Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics/
Oracle Data Miner Hands on Lab

Charlie Berger, MS Engineering, MBA, Sr. Director Product Management, Advanced Analytics and Machine Learning
charlie.berger@oracle.com
www.twitter.com/CharlieDataMine

Tim Vlamis, Consultant, Vlamis Software Solutions, Inc.

Karl Rexer, President, Rexer Analytics
Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

• **Lessons—Novices**
  1. OAA Quick Overview
  2. Data Mining Concepts Briefly
  3. Quick Oracle Data Miner GUI Demo
  4. Take off! Do as many Tutorials as you can in the 2 hours HOL
  5. Ask questions! We’re all here to help and discuss use cases!

• **Take off!—Intermediate/Experts**
  1. Environment
     - Oracle 12c on the Oracle Database Cloud
     - Will be using SQL Developer 4.2 EA
  2. Do 3-5 Tutorials
     - Instructors will walk around helping
Oracle’s Advanced Analytics
Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

- Parallel, scalable data mining algorithms and R integration
- In-Database + Hadoop—Don’t move the data
- Data analysts, data scientists & developers
- Drag and drop workflow, R and SQL APIs
- Extends data management into powerful advanced/predictive analytics platform
- Enables enterprise predictive analytics deployment + applications
Oracle Data Mining/ Machine Learning/Predictive Analytics

Data Preparation & Adv. Analytical Process Runs In-Database

Oracle Database 12c

Additional relevant data and “engineered features”

Historical data

Assembled historical data

Build Predictive Model

Historical or Current Data to be “scored” for predictions

Make Predictions

Predictions & Insights

Sensor data, Text, unstructured data, transactional data, spatial data, etc.
Oracle Advanced Analytics 12.2
Model Build Time Performance

<table>
<thead>
<tr>
<th>OAA 12.2 Algorithms</th>
<th>Rows (Ms)</th>
<th>T7-4 (Sparc &amp; Solaris)</th>
<th>X5-4 (Intel and Linux)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes Importance</td>
<td>640</td>
<td>28s / 512</td>
<td>44s / 72</td>
</tr>
<tr>
<td>K Means Clustering</td>
<td>640</td>
<td>161s / 256</td>
<td>268s / 144</td>
</tr>
<tr>
<td>Expectation Maximization</td>
<td>159</td>
<td>455s / 512</td>
<td>588s / 144</td>
</tr>
<tr>
<td>Naive Bayes Classification</td>
<td>320</td>
<td>17s / 256</td>
<td>23s / 72</td>
</tr>
<tr>
<td>GLM Classification</td>
<td>640</td>
<td>154s / 512</td>
<td>363s / 144</td>
</tr>
<tr>
<td>GLM Regression</td>
<td>640</td>
<td>55s / 512</td>
<td>93s / 144</td>
</tr>
<tr>
<td>Support Vector Machine (IPM solver)</td>
<td>640</td>
<td>404s / 512</td>
<td>1411s / 144</td>
</tr>
<tr>
<td>Support Vector Machine (SGD solver)</td>
<td>640</td>
<td>84s / 256</td>
<td>188s / 72</td>
</tr>
</tbody>
</table>

Prelim/Unofficial

The way to read their results is that they compare 2 chips: X5 (Intel and Linux) and T7 (Sparc and Solaris). They are measuring scalability (time in seconds) with increase degree of parallelism (dop). The data also has high cardinality categorical columns which translates to 9K mining attributes (when algorithms require exploration). There are no comparisons to 12.1 and it is fair to say that the 12.1 algorithms could not run on data of this size.

Wow! That’s Fast!
Oracle’s Advanced Analytics and Machine Learning Platform
Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

Information Producers
- R Client
- SQLDEV/Oracle Data Miner

Information Consumers
- OBIEE/DV
- Applications

Users
- R programmers
- Data & Business Analysts

Business Analysts/Mgrs
- Domain End Users (HCM, CRM)

Platform

Hadoop
- ORAAH
  Parallel, distributed algorithms

Oracle Database Enterprise Edition
- Oracle Advanced Analytics - Database Option
  SQL Data Mining, ML & Analytic Functions + R Integration
  for Scalable, Distributed, Parallel in-DB ML Execution

Oracle Cloud
Oracle Data Miner GUI
Easy to Use for “Citizen Data Scientist”

- Easy to use to define analytical methodologies that can be shared
- SQL Developer Extension
- Workflow API and generates SQL code for immediate deployment
Take off!—Intermediate/Experts

Quick Set up Overview
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab

• We’re using the Oracle Database Cloud!
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab

• We’re using the Oracle Database Cloud!

