Practical Customer Analytics using Predictive Approaches

A Dell Statistica® Webinar

Dean Abbott
SmarterHQ, Inc. and Abbott Analytics
October 8, 2015

Email: dean@abbottanalytics.com
Blog: http://abbottanalytics.blogspot.com
Twitter: @deanabb
Overview & History

Established in 1984
Deep understanding of analytics

Over 1,000,000 users
Proven solution with built-in expertise

Over 16,000 functions
Integrated & comprehensive

Flexible & extendible
Not a black-box, fits in your environment

Process Monitoring & Quality Control
Largest manufacturer or enterprise-wide solutions

Open standards
Avoids vendor lock

Dell Statistica
Embed Analytics Everywhere with Statistica

- Easy to use predictive analytics with built-in smarts to enable your workforce
- Data blending on any data, anywhere using databases, cloud, and Hadoop sources
- Data discovery & visualization to distribute and share relevant insights
- Real-time analytics to process high-volume, streaming data
- Collaborate & share insights and best practices across geographies
- Combine rules & analytics to make prescriptive business decisions
- Security & governance for a well-managed approach to analytics
- Open, flexible, & extensible to adapt to specific use cases and fit with existing investments

More than 1 million users worldwide
Intuitive and easy to use
Data Discovery & Visualization
Why Statistica?

**Pharma**
7 of 10 largest companies use Statistica

**Marketing**
for churn at one of the largest telecoms in EMEA

**Finance**
used by largest banks for real-time credit scoring

**Manufacturing**
used by the largest semiconductor fabs

**Healthcare**
real-time analytics saving lives

**University**
used for leading edge research world-wide

**#1 Cloud**
solution used for cloud-analytics projects

**1,000,000**
users worldwide

**16,000**
functions available from one user interface

**1,000,000**
users worldwide
Who am I?

• Applying predictive analytics to real applications since 1987

• Independent consultant (Abbott Analytics) since 1999.
  • Taught 1000s of professionals data mining and predictive analytics
  • Frequent speaker at conferences: Predictive Analytics World, TDWI, PASS Business Analytics, Open Data Science Conference, more

• Co-Founder and Chief Data Scientist SmarterHQ
  • Behavioral marketing SaaS powered by machine learning
Motivating Examples: Retail

- Retail
  - Millions of customers participating in tens of millions visits and purchases

- Objectives
  - Increase engagement with customers
  - Understand intent of visits: exploring? Aspirational shopping? Poised for a purchase?

- Method
  - Augment behavioral segmentation with purchase propensity models (# days to next purchase predictions)
Motivating Examples: Fraud Detection

- Federal Computer Week, Jan 24, 2005
- Fraud detection using data mining.
  - Identify potential misuse of government purchase cards.
  - Data mining applied to purchase cards started in 1996 with Defense Finance and Accounting Service.
  - Data mining models identified 1,357 cardholder to investigate. After review, 182 flagged for investigation.
  - Data mining here was used as a filter to flag cardholders for further investigation, rather than to provide decisions on who made fraudulent purchases

DOD mines data to detect abuse

Officials hope to use data to improve government purchase card policies
BY Michael Hardy
Published on Jan. 24, 2005

Defense Department officials have gathered data on the misuse of government purchase cards, and they hope to create policies based on that data.

Col. William Kelley, program director of the Army's data-mining division in the inspector general's office, said government credit card users could be misspending millions of taxpayer dollars.

Agency officials issue the cards, which look and function like ordinary credit cards. Employees use them for purchases less than $2,500. Most card purchases are for office supplies or building materials, Kelley said.

Studies that Kelley's office conducted in 2002 and 2003 looked at 1,357 cardholders flagged by data-mining tools. The cardholders accounted for 13,052 flagged transactions out of 562,499 transactions available for review.

Officials identified 182 cardholders who had used their cards fraudulently or who warranted further investigation.
Motivating Examples: Non-Profit Donation Models

- KDD Cup Competition, 1998
- Lapsed Donor Identification.
  - Test mailing to lapsed donors, 191K of them
  - Observe who responded
  - Build predictive models that predict
    - Likelihood to respond
    - Amount of gift
  - Rank-order population according to metric Cumulative Net Revenue
Predictive Analytics Projects

CRISP-DM
What do Predictive Modelers do?

