MicroStrategy Advanced and Predictive Analytics

Seamlessly inject sophisticated data science into real-world applications

Rick Pechter, Solutions Architect
Copyright and Safe Harbor Notice

This presentation may include statements that constitute “forward-looking statements” for purposes of the safe harbor provisions under the Private Securities Litigation Reform Act of 1995, including descriptions of technology and product features that are under development and estimates of future business prospects. Forward-looking statements inherently involve risks and uncertainties that could cause actual results of MicroStrategy Incorporated and its subsidiaries (collectively, the “Company”) to differ materially from the forward-looking statements.

Factors that could contribute to such differences include: the Company’s ability to meet product development goals while aligning costs with anticipated revenues; the Company’s ability to develop, market and deliver on a timely and cost-effective basis new or enhanced offerings that respond to technological change or new customer requirements; the extent and timing of market acceptance of the Company’s new offerings; continued acceptance of the Company’s other products in the marketplace; the timing of significant orders; competitive factors; general economic conditions; and other risks detailed in the Company’s Form 10-K for the year ended December 31, 2016 and other periodic reports filed with the Securities and Exchange Commission. By making these forward-looking statements, the Company undertakes no obligation to update these statements for revisions or changes after the date of this presentation.
About Me
Rick Pechter

Education
- 1985: UC Irvine, BS, Electrical Engineering
- 1992: National Technological University, MS, Engineering Management

Career Path
- 1985-1999: NCR → Teradata – San Diego, CA
  - Design Engineer – Mainframe Class Computer Systems
  - Project Leader – Major Computing Platform Releases
  - Manager – Logic Design, ASIC Development, Device Drivers, Software
- 1999-Present: MicroStrategy – San Diego, CA
  - Director – Pacific Technology Center – San Diego, CA
    - MicroStrategy OLAP Provider/MDX Adapter
    - MicroStrategy Office
    - MicroStrategy Web Services
    - C++ Function Plug-in Wizard
  - Sr. Director – Advanced Analytics and Data Mining
    - Native Data Mining Algorithms, Models Persisted as PMML
    - Predictive Model Import for SAS, SPSS, KXEN, R and many more
    - Member of PMML Industry Consortium , Vice-Chair Data Mining Group
    - R Function Plug-ins, R Integration Pack, Author of MicroStrategyR, package
    - Professional Services Director
    - Sales Engineer
    - Solutions Architect

- Member of ACM and IEEE
Session Synopsis

Track 5 – Session 1 & 8:
Advanced and Predictive Analytics: Seamlessly inject sophisticated data science into real-world applications
Session 1: Tuesday January 16 – 11:30am – 12:30pm
Session 8: Wednesday January 17 – 4:45am – 5:45pm

Whether you’re new to data science or a seasoned veteran, this session is the perfect refresher on everything in the world of advanced analytics. In the session, you'll learn how to leverage MicroStrategy’s extensive built-in function library, out-of-the-box data mining algorithms, and native R analytical engine to expand the depth and scope of your analysis. Topics will include:

• MicroStrategy’s organic approach to advanced and predictive analytics
• The powerful capabilities delivered by the unique MicroStrategy platform architecture, which is designed for maximum extensibility
• MicroStrategy's industry-leading, native R analytical engine that lets you deploy powerful R analytics just like all other MicroStrategy metrics
• Integrating models from third-party tools like SAS and IBM/SPSS with MicroStrategy’s support of PMML, the Predictive Model Mark-up Language
• Exciting changes for the MicroStrategy R Integration Pack in 2018
Welcome to the 16th Annual Presentation on Advanced & Predictive Analytics
Welcome to the 16th Annual Presentation on Advanced & Predictive Analytics
Welcome to the 16th Annual Presentation on Advanced & Predictive Analytics
Topic for 2018? – Option #1
Focus on the Latest Trends & Jump on the AI Bandwagon

Gartner Top 10 Strategic Technology Trends

Advanced Analytics: Using Spark and R to explore data and leverage advanced and predictive analytics by Tony Maresco
Track 5, Session 5
Room: Gracia 6
Wednesday 11:30am – 12:30pm