![Image of Remote Desktop Connection window with Oracle Cloud and Oracle Advanced Analytics diagram]
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab
Check your connection! Service name should be simply “DEMOS”

1. Change & Save Connection
2. Restart SQLDEV & ODMr
Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

• Step 1—Install SQLDEV 4.1.3
• Step 2—Connect to Oracle Database Cloud
  – 1. Go to Oracle Data Miner & create a NEW Connection e.g. HOL”N”
  – Select HOL”N” from drop down menu
  – Optionally may need to upgrade older Data Mining repository (may take 3 mins)
  – You are done!
Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

• Step 1—Install SQLDEV 4.1.3
• Step 2—Connect to Oracle Database Cloud
  – Connect as SYS/Welcome#1
  – Start to run task
  – Running task (may take 3 mins)
Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

• Step 1—Install SQLDEV 4.1.3
• Step 2—Connect to Oracle Database Cloud
• Step 3—Start HOL!
  – dmuser/dmuser
  – Demo data for learning
  – Follow 3-5 OBE Online Tutorials
    1. Using Oracle Data Miner 4.1
    2. Star Schema Mining Using Oracle Data Miner 4.1
    3. Text Mining with an EM Clustering Model Using Data Miner 4.1
    4. Anomaly Detection (CLAIMS) See Instructor for assistance
    5. Market Basket Analysis (SH.SALES) See Instructor for assistance
OAA/Oracle Data Miner 4.2 HOL
We’re Using the Oracle by Example Free Online Tutorials

• Google “Oracle Data Miner”
• Scroll down to bottom of page & launch tutorials
    NO::P24_CONTENT_ID,P24_PREV_PAGE:11925,2

Demos and Training

- 2-Day Oracle University Course
- Oracle Data Miner Workflow GUI Tutorials
- Oracle Data Miner on Big Data Lite VM Tutorial
- OAA YouTube demos
- Sample Programs
There are 6 Tutorials

- The first tutorial is already done for you
- Recommend doing 3-5 Tutorials

1. Using Oracle Data Miner 4.1
2. Star Schema Mining Using Oracle Data Miner 4.1
3. Text Mining with an EM Clustering Model Using Data Miner 4.1
4. Anomaly Detection (CLAIMS) See Instructor for assistance
5. Market Basket Analysis (SH.SALES) See Instructor for assistance
Great book on Oracle Advanced Analytics
Available on Amazon or from Author

• Predictive Analytics Using Oracle Data Miner: Develop for ODM in SQL & PL/SQL
OAA/Oracle Data Miner 4.1 HOL
Setting Up Oracle Data Miner

Setting Up Oracle Data Miner 4.1

Done

New / Select Database Connection

Connection Name: admin
Username: SYS
Password: ********

Oracle
Connection Type: Basic
Hostname: localhost
Port: 1521
Service name: orcl
Role: SYSDBA

Save, Clear, Test, Connect, Cancel
Setting Up Oracle Data Miner 4.1

Done

Create User

User Name: dmuser
New Password: ********
Confirm Password: ********

Default Tablespace: USERS
Temporary Tablespace: TEMP

Connections

- Create a Data Miner User Account
- Create a SQL Developer Connection for the Data Miner User
- Install the Data Miner Repository
- Want to Learn More?
Introducing the Data Miner Interface
Examining Oracle Data Miner Nodes

Data
- Create Table or View
- Data Source
- Explore Data
- Update Table

Transforms
- Aggregate
- Filter Columns
- Filter Columns Details
- Filter Rows
- Join
- Sample
- Transform

Text
- Apply Text
- Build Text
- Text Reference
Examining Oracle Data Miner Nodes

Models

Evaluate and Apply

Linking
Previewing a Data Miner Workflow
Rapidly Build, Evaluate & Deploy Analytical Methodologies
Leveraging a Variety of Data Sources and Types