The CRISP-DM Process Model

- **CRoss-Industry Standard Process Model for Data Mining**
- Describes Components of Complete Data Mining Cycle from the Project Manager’s Perspective
- Shows Iterative Nature of Data Mining
CRISP-DM: Business Understanding Steps

- Ask Relevant Business Questions
- Determine Data Requirements to Answer Business Question
- Translate Business Question into Appropriate Data Mining Approach
- Determine Project Plan for Data Mining Approach
Characterize data available for modeling. Provides assessment and verification of data.
CRISP-DM Step 2: Data Understanding Steps

- Collect initial data
  - Internal data: historical customer behavior, results from previous experiments
  - External data: demographics & census, other studies and government research
  - Extract superset of data (rows and columns) to be used in modeling
  - Identify form of data repository: multiple vs. single table, flat file vs. database, local copy vs. data mart

- Perform Preliminary Analysis
  - Characterize Data (describe, explore, verify)
  - Condition Data
Condition existing data and construct new data to aid in model predictions.
CRISP-DM Step 3: Data Preparation (Conditioning) Steps

Fix Data Problems

Create Features

- Select Data
- Clean Data
- Construct Data
- Integrate Data
- Format Data
- Rationale for Inclusion/Exclusion
- Data Cleaning Report
- Derived Attributes
- Merged Data
- Reformatted Data
- Generated Records
The CRISP-DM Process Model

Build Predictive Models

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Deployment
- Evaluation
CRISP-DM Step 4: Modeling Steps

Algorithm Selection
- Select Modeling Techniques
  - Model Ranking
  - Model Assessment
  - Model Description

Sampling
- Generate Test Design
  - Test Design
  - Parameter Settings

Algorithms
- Build Model
  - Parameter Settings
  - Models

Model Ranking
- Assess Model
  - Revised Parameter Settings
The CRISP-DM Process Model

Evaluate
Models
CRISP-DM Step 5: Evaluation Steps

- Score models (assess results)
  Is model good enough?

- Review model
  Did we miss anything?
  Any assumptions violated?

Next Step
Deploy vs. recreate models

Diagram:
1. Evaluate Results
2. Assessment of Data Mining Results
3. Approved Models
4. Review Process
5. Review of Process
6. Determine Next Steps
7. List of Possible Actions
8. Decisions
The CRISP-DM Process Model

Deploy Model
CRISP-DM Step 6: Deployment Steps

How to deploy model?
Software, source code, in database

How often, when to update model

Report results

Lessons learned
Data Preparation

What do we need to do?
Do for algorithms what they can’t do for themselves

- Get the data right
- Understand how algorithms can be fooled with “correct” data – flag potential data problems
  - Missing Values
  - Outliers and Skew
  - High Cardinality
- Improve data by building features
Dependence on Algorithms

- Neural Networks
- Linear Regression
- Logistic Regression
- K Nearest Neighbor
- PCA
- Nearest Mean
- Kohonen Self-Organizing Maps
- Support Vector Machines
- Radial Basis Function Networks
- Discriminant Analysis

- MUST!!
  - Fill missing values
  - Explode categorical variables
- Sometimes
  - automatic in software; beware!
  - software fails: error
  - Algorithm assume distributions so beware of skew, kurtosis

- Categorical variables are fine
- Numeric data must be binned (except some decision trees)
- Outliers don’t matter
- Missing values often treated as a separate category
Clean Data: Missing Values

• Missing data can appear as
  • blank, NULL, NA, or a code such as 0, 99, 999, or -1.

• Fixing Missing Data:
  • Delete the record (row), or delete the field (column)
  • Replace missing value with mean, median, or distribution
  • Replace with the missing value with an estimate
    • Select value from another field having high correlation with variable containing missing values
    • Build a model with variable containing missing values as output, and other variables without missing values as an input

• Other considerations
  • Create new binary variable (1/0) indicating missing values
  • Know what algorithms and software do by default with missing values
    • Some do listwise deletion, some recode with “0”, some recode with midpoints or means
Clean Data: Missing Data

- How much can missing data effect models?
- Example at upper right has 5300+ records, 17 missing values encoded as “0”
- After fixing model with mean imputation, R^2 rises from 0.597 to 0.657
- Why? Missing was recoded with “0” in this example, which was a particularly bad imputation for this data
Constructing Data: Feature Creation

• What is a feature?
  • New version of one or more attributes; derived attributes

• Why create features?
  • Improved Classifier Accuracy and Robustness
    • Provide more predictive variables
    • Create variables difficult or impossible for classifiers to construct themselves.
  • Possibly reduce complexity of data mining models
  • Understandability and Insight

Represent Information with Most Descriptive Versions of Variables
What do top Kaggle competitors focus on?
To do well in a competition, clearly many aspects are important. But, what do you think helped you (or a top competitor you know) do better than others?

