamp in Knowledge Studio or Knowledge Seeker before training the model and exporting it to Knowledge Hub.
- The Date and Time data types are not supported by Knowledge Hub at this time. Columns of these types in Altair datasets are exported to Knowledge Hub as Text.
- Datasets exported to Knowledge Hub can have no more than 512 columns. Altair datasets with more than 512 columns are accepted as inputs to the Knowledge Hub Export node, but the user needs to select no more than 512 columns in the Variable Selection page of the wizard.
- The size of the data exported to Knowledge Hub should not exceed 2 GB. The error "Connection aborted" may occur if the text format equivalent of the uploaded Angoss dataset exceeds 2 GB.

### Modeling, Model Evaluation, and Deployment

- Dependent Variable is not shown in the Linear regression results in some cases, and the model cannot be scored or validated.
- Scoring and validation of Deep Learning (Neural Network) models imported from PMML is not supported.
- The training of a Deep Learning (Neural Network) model with continuous dependent variable does not converge in some cases and runs indefinitely.
- The Tree training measure setting in Preferences does not apply to Strategy Trees, whose default measure is Entropy Variance. The workaround is to set the desired measure in Options after creating a Strategy Tree.

- Field mapping in the Scoring wizard does not apply in the SAS code generated from the scoring node.

### **Optimization**

- If the derivative of a function is 0, the optimizer will stop at this point, even if it doesn't solve the problem.
- Constraints are being ignored in some cases.
- Constraints with conditions such as “WHEN ([Channel] is null)..” cause errors.
- \$function\_value\$ is not populated with the correct data in the dataset level node output.
- Scenarios that run fine as individual optimization (Dataset Level) scenarios fail in Multiple Scenarios Optimization scheduled to run immediately after setup.

## 5. Fixed Issues

### 5.1 Issues in v2020.0 fixed in v2020.1

#### Installation, Configuration, Licensing, and KS Server Connectivity

- Each failed connection from KS client to KS Server is recorded as two failed logons on the server OS.
- An object reference exception occurs after specifying an invalid Working Directory during client/server connection and trying to create a new folder.

#### Data Import, Data Preparation and Profiling

- ODBC Import node does not work with PostgreSQL database.
- An exception "Index was out of range" occurs in File Import if the source text file contains a field with Date values in an invalid format.
- Cannot import expressions in Variable Transformations if there are two fields with identical expressions. Error: "Column already belongs to this DataTable".
- Multiple row selection does not work in the grid of the Multicollinearity Analysis wizard.
- The *Optimize All* function in the WOE node wizard forces the *Monotonic* and *No Pure Nodes* optimizations for all variables regardless of the user's selection.
- "OLE DB: Unexpected error occurred." when refreshing Data Preview in the Append node wizard if the input datasets are large (~ 1 million records or more).
- Histograms in Dataset Chart and other views do not work properly for integer variables, have issues with equal-width bins.

#### Decision Trees and Strategy Trees

- The Bisection Samples value in the Forward Exhaustive method affects splits on nominal variables.
- Using the Ignore Variable command in a Strategy Tree may cause node calculations to become *null*.
- Split statistics in the Split Report of Decision Trees have default numeric precision 2 instead of 5 in some cases.
- The P-Value column is missing in the Attribute Editor of Strategy Trees if the Split Search Method is set to "Exhaustive".

#### Other Models

- Weight field is not forced to be treated as continuous regardless of its cardinality when creating a model, which causes wrong model results.
- Start point refinement selection is not saved by the Cluster Analysis wizard.

- Infinite values setting in Logistic Regression does not propagate to Scorecard.
- The missing value substitution setting for continuous variables in Logistic Regression does not propagate to Scorecard based on it. This causes inconsistent scores and different curves in Model Analyzer for scorecard and its parent model.
- Dependent Variable list in the Logistic Regression wizard does not save the DV selected by the user in Attribute Editor to override the default DV cardinality restriction
- Setting format in the Options of the Market Basket Analysis model result views has no effect.
- Opening a model instance created from a Cluster Analysis model causes an exception if the source model no longer exists.
- Wrong Segment Viewer charts for ordinal variables in Cluster Analysis model results.
- Incorrect date of model creation in the Market Basket Analysis model results report.

### **Scoring, Validation, Model Analyzer, and Code Generation**

- Running a Stability & Characteristic Reports node on models with variables derived from fields with infinite values {INF} causes an exception.
- Generating LOS code for the Filter node causes an exception in any expression in the filter contains a numeric constant in Scientific format.
- Model Validation Report has wrong values of Valid Records and Entropy Explained or Variance Explained for models where the missing value option for the dependent variable or any independent variables in Attribute Editor is set to “ignore”.