SQL Joins and arbitrary SQL transforms & queries – power of SQL

Consider:
• Demographics
• Past purchases
• Recent purchases
• Comments & tweets

Unstructured data also mined by algorithms

Generates SQL scripts and workflow API for deployment

Transactional POS data

Modeling Approaches
Previewing a 4.2 Feature

Workflow Scheduler
Previewing a 4.2 Feature Workflow Scheduler
Previewing a 4.2 Feature

Workflow Scheduler
Previewing a 4.2 Feature

Workflow Scheduler
Previewing a 4.2 Feature

Workflow Scheduler
Previewing a 4.2 Feature

Workflow Scheduler—Email Notifications
The Data Mining Sample Programs

• The Data Mining Sample Programs 12c Documentation
  – You can learn a great deal about the Oracle Data Mining API from the data mining sample programs. The programs illustrate typical approaches to data preparation, algorithm selection, algorithm tuning, testing, and scoring.
  – The programs are easy to use. They include extensive inline comments to help you understand the code. They delete all temporary objects on exit; you can run the programs repeatedly without setup or cleanup.
  – The data mining sample programs are installed with Oracle Database Examples in the demo directory under Oracle Home. The demo directory contains sample programs that illustrate many features of Oracle Database. You can locate the data mining files by doing a directory listing of dm*.sql.
The Data Mining Sample Programs
Attribute Importance Sample Code

BUILD THE MODEL

-- Cleanup old output table for repeat runs
BEGIN EXECUTE IMMEDIATE 'DROP TABLE a1_explain_output';
END;
/

-- Run the EXPLAIN routine to get attribute importance results
BEGIN
DEMS_PREDICTIVE_ANALYTICS.EXPLAIN(
  data_table_name => 'mining_data_build_v',
  explain_column_name => 'affinity_card',
  result_table_name => 'a1_explain_output');
END;
/

-- DISPLAY RESULTS

<table>
<thead>
<tr>
<th>ATTRIBUTE_NAME</th>
<th>EXPLANATORY_VALUE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSEHOLD_SIZE</td>
<td>.195</td>
<td>1</td>
</tr>
<tr>
<td>CUST_MARITAL_STATUS</td>
<td>.194</td>
<td>2</td>
</tr>
<tr>
<td>YRS_RESIDENCE</td>
<td>.115</td>
<td>3</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>.106</td>
<td>4</td>
</tr>
<tr>
<td>AGE</td>
<td>.102</td>
<td>5</td>
</tr>
<tr>
<td>OCCUPATION</td>
<td>.082</td>
<td>6</td>
</tr>
<tr>
<td>Y_BOX_GAMES</td>
<td>.077</td>
<td>7</td>
</tr>
<tr>
<td>HOME_THEATER_PACKAGE</td>
<td>.069</td>
<td>8</td>
</tr>
<tr>
<td>CUST_GENDER</td>
<td>.043</td>
<td>9</td>
</tr>
<tr>
<td>BOOKKEEPING_APPLICATION</td>
<td>.024</td>
<td>10</td>
</tr>
<tr>
<td>BULK_PACK_DISKETTES</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>COUNTRY_NAME</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>CUST_ID</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>CUST_INCOME_LEVEL</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>FLAT_PANEL_MONITOR</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>GS_DOC_SET_KANJI</td>
<td>.000</td>
<td>11</td>
</tr>
<tr>
<td>PRINTER_SUPPLIES</td>
<td>.000</td>
<td>11</td>
</tr>
</tbody>
</table>

17 rows selected
Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

• READY, SET, GO!!!!

• Recommend doing 3-5 Tutorials
  1. Using Oracle Data Miner 4.1
  2. Star Schema Mining Using Oracle Data Miner 4.1
  3. Text Mining with an EM Clustering Model Using Data Miner 4.1
  4. Anomaly Detection (CLAIMS) See Instructor for assistance
  5. Market Basket Analysis (SH.SALES) See Instructor for assistance
Novice/Introductory/Overviews

Quick Overview of Concepts, Process and Use Cases
What is Machine Learning, Data Mining & Predictive Analytics?