Vik Paruchuri, Founder at dataquest.io
19.2k Views • Upvoted by Anthony Goldbloom, Founder and CEO of Kaggle
Vik has 10+ answers in Machine Learning.

The Rest of the Factors

Now that I have addressed what I think is in the single most important factor (persistence), I will address the rest of your question:

1. The most important data-related factor (to me) is how you prepare the data, and what features you engineer. Algorithm selection is important, but much less so. I haven’t really seen the use of any proprietary tools among top competitors, although a couple of first place finishers have used open-source tools that they coded/maintain.
Clean Data: Outliers

- Are the outliers problems?
  - Some algorithms: “yes”
    - Linear regression, nearest neighbor, nearest mean, principal component analysis
    - In other words, algorithms that need mean values and standard deviations
  - Some algorithms: “no”
    - Decision trees, neural networks

- If outliers are problems for the algorithm
  - Are they key data points?
    - Do not remove these
    - Consider “taming” outliers with transformations (features)
  - Are they anomalies or otherwise uninteresting to the analysis
    - Remove from data so that they don’t bias models
Effect of Skew and Outliers on Correlations (and Regression)

- 4,843 records

Corresponds to $R^2$ increase from 0.42 to 0.53
Why Skew Matters (In Regression Modeling)

- Obscures information in plot
  - Spaced in scatterplot taken up by empty space in upper (or lower) end of skewed values

- Regression models fit worse with skewed data
  - In example at right, by simply applying the log transform, performance is improved from $R^2=0.566$ to 0.597
Taming Skew with Log10

TimeLag_log10 = log10(1 + TIMELAG)
Effect of Distance on Clusters
Effect of Distance on Clusters
Effect of Distance on Clusters
Effect of Distance on Clusters
Sample Transformations

- Maybe none; close enough to uniform.
- Primarily positive skew; consider log10 transform.
- Consider binning to capture four or five regions centered on spikes.
- Consider binning into two bins (for peaks) or four bins (two peak and two trough regions).
- Negative skew; consider power transform or flip transform and
- Consider dummy variables for spikes on left and right.
- Positive skew; consider a log transform.
- Spike in distribution, consider dummy indicator of the spike vs. non-spike. If spike generated through mean imputation, consider imputing with a distribution.
- Classic positive skew. Log10 transform.
Data Preparation

Sampling
Partitioning Data

- **Objective of Creating Model:** Generalization

- **Split Data into Distinct Data Sets (Partitions):**
  - Training Subset: Used to create model
  - Testing Subset: Used to assess model, then a decision is made whether to retrain model
  - Validation Subset: Used to provide final assessment of model