Find out how to integrate Spark with MicroStrategy, allowing analysts to know what your data looks like before visualizing data. This session will cover:
• How to run scripts against tables to summarize and analyze data relationships
• How analytics can run on the cluster vs. in-memory on the client
• Visualizing these exploratory scripts
• Extending this technique to other predictive and machine learning scenarios

Topic for 2018? – Option #2

Focus on What’s New with MicroStrategy & R

- Installers
- User Guide
- Off-the-Shelf R Scripts
- Source Code

http://github.com/MicroStrategy/RIntegrationPack
Topic for 2018? – Option #2
Focus on What’s New with MicroStrategy & R

RIP CodePlex 2013-2017
Thanks for all 40,309 Downloads

Option to Install
R and R Integration Pack
Incorporated into MicroStrategy Installers
Platform: Starting with 10.8
Desktop/Workstation: 10.11

- Installers
- User Guide
- Off-the-Shelf R Scripts
- Source Code

http://github.com/MicroStrategy/RIntegrationPack
Topic for 2018? – Option #2

Focus on What’s New with MicroStrategy & R

Predictive Analytics and R: How to incorporate advanced analytics and R into your dossiers
by Josef Timchenko
Track 7, Session 4
Room: Gracia 4
Tuesday 4:45pm – 5:45pm

This hands-on workshop is a deep dive into advanced data discovery features and capabilities, with a focus on how to leverage MicroStrategy’s native library of advanced analytics and data mining capabilities, as well as how to include R models into your analysis. This session is designed for IT and business users who are looking to get familiar with the out-of-the-box advanced analytical functions in MicroStrategy and want to learn about MicroStrategy’s integration with open source tools like R Advanced Analytics.

Option to Install R and R Integration Pack Incorporated into MicroStrategy Installers

Platform: Starting with 10.8
Desktop/Workstation: 10.11

- Installers
- User Guide
- Off-the-Shelf R Scripts
- Source Code

http://github.com/MicroStrategy/RIntegrationPack
Focus on What’s REALLY New with MicroStrategy & R

New Push API

Intelligence Server(s)

PUSH API

JSON Data

Persist Data as Cubes

Machine Learning with Python: Train models on trusted data to equip users with predictive analytics
By Scott Rigney
Track 7, Session 4
Room: Gracia 6
Wednesday 3:30pm – 4:30pm

MicroStrategy integrates with cutting-edge machine learning technologies like Python. Attend this session to learn how to train models on trusted data so that business decisions based on predictive technologies can be made with confidence. This session will cover:

• Why a governed and scalable environment is essential for successful data science projects
• Using MicroStrategy APIs to export and import governed data for machine learning
• Using Python to train advanced machine learning models
• How MicroStrategy enables organizations to put predictive analytics in the hands of decision makers
Machine Learning with Python: Train models on trusted data to equip users with predictive analytics

By Scott Rigney
Track 7, Session 4
Room: Gracia 6
Wednesday 3:30pm – 4:30pm

MicroStrategy integrates with cutting-edge machine learning technologies like Python. Attend this session to learn how to train models on trusted data so that business decisions based on predictive technologies can be made with confidence. This session will cover:

- Why a governed and scalable environment is essential for successful data science projects
- Using MicroStrategy APIs to export and import governed data for machine learning
- Using Python to train advanced machine learning models
- How MicroStrategy enables organizations to put predictive analytics in the hands of decision makers
Topic for 2018? – Option #3
Focus on What’s REALLY New with MicroStrategy & R

New Push API
- Intelligence Server(s)
- JSON Data
- Push API
- Persist Data as Cubes

Text Analytics
- Off-the-shelf R scripts
- These ready-to-use R scripts can easily be added to your R-enabled MicroStrategy desktop or Intelligence Server and reference them with a MicroStrategy metric.
- Contents:
  - Forecasting, Classification, Descriptive
  - ARIMA, k-Nearest Neighbors, k-Means Clustering
  - Seasonal Forecasting, Neural Network, k-Medoids Clustering
  - Stepwise Regression, Naive Bayes
  - Survival Analysis, Random Forest
  - Stepwise Logistic Regression
- Sentiment Analysis

Natural Language Generation
- Wordsmith by Automated Judgment
Sentiment Analysis Example
Take Social Media Analysis to the next level

