Data Community

SQL Day

PLATINUM SPONSOR

TECHNOLOGY INNOVATION DATA KNOWLEDGE tidk

STRATEGIC PARTNER

Microsoft

GOLD SPONSORS

CLOUDS ON MARS

Future Processing

SPYROSOFT

SILVER SPONSOR

dbWatch DATABASE CONTROL

elitmind think bright

VEEAM

BRONZE SPONSOR

Największa księgarnia IT w Polsce

NOVATECH www.novatech.com.pl

Failure prediction for manufacturing industry

Andrea Martorana Tusa

Speaker info

- Microsoft **MVP** Data Platform
- **Team Leader** for **Business Intelligence** in Würth-Phoenix (Bolzano, Italy). Deep knowledge on ERP and CRM processes. Implementing BI solutions based on Microsoft Dynamics AX and 365.
- Previously worked by Widex a danish company as **BI Specialist**
- Speaker at SQL Saturdays, and other community-driven events in Europe
- Author for sqlservercentral.com, sqlshack.com, UGISS (User Group Italiano SQL Server)

- Twitter: @bruco441 andrea.martoranatusa@gmail.com

Agenda

- R in SQL Server 2016
- R for predictive maintenance
- Predicting failure. Build predictive models
- Output of R in SSMS, SSRS
- R in Power BI for data analysis and visualization

SQL Server Machine Learning Services

SQL Server Machine Learning Services is an engine embedded into a SQL Server database, to deliver advanced analytics calculations and perfom data science operations.

SQL Server Machine Learning Services is available from SQL Server 2017. Into the 2016 release it has been introduced the capability of executing **R** scripts. From 2017 release SQL Server can execute **Python** code too.

R

R is the most common open source programming language for data analysis and statistics.

R in SQL Server

Available from 2016 release.

- Two flavors. Options available during installation:
  - **R Services (In-Database)** integrated directly into SQL Server engine.
  - **Microsoft R Server** a standalone R server installed alongside with the database but externally.

- In addition there is a community edition **Microsoft R Open** and a community client **Microsoft R Client**.

R in SQL Server

Scalability and performance

Any R script is executed by calling a stored procedure, and the outputs are returned as tabular results directly to SQL Server.

Performance optimization is based on scalability and parallelization.
Two packages especially are available for handling resource governance inside SQL Server:

- **RevoScaleR** a set of functions for processing, analyzing, querying data with fast perfomances and scalability.
- **RevoPemarR** for writing custom parallel algorithms.

R in SQL Server

Execute R script within SQL Server

R scripts are invoked by a new system stored procedure **`sp_execute_external_script`** to embed R (or Python) code into a format suitable for SQL Server. The stored procedure executes the scripts and retrieves the output.

Some parameters are passed to the sp in order to define, language, context, input data, etc …

R in SQL Server

```
EXEC sp_execute_external_script
@language = N'R',
@script= N'OutputDataSet<-InputDataSet',
@input_data_1 = N'SELECT 1 AS hello‘
WITH RESULT SETS (([hello] int not null));
GO
```

Language declaration

R script

Input data

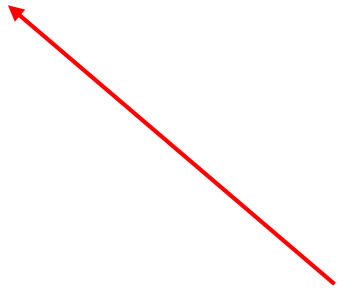
Resultset from R

The above R script generates the following outcome

	hello
1	1

R in SQL Server

```
EXEC sp_execute_external_script
  @language = N'R',
  @script= N'OutputDataSet<-InputDataSet',
  @input_data_1 = N'SELECT 1 AS hello'
  WITH RESULT SETS (([hello] int not null));
```



N.B.! All the output from R to SQL Server, must return the data as a **data.frame**!
Any other type of object that you generate in your script must be converted to a data frame if you want to output it as part of the stored procedure outputs.

Predictive maintenance for manufacturing industry

*”Predictive modeling is the process of using advanced statistics and probability algorithms to predict outcomes, based on a pretrained and built model or function.”**

Predictive maintenance solutions can provide key performance indicators about health state and lifecycle for products in manifacturing industries. Some use cases could be:

- Aircraft component failure
- Wind turbine failure
- Circuit breaker failure

- And, of course ……. Hearing aids

* Tomaz Kastrun, Julie Koesmarno. SQL Server 2017 Machine Learning Services with R. Packt Publisher

Predictive maintenance for manufacturing industry

R & SQL Server can be used together for building a predictive model to address your business needs.

For example, we want to predict the probability of failure for a bicycle component produced by the company **Adventure Works**.