- **Each subset should be representative of the entire data set**
Model Overfit

Simple Model
Decision Boundary

Training Data

Testing Data

Complex Model
Decision Boundary

Outlier that influences decision boundary
Random Sampling into Subsets

Entire Dataset
Randomly select inclusion in Train, Test, and Validate subsets*

<table>
<thead>
<tr>
<th>Split</th>
<th>ID</th>
<th>AcctAge</th>
<th>City_clean</th>
<th>State_cln</th>
<th>Zip</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>1512</td>
<td>124</td>
<td>XENIA</td>
<td>OH</td>
<td>45385</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1562</td>
<td>157</td>
<td>DAYTON</td>
<td>OH</td>
<td>45424</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4427</td>
<td>94</td>
<td>PONTE V BEACH</td>
<td>FL</td>
<td>32082</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4902</td>
<td>268</td>
<td>WEST POINT</td>
<td>KY</td>
<td>40177</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3861</td>
<td>113</td>
<td>PASADENA</td>
<td>MD</td>
<td>21122</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5163</td>
<td>147</td>
<td>CLEVELAND</td>
<td>OH</td>
<td>44120</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2656</td>
<td>12</td>
<td>BEAUMONT</td>
<td>TX</td>
<td>77707</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>781</td>
<td>23967</td>
<td>AKRON</td>
<td>OH</td>
<td>44333</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>218</td>
<td>31</td>
<td>MCKINLEYVILLE</td>
<td>CA</td>
<td>95519</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3565</td>
<td>23967</td>
<td>SALT LAKE CITY</td>
<td>UT</td>
<td>84102</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3707</td>
<td>90</td>
<td>TOPSFIELD</td>
<td>MA</td>
<td>01983</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4603</td>
<td>150</td>
<td>TUSCALOOSA</td>
<td>AL</td>
<td>35403</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4079</td>
<td>167</td>
<td>BOWLING GREEN</td>
<td>KY</td>
<td>42103</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>565</td>
<td>32</td>
<td>BARBERTON</td>
<td>OH</td>
<td>44203</td>
<td>1</td>
</tr>
</tbody>
</table>

Training subset

<table>
<thead>
<tr>
<th>Split</th>
<th>ID</th>
<th>AcctAge</th>
<th>City_clean</th>
<th>State_cln</th>
<th>Zip</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>1512</td>
<td>124</td>
<td>XENIA</td>
<td>OH</td>
<td>45385</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1562</td>
<td>157</td>
<td>DAYTON</td>
<td>OH</td>
<td>45424</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4427</td>
<td>94</td>
<td>PONTE V BEACH</td>
<td>FL</td>
<td>32082</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4902</td>
<td>268</td>
<td>WEST POINT</td>
<td>KY</td>
<td>40177</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5163</td>
<td>147</td>
<td>CLEVELAND</td>
<td>OH</td>
<td>44120</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>218</td>
<td>31</td>
<td>MCKINLEYVILLE</td>
<td>CA</td>
<td>95519</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2636</td>
<td>12</td>
<td>BEAUMONT</td>
<td>TX</td>
<td>77707</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>781</td>
<td>23967</td>
<td>AKRON</td>
<td>OH</td>
<td>44333</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3861</td>
<td>113</td>
<td>PASADENA</td>
<td>MD</td>
<td>21122</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5163</td>
<td>147</td>
<td>CLEVELAND</td>
<td>OH</td>
<td>44120</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2656</td>
<td>12</td>
<td>BEAUMONT</td>
<td>TX</td>
<td>77707</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>781</td>
<td>23967</td>
<td>AKRON</td>
<td>OH</td>
<td>44333</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3861</td>
<td>113</td>
<td>PASADENA</td>
<td>MD</td>
<td>21122</td>
<td>0</td>
</tr>
</tbody>
</table>

Testing subset

<table>
<thead>
<tr>
<th>Split</th>
<th>ID</th>
<th>AcctAge</th>
<th>City_clean</th>
<th>State_cln</th>
<th>Zip</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>3861</td>
<td>113</td>
<td>PASADENA</td>
<td>MD</td>
<td>21122</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3815</td>
<td>261</td>
<td>SALT LAKE CITY</td>
<td>UT</td>
<td>84102</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3707</td>
<td>90</td>
<td>TOPSFIELD</td>
<td>MA</td>
<td>01983</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4603</td>
<td>150</td>
<td>TUSCALOOSA</td>
<td>AL</td>
<td>35403</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4079</td>
<td>167</td>
<td>BOWLING GREEN</td>
<td>KY</td>
<td>42103</td>
<td>0</td>
</tr>
</tbody>
</table>

Validation subset

<table>
<thead>
<tr>
<th>Split</th>
<th>ID</th>
<th>AcctAge</th>
<th>City_clean</th>
<th>State_cln</th>
<th>Zip</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validate</td>
<td>3861</td>
<td>113</td>
<td>PASADENA</td>
<td>MD</td>
<td>21122</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>3707</td>
<td>90</td>
<td>TOPSFIELD</td>
<td>MA</td>
<td>01983</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>3815</td>
<td>261</td>
<td>SALT LAKE CITY</td>
<td>UT</td>
<td>84102</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>4603</td>
<td>150</td>
<td>TUSCALOOSA</td>
<td>AL</td>
<td>35403</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>4079</td>
<td>167</td>
<td>BOWLING GREEN</td>
<td>KY</td>
<td>42103</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>4677</td>
<td>92</td>
<td>MOBILE</td>
<td>AL</td>
<td>36617</td>
<td>0</td>
</tr>
<tr>
<td>Validated</td>
<td>565</td>
<td>32</td>
<td>BARBERTON</td>
<td>OH</td>
<td>44203</td>
<td>1</td>
</tr>
</tbody>
</table>
Modeling

Key Classification Algorithms
Regression, decision trees, and cluster analysis continue to form a triad of core algorithms for most data miners. This has been consistent since the first Data Miner Survey in 2007.