Commenters who get the most likes
Sentiment Analysis Example
Delivering Insights into What People are Saying

Who’s Nice?  
Who’s Not?
Sentiment Analysis Example
Delivering Insights into What People are Saying

“Good Movie” got the most Likes
Sentiment Analysis Example
Delivering Insights into What People are Saying

“Something ain’t right here” is #3
Topic for 2018? – Option #3
Focus on What’s REALLY New with MicroStrategy & R

New Push API
Intelligence Server(s)
Persist Data as Cubes

Text Analytics
Off-the-shelf R scripts
- Forecasting
- Classification
- Descriptive
- ARIMA
- k-Nearest Neighbors
- k-Means Clustering

Sentiment Analysis

Natural Language Generation

PUSH API
JSON Data

https://github.com/MicroStrategy/RIntegration/tree/master/scripts#off-the-shelf-r-scripts
Topic for 2018? – Option #3
Focus on What’s REALLY New with MicroStrategy & R

Social Media
Influencers
Tone
Role Models
Tenor
Trolls
Sentiment Analysis

Smallish
“Brevity is the Soul of Wit”

It’s about the words
Words Matter
Lincoln’s Gettysburg Address

Sentiment Analyzed under Two Minutes!

on YouTube at http://tinyurl.com/MSTR-SA
Lincoln’s Gettysburg Address

Sentiment Analyzed under Two Minutes!

“Truth in Advertising” Notes:

• The default name for the input in the metric expressions is “Text”

RScript<RScriptFile="SentimentAnalysis.R", _InputNames="Text",
_Params="FileName='SA_mstr', PlotWordCloud=TRUE,
PlotHistogram=TRUE, RemoveRetweets=TRUE, SaveCSV=TRUE">(Text)

• If your metric is named different, just replace “Text” with it’s name

RScript<RScriptFile="SentimentAnalysis.R", _InputNames="MyText",
_Params="FileName='SA_mstr', PlotWordCloud=TRUE,
PlotHistogram=TRUE, RemoveRetweets=TRUE, SaveCSV=TRUE">(MyText)

• “Text” has to be a Metric... Why?
  • Metrics always have a Level and an Aggregation function
  • Essential to the Algebra of the SQL Engine

• If “Text” is an Attribute, simply convert it to a metric
  • Wrap it in an Agg Function and Give it a Level

RScript<RScriptFile="SentimentAnalysis.R", _InputNames="MyText",
_Params="FileName='SA_mstr', PlotWordCloud=TRUE,
PlotHistogram=TRUE, RemoveRetweets=TRUE, SaveCSV=TRUE">(Max(MyText))(~)
Have you noticed our President likes to use Twitter?
Many Have Noticed an Irony in Trump’s Tweets

Day after the Election

9 Hours Later

Day after the Inauguration

1½ Hours Later
2016 Presidential Campaign
Trump uses more than one device

Apple never did

Donald J. Trump 📱 @realDonaldTrump 19 Feb 2016
I use both iPhone & Samsung. If Apple doesn’t give info to authorities on the terrorists I’ll only be using Samsung until they give info.
2017 as President Trump
No longer uses an Android device
How to Tell Which Tweets Are Really from The Donald Himself

Day after the Election

Before Inauguration
Different devices ✔
Different people ?

9 Hours Later

Let’s Predict Which Tweets are Really from President Trump?

Day after the Inauguration

After Inauguration
Different devices ☠
Different people ✔
Demo

Trump’s tweets dashboard

Twitter Data Discovery combined with Advanced Analytics and Predictive Models

Try it yourself:

Dashboard: http://tinyurl.com/TrumpTweets-Dossier

1. Go to the bottom where it says “Or, you can enter your Credentials.”
2. Click on “Credentials.” to get to a new login window
3. User name = “tt” and Password = “tt” (use Standard Authentication)

Check out the video:
Video: http://tinyurl.com/Video-TrumpTweets
The Explanation

David Robinson’s Blog
VarianceExplained.org
has all the details!
The Gartner Analytic Continuum
Classifying Analytics, from Basic to Advanced, Predictive to Prescriptive
## Analytics Used to Analyze Trump’s Tweets

