Healthcare predictive analytics today requires the processing of big data relating to hospital and patients administrative data, clinical and non-clinical data including patient demographics, disease diagnoses and procedures, patient charges, medical health records, discharge status. This big data needs to be processed and analyzed to extract knowledge for decision-making and cost-saving. Data mining techniques provide a set of tools that can be applied to detect patterns, classifications, hospital transfers, and mortality. In this session we demonstrate data mining techniques including Decision Trees, Logistic Regression, Neural Networks, and Survival Data Mining using an example with hospital data, to identify at which medical state a patient should be transferred to tertiary hospital, and to predict the probability of mortality, and when the next event will occur.
This presentation is a Practitioner’s Guide on Data Mining Techniques as Healthcare Analytics for Big Healthcare data. We explain:

- How can we implement Healthcare Analytics Data Mining Techniques?
- Example Applications
What is Healthcare Analytics Data Mining?

• What is Healthcare data? – Large data Management
  • Huge amount of data generated by healthcare transactions (too complex and voluminous to be processed and analysed by traditional methods)
  • New methods needed – Data Mining provides the methodology and technology to transform these mounds of data into useful information for decision making.

• Goal – improve the quality and cost of healthcare
  • Healthcare insurers detect fraud and abuse
  • Healthcare organizations make customer relationship management decisions
  • Physicians identify effective treatments and best practices
  • Patients receive better and more affordable healthcare services.
IMPACTS...

Lower Costs

Better Health Outcomes

Improved Quality of Life

Through Prediction and Prevention
Step 1: Data Checks, Segmentation Suites

Description: Cleaned data for modeling

Data checks, Feasibility Analysis

EM Input data, StatExplore, Multiplot, Impute nodes

Output Clean data for modeling

Step 2: Variable Reduction

Identify Patient risk drivers

Fine/Coarse Classing, Variable Clustering, Correlation, VIF Analysis

Candidate variables and Shortlist variables

Step 3: Segmentation Variable Selection

Discrete & Continuous Segmentation variables

Decision Tree, Segment Profiling Analysis

Optimum Segmentation split on segmentation variables

Step 4: Model Build

Build both a whole patient level and segment level models

Fine/Coarse Classing, Correlation, VIF, Logistic Regression

EM Decision Tree, Clustering, Logistic Regression, Segment Profiling nodes

EM Interactive Grouping, Scorecard, SAS code, Variable Clustering, Segment Profiling nodes

Model Estimates, Parent Gini & System Gini’s

Step 5: Model Comparison

Compare the best System Gini with the Parent Gini

Compare Gini’s, consider Benefits of segmentation

EM SAS code, Model Comparison nodes

Identify if the Segmentation is effective and required
COMPREHENSIVE APPROACH
ANALYTICS TOOLS

Discovery

Modeling

Deployment

Operationalize

Automated In-Database
Real Time
Embedded in Applications

1. Development

2. Test/Stage

3. Deploy

4. Track

5. Retire
MODEL DEVELOPMENT PROCESS

S - Sample
   - Input Data
   - Data Partition
   - Sample
   - Filter
   - Append
   - Merge
   - Time Series

E - Explore
   - Association
   - Market Basket
   - Path Analysis
   - Cluster
   - SOM/Kohonen
   - DMDB
   - Graph Explore
   - Multiplot

M - Modify
   - StatExplore
   - Variable Clustering
   - Variable Selection
   - Interactive Binning
   - Principal Components
   - Replacement
   - Rules Builder
   - Transform Variables
   - Gradient Boosting
   - LARS
   - MBR

M - Model
   - AutoNeural
   - Neural Network
   - Partial Least Squares
   - Regression
   - Decision Tree
   - Random Forest
   - Coefficient Regression
   - Decision Trees
   - Rule Induction
   - Cutoff

A - Assess
   - Model Comparison
   - Model Import
   - Score
   - Decisions
   - Profile
   - Segment Profile

Utility
   - Control Point
   - SAS Code
   - Metadata
   - Reporter
Example using SAS Enterprise Minter

- Identify Transfer to Tertiary Hospital
- Predicting Probability of Mortality
## Scorecard