**Automatically** sifting through large amounts of data to create models that find previously hidden patterns, discover valuable new insights and make predictions

- Identify most important factor (*Attribute Importance*)
- Predict customer behavior (*Classification*)
- Predict or estimate a value (*Regression*)
- Find profiles of targeted people or items (*Decision Trees*)
- Segment a population (*Clustering*)
- Find fraudulent or “rare events” (*Anomaly Detection*)
- Determine co-occurring items in a “baskets” (*Associations*)

Copyright © 2016 Oracle and/or its affiliates. All rights reserved.
Predictive Analytics & Data Mining

Typical Use Cases

• Targeting the right customer with the right offer
• How is a customer likely to respond to an offer?
• Finding the most profitable growth opportunities
• Finding and preventing customer churn
• Maximizing cross-business impact
• Security and suspicious activity detection
• Understanding sentiments in customer conversations
• Reducing medical errors & improving quality of health
• Understanding influencers in social networks
Data Mining Provides Better Information, Valuable Insights and Predictions

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

Segment #3
IF CUST_MO > 7 AND INCOME < $175K, THEN Prediction = Lease 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 Advanced Analytics DB Option

In-Database Machine Learning Algorithms* — SQL & R & GUI Access

**Classification**
- Decision Tree
- Logistic Regression (GLM)
- Naïve Bayes
- Support Vector Machine (SVM)
- Random Forest

**Regression**
- Multiple Regression (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear Regression
- Linear Model
- Generalized Linear Model
- Multi-Layer Neural Networks

**Anomaly Detection**
- 1-Class Support Vector Machine

**Clustering**
- Hierarchical k-Means
- Orthogonal Partitioning Clustering
- Expectation-Maximization

**Market Basket Analysis**
- Apriori – Association Rules

**Predictive Queries**
- Clustering
- Regression
- Anomaly Detection
- Feature Extraction

**Attribute Importance**
- Minimum Description Length
- Unsupervised pair-wise KL div.

**Feature Extraction & Creation**
- Nonnegative Matrix Factorization
- Principal Component Analysis
- Singular Value Decomposition

**Text Mining**
- All OAA/ODM SQL ML support
- Explicit Semantic Analysis

**Time Series**
- Single & Double Exp. Smoothing

**Open Source R Algorithms**
- Ability to run any R package (9,000+) via Embedded R mode

* Ability to Mine Unstructured, Structured & Transactional data
+ Partitioned Models

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
## Oracle University’s Learn Predictive Analytics Using Oracle Data Mining Course Agenda

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>7. Using Clustering Models</td>
</tr>
<tr>
<td>2. Predictive Analytics and Data Mining Concepts</td>
<td>8. Performing Market Basket Analysis</td>
</tr>
<tr>
<td>3. The Data Mining Process</td>
<td>9. Performing Anomaly Detection</td>
</tr>
<tr>
<td>5. Using Classification Models</td>
<td>11. Using Predictive Queries</td>
</tr>
<tr>
<td>6. Using Regression Models</td>
<td>12. Deploying Data Mining Results</td>
</tr>
</tbody>
</table>
The Data Mining Process

Data Selection

Preparation
- Define variable pool
- Perform ETL

Preprocessing
- Normalization
- Binning
- Outlier Detection
- Missing Value Cleanup
- Transformations
- Derived Attributes

Model building

Deploying results
Data Mining Attributes

Data mining attributes can be:

- Categorical
- Numerical
- Text
- Target
- Active or Inactive
Building and Evaluating Models

Data

Preparation
- Define variable pool
- Perform ETL

Preprocessing
- Normalization
- Binning
- Outlier Detection
- Missing Value Cleanup
- Transformations
- Derived Attributes

Model Building
- Build
- Test
- Evaluate
- Apply

Deploying Results
Model Train & Test: Supervised Learning
Model Evaluation: Supervised Learning
Applying the Selected Model(s)

(* Output settings may be specified by the user.)
<table>
<thead>
<tr>
<th>CUST_ID</th>
<th>N_TRANS_ATM</th>
<th>CHECKING.Amount</th>
<th>MARRITAL_STATUS</th>
<th>SEX</th>
<th>N_TRANS_KING</th>
<th>STATE</th>
<th>HOUSE_OWNERSHIP</th>
<th>MONTHLY_CHECKS_WRITTEN</th>
<th>LTV</th>
<th>SALARY</th>
<th>BANK_FUNDS</th>
<th>RMU_INS</th>
<th>CREDIT_BALANCE</th>
<th>N_MORTGAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12464</td>
<td>6</td>
<td>25 WIDOWED</td>
<td>F</td>
<td>9 DC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12466</td>
<td>5</td>
<td>25 WIDOWED</td>
<td>M</td>
<td>2 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12467</td>
<td>2</td>
<td>25 DIVORCED</td>
<td>F</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12468</td>
<td>2</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12469</td>
<td>2</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12470</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>F</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12471</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12472</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>12473</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12474</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>12475</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12476</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12477</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12478</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12479</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12480</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12481</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>12482</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>12483</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>12484</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>12485</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>12486</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>12487</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>12488</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>12489</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>12490</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>12491</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>12492</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>12493</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>12494</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>12495</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>12496</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>12497</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>12498</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>12499</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>12500</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>12501</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>12502</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>12503</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1 MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>12504</td>
<td>1</td>
<td>25 DIVORCED</td>
<td>M</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If BANK_FUNDS > 246
And CHECKING_AMOUNT <= 282
And CREDIT_BALANCE <= 2445
And MONEY_MONTHLY_OVERDRAWN > 54.095
And T_AMOUNT_AUTOM_PAYMENTS <= 14993
Then Yes

Support: 0.034248323636363636
Confidence: 88.22%
<Details algorithm="Naive Bayes" class="Yes">
  <Attribute name="LIV_BIH" actualValue="MEDIUM" operator="in" range="LOW,MEDIUM,VERY HIGH" weight=".587" rank="1"/>
  <Attribute name="N_TRANS_WER_RANK" actualValue="2300" operator="greaterThan" value="1419.5" weight=".963" rank="2"/>
  <Attribute name="LIV" actualValue="21389.75" operator="greaterThan" value="6861.625" weight=".848" rank="3"/>
  <Attribute name="CAR_OWNERSHIP" actualValue="1" operator="between" range="(0,100]" weight=".313" rank="4"/>
  <Attribute name="AGZ" actualValue="31" operator="between" range="(17.5:53.5]" weight=".8" rank="5"/>
</Details>
PREDICTION_COUNT 731.00
FULL_SIMPLE_RULE (BANK_FUNDS > 248) AND (CHECKING_AMOUNT <= 283) AND (CREDIT_BALANCE <= 3445) AND (MONEY_MONTHLY_OVERDRAWN > 54,295) AND (T_AMOUNT_AUTOM_PAYMENTS <= 14963)
PREDICTION Yes
Sharing, Automation and Deployment
Immediately Go to “Productionization” of Analytical Methodologies

- Share ODMr workflows
- Workflow API for 100% automation
  - Immediate deployment of data analyst’s methodologies
- SQL Script Generation
  - Deploy methodology as SQL scripts
Fraud Prediction Demo

Automated In-DB Analytical Methodology

```sql
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values ('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO','ON');
commit;
begin
    dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION', 'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/
```

-- Top 5 most suspicious fraud policy holder claims
select * from
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,
    rank() over (order by prob_fraud desc) rnk from
(select POLICYNUMBER, prediction_probability(CLAIMSMODEL, '0' using *) prob_fraud from CLAIMS
where PASTNUMBEROFCLAIMS in ('2to4', 'morethan4'))
where rnk <= 5
order by percent_fraud desc;

Automated Monthly “Application”!

Just add:
Create View CLAIMS2_30
As
Select * from CLAIMS2
Where mydate > SYSDATE – 30

Time measure: set timing on;
Oracle Advanced Analytics
Real-Time Scoring, Predictions and Recommendations

• On-the-fly, single record apply with new data (e.g. from call center)

```
Select prediction_probability(CLAS_DT_1_64, 'Yes'
    USING 7800 as bank_funds, 125 as checking_amount, 20 as credit_balance, 55 as age, 'Married' as marital_status, 250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```

Likelihood to respond:

Query Result

All Rows Fetched: 1 in 0 seconds

<table>
<thead>
<tr>
<th>PREDICTION_PROB...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8382936507936...</td>
</tr>
</tbody>
</table>
OAA/Oracle Data Miner 4.1 HOL
Uses Oracle by Example Free Online Tutorials

• There are 6 Tutorials
  – The first tutorial is already done for you
  – Recommend doing 3-5 Tutorials
  1. Using Oracle Data Miner 4.1
  2. Star Schema Mining Using Oracle Data Miner 4.1
  3. Text Mining with an EM Clustering Model Using Data Miner 4.1
  4. Anomaly Detection (CLAIMS) See Instructor for assistance
  5. Market Basket Analysis (SH.SALES) See Instructor for assistance