### Basic, Advanced and Predictive Analytics Delivered in a Single Dashboard

<table>
<thead>
<tr>
<th>Tweets &amp; Behaviors</th>
<th>Words &amp; Sentiments</th>
<th>Predict @POTUS</th>
<th>Bot or Not</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweet Count</td>
<td>Word Count</td>
<td>Predictive Models</td>
<td>Model Explanation</td>
</tr>
<tr>
<td>• By Time Frame</td>
<td>• By Time Frame</td>
<td>• ADA Boost</td>
<td>• Simulation</td>
</tr>
<tr>
<td>• By Account</td>
<td>• By Device</td>
<td>• Decision Tree</td>
<td>• Intertweet Timings</td>
</tr>
<tr>
<td>• By Device</td>
<td>• By Day</td>
<td>• GLM (Regression)</td>
<td>• Seconds Since Last Tweet</td>
</tr>
<tr>
<td>• By Hour of Day</td>
<td>• By Day of Week</td>
<td>• Neural Network</td>
<td>• Seconds Until Next Tweet</td>
</tr>
<tr>
<td>• By Day</td>
<td>• Is Retweet</td>
<td>• Random Forests</td>
<td></td>
</tr>
<tr>
<td>• By Day of Week</td>
<td>• Has Pic/Link</td>
<td>• SVM, Consensus</td>
<td></td>
</tr>
<tr>
<td>• Is Retweet</td>
<td>• Has Hashtag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Popularity</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Tweets & Behaviors
- Tweet Count
- By Time Frame
- By Account
- By Device
- By Hour of Day
- By Day
- By Day of Week
- Is Retweet
- Has Pic/Link
- Has Hashtag
- Popularity

### Words & Sentiments
- Word Count
- By Time Frame
- By Device
- By Sentiment

### Predict @POTUS
- Predictive Models
  - ADA Boost
  - Decision Tree
  - GLM (Regression)
  - Neural Network
  - Random Forests
  - SVM, Consensus

### Bot or Not
- Model Explanation
  - Lift Charts
  - Confusion Matrices
  - Variable Importance
  - PMML

- Simulation
- Intertweet Timings
  - Seconds Since Last Tweet
  - Seconds Until Next Tweet
MicroStrategy: A Unified Platform of Integrated Capabilities
MicroStrategy: A Unified Platform of Integrated Capabilities
MicroStrategy: A Unified Platform of Integrated Capabilities

Users

Focus on Analytics

Data
Advanced and Predictive Analytics

MicroStrategy offers three easy methods to include advanced analytics into your applications.

- **Native Statistical Library of 350+ functions for self-service**
- **MicroStrategy’s Native Data Mining and PMML capabilities for predictive analytics**
- **The industry’s best integration with the world’s most popular analytic toolset**

MicroStrategy offers an advanced analytics library that is built-in to the product. From basic summarization to optimization to even forecasting and predictions, users can embed statistical metrics into the analytical workflow very easily.

MicroStrategy offers over 350 native functions, out of the box, so that business users are able to perform data mining and create highly advanced statistical algorithms, on their own, with minimal support from IT.

MicroStrategy includes native functions implementing the most commonly used predictive algorithms so that any organization can create and deploy data mining models without any third party tools.

MicroStrategy allows organizations to leverage existing investments in other technologies by seamlessly integrating with data mining models from other applications by importing Predictive Model Markup Language (PMML) into the metadata repository.

MicroStrategy can seamlessly integrate your advanced analytics into reports, documents, dashboards and alert driven apps.

76% of analytic professionals report using R
- 2015 Rexer Data Science Survey (www.RexerAnalytics.com)
MicroStrategy Functions Library