<table>
<thead>
<tr>
<th>AccomCharges</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>accomcharges&lt; 2700</td>
<td>-3</td>
</tr>
<tr>
<td>2700 &lt;= accomcharges&lt; 3366</td>
<td>5</td>
</tr>
<tr>
<td>3366 &lt;= accomcharges&lt; 17452, <em>MISSING</em></td>
<td>10</td>
</tr>
<tr>
<td>17452 &lt;= accomcharges&lt; 25488</td>
<td>28</td>
</tr>
<tr>
<td>25488 &lt;= accomcharges</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Admit-Dx-Code</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>29300, 311, 40391, 41011, 41041, 41091, 4111, 41199, 4240, 42721, 43141, 44021, 44023, 44422, 4539, 49122, 514, 51891, 5715, 5722, 5762, 5780, 5781, 6822, 7071, 71599, 7197, ...</td>
<td>-8</td>
</tr>
<tr>
<td>25000, 2592, 2599, 20402, 40291, 41071, 41090, 4139, 4241, 42731, 42732, 4275, 42789, 4279, 4280, 43491, 4539, 4549, 4538, 5609, 5750, 5865, 6827, 70809, 70809, 7840, 78606, 78609, 78650, ...</td>
<td>5</td>
</tr>
<tr>
<td>2761, 2859, 389, 41400, 41401, 438, 486, 49121, 496, 6070, 5119, 5849, 588, 5890, 5897, 7802, 7804, 7808, 78079, 78099, 78099, 8208</td>
<td>10</td>
</tr>
<tr>
<td>1619, 1638, 1650, 1629, 1748, 1820, 1830, 186, 1977, 20280, 20300, 20500, 2113, 25002, 25011, 25012, 25013, 25070, 25080, 25082, 2765, 27651, 2786, 2787, 2788, 2769, 2800, 28282, ...</td>
<td>84</td>
</tr>
<tr>
<td>1104, 1190, 126, 1369, 1431, 1459, 1508, 1508, 1510, 1518, 1520, 1530, 1522, 1633, 1634, 1635, 1537, 1540, 1542, 1548, 1570, 1578, 1619, 1623, 1628, 1729, 1737, 1745, 1749, 1808, ... , <em>MISSING</em>, ...</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>age&lt;_74, <em>MISSING</em></td>
<td>5</td>
</tr>
<tr>
<td>74 &lt;= age&lt; 76</td>
<td>1</td>
</tr>
<tr>
<td>76 &lt;= age&lt; 79</td>
<td>0</td>
</tr>
<tr>
<td>79 &lt;= age&lt; 89</td>
<td>15</td>
</tr>
<tr>
<td>89 &lt;= age</td>
<td>35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CMC1</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1603, 1707, 3013, 303, 303, 304, 307, 3308, 3506, 3625, 3738, 3807, 3810, 4143, 4502, 4601, 5008, 5204, 602</td>
<td>-5</td>
</tr>
<tr>
<td>1701, 1702, 2211, 3408, 3803, 3820, 3922, 403, 4101, 4106, 4138, 4142, 4144, 4228, 4491, 4606, 4808, 5001, 5013, 5203</td>
<td>4</td>
</tr>
<tr>
<td>1001, 101, 105, 106, 107, 108, 1102, 1103, 1202, 1301, 1401, 1404, 1406, 1409, 1409, 1501, 1502, 1511, 1524, 1526, 1567, 1608, 1706, 1801, 1803, 207, 208, 209, 2102, 211, ... , <em>MISSING</em>, ...</td>
<td>10</td>
</tr>
<tr>
<td>1601, 1704, 2401, 3605, 3620, 3906, 3918, 401, 406, 4135, 4322, 4371, 5004, 5005, 5012, 5412, 703, 704</td>
<td>19</td>
</tr>
<tr>
<td>102, 105, 1201, 1203, 1205, 1207, 1208, 1605, 1606, 1703, 203, 2203, 2304, 2403, 2404, 2410, 2413, 2602, 2503, 2504, 2912, 3022, 306, 3303, 3502, 3505, 3615, 3622, 3623, 3626, ...</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CMC2</th>
<th>Scorecard Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>103, 1603, 1604, 2404, 301, 302, 307, 390, 4008, 4007, 4123, 4127, 4142, 4143, 4319, 4402, 4609, 4610, 5008, 5203, 5302, 5407, 5418, 712</td>
<td>-3</td>
</tr>
<tr>
<td>0, 105, 106, 1201, 1202, 1205, 1301, 1401, 1402, 1406, 1408, 1410, 1601, 1607, 1701, 1703, 1707, 1702, 2001, 2006, 201, 206, 2202, 2206, 2304, 2308, 2307, 2308, 2310, 2402, 2410, ... , <em>MISSING</em>, ...</td>
<td>7</td>
</tr>
<tr>
<td>2401, 2403, 3191, 3608, 3803, 3915, 3925, 3930, 4002, 4003, 4048, 4102, 4222, 4371, 4502, 4506, 4511, 5005, 5012, 5412, 707</td>
<td>15</td>
</tr>
<tr>
<td>3817</td>
<td>28</td>
</tr>
</tbody>
</table>
## Model Event Classification Table

<table>
<thead>
<tr>
<th>Model</th>
<th>Data</th>
<th>Role</th>
<th>Target</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorecard2</td>
<td>Scorecard</td>
<td>TRAIN</td>
<td>transfer</td>
<td>310</td>
<td>9568</td>
<td>41</td>
<td>79</td>
</tr>
<tr>
<td>Scorecard2</td>
<td>Scorecard</td>
<td>VALIDATE</td>
<td>transfer</td>
<td>263</td>
<td>7167</td>
<td>41</td>
<td>29</td>
</tr>
<tr>
<td>Scorecard</td>
<td>Scorecard</td>
<td>TRAIN</td>
<td>transfer</td>
<td>338</td>
<td>9579</td>
<td>30</td>
<td>51</td>
</tr>
<tr>
<td>Scorecard</td>
<td>Scorecard</td>
<td>VALIDATE</td>
<td>transfer</td>
<td>273</td>
<td>7176</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>Tree</td>
<td>Decision</td>
<td>TRAIN</td>
<td>transfer</td>
<td>354</td>
<td>9609</td>
<td>.</td>
<td>35</td>
</tr>
<tr>
<td>Tree</td>
<td>Decision</td>
<td>VALIDATE</td>
<td>transfer</td>
<td>256</td>
<td>7204</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>Reg</td>
<td>Regression</td>
<td>TRAIN</td>
<td>transfer</td>
<td>357</td>
<td>9607</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>Reg</td>
<td>Regression</td>
<td>VALIDATE</td>
<td>transfer</td>
<td>265</td>
<td>7198</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>Neural</td>
<td>Neural</td>
<td>TRAIN</td>
<td>transfer</td>
<td>389</td>
<td>9609</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neural</td>
<td>Neural</td>
<td>VALIDATE</td>
<td>transfer</td>
<td>292</td>
<td>7208</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>