Getting Started with Advanced Analytics in Finance, Marketing, and Operations

East Coast Oracle Users Conference
Tim Vlamis
November 7, 2017
@VlamisSoftware
Vlamis Software Solutions

- Vlamis Software founded in 1992 in Kansas City, Missouri
- Developed 200+ Oracle BI and analytics systems
- Specializes in Oracle-based:
  - Enterprise Business Intelligence & Analytics
  - Analytic Warehousing
  - Data Mining and Predictive Analytics
  - Data Visualization
- Multiple Oracle ACEs, consultants average 15+ years
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Partner
- Oracle Gold Partner
Vice President & Analytics Strategist

- 30+ years in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE
- Joined Vlamis in 2007
- Instructor for Oracle University’s Predictive Analytics, Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University
Presentation Agenda

▪ Background on Analytic Options to the Oracle DB
▪ Oracle Advanced Analytics
  ▪ Oracle Data Mining
  ▪ Oracle R Enterprise
▪ How to start with OAA – comparison of options
▪ Switch to “Data Visualization for Oracle BI 12c”?
"Data! Data! Data!" he cried impatiently. "I can’t make bricks without clay."

Sherlock Holmes in “The Adventure of the Copper Beeches” by Sir Arthur Conan Doyle
But we have plenty of data...

- Est. Global IP traffic/month is **89 Exabytes** (89 Billion Gigabytes)
- Est. Global data stores total **4.4 Zetabytes** (44 Trillion Gigabytes)
▪ Start with a general plan
Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
- Prepare one bed at a time
Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
- Prepare one bed at a time
- Mulch and weed early on
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Time is a Great Designer

Pictures Courtesy of National Trust Sissinghurst
Four Realms of Analytics

<table>
<thead>
<tr>
<th>Probability Based</th>
<th>Rules Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Analytics</td>
<td>Descriptive Analytics</td>
</tr>
<tr>
<td>Predictive Analytics</td>
<td>Prescriptive Analytics</td>
</tr>
</tbody>
</table>

Past | Future
Good Questions/Hypotheses are Needed

What behaviors in the past year are most significant in terms of segmenting our customers?

What’s the Life Time Value of each customer? What’s a potential new customer worth?

Which products are purchased together most often? Which products are purchased with our most profitable products?
Listen to Data

- Relative importance
- Natural relationships
- Similarities/differences
- Predictions
Analytics in the Real World

- Finance – forecasting is a messy business
- Marketing – clustering should be behavior-based, not demographic-based
- Operations – Think broadly on how to use analytics
- Do the analytics near the data, not in the application layer
What is Data Mining?

- Automatically sifts through data to find hidden patterns, discover new insights, and make predictions

Data Mining can provide valuable results:
  - Predict customer behavior *(Classification)*
  - Predict or estimate a value *(Regression)*
  - Segment a population *(Clustering)*
  - Identify factors more associated with a business problem *(Attribute Importance)*
  - Find profiles of targeted people or items *(Decision Trees)*
  - Determine co-occurrences and “market baskets” within an event set *(Associations)*
  - Find fraudulent or “rare events” *(Anomaly Detection)*
Oracle Data Mining

- Oracle Data Mining is an option for the Enterprise Edition of the Oracle Database.
- A collection of APIs and specialized SQL functions.
- Includes a large number of specialized algorithms and built-in procedures.
- Makes use of many built-in capabilities of the Oracle Database.
- ODM typically refers to “Oracle Data Mining”
In-Database Data Mining

Traditional Analytics

- Data Import
- Data Mining Model “Scoring”
- Data Preparation and Transformation
- Data Extraction

Hours, Days or Weeks

Oracle Data Mining

- Model “Scoring”
- Data Preparation and Transformation
- Model Building

Secs, Mins or Hours

Results

- Faster time for “Data” to “Insights”
- Lower TCO—Eliminates
- Data Movement
- Data Duplication
- Maintains Security

Savings

- Model “Scoring”
- Data remains in the Database
- Embedded data preparation
- Cutting edge machine learning algorithms inside the SQL kernel of Database
- SQL—Most powerful language for data preparation and transformation
- Data remains in the Database

