OTUG Kscope17
SAN ANTONIO, TEXAS • JUNE 25-29

PLEASE FILL OUT YOUR EVALUATIONS
Predictions and Seeing the Future Clearly: Tales from the Trenches of Advanced Analytics

Kscope 17
Tim Vlamis
Tuesday, June 27, 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
- Instructor for Oracle University’s Predictive Analytics, Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- Adjunct Professor of Business Benedictine College
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University
Presentation Agenda

- Three Quick Case Studies
- Grow your analytics projects, don’t build them
- Scale your projects
- Single seeds may or may not succeed
- In-database vs. BI vs. application

- Oracle Advanced Analytics
  - Oracle Data Mining
  - Oracle R Enterprise
- How to start with OAA – comparison of options
Three Quick Cases

- Warehouse layout using Association Rules/Market Basket
- Influencing specialty drink sales through associated sandwiches
- Predicting profitability through map views of concession sales
Evidence-based analysis requires data

“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)
Good Questions/Hypotheses are Needed

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

What is the Life Time Value of our customers? What’s a potential new customer worth?

Which products are purchased together most often? Which products are purchased with our most profitable products?
Start with a general plan
Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
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
Oracle Business Analytics Offerings

- Many Oracle applications have adaptive intelligence built in
  - Black box analytics
- OAC, OBIEE 12c, DVD all can use R
- R can either run on BI Server or in a warehouse using ORE
- Oracle managed cloud services do not involve R (BICS, DVCS)
  - R packages must be downloaded and managed externally.
- Oracle Data Mining and Oracle R Enterprise are faster and more capable than BI server based R.
Oracle Advanced Analytics (OAA) DB Option

Oracle Data Mining + Oracle R Enterprise

- Powerful in-database algorithms for Data Mining and Statistical Analysis
- Easy to add predictive analytics to enterprise applications and BI
- Fastest way to deliver scalable, enterprise-wide predictive analytics
- ORE eliminates R’s limitations (memory and speed) for Enterprise-scale analytics
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 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

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

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
# 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
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](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
R now integrated into OBIEE 11g and 12c
Oracle Advanced Analytics & Spatial

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
Dos and Don’ts

▪ Do
  ▪ See analytics as a continuous process
  ▪ Invest in an overall strategy, not in isolated tools and projects
  ▪ Put enough resources in place
  ▪ Grow your own talent and systems and involve IT
  ▪ Seek guidance and expertise early on

▪ Don’t
  ▪ Export your internal data and outsource to “experts”
  ▪ Try to develop predictive analytics “on the cheap”
  ▪ Anoint then Isolate internal experts
  ▪ Build multiple, parallel infrastructure systems without IT
  ▪ Struggle with basics and then reinvent the wheel
Open Source Software is like a Free Puppy

- You still have to feed it.
- You still have to train it.
- You still have to clean up after it.
- You still have to give it shots.
- You may not know what it will be like when it grows up.
- Mutts can be great dogs.
Forecasting for 30,000 Products

- A medical equipment company had complex forecasting needs
- International demand and distribution
- International production facilities
- High degree of seasonality (sports injuries)
- 445 pattern of months was creating problems for exponential smoothing algorithm
- Automated algorithms are powerful, but need experienced implementers.
Behavioral Segmentation

- Clustered approx 1000 restaurant locations into natural groups
- Rely on behavior, not traditional attributes for segmentation
- Many expected and unexpected patterns revealed
- Segmentation informed the development of marketing programs
Acting like a Venture Capitalist

- Place small bets on dozens of projects
- Double down and support most promising projects
- Hold a summit to teach predictive analytics and develop cases
- 2-5 days of hands on learning and case development
- Have participants bring business case and data set
- Group participants in functional roles to assist one another (table of operations, table of service, table of marketing, etc.)
- Have promising cases present to overall group to showcase new techniques and foster cross pollination and collaboration
- Requires a lot preplanning and onsite support
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
- Conduct a “predictive analytics summit” BYOD workshop
  - Two to Five days
- Conduct ODM and ORE training classes with 1-day workshop
- Use a defined Quick Start program (two weeks)
- Do a Proof of Concept project with experts (two to six weeks)
ODM Quick Start Overview

- Oracle Database Cloud Service
  - High Performance or Extreme Performance
- 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
Bring Your Own Data BYOD Workshop

- Requires preparation and pre-workshop activities by participants and organizers
- Develop business use cases and data sets in advance
- Prepare environments and verify data sets in advance
- Determine depth of predictive/advanced analytics to be used in workshop
- Include participants from across the organization
- Have experts/proctors with expertise in interface and analytics
- Spread risk across many business cases (VC model)
Oracle Data Mining Training (2 days)

- Introduction
- Data Mining Concepts and Terminology
- The Data Mining Process
- Introducing Oracle Data Miner
- 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
Oracle Test Drive

- Free to try Oracle BI, Advanced Analytics Go to [www.vlamis.com/td](http://www.vlamis.com/td)
- Runs off of Oracle Cloud
- Test Drives for:
  - Oracle BI
  - Oracle Advanced Analytics
- Once sign up, you have private instance for one day
- Available now
Add business card to basket or fill out card
Thank You!

Predictions and Seeing the Future Clearly

Tim Vlamis
tvlamis@vlamis.com
@TimVlamis
@VlamisSoftware
www.vlamis.com
ODTUG Kscope17
SAN ANTONIO, TEXAS • JUNE 25-29

PLEASE FILL OUT YOUR EVALUATIONS