A predictive maintenance solution helps reducing repair costs and increasing the lifecycle of our cycles. These models also give insights on correlations among data, helping to reveal patterns or outliers in the production process.

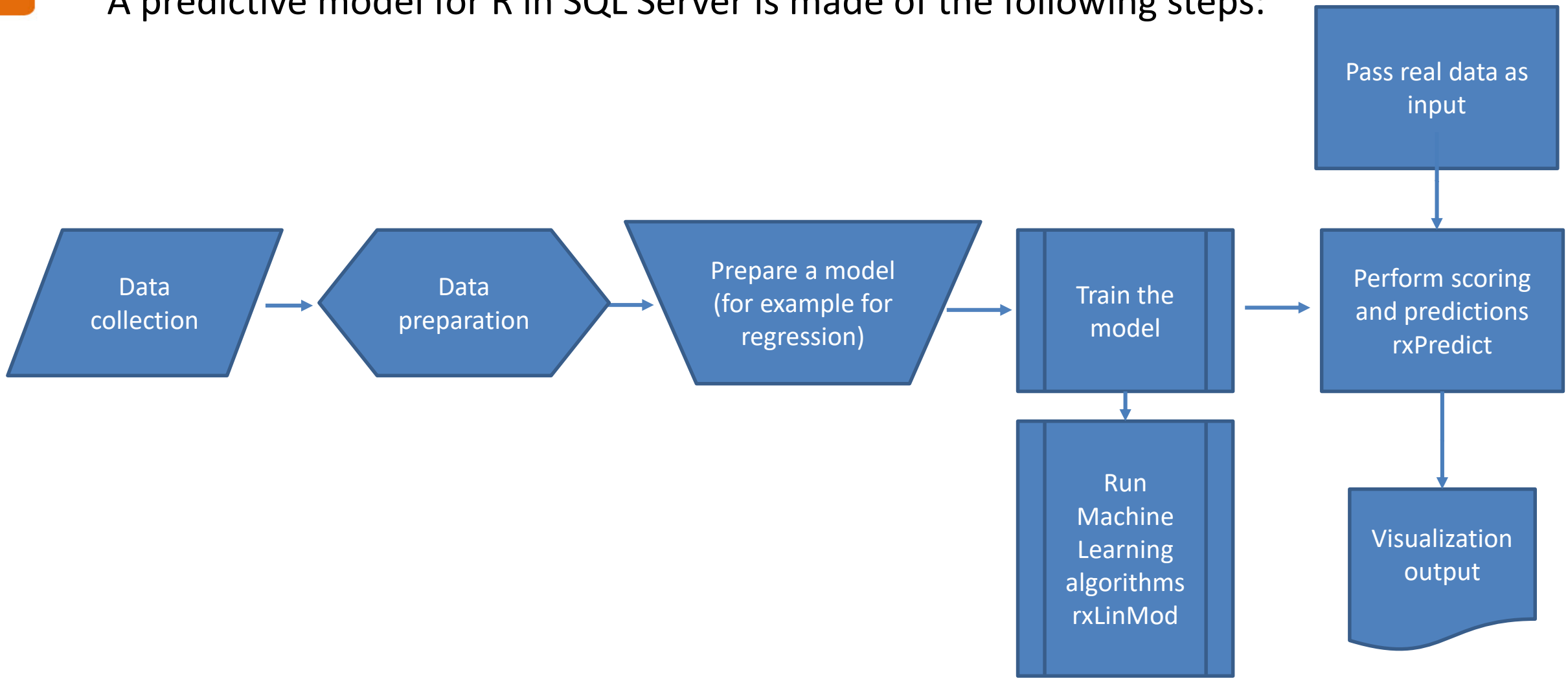
Predictive maintenance for manufacturing industry

Tipical problems addressed by a predictive maintenance solution:

- Predict Time to Failure (TTF) for a device
- Predict which component has the highest likelihood to fail
- Predict if an asset will fail in different time windows
- …

Predictive model for R in SQL Server

A predictive model for R in SQL Server is made of the following steps:



Predictive model for R in SQL Server

RevoScaleR package

01/29/2018 • 9 minutes to read • Contributors 

The **RevoScaleR** library is a collection of portable, scalable, and distributable R functions for importing, transforming, and analyzing data at scale. You can use it for descriptive statistics, generalized linear models, k-means clustering, logistic regression, classification and regression trees, and decision forests.

