Predicting Hospital Length of Stay (in days) using SQL Server with R Services

Xinwei Xue, PhD
Principal Data Scientist Manager
Microsoft, Azure Data Science Group
xixue@Microsoft.com

Carl Saroufim, MSc
Data Scientist
Microsoft, Azure Data Science Group
casarouf@Microsoft.com

SQL Server 2016 with R Services

Develop a prediction model in SQL Server from R IDE

Data Scientist Workstation

R IDE

Microsoft R Open

Microsoft R Server

```
cs <- "Driver=SQL Server;
Server=[Server Name];Database=[DB
Name];UID=[User ID];PWD=[Password]"
sqlCompute <- RxInSqlServer(cs)
rxSetComputeContext(sqlCompute)
forest_model <- rxDForest()
```

1 Script

3 Results

2 Execution

SQL Server 2016

Advanced Analytics Extension

Microsoft R Open

Microsoft R Server

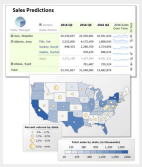
Operationalize the model with SQL R Services

2 **Stored Procedure with R code, executing in-database**

```
exec
sp_execute_external_script
   @language = 'R'
,  @script =
-- R code --
```

1 Call System Stored Procedure

3 Results (scores, plots)



Application

Data Science Process

Import the Input Dataset to SQL

LengthOfStay

29 variables including:
- Encounter ID (Key)
- Length of Stay (Target variable)
- Continuous health measurements (glucose levels, pulse rate, etc.)
- Flags for an illness (diabetes, asthma, etc.)

Preprocessing & Feature Engineering

- **Missing Value Replacement:**
  - Global mean (continuous var.)
  - Global mode (discrete var.)
- **Feature Engineering:**
  - Standardize continuous var.
  - Number of issues: how many illnesses flags does the patient have?

Splitting

70% Training Set

30% Testing Set

Training & Testing

Random Forest (rxDForest)

Gradient Boosted Trees (rxFastTrees)

Evaluation

Predictions on Test Set

Predictions on Test Set

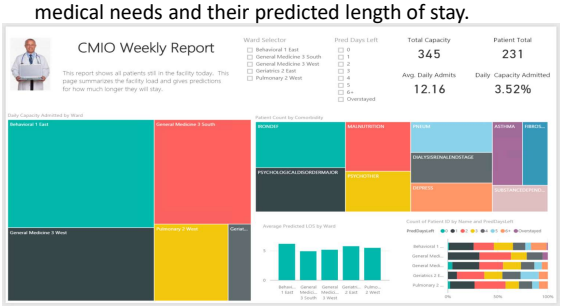
Evaluation & Model selection based on mean absolute error (for e.g.)

Business Manager Decision Making: Predictions & Visualization in PowerBI

Predictions & Medical data for new patients

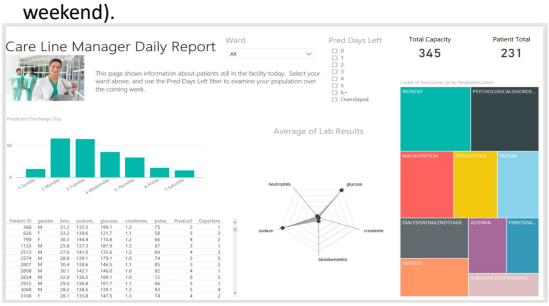
Chief Medical Information Officer of the hospital

Provides recommendations on the:
- **Transfer or re-routing of incoming patients** to less populated facilities, using the predicted length of stay in each facility.
- **Resources allocation** in facilities based on patients medical needs and their predicted length of stay.



Care Line Manager for a given ward

Monitors the:
- **Patients individual status.**
- **Staff availability for the ward** based on patients care requirement and their predicted length of stay (especially when the predicted discharge date is during the weekend).



Code & Deployment

Github Code: https://github.com/Microsoft/r-server-hospital-length-of-stay
Solution Template: One click deployment: http://aka.ms/hospital-los