Copyright © 2017, Vlamis Software Solutions, Inc.
Data Mining Provides  
Better Information, Valuable Insights and Predictions

Segment #1:
IF CUST_MO > 14 AND INCOME < $90K, THEN
Prediction = Cell Phone Churner,
Confidence = 100%
Support = 8/39

Segment #3:
IF CUST_MO > 7 AND INCOME < $175K, THEN
Prediction = Cell Phone Churner
Confidence = 83%
Support = 6/39

Source: Inspired from Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management by Michael J. A. Berry, Gordon S. Linoff
# Oracle Data Mining Algorithms

<table>
<thead>
<tr>
<th>Problem</th>
<th>Algorithm</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>Logistic Regression (GLM)</td>
<td>Classical Statistical Technique</td>
</tr>
<tr>
<td></td>
<td>Decision Trees</td>
<td>Popular/Rules/Transparency</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
<td>Embedded app</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
<td>Wide/Narrow Data or Text</td>
</tr>
<tr>
<td>Regression</td>
<td>Linear Regression (GLM)</td>
<td>Classical Statistical Technique</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
<td>Wide/Narrow Data or Text</td>
</tr>
<tr>
<td>Anomaly Detection</td>
<td>One Class SVM</td>
<td>Unknown fraud cases or anomalies</td>
</tr>
<tr>
<td>Attribute Importance</td>
<td>Minimum Description Length</td>
<td>Attribute reduction</td>
</tr>
<tr>
<td></td>
<td>Principal Component Analysis</td>
<td>Reduce data noise</td>
</tr>
<tr>
<td>Association Rules</td>
<td>Apriori</td>
<td>Market Basket Analysis</td>
</tr>
<tr>
<td>Clustering</td>
<td>Hierarchical K-Means</td>
<td>Market Segmentation</td>
</tr>
<tr>
<td></td>
<td>Orthogonal Partitioning</td>
<td>Product / Location Groupings</td>
</tr>
<tr>
<td></td>
<td>Expectation Maximization</td>
<td>Text analysis</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>Non-negative Matrix Factorization</td>
<td>Feature Reduction</td>
</tr>
<tr>
<td></td>
<td>Singular Value Decomposition</td>
<td>Text Analysis</td>
</tr>
</tbody>
</table>
Oracle Data Miner

- Easy to Use
  - Oracle Data Miner GUI for data analysts
  - “Work flow” paradigm

- Powerful
  - Multiple algorithms & data transformations
  - Runs 100% in-DB
  - Build, evaluate and apply models

- Automate and Deploy
  - Save and share analytical workflows
  - Generate SQL scripts for deployment
Understand Model Details

- Interactive model viewers
Oracle Data Mining & OBI 11g

ODM’s predictions & probabilities are available in the Database for reporting using Oracle BI EE and other tools.
Dynamically Using ODM From Oracle BI
What is R?

- Popular alternative to **SAS, SPSS** & other proprietary statistical environments
- 2 million+ users worldwide and growing
- Thousands of R packages available
- Taught extensively in higher education

**R Usage**

- 76% of data miners report using R
- 36% of data miners select R as their primary tool

*2015 Rexer Analytics Data Miner Survey [http://www.rexeranalytics.com](http://www.rexeranalytics.com)*
R is extensively used by Statisticians, Data Analysts, Students

- Free (Open source)
- Graphical
- Powerful
- Extensible
- Ease to install and use
- Industry/subject specific packages
- Out-of-the-box functionality with many ‘knobs’, but smart defaults
Oracle’s R Technologies

- Oracle R Distribution
- ROracle
- Oracle R Enterprise (ORE)
- Oracle R Advanced Analytics for Hadoop (ORAAH)
- Oracle R Connector for Hadoop (ORCH)

Open Source Software available to R Community for free
Oracle R Enterprise (ORE) is a component of the Oracle Advanced Analytics (OAA) option to Oracle Database EE.