Functions run on the **RevoScaleR** interpreter, built on open-source R, engineered to leverage the multithreaded and multinode architecture of the host platform.

https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revoscaler/revoscaler

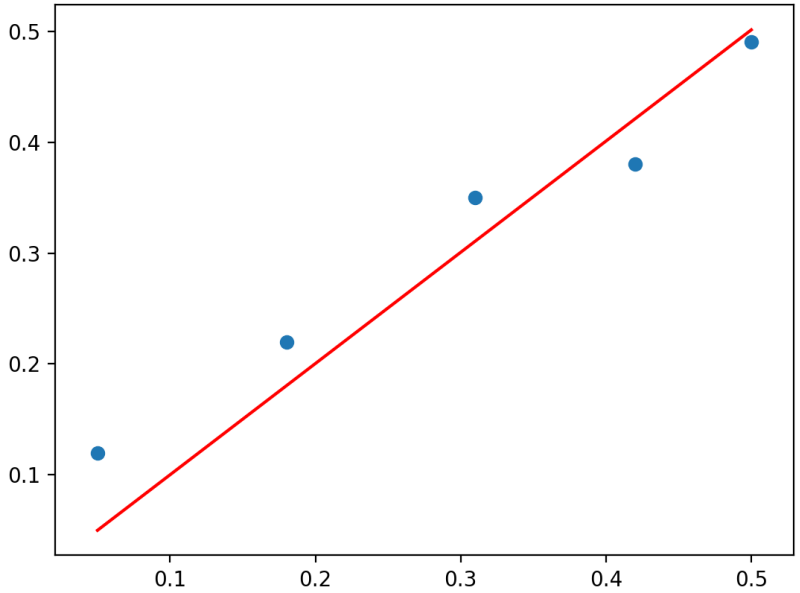
RevoScale R Functions and Machine Learning algorithms

rxLinMod: Linear Models

01/24/2018 • 10 minutes to read • Contributors

Description

Fit linear models on small or large data.



Simple regression model

https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revoscaler/revoscaler

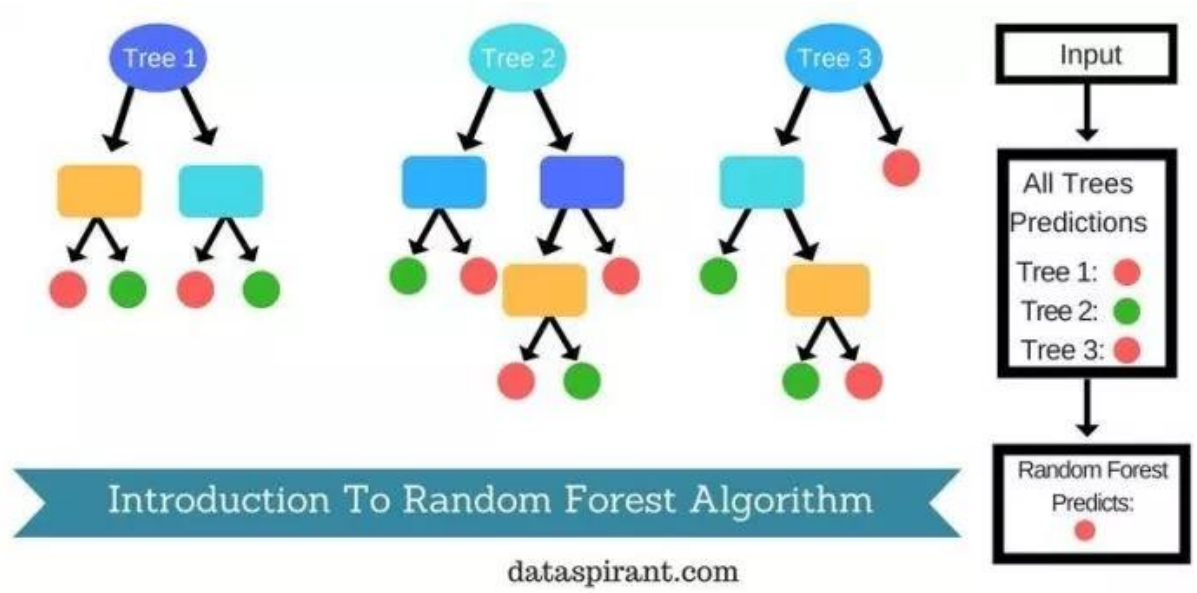
RevoScale R Functions and Machine Learning algorithms

rxDForest: Parallel External Memory Algorithm for Classification and Regression Decision Forests

01/24/2018 • 12 minutes to read • Contributors

Description

Fit classification and regression decision forests on an .xdf file or data frame for small or large data using parallel external memory algorithm.