The average respondent reports typically using 12 algorithms. People with more years of experience use more algorithms, and consultants use more algorithms (13) than people working in other settings (11).

The number of algorithms used varies by the labels people use to describe themselves, with Data Miners (14) and Data Scientists (14) using the most, and Software Developers (9) and Programmers (8) the fewest.

Question: What algorithms / analytic methods do you TYPICALLY use? (Select all that apply)
Why Learn Multiple Algorithms

- Hard to know in advance which algorithm will ‘win’
- Each algorithm has its own strengths and weaknesses
- Algorithms provide different interpretations of the data
Classifiers Find Different Decision Boundaries

Actual Data

11-Nearest Neighbor

Neural Network

Naïve Bayes

Logistic Regression

Decision Tree
Logistic Regression

- Creates linear decision boundaries (just like linear regression)
- More appropriate for classification
- Finding weights (the model) more complex: requires an iterative algorithm
- Understanding the weights also more complex (they represent the relative change in the log odds of the output)
- Like linear regression, requires numeric inputs—categorical variables are recoded as binary (dummy)
Decision Trees: Sample Tree

If $X_1 \leq 5$
- Then data point is “Class 1”

Else
- if $X_2 \leq 3$
  - Then data point is “Class 2”
  - Else data point is “Class 3”
Benefits of Decision Trees

- Fast Training
  - Even when large numbers of attributes (inputs) or rows (records)

- Insensitive to outliers

- Usually easy to understood the model
  - Creates rules
  - Most important variables usually are at the top of the tree

- Fast deployment
Decision Trees: Weaknesses

- Can be hard to understand and interpret (if many branches, a complex tree)
- Typically less accurate than neural networks and support vector machines
  - Decision boundaries are constants aligned to variable axes
  - Only one attribute at a time used in splits; more difficult to account for interactions between variables
  - Greedy, forward selection search strategy can be fooled
  - Unstable—small changes in inputs can produce large changes in decision tree
CHAID: Example Split with RFA_2A

TARGET_B

Node 0
Category   %       n
0.000     94.976   45298
1.000     5.024    2396
Total     100.000  47694

RFA_2A
Adj. P-value=0.000, Chi-square=278.473, df=3

D
Node 1
Category   %       n
0.000     90.656   3357
1.000     9.344    346
Total     7.764    3703

E
Node 2
Category   %       n
0.000     93.417   10160
1.000     6.583    716
Total     22.804   10876

F
Node 3
Category   %       n
0.000     95.736   22496
1.000     4.264    1002
Total     49.268   23498

G
Node 4
Category   %       n
0.000     96.548   9285
1.000     3.452    332
Total     20.164   9617
Neural Networks: Sample Network

- Neurons typically “fully connected”
- Architecture typically “pyramid-like”
Neural Networks: Neuron

- Linear weighted sum of inputs
- Sum output is “squashed” by the sigmoid
  - Max. value of neuron output is 1
  - Min. value of neuron output is 0
- All inputs are used in model
- Output layer may omit sigmoid for non-[0,1] outputs
- This is like a hybrid of a linear regression and logistic regression model
  - Linear in inputs before squashing
  - Squashing function is logistic in shape
Neural Networks: Benefits

- Universal approximator: Given suitable architecture, *any* real-valued function can be approximated perfectly
- Very flexible decision boundary formation
- Great name
  - Biological analogy
  - Looks “complicated”
Neural Networks: Pitfalls

- Training is Relatively Slow
- Finding a Good Model Takes Many Tries
  - Reason: get stuck in local minima
  - Never know if and when you have found the “best” solution
- Cannot Have Any Missing Data
  - Recode to numeric, remove record or variable
- No rule for determining architecture
  - Try several to determine which works best
Revisiting Case Studies

Key Classification Algorithms
Motivating Examples: Retail

• Retail
  • Multiple million customers creating tens of millions visits and purchases online

• Objectives
  • Increase engagement with customers
  • Understand intent of visits: exploring? Aspirational shopping? Poised for a purchase?