- Provides transparent access to database-resident data from R.
- Execute R scripts at the database machine managed by Oracle Database with data and task parallelism.
- Execute R scripts from SQL.
- Integrates R into the IT software stack.
- Extends and enhances open source R.
Oracle R Enterprise

- A comprehensive, database-centric environment for end-to-end analytical processes in R, with immediate deployment to production environments
- Operationalize entire R scripts in production applications – eliminate porting R code
- Seamlessly leverage Oracle Database as an HPC environment for R scripts, providing data parallelism and resource management
- Avoid reinventing code to integrate R results into existing applications
- Transparently analyze and manipulate data in Oracle Database through R using versatile and customizable R functions
- Eliminate memory constraint of client R engine
- Score R models in Oracle Database
- Execute R scripts through Oracle Database server machine for scalability and performance
- **Get maximum value from your Oracle Database and Exadata**
- Enable integration and management through SQL
- Integrate R into the IT software stack, e.g. OBIEE
R now integrated into OBIEE 11g and 12c
require(gplots)
set.seed(123)
# simulate an AR(1) process
coeff <- 0.55
series <- arima.sim(list(ar=coeff), n=250)

# fit AR(1) with the 200 first data
model <- arima(series[1:200], order=c(1, 0, 0))

# make forecast from the model
forecast <- predict(model, n.ahead=50)

# compute the limits of the graph
ylim <- c(min(series[1:200]), max(series[1:200]) - 1.96 * forecast$se) +
        c(1.96 * forecast$se, 1.96 * forecast$se)

# prepare the space where to plot
par(mar=c(4,4,2,2), las=1)
plot(series, ylim=ylim, type="b", main='Time Series Analysis using R', xlab='Time', ylab='Series', col=c(1,260))

# split the figure into two parts
# - the part used to fit the model
rect(user[2], user[3], user[4], border="NA", col="lemonchiffon")
# - the part used to make the forecast
rect(user[2], user[3], user[4], border="NA", col="lavender")
abline(h = c(-3:3)*2, col = "gray", lty =3)

# draw a 95% confidence band
polygon(c(forecast$pred - 1.96*forecast$se, rev(forecast$pred + 1.96*forecast$se)),
        col = "orange", lty=2, border="NA")

lines( series, lty=1)
lines( series, lty=1)

Copyright © 2017, Vlamis Software Solutions, Inc.
Customer “most likely” be HIGH and VERY HIGH value customer in the future
5 Common use cases for predictive analytics

1) Customer Segmentation using Clustering algorithms
   ▪ Discovered patterns can be extremely meaningful
   ▪ Able to include hundreds of dimensions
   ▪ Great first project

2) Predict Lifetime Customer Value
   ▪ Measure impact of different product purchases on LCV
   ▪ Promote and incentive profitable purchases
5 Common use cases for predictive analytics

3) Market Basket Analysis for retailers and warehouses
   ▪ Understand purchasing and picking patterns

4) Employee Retention analysis
   ▪ Classify employees into basic categories
   ▪ Understand impact of different incentives and rewards

5) Optimize Customer Service and Next Best Offer
   ▪ Use decision trees to determine rules for customers
   ▪ Dramatically increase effectiveness of offers
Basic Ways to Get Started

- Do a POC project on your own
- Conduct a workshop for key stakeholders to build support
  - One hour to one day
  - Half-day works great
- Conduct ODM and ORE training classes with 1-day workshop
- Use a defined Quick Start program (2 weeks)
ODM Quick Start Overview

- **Hardware or Cloud**
  - Oracle Database Appliance/Oracle Database Cloud Service

- **Software**
  - Oracle Database 12c (with options)
  - Oracle Advanced Analytics Option including Oracle Data Mining
  - Oracle SQL Developer: Data Miner Add-in (free download)

- **Services**
  - Implementation and configuration from Vlamis Software Solutions (Oracle Gold Partner)
  - Oracle University Oracle Data Mining Techniques course (taught by Vlamis Software Solutions)
  - Market Basket Analysis Project performed on company data