350+ Analytical Functions available Out-of-the-Box

<table>
<thead>
<tr>
<th>Basic aggregations</th>
<th>Mathematical</th>
<th>Date and time</th>
<th>Statistical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Absolute</td>
<td>Add Days</td>
<td>Average Deviation</td>
</tr>
<tr>
<td>Mean</td>
<td>A-cosine</td>
<td>Add Months</td>
<td>AvgDev</td>
</tr>
<tr>
<td>Count</td>
<td>Hyx A-cos</td>
<td>Current Date</td>
<td>Beta CDF</td>
</tr>
<tr>
<td>Sum</td>
<td>A-sine</td>
<td>Current Date &amp; Time</td>
<td>Binomial PDF/CDF</td>
</tr>
<tr>
<td>Minimum</td>
<td>A-tan</td>
<td>Current Time</td>
<td>ChiSquare PDF/CDF</td>
</tr>
<tr>
<td>Median</td>
<td>Hyx A-tan</td>
<td>Day of Month</td>
<td>ChiSquareTest</td>
</tr>
<tr>
<td>Mode</td>
<td>Ceiling</td>
<td>Day of Week</td>
<td>Correlation</td>
</tr>
<tr>
<td>Product</td>
<td>Combine</td>
<td>Day of Year</td>
<td>Covariance</td>
</tr>
<tr>
<td>Rank</td>
<td>Cosine</td>
<td>Days Between</td>
<td>CritBinomial</td>
</tr>
<tr>
<td>Percentile</td>
<td>Hyx Cosine</td>
<td>Month Start Date</td>
<td>Exponential PDF/CDF</td>
</tr>
<tr>
<td>&quot;N&quot;-Tile</td>
<td>Degrees</td>
<td>Month End Date</td>
<td>F PDF/CD/Inv</td>
</tr>
<tr>
<td>N-tile by Step</td>
<td>Exponent</td>
<td>Months Between</td>
<td>Fisher PDF/CDI</td>
</tr>
<tr>
<td>N-tile by Value</td>
<td>Factorial</td>
<td>Year Start Date</td>
<td>Forecast</td>
</tr>
<tr>
<td>N-tile by Step and Value</td>
<td>Floor</td>
<td>Year End Date</td>
<td>ForecastV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FTest</td>
</tr>
</tbody>
</table>

Financial functions

Accrued Interest
Accrued Interest Maturity
Amount Received at Maturity
Bond-equivalent Yield for T-BILL
Convert Dollar Price from Fraction to Decimal
Convert Dollar Price from Decimal to Fraction
Cumulative Interest Paid on Loan
Cumulative Principal Paid on Loan
Depreciation for each Accounting Period
Days In Coupon Period to Settlement Date
Days In Coupon Period with Settlement Date
Days from Settlement Date to Next Coupon
Double-Declining Balance Method
Discount Rate For a Security
Effective Annual Interest Rate
Fixed-Declining Balance Method
Future Value
Future Value of Initial Principal with Compound Interest Rates
First Period Payment
Price of Principal
Price on Principal
Interest Rate
Interest Payment
Internal Rate of Return
Interest Rate per Annuity
Macaulay Duration
Modified Duration
Modified Internal Rate of Return
Next Coupon Date After Settlement Date
No of Coupons Settlement and Maturity Date
Nominal Annual Interest Rate
No of Investment Periods
Net Present Value
Odd Last Period / Yield
Payoff of Option
 Prev Coupon Date Before Settlement Date
Price Per $100 Face Value w Odd
Price at Maturity
Present Value
Prorated Depreciation for each Period
Straight Line Depreciation
Sum-Of-Years’ Digits Depreciation
T-BILL Price
T-BILL Yield
Variable Declining Balance
Yield
Yield at Maturity
Yield for Discounted Security
OLAP functions

Running Total
Running Std Deviation
Running Std Deviation of Population
Running Minimum
Running Maximum
Running Count
Moving Difference
Moving Maximum
Moving Minimum
Moving Average
Moving Sum
Moving Count
Moving Std Deviation
Moving Std Deviation
First /Last in Range
Exponential Weight Moving Avg
Exponential Weight Running Avg

Mathematical

Integer
Ln
Log
Log10
Mod
Power
Quotient
Radians
Round
Sin
Sine
Sin-1
Square Root
Tan
Tan-1
Tanh
Trunc

Statistical

Gamma PDF/CDF/lnv
Geometric Mean
Growth
GrowthV
HeteroscedasticTTest
HomoscedasticTTest
Hypergeometric PDF
Intercept
Kurtosis
Lognormal PDF/inv
MeanTest
NegativeBinomial PDF
Normal PDF/CD/Inv
PairedTTest
Pearson
Permut