`rxDForest` is a parallel external memory decision forest algorithm targeted for very large data sets. It is modeled on the random forest ideas of Leo Breiman and Adele Cutler and the randomForest package of Andy Liaw and Matthew Weiner, using the tree-fitting algorithm introduced in rxDTree.

https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revoscaler/revoscaler

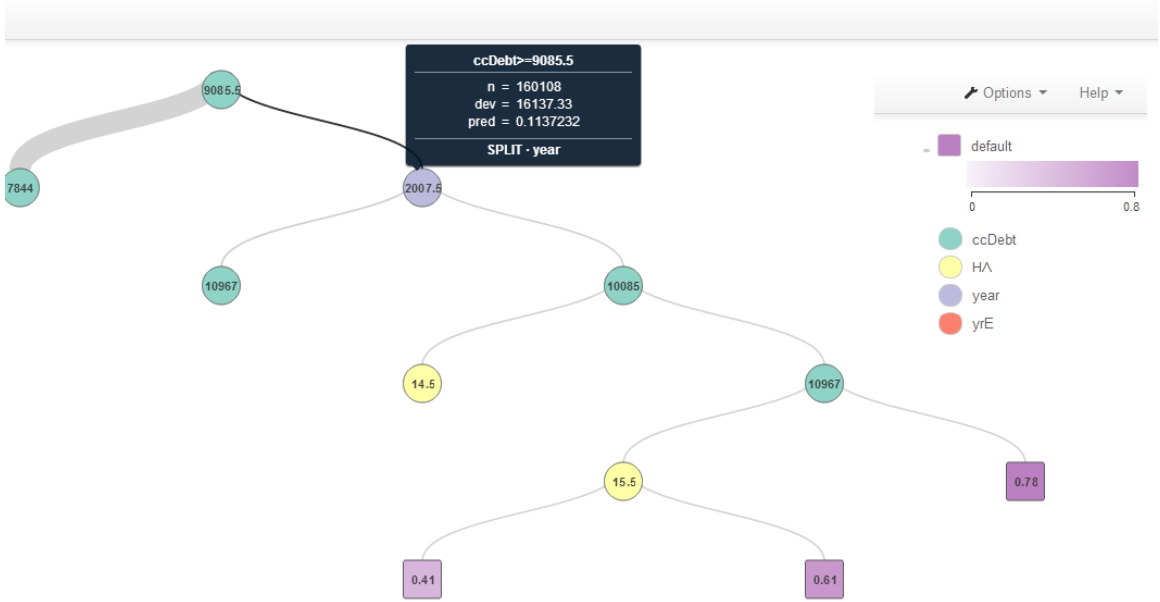
RevoScale R Functions and Machine Learning algorithms

rxBTrees: Parallel External Memory Algorithm for Stochastic Gradient Boosted Decision Trees

01/24/2018 • 13 minutes to read • Contributors

Description

Fit stochastic gradient boosted decision trees on an .xdf file or data frame for small or large data using parallel external memory algorithm.



`rxBTrees` is a parallel external memory algorithm for stochastic gradient boosted decision trees targeted for very large data sets. It is based on the gradient boosting machine of Jerome Friedman and Trevor Hastie and Robert Tibshirani and modeled after the gbm package of Greg Ridgeway with contributions from others, using the tree-fitting algorithm introduced in rxDTree.

https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revoscaler/revoscaler

Functions and Machine Learning algorithms

rxPredict: Predicted values and residuals for model objects built using RevoScaleR

01/24/2018 • 8 minutes to read • Contributors

Description

Compute predicted values and residuals using the following objects: rxLinMod, rxGlm, rxLogit, rxDTree, rxBTrees, and rxDForest.

`rxPredict` computes predicted values and/or residuals from an existing model type. The most common way to call rxPredict is `rxPredict(modelObject, data, outData)`. Typically, all the other arguments are left at their defaults.

https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revoscaler/revoscaler

DEMO

- Linear regression model in R studio
- Descriptive statistics with R in SSMS
- Linear regression model with R in SSMS
- Multivariate analysis
- Analysis of Variance
- Binary and multi class models
- Plotting output in Reporting Services
- Models and output in Power BI

Data Community

SQL Day

PLATINUM SPONSOR

TECHNOLOGY INNOVATION DATA KNOWLEDGE tidk

STRATEGIC PARTNER

Microsoft

GOLD SPONSORS

CLOUDS ON MARS

Future Processing

SPYROSOFT

SILVER SPONSOR

dbWatch DATABASE CONTROL

elitmind think bright

VEEAM

BRONZE SPONSOR

Największa księgarnia IT w Polsce

NOVATECH www.novatech.com.pl

References

- Tomaz Kastrun, Julie Koesmarno. ”SQL Server 2017 Machine Learning Services with R”. Packt
- https://docs.microsoft.com/en-us/sql/advanced-analytics/what-is-sql-server-machine-learning?view=sql-server-2017
- https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/cortana-analytics-playbook-predictive-maintenance
- https://gallery.azure.ai/Tutorial/Predictive-Maintenance-Template-with-SQL-Server-R-Services-1
- https://github.com/Microsoft/SQL-Server-R-Services-Samples/blob/master/PredictiveMaintenance/SQLR/README.md