• Method
  • Augment behavioral segmentation with purchase propensity models (# days to next purchase predictions)
Setting up the Data

• Unit of analysis: Records
  • Visit (customer_id + session_id)

• Context: Fields + Derived Fields
  • Behavioral data captured upon session teardown
  • Derived attributes include binning, interactions, time-series summaries (historic data rolled up to “today”)

• Scoring metrics
  • Want scores apply to all visitors so need global metric like ROC AUC
Setting Up Data for Predictive Analytics Models

Modeler Builds DaystoNextPurchase model on data collected here

Date of data pull for DTNP Model

Time
Modeler Builds DaystoNextPurchase model on data collected here

Extract all visits in 90 days prior to data extract for Model

Time
Modeler Builds DaystoNextPurchase model on data collected here

Visit here:
- Visit Quality
- Channel Engagement
- Purchase
- Days Since Last Visit
- Days Since Last Purchase
- Etc.

Purchase within this time window?

7 days
Setting Up Data for Predictive Analytics Models

Modeler Builds DaystoNextPurchase model on data collected here

Visit here:
- Visit Quality
- Channel Engagement
- Purchase
- Days Since Last Visit
- Days Since Last Purchase
- Etc.

Purchase within this time window?

Time
Modeler Builds DaystoNextPurchase model on data collected here

Visit here:
Visit Quality
Channel Engagement
Purchase
Days Since Last Visit
Days Since Last Purchase
Etc.

7 days
Purchase within this time window?

Time
Setting Up Data for Predictive Analytics Models

Modeler Builds
DaystoNextPurchase model on
data collected here

Visit here:
Visit Quality
Channel Engagement
Purchase
Days Since Last Visit
Days Since Last Purchase
Etc.

Purchase within this
time window?

7 days

Time
Setting Up Data for Predictive Analytics Models

Modeler Builds
DaystoNextPurchase model on
data collected here

Apply (deploy) the Model
anytime

Apply model.
How likely will a visitor purchase in this window?

For visitor here, at this time, we know:
Visit Quality
Channel Engagement
Purchase
Days Since Last Visit
Days Since Last Purchase
Etc.

Time
Days To Next Purchase: 
Set Of Binary Classification Models

Single customer

Probability will Purchase In Day Range

- 3-7 days
- 7-15 days
- 15-30 days
- 30 days

- 2-3 days
- 1 day

© Abbott Analytics 2001-2015
Days To Next Purchase: Set Of Binary Classification Models

- Days To Next Purchase:
  - 15 days
  - 1 day
  - 7 days
  - 30 days
  - 3 days
  - 15-30 days
  - 7-15 days
  - 3-7 days
  - 2-3 days

Single customer

Probability of Purchase In Day Range:
- 1 day
- 3 days
- 7 days
- 15 days
- 30 days

© Abbott Analytics 2001-2015
Days To Next Purchase: Set Of Binary Classification Models

Probability will Purchase In Day Range

- tomorrow
- 2-3 days
- 3-7 days
- 7-15 days
- 15-30 days
- 30 days

Single customer
Conclusion

• Think of predictive analytics as an iterative process rather than a commodity

• Think of predictive analytics approaches as a set of flexible principles rather than recipes to be followed indiscriminately

• Leverage mature software to help build robust models quickly rather than building everything from scratch
More information and resources

- Visit the webpage to learn about Statistica’s analytics capability
  - Case studies, whitepapers, and webinars
- Download a Free Trial
- Questions?
  - David Sweenor – Analytics Product Marketing
    - Twitter: @DavidSweenor
    - David.Sweenor@software.dell.com

Information Management Solutions

Download Free Trial  Buy Online  Request Pricing
Applied Predictive Analytics

From Amazon.com

- Paperback: 456 pages
- Publisher: Wiley; 1 edition (April 7, 2014)
- Language: English
- ISBN-10: 1118727967
- Product Dimensions: 9.8 x 5.9 x 0.6 inches