- **Time frame**: 9 business days (less than 2 weeks)
Quick Start Compressed Schedule

- **Day 1:**
  - Two consultants meet with client team to review project plan, review data sources, identification of best data to start with, set technical objectives for project (basic market basket analysis deliverable)

- **Day 2:**
  - Consultant One: Install ODA and configure to network (need support from client tech staff)
  - Consultant Two: Conduct first day of ODM class with client team

- **Day 3:**
  - Consultant One: Install new pluggable Database, SQL Developer
  - Consultant Two: Conduct second day of ODM class with client team

- **Day 4:**
  - Two consultants establish data plan for project with client and import data

- **Day 5:**
  - Consultant One: Prepare tables for mining (add keys, new tables, transforms, etc.)
  - Consultant Two: Document data plan

- **Day 6:**
  - Consultant Two: Build market basket workflow

- **Day 7:**
  - Consultant Two: Conduct market basket analyses

- **Day 8:**
  - Consultant Two: Prepare presentation of findings from market basket analyses

- **Day 9:**
  - Consultant Two: Deliver presentation with client
Oracle Data Mining Training (2 days)

- Introduction
- Data Mining Concepts and Terminology
- The Data Mining Process
- Introducing Oracle Data Miner 11g Release 2
- Using Classification Models
- Using Regression Models
- Using Clustering Models
- Performing Market Basket Analysis
- Performing Anomaly Detection
- Deploying Data Mining Results
Oracle R Enterprise Training (2 days)

- Oracle R Enterprise technologies introduction
- Introduction to R hands-on
- ORE transparency layer with hands-on exercises
- ORE embedded R execution with hands-on exercises
- ORE predictive analytics with hands-on exercises
- Using ROracle
- Overview of ORE with OBIEE
Comparison of Training Courses

Oracle Data Mining
- Organized by algorithm
- Intro to data mining
- MBAs, BI Admin, DBAs
- Focused on business issues
- Uses GUI
- Approachable for new users

Oracle R Enterprise
- Organized by process
- Intro to Oracle R Enterprise
- Data Scientists, BI Admin, DBAs
- Focused on executing R in Oracle Database
- Uses R scripts
- Technical
Important Factors in Getting Started

- Lots of internal experts and people who would like to be involved and learn
- Lots of people intimidated by what they don’t know
- Start by “level setting” and establishing a strong foundation
  - Bring people along on the journey, establish culture
  - Everyone shares a minimum common knowledge base
- Use workshops (JAD style session) for investigation of possibilities
  - Evaluation of data sources and data sets
  - Recognition of major business issues
  - Review of basic algorithms
  - Identification of potential PoC projects (plusses and minuses)
- Decide on pilot projects and who works on it
- Start simple and return value quickly
Analytics and Data Summit

All Analytics. All Data. No Nonsense.

March 20 – 22, 2018

Formerly called the BIWA Summit with the Spatial and Graph Summit

Same great technical content...new name!

www.AnalyticsandDataSummit.org

Call for speakers is now open with rolling acceptances until December 3, 2017
Add business card to basket or fill out card
Oracle Machine Learning & Advanced Analytics on Oracle Technology Network:

Oracle's ML & Advanced Analytics Overview presentation:

Oracle's Machine Learning and Advanced Analytics Data Management Platforms: Move the Algorithms; Not the Data white paper

Oracle ML & AA Product Management Internal Wiki/Workspace:
https://stbeehive.oracle.com/teamcollab/library/st/Oracle+Advanced+Analytics+PM+Workspace/Public+Documents

YouTube recorded Presentations and Demos:
https://blogs.oracle.com/datamining/entry/oracle_advanced-analytics_and_data


Analytics and Data Summit 2017, March 20-22 at Oracle HQ, Redwood Shore All Analytics. All Data. No Nonsense User Community. www.analyticsanddatasummit.org
Thank You!

Getting Started with Advanced Analytics in Finance, Marketing, and Operations

Tim Vlamis
tvlamis@vlamis.com
www.vlamis.com
@timvlamis