POisson PDF
RSquare
Slope
Standard Deviation
Standardize
StandardNormal PDF/inv
SteYX
T Dist CDF/Inv
Trend
TrendV
Variance
VarTest
Weibull PDF/CDF

Date and time

Add Days
Add Months
Average Deviation
Current Date
Current Date & Time
Current Time
Day of Month
Day of Week
Day of Year
Days Between
Month Start Date
Month End Date
Years Between
Year Start Date
Year End Date

MicroStrategy Analytics Library
Predictive Analytics We Support
Deploy Analytics using virtually Any and All Techniques

"Which Techniques Do You Use Most"

Three work horses of data mining

4th

Source: 2015 Rexer Data Miner Surveys
www.RexerAnalytics.com
Over 1,220 Data Miners from 72 Countries

® = via R
○ = via PMML
✓ = MicroStrategy Native
Evolution of MicroStrategy Advanced Analytics
All MicroStrategy Analytics are Always Deployable Anyone to Anywhere

MicroStrategy Platform Architecture

MicroStrategy Function Plug-ins

C++ FP

Linear Regression
Decision Tree
K-Means Clustering
Logistic Regression

Association Rules
Time Series
Ensembles of Models
Trendlines

Neural Network
SVM
Ruleset

Create

Developer

• Import functions
• Training metric wizard
• Import predictive model
• Metric editor

Desktop

• Metric editor

Web

• Metric editor

MicroStrategy Intelligence Server

Query Engine

Visual C++

300+ functions

Copyright © 2018 MicroStrategy Incorporated. All Rights Reserved.
Evolution of MicroStrategy Advanced Analytics
All MicroStrategy Analytics are Always Deployable Anyone to Anywhere

MicroStrategy Platform Architecture

MicroStrategy Intelligence Server
Query Engine
Visual C++
300+ functions

C++ Function Plug-ins

C++ FP
Linear Regression
Decision Tree
K-Means Clustering
Logistic Regression

Neural Network

SVM
Association Rules
Time Series
Ensembles of Models
Trendlines

Ruleset

C++ FP

Create

Developer
• Import functions
• Training metric wizard
• Import predictive model
• Metric editor

 Deploy anywhere and everywhere

Desktop
• Metric editor

Web
• Metric editor

Copyright © 2018 MicroStrategy Incorporated. All Rights Reserved.
Advantages of MicroStrategy’s Architecture and Native R Analytical Engine

“MicroStrategy’s Approach is the Best R Integration in the Industry” --VP of Products at Revolution Analytics

Most Vendors, including Tableau, SAP and Qlik, require using an R Server:

- Acquire new server
- Acquire user privileges and/or $ Licenses
- Insert R code in-line
- Special handling for user parameters
- Performance is inversely proportional to security
- Deploying results can require special handling

MicroStrategy’s unique approach avoids additional costs and overhead required by R Servers:

- Not required
- No Licenses, no special Privileges
- Reference stand-alone R script
- User Parameters controlled as usual
- Native high performance engine with all security features
- Results deployed same as usual

Plus MicroStrategy supports all R packages, handles/logs errors, full access to R Environment/Objects, and more
Advantages of MicroStrategy’s Architecture and Native R Analytical Engine

“MicroStrategy’s Approach is the Best R Integration in the Industry” --VP of Products at Revolution Analytics

Most Vendors, including Tableau, SAP and Qlik, require using an R Server:

Of course, MicroStrategy can optionally support an R Server, especially when performance or data scalability demand one – But an R Server is not required for the vast majority of applications

Plus MicroStrategy supports all R packages, handles/logs errors, full access to R Environment/Objects, and more
R Allows Applications to do More Than Just Crunch Numbers