### **Optimization**

- The error "01/01: E00001E: INTERNAL: 'bad allocation'" occurs in Multiple Scenario Optimization task in some cases.

### **Integration with R and Python**

- R Code node does not work if the Working Directory is on a network drive specified as a UNC path.
- R Code node does not work after changing the Working Directory. It still assumes the old location and fails to connect to R.
- R Import node does not support data import from RData files created in R 3.6.0 or higher.
- A memory leak occurs in some cases in the Python Code node during the execution of AngossUtilities.py.

## **5.2 Issues in v10.9 fixed in v2020.0**

### **Installation, Configuration, Licensing, and KS Server Connectivity**



- The "Set Working Directory" command in the client/server mode on Windows causes exception when specifying an invalid path.
- Usage monitoring utility Usage.bat does not work on Windows server.
- KS Server on Linux unnecessarily requires that the ODBCINI environment variable be defined.
- An exception occurs and the application shuts down when deleting a project with an R Code node.
- KS Server uninstaller on Windows does not remove the service KSWinLoginService.

### **Data Preparation and Profiling**

- Cannot overwrite the output dataset after changing an input dataset in the Join wizard.
- The Filter node wizard changes the order of fields in the dataset to alphabetical.
- For any node from the Manipulate palette, all settings specified in the node wizard are lost if the output dataset is disconnected on the workflow.
- Unable to modify or run a WOE node after changing the data type of an independent variable in the input dataset.
- Small numeric constants (e.g., 0.00001) specified in a condition expression in the Filter node becomes NaN (undefined).
- Error "Null reference encountered" occurs in Filter node if a field in the input dataset is renamed.
- A parse error may occur when running Variable Transformations node using the Run command in the context menu in projects created in older versions.
- Data Preview in the Filter node shows only the Date part of Timestamp fields.

### **Decision Trees and Strategy Trees**

- Changing the threshold (percent error) and Bonferroni adjustment value in the Decision Tree wizard does not affect the Split Report page.
- The Copy/Paste Split option "Paste Unknown Values" is ignored.
- In Attribute Editor, the Max Branches option is enabled even if Usage is "no binning".
- Decision Tree is not updated after modifying the input Schema node.
- The "Percentage of Training data" parameter in Preferences does not apply to Ensemble Tree models.

### **Other Models**

- Cluster Analysis model fails if the option "Start from the cluster centers of the template model" is selected.

- For variables that contain infinities {INF}, setting Usage to 'discrete' in Predictive model Attribute Editor causes a parse error.
- Null values are displayed in two different ways in the same Optimization output dataset (null and blank).

### **Scoring, Validation, Model Analyzer, and Code Generation**

- Model Validation and Scoring wizards do not respect the missing value substitution options specified in models and fail with the error "Division by zero".
- Errors in LOS code generated from Filter node.
- Data type changes in the Variable Transformations are not applied to the LOS code generated from it.
- If the weight field is a calculated column, the Model Validation and Scoring wizards unnecessarily require that Weight be included in their input.
- Slow performance of scoring for Ensemble Trees.
- Wrong field mapping in the Cross Validation node wizard if the input model uses calculated fields.

### **Knowledge Studio for Apache Spark**

- Slow Automatic Grow performance in Spark decision trees based on wide datasets with thousands of variables.
- Unnecessary mapping of the dependent variable in the Scoring wizard for decision trees on Spark.
- Variable Selection node fails on large dataframes if the dependent variable is continuous.

## **5.3 Issues in v10.8.3 and v10.8.4 fixed in v10.9**

### **Installation, Configuration, and Licensing**

- Unable to change license location if initial license path is invalid.
- The function "Acquire License by Email" does not work on Windows 10
- In the client/server mode, the client connects to the root directory if the user specifies an invalid Working Directory. No error message and no file browser is provided.

### **Data Preparation and Profiling**

- Function names not alphabetically ordered in Expression Editor in some cases.
- Data grid in Dataset Chart shows integers instead of decimal values in some cases.
- Exception when using "median absolute deviation" scale measure in Outlier Detector with Univariate Distance.

- Native KDD file import convert Date type to Timestamp.
- Multiple selection is not supported in the last page of the Optimal Binning wizard.
- In the Aggregate node wizard, the field of the aggregated values cannot have the same name as the original field being aggregated.
- Variable Transformation node retains deleted WOE fields no longer present its input dataset.
- Display issues in Segment Viewer when it is in the docked state.

