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

**Data Science Process**

1. Import the Input Dataset to SQL
2. 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?
3. Splitting
4. Training & Testing
   - Gradient Boosted Trees (rxFastTrees)
5. Evaluation & Model selection based on mean absolute error (for e.g.)

**Business Manager Decision Making: Predictions & Visualization in PowerBI**

- **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](https://github.com/Microsoft/r-server-hospital-length-of-stay)
- Solution Template: One click deployment: [http://aka.ms/hospital-los](http://aka.ms/hospital-los)