R Analytical Engine breaks out of the grid

Users

In-Band Analytics
Scalar Out
Vector Out

Out-of-Band Analytics

Data

MicroStrategy
R Analytical Engine Breaks Out of the Grid

In-Band Analytics
All MicroStrategy Functions return Vectors or Scalars

Vectors in

Scalar in

Vector out

 Scalars in

Vector out

 Scalars out

* = Sorted and unsorted

In-Band Analytics

Generic* Table-at-a-Time

Simple Row-at-a-Time

Aggregation*
Group by

 Scalars in

Vector in

 Scalars out

 Scalars out

 Scalars out
R Analytical Engine Breaks Out of the Grid

In-Band Analytics
All MicroStrategy Functions return Vectors or Scalars

Generic*
Table-at-a-Time

Simple
Row-at-a-Time

Aggregation*
Group by

* = Sorted and unsorted

Vectors in

Vector out

Scalars in

Scalar out

Vector in

Scalar out

Analytics can optionally access internet assets, read/persist files, publish cubes, and even leverage R servers
MicroStrategy: A Unified Platform of Integrated Capabilities

In-Band Analytics
- Scalar Out
- Vector Out

Out-of-Band Analytics
- R Studio
- File System
- R Data

Users

Data
What can we do with all these analytics?
Predictive Analytics in Action
Forecasting healthcare reimbursements

HMSA: Engineering the future with state-of-the-art forecasting powered by MicroStrategy
By Charlotte Marcil & Joseph Hawkins
Track 2, Session 2b
Room: Brera 5
Tuesday 2:50pm – 3:15pm

With MicroStrategy, you can combine sophisticated data mining capabilities with easy-to-use data discovery for predictive and prescriptive analytics. In this session, learn how long-time customer Hawaii Medical Service Association (HMSA) has worked with MicroStrategy to take forecasting to the next level. Discover how HMSA was able to use historical data combined with machine learning algorithms from R to deliver the most accurate forecasts possible.
Forecasting Healthcare Reimbursements
State-of-the-art Time Series Analysis in action

Past approaches:
- Relying on Intuition / "Rules of Thumb"
- Using Straight Line Projections
- No standards for forecast accuracy

Business Needs:
- Estimate future costs and utilization
- Quantify forecast accuracy and predictability
- Better understand trends over time

Goal:
Forecast future reimbursements

For Each Billion Dollars in Reimbursements, even a 1% improvement in forecasts drops $10 Million from the Top Line
Seven State-of-the-Art Forecasting Features

Automated machine learning

- Bottom-Up Forecasting using non-overlapping Segments
- Predicting IBNR Claims
- Automatic Outlier Handling
- Predictive Algorithms
- Hold-Out Validation
- Champion-Challenger
- Back-Testing and Trim
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. **Bottom-Up Forecasting**

2. Predicting IBNR Claims

3. Automatic Outlier Handling

4. Predictive Algorithms

5. Hold-Out Validation

6. Champion-Challenger

Bottom-Up Forecasting
“Top Line is for Show, Bottom-Up is for Dough”

• You cannot manage a business from the top line

• **The Devil is in the details** so divide-and-conquer

• Break the data into Segments
Bottom-Up Forecasting
“Top Line is for Show, Bottom-Up is for Dough”

- You cannot manage a business from the top line
- The Devil is in the details so divide-and-conquer
- Break the data into Segments by Plan
Bottom-Up Forecasting
“Top Line is for Show, Bottom-Up is for Dough”

- You cannot manage a business from the top line
- **The Devil is in the details** so divide-and-conquer
- Break the data into Segments by **Category**

```
Surgery-OP
Radiology
Other Services
Mental Health
Medical Services
Drug
ER
Hospital-IP
Lab
```
Bottom-Up Forecasting
“Top Line is for Show, Bottom-Up is for Dough”

- You cannot manage a business from the top line
- **The Devil is in the details** so divide-and-conquer
- Break the data into Segments by Plan and Category
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger

Charts are also Intuitive Selectors
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. **Predicting IBNR Claims**
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. **Automatic Outlier Handling**
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. **Predictive Algorithms**
5. Hold-Out Validation
6. Champion-Challenger
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. **Hold-Out Validation**
6. Champion-Challenger
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger

Challengers Compete
Best Forecast Accuracy → Champion

7 Predictive Models per Hold-Out:
- IBNR, ets (Test and Full)
- arima (Test and Full)
- stlm (Test and Full)

49 Predictive Models Built
Health Insurance Claims Take Time to Process

Claims are reported sometime after they occur and reimbursements can lag by weeks or months.