### **Models (including Decision Trees and Strategy Trees)**

- The Target Category is not preserved when reopening the Model Selector node.
- Ensemble Tree wizards do not inherit settings from Preferences
- Wrong histogram in Force Split Range Editor if the "####" (OTHER) category is present.
- If the usage of an independent variable in a Regularization model is changed from 'ordinal' to 'discrete', the model training fails with an error "Category is missing from dataset".
- An XML document error occurs when copying a constrained regression model.
- Node color in a Strategy tree does not match the setting in Options if Node Color Source is "Treatment".

### **Scoring, Validation, Model Analyzer, and Code Generation**

- Lift calculation in Lift Report is calculated from Estimated Target Values column instead of Actual Target Values column.
- Scorecard models imported from PMML cannot be validated.
- Wrong scoring and validation output of linear Regularization model if the DV Usage type is ordinal.

### **Optimization**

- Optimization node gets invalidated after being disconnected from another node it's a template for.
- Wizard navigation in Optimization (Dataset Level) node is blocked by error message about insufficient input for constraints.

### **Integration with R and Python**

- Large integers become null when using R code.
- Python Code node produces incorrect predictions for decision trees with dependent variable of type Date.

## Documentation

- A script error occurs in the Help page “Dataset Views: Summary”.

## 5.4 Issues in v10.8.2 Fixed in v10.8.3

### Installation and Configuration

- Custom port specified in Angoss Server installation wizard does not persist. It is reset to 5470 after restarting the service.

### Integration with Knowledge Hub™

- Failure to publish weighted Angoss models to Knowledge Hub.

### Data Preparation and Profiling

- The results of Outlier Detector with Mahalanobis distance method may depend on the variable selection order.

### Decision Trees, Strategy Trees, and Regression Models

- The Scoring node wizard for Strategy Trees does not save the selection of the *null* treatment.
- An exception “error in XML document” occurs when creating or copying a Constrained Regression model in some cases.

### Code Generation

- Various fixes in PMML code generation for all model types: handling missing values, dependent variables of date and datetime type, etc.

## 5.5 Issues in v10.8.1 Fixed in v10.8.2

### Decision Trees

- Wrong Bonferroni value when Forward Exhaustive split search method is selected.
- Changing the Measure in the Training Parameters page of Decision Tree wizard causes the Split Report page to display wrong header for the Info parameter.

### License Manager

- Host OS value in License Manager is not displayed correctly on Windows 8.x and Windows 10.

## 5.6 Issues in v10.8 Fixed in v10.8.1

### Analytics on Apache Spark

- Inadequate error handling in the Spark Log In dialog if wrong Authentication method is selected.
- Wrong column order on Spark model output if spark connecting to Spark on Azure HDInsight.
- The error "Spark execution timed out" occurs if the connection to Spark on Azure HDInsight is idle for a long period of time.
- Unable to open Jupyter Notebook after connecting to Spark on Azure HDInsight.

#### **Integration with Monarch Swarm**

- Slow performance of data import from Swarm compared to Text file import.
- In the wizard for the node *Import Dataset from Swarm*, if a numeric field type is changed to Boolean, clicking "Preview" or running the node results in an error.
- In the wizard of the node *Export Dataset to Swarm*, the Variable Selection page search boxes do not show all available columns.
- In the wizard of the node *Export Dataset to Swarm*, the Save and Run buttons become disabled after going to the next page and back.
- The "Stop Run" function is not available while importing data from Swarm.

#### **Data Import and Export; Server Connectivity**

- ODBC Import from Hadoop Hive database does not work with the default option for new versions of Hortonworks and Cloudera ODBC drivers for Apache Hive, such as Hortonworks ODBC driver v2.1.16 and higher, Cloudera ODBC driver v2.6 and higher.
- Failure to export data to Hive table from Angoss Server on Linux.
- An exception "Object reference not set to an instance of object" occurs in the Angoss client application if network connection is lost.

#### **Decision Trees, Strategy Trees, and Ensemble Trees**

- Random Forest model training fails to complete in some cases in the client/server mode with Angoss Server on Red Hat Linux 6.x and CentOS 6.x.

### **5.7 Issues in v10.7.3 Fixed in v10.8**

#### **Analytics on Apache Spark**

- Slow performance of the Analyze function in the decision tree wizard.
- Failure to insert a decision tree on a wide dataframe with thousands of variables.
- Angoss process listening on WAMP ports 6060-6090 (both on client and server) allows TLS protocol version 1.0.

### **Decision Trees, Strategy Trees, and Ensemble Trees**

- Error in Computed Conditional Treatment Helper in Strategy Trees: "An item with the same key has already been added."
- Force Split Range Editor loses records in some cases if you switch from Equal Width to Equal Height binning.

