Anomaly and Fraud Detection with Oracle Data Mining 11g Release 2

ORACLE

Charlie Berger
Sr. Director Product Management, Data Mining Technologies
Oracle Corporation
charlie.berger@oracle.com
www.twitter.com/CharlieDataMine

20 Ways to Detect Fraud

http://www.auditnet.org/testing_20_ways.htm

1. Unusual Behavior

The perpetrator will often display unusual behavior, that when taken as a whole is a strong indicator of fraud. The fraudster may not ever take a vacation or call in sick in fear of being caught. He or she may not assign out work even when overloaded. Other symptoms may be changes in behavior such as increased drinking, smoking, defensiveness, and unusual irritability and suspiciousness.

2. Complaints

Frequently tips or complaints will be received which indicate that a fraudulent action is going on. Complaints have been known to be some of the best sources of fraud and should be taken seriously. Although all too often, the motives of the complainant may be suspect, the allegations usually have merit that warrant further investigation.

Pretty Easy? Huh?

3. Stale Items in Reconciliations

In bank reconciliations, deposits or checks not included in the reconciliation could be indicative of theft. Missing deposits could mean the perpetrator absconded with the funds; missing checks could indicate one made out to a bogus payee.

4. Excessive Voids

Voided sales slips could mean that the sale was rung up, the payment diverted to the use of the perpetrator, and the sales slip subsequently voided to cover the theft.

5. Missing Documents

Documents which are unable to be located can be a red flag for fraud. Although it is expected that some documents will be misplaced, the auditor should look for explanations as to why the documents are missing, and what steps were taken to locate the requested items. All too often, the auditors will select an alternate item or allow the auditee to select an alternate without determining whether or not a problem exists.

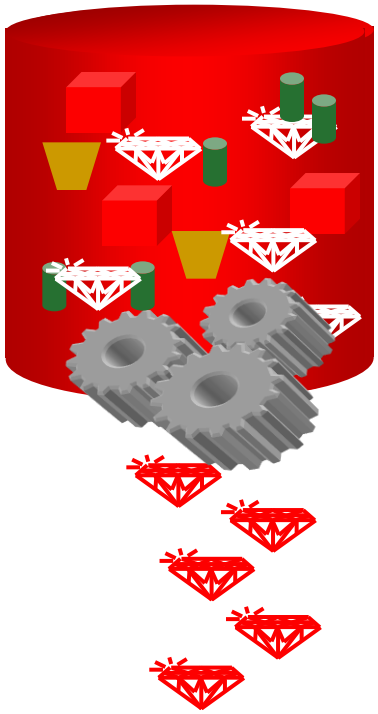
6. Excessive Credit Memos

Similar to excessive voids, this technique can be used to cover the theft of cash. A credit memo to a phony customer is written out, and the cash is taken to make total cash balance.

http://www.fraudiscovery.com/detect.html

What is Data Mining?