Typical Healthcare Claims Processing Workflow
Forecasting using Trendlines
Upward long-term trend breaks downward in the short-term

• Forecasting Reimbursements presents challenges

• Long-term trend is up

• But short-term trend is down because Reimbursements always lag Claims Events
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger

Depending on the segment, claims processing can lag by days, weeks or months.
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger

Algorithms x Hold-Outs x Trims x Back-Tests  2,548 Predictive Models Built
Seven State-of-the-Art Forecasting Features
Automated Machine Learning

1. Bottom-Up Forecasting
2. Predicting IBNR Claims
3. Automatic Outlier Handling
4. Predictive Algorithms
5. Hold-Out Validation
6. Champion-Challenger

Algorithms x Hold-Outs x Trims x Back-Tests 80,136 Predictive Models Built
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

Claims Data

Train Mode True

For each Segment
Build ensemble of models

For each BackTest and Trim
Train Models

Segment Dataset
IBNR
Outlier Handling
Champion Challenger
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

For each Segment
Build ensemble of models
TrainMode True
For each BackTest and Trim
Train Models
Persist All Models

Claims Data
Break-by Segment
Model Repository

IBNR Segment Dataset
Outlier Handling
Champion Challenger

Persist All Models

Copyright © 2018 MicroStrategy Incorporated. All Rights Reserved.
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

- Claims Data
- Break-by-Segment
- TrainMode: True
- For each Segment: Build ensemble of models
  - For each BackTest and Trim: Train Models
  - Segment Dataset
  - IBNR
  - Outlier Handling
  - Champion
  - Challenger

- Model Repository
- Persist All Models
- Score Models
- Deploy Forecasts
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

TrainMode: True

Build ensemble of models

For each BackTest and Trim

Train Models

Persist All Models

For each Segment

Outlier Handling

IBNR

Champion

Challenger

Train Data

Break-by-Segment

Model Repository

Persist All Models

Trained Models are Persisted for Fast Scoring!
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

Claims Data → Break-by-Segment → False → Model Repository

TrainMode

Trained Models are Persisted for Fast Scoring!
Predictive Analytics Workflow
Optimized for high-volume model building and high-performance forecast deployment

![Diagram showing the workflow process]

- Claims Data
- Break-by-Segment
- False
- Model Repository
- Score Models
- Deploy Forecasts

TrainMode

Trained Models are Persisted for **Fast Scoring!**
Predictive Analytics Workflow

Optimized for high-volume model building and high-performance forecast deployment

Claims Data → Break-by-Segment → Model Repository → Score Models

TrainMode → False

Build ensemble of models

For each Segment → For each BackTest and Trim → Train Models

Persist All Models

Deploy Forecasts

Deploy Powerful Analytics to Anyone Anywhere
Demo
Forecasting healthcare reimbursements

Try it yourself!

Dashboard: https://tinyurl.com/Forecasting-Dossier

Login = tt, Password=tt
Go to the bottom where it says “Or, you can enter your Credentials.”
Click on “Credentials.” to get to a new login window
User name = “tt” and Password = “tt” (use Standard Authentication)

Check out the Video: http://tinyurl.com/Video-Forecasting
Next Steps

• As MicroStrategy users, you have already have the tools you need!

• Advanced Analytics today isn’t a technology problem, it’s a people problem – find or train good people!

• Best practice: Collaboration between business and data science teams

• Best practice: It’s an iterative process so iterate frequently, iterate fast

• Best practice: Validate models against data not used in training

• Be patient: If this was easy, everyone would be doing it!
Try It Yourself

Sentiment Analysis:

Video:  http://tinyurl.com/MSTR-SA

Trump Tweets:

Dashboard:  http://tinyurl.com/TrumpTweets-Dossier
Video:  http://tinyurl.com/Video-DonaldTweets

Forecasting Healthcare Reimbursements

Dashboard:  https://tinyurl.com/Forecasting-Dossier
Video:  http://tinyurl.com/Video-Forecasting

Login = tt, Password=tt
Go to the bottom where it says “Or, you can enter your Credentials.”
Click on “Credentials.” to get to a new login window
User name = “tt” and Password = “tt” (use Standard Authentication)
Thank You!

Rick Pechter
rpechter@microstrategy.com