### **Data import, Export, Data Profiling, Data Manipulation**

- The application crashes and shuts down when importing Excel files exported from SAS.
- Different representation of missing values in numeric and integer fields in the Data tab (null vs blank values).
- The Analyze button clears existing selection (checkmarks) in the Variable Selection node wizard.
- "KSValue - Type mismatch" error when running WOE node after changing a field type in an input dataset.
- Export of Timestamp values to R adds a number of hours to each value
- SAS Export wizard does not automatically name the output file according to SAS naming conventions or warn the user to name it properly.

### **Modeling, Model Evaluation, and Scoring**

- Scorecard is incorrect if built on model with ordinal IV whose cardinality exceeds discrete threshold.
- Constrained Regression fails with the error message "Cannot find categories for discrete variable..."
- Exception in Scorecard Editor if the referenced regression model and scorecard have been deleted.
- Validation and Scoring nodes corrupt expressions by substituting "GB\_VALUE" for aggregate functions.
- Failure to validate a Logistic regression model on large dataset in some cases.
- Scorecard instances cannot be connected to scorecard editor nodes.
- Order of variables is not consistent across the clusters in the Summary tab of the Clustering output.
- An exception may occur in Model Validation after the tree has been modified: "MODELEVAL(): [Project].[tree]: model expected 8 variables; got 7 from source"
- "Parse error" when training certain a Linear regression model on a very large dataset.
- A newly added Node Calculation in a Strategy Tree does not show up in the Strategy Validation wizard.

- Slow scoring and validation on very large datasets.

#### **Code Generation**

- R code integration does not work with R versions 3.5.x
- Python Code Generation truncates variable names at 32 characters.
- Various problems with PMML code generation make the code invalid and not suitable for scoring in a number of contexts.
- Error "Cannot be resolved or is not a field" when running generated Java code for tree with open-ended attributes.
- LOS code generated for Cluster Analysis has mismatching indices of output variables and other issues.
- Incorrect LOS code generated for decision trees with continuous dependent variable.

#### **Server / Connectivity**

- Error: "server did not start properly" when starting Angoss Server on Linux.
- Misleading error message "*Not enough memory to complete the operation*" when trying to open a project without write access to the project or working directory.
- User interface issues in model wizards on Windows 10 (e.g., no frame or header separator in grids).
- The Stop button (the hand icon) in the toolbar does not work in some cases, while the Stop command in the context menu does.

### **5.8 Issues in v10.7 Fixed in v10.7.3**

#### **Decision Trees and Ensemble Trees**

- Random Forest and Bagging models on Linux consume too much memory and may crash.

#### **Integration with R**

- R language integration does not work with R versions 3.5.0 and higher.

#### **Modeling, Model Evaluation, and Deployment**

- Stepwise Logistic Regression fails on large datasets in some cases. Error: "No significant model found {Domain error}".
- Model Validation and Scoring nodes corrupt the expressions in calculated fields of datasets by substituting "GB\_VALUE" for aggregate functions such as AVG.

### **5.9 Issues in v10.7 Fixed in 10.7 Update of April 26, 2018**

- Inadequate error message and connection loss if wrong credentials are specified for the client / server connection.
- Empty WOE expressions in the WOE node for wide datasets with hundreds of variables.
- The WOE node (Weight of Evidence transformation) consumes too much memory and may run out of memory even for medium-sized datasets.
- Logistic Regression and Multi-Level Neural Network models fail with the error: "all DV values are excluded" if the DV is of *Number* type. The workaround is to convert the DV to the *Integer* data type.
- An invalid string enclosed in question marks may appear in the Multi-Level Neural Network model output in the case of a discrete DV of the *Number* type.
- For Multi-Level Neural Network models with discrete DV of the *Number* type, Model Validation report may contain invalid strings *NaN%* in the confusion matrix, and the Scoring report an incorrect grid of predicted values. The workaround is to convert the dependent variable to the *Integer* type.

## 5.10 Issues in v10.4 Fixed in v10.7 and 10.7 Update of April 26, 2018

### Installer

- Installation issue in the Silent Install mode: Unless the Working Directory (*CWDGROUP*) is specified in the Silent install command, it is set to an invalid default (0).

### Client/Server Connection and Cryptographic Protocols

- Angoss client/server connection uses TLS version 1.0, while the recommended version is TLS 1.2 or 1.1.
- Redundant connections for license check at login time.