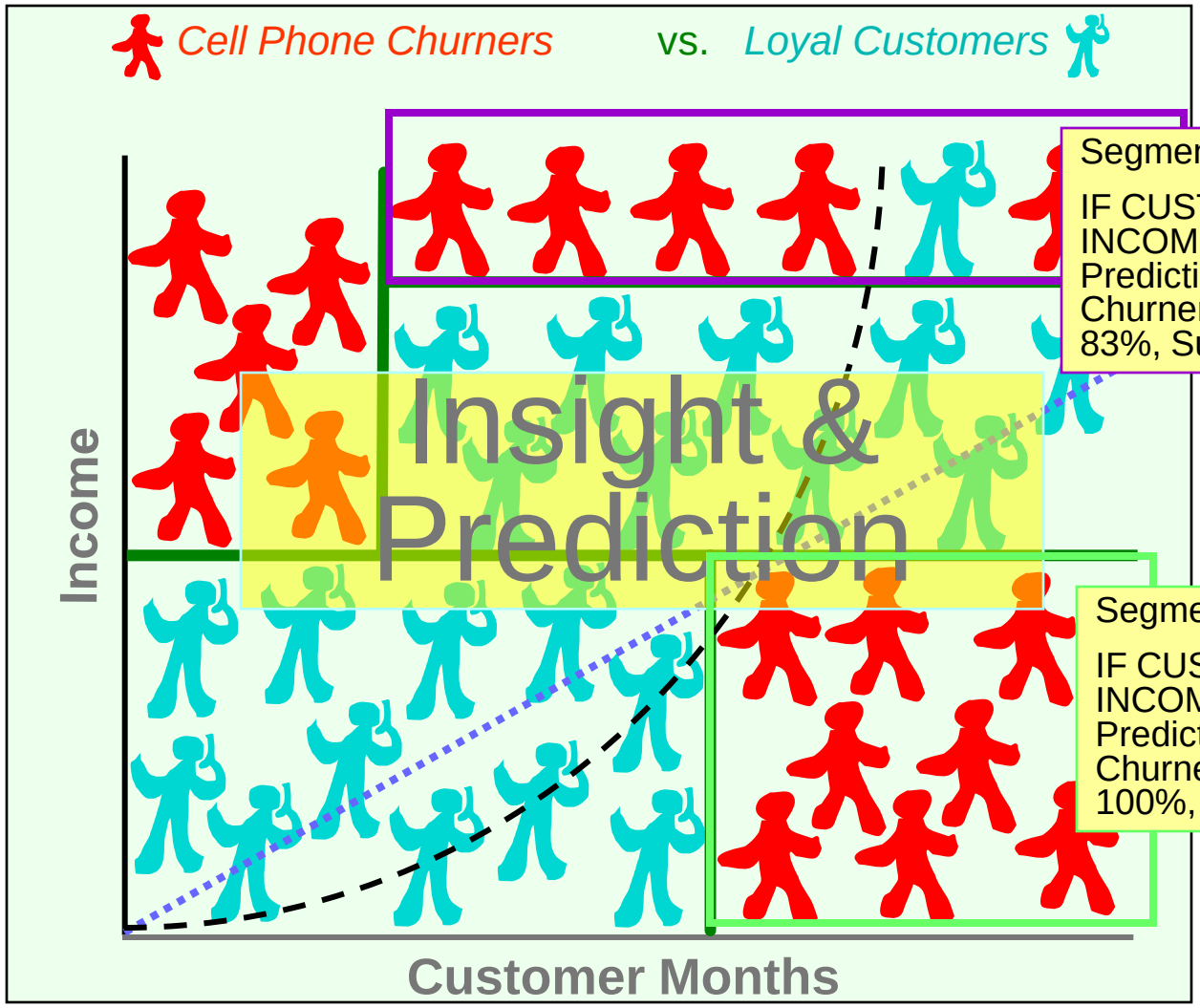
ORACLE DATABASE 11g



- 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 important relationships and "market baskets" within the population *(Associations)*
  - Find fraudulent or "rare events" *(Anomaly Detection)*

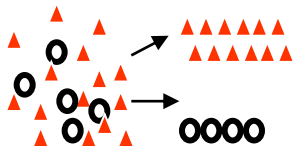
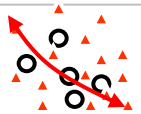
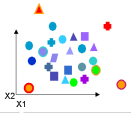
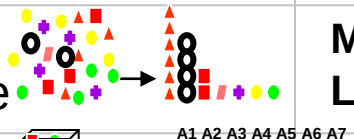
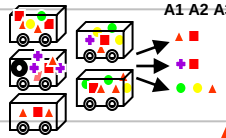
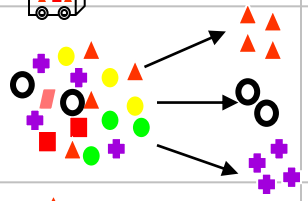
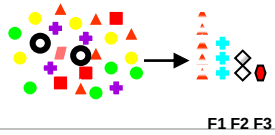
Data Mining Provides

Better Information, Valuable Insights and Predictions



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

Problem	Algorithm	Applicability
Classification 	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression 	Multiple Regression (GLM) Support Vector Machine	Classical statistical technique Wide / narrow data / text
Anomaly Detection 	One Class SVM	Lack examples
Attribute Importance 	Minimum Description Length (MDL)	Attribute reduction Identify useful data Reduce data noise
Association Rules 	Apriori	Market basket analysis Link analysis
Clustering 	Hierarchical K-Means Hierarchical O-Cluster	Product grouping Text mining Gene and protein analysis
Feature Extraction 	NMF	Text analysis Feature reduction

You Can Think of It Like This…

Traditional SQL

- "Human-driven" queries
- Domain expertise
- Any "*rules*" must be defined and managed

- SQL Queries
  - `SELECT`
  - `DISTINCT`
  - `AGGREGATE`
  - `WHERE`
  - `AND OR`
  - `GROUP BY`
  - `ORDER BY`
  - `RANK`

+

Oracle Data Mining

- Automated knowledge discovery, model building and deployment
- Domain expertise to assemble the "*right*" data to mine

- ODM "Verbs"
  - `PREDICT`
  - `DETECT`
  - `CLUSTER`
  - `CLASSIFY`
  - `REGRESS`
  - `PROFILE`
  - `IDENTIFY FACTORS`
  - `ASSOCIATE`

Fraud Prediction

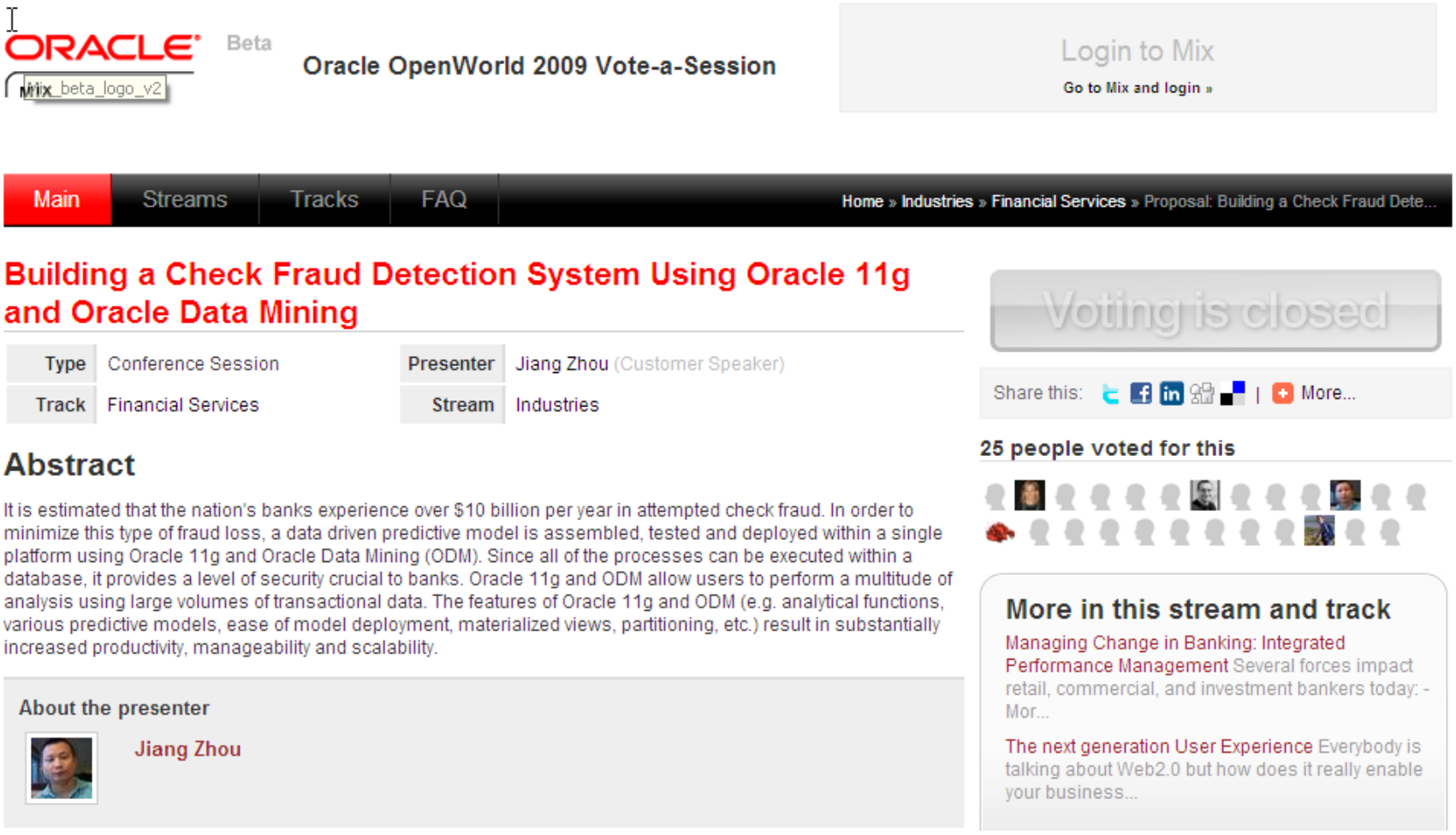
```
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values
('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO','ON');
commit;
```

```
begin
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
  'CLAIMS2', '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 CLAIMS2
where PASTNUMBEROFCLAIMS in ('2 to 4', 'more than 4')))
where rnk <= 5
order by percent_fraud desc;
```

POLICYNUMBER	PERCENT_FRAUD	RNK
6532	64.78	1
2749	64.17	2
3440	63.22	3
654	63.1	4
12650	62.36	5

Building a Check Fraud Detection System Using Oracle 11g and Oracle Data Mining



ORACLE®

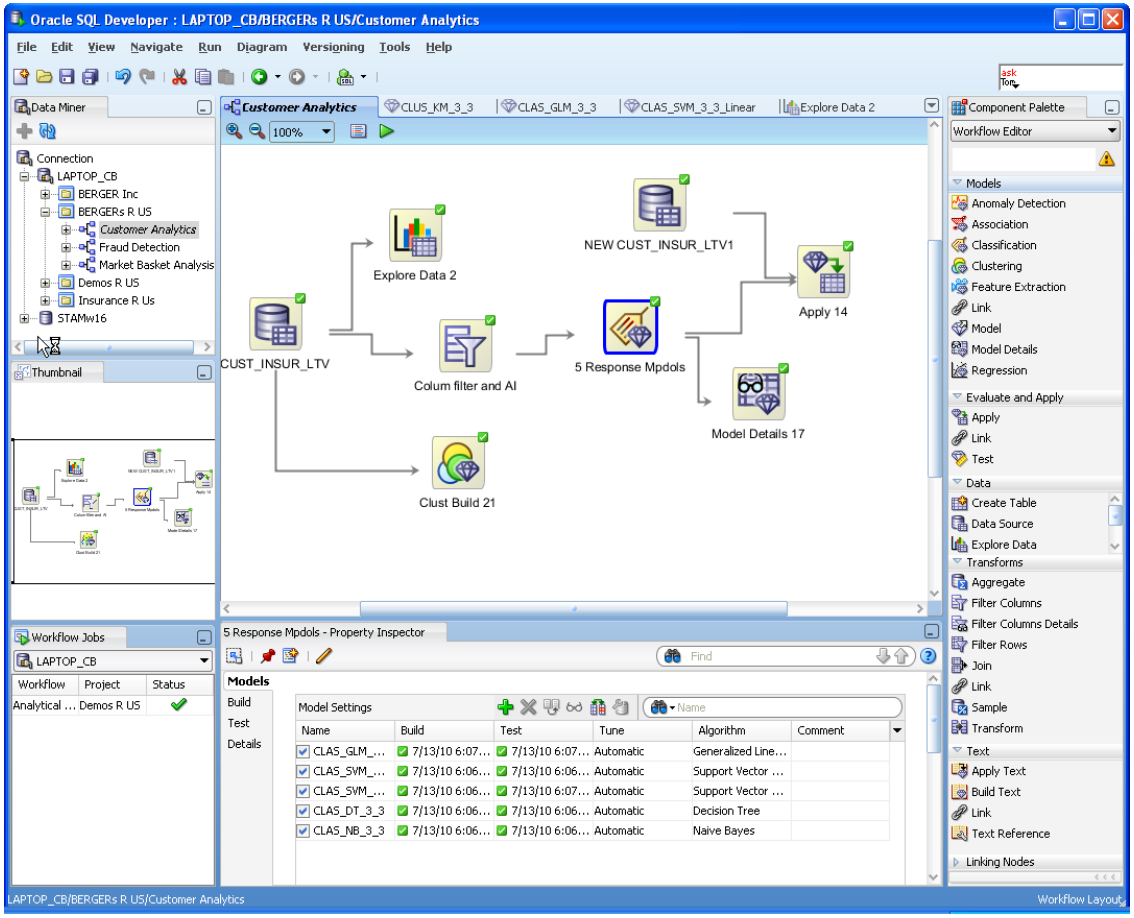
- 11 years “stem celling analytics” into Oracle
  - Designed advanced analytics into database kernel to leverage relational database strengths
  - Naïve Bayes and Association Rules—1st algorithms added
  - Leverages counting, conditional probabilities, and much more
- Now, analytical database platform
  - 12 cutting edge machine learning algorithms and 50+ statistical functions
  - A data mining model is a schema object in the database, built via a PL/SQL API and scored via built-in SQL functions.
  - When building models, leverage existing scalable technology
    - (e.g., parallel execution, bitmap indexes, aggregation techniques) and add new core database technology (e.g., recursion within the parallel infrastructure, IEEE float, etc.)
  - True power of embedding within the database is evident when scoring models using built-in SQL functions (incl. Exadata)

```
select cust_id
from customers
where region = 'US'
and prediction probability(churnmod, 'Y' using *) > 0.8;
```

Oracle Data Miner 11g Release 2 GUI

Free product on OTN for Oracle Data Mining Option

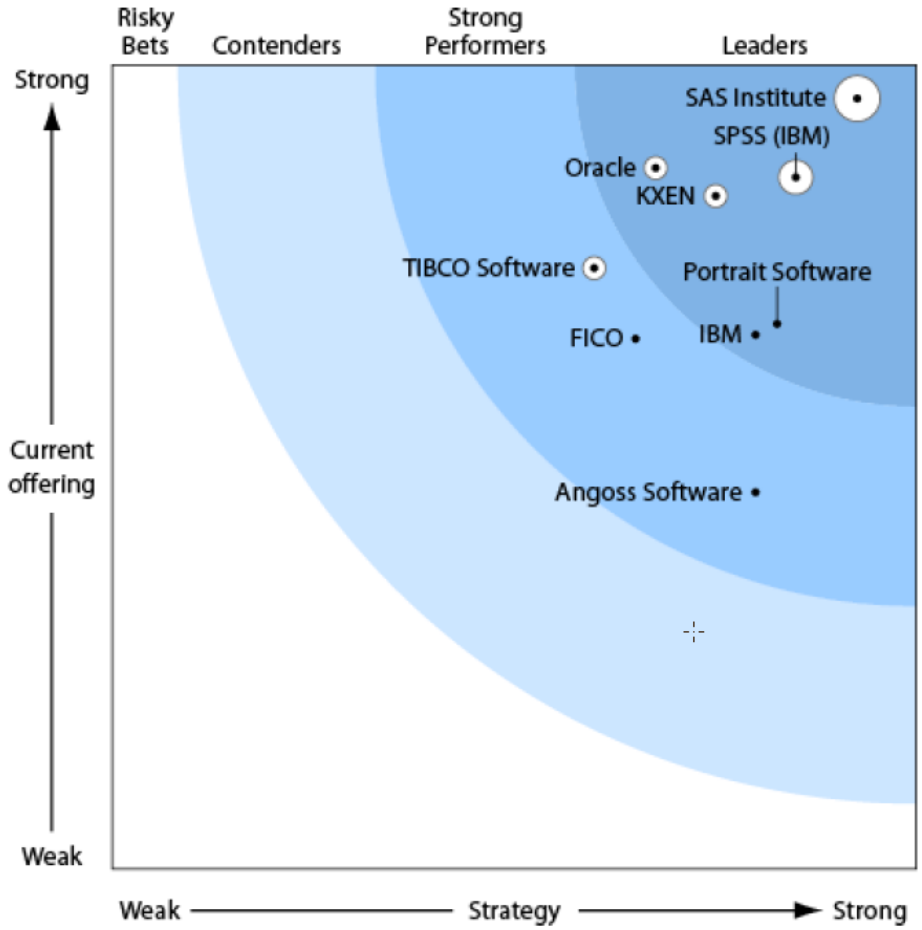
- Graphical User Interface for data analyst
- SQL Developer