### Data Preparation and Profiling

- The Weight of Evidence node (WOE) runs out of memory in some cases.
- In the Aggregate node output, the Aggregated record count in the report does not match the actual number in the output dataset
- The Gini value in the Measure of Predictive Power report is not formatted properly – has too many digits after the decimal point (15). Four digits are sufficient.
- Inconsistent handling of function compositions in the Variable Transformations node. Expression validation succeeds, but adding the new column fails with a file access error. Example: Adding a derived variable with expression such as `AVG(AVG[age])` fails with the error "`ALTER TABLE: ... could not access file`".
- The Variable Importance node fails if the selected dependent variable is a calculated column, e.g., a derived variable created by the Variable Transformations node. The workaround is to use the *Instantiate* option in the Expression Editor of the Variable Transformations wizard.



- The "*Highlight Values*" filter in the Correlations tab of datasets allows entering values outside of [-1,1] interval.

### Strategy Trees

- Node calculations cannot be removed in the Strategy Tree node wizard in some cases.
- In some cases, the Scoring wizard for Strategy Trees shows treatments that were previously deleted from the input tree.

### Predictive Models and Scorecards

- Logistic Regression and Multi-Level Neural Network models fail with the error: "all DV values are excluded" if the DV is of *Number* type. The workaround is to convert the DV to the *Integer* data type.
- An invalid string enclosed in question marks may appear in the Multi-Level Neural Network model output in the case of a discrete DV of the *Number* type.
- For Multi-Level Neural Network models with discrete DV of the *Number* type, Model Validation report may contain invalid strings *NaN%* in the confusion matrix, and the Scoring report an incorrect grid of predicted values. The workaround is to convert the dependent variable to the *Integer* type.
- The "Compute Errors" option for Logistic Regression is redundant. Standard errors and the related numbers are always computed and displayed anyway.
- Scorecard Editor cannot be applied to a Scorecard built on a regression model instance from a different workflow.
- Reject Inference not iterating properly in some cases. The Proportional Assignment or Hard Cut-Off methods may yield 1 iteration while not converging according to the RMSE value criterion.
- If one of its inputs is a regression model with DV of *Number* type, the Model Selector node reports a DV data type mismatch between the model and the parent dataset.
- Incorrect Scorecard based on a regression model with numeric independent variables whose *Usage* in Attribute Editor is set to "ordinal". The ScorePoints expression for ordinal variables do not take into account the discrete character of the variable.

### Reports

- The "Copy to PDF" command is not available for the Ranking report of the Variable Importance node output.
- Various formatting issues in reports: KS-7487: Grid formatting issues in the Ranking tab of the Variable Importance node output; KS-7236: Unnecessary double lines in some model reports; KS-7620: Number formatting issues in the Logistic Regression results: all values in the columns -95% / Odds / +95% end with three redundant zeroes; KS-7360: The CTRL F shortcut for Searching on page does not work in HTML reports. The workaround is to right-click and select "Find".

- The "Page Setup" function when printing reports is slow to open and ignores the Landscape option.
- The "*Copy Layout*" / "*Paste*" commands do not paste all copied nodes when copying the SuperNode layout to a new workflow.

## 6. Client Service and Support: Contacting Altair

### 6.1 Additional Resources

The following resources are available for our customers on the product download site and elsewhere on the Altair website:

- **User Guides** in PDF.
- **Deployment Guide** - Installation guide for the standalone and client/server configurations of Knowledge Works software for all supported platforms.
- **Angoss Server Administration Utilities** - A user guide to simple utilities for usage monitoring and user process administration for the client/server configuration.
- Brochures, Videos, Case Studies, etc.:  
<https://www.altair.com/resourcelibrary/?brand=Knowledge%20Works>
- User Community: <https://community.datawatch.com/home>

### 6.2 Reporting Issues and Requesting Product Enhancements

If you have any questions or problems with the software or documentation, please contact Customer Support:

- Customer Support Portal: <https://www.altair.com/knowledge-works-support/>
- E-Mail: [support@datawatch.com](mailto:support@datawatch.com)
- Telephone: 1-888-715-7391 (toll-free within North America) or 1-416-593-2417.

We welcome your suggestions on how to make our application easier to use and understand. Please send your comments and suggestions via the Customer Support Portal or email [support@datawatch.com](mailto:support@datawatch.com).



## About Altair | Knowledge Works

Altair is a global leader in delivering advanced analytics to businesses looking to improve performance across risk, marketing and sales.

Altair removes the complexity inherent with predictive analytics and machine learning with a platform that is intuitive to use and rich in features. It is designed so that Data Scientists and Data Citizens alike can build models used to help solve business problems.

<https://www.altair.com/>

### **Altair Engineering, Inc. World Headquarters**

1820 E. Big Beaver Rd.  
Troy MI 48083  
United States  
Phone: (248) 614-2400  
Fax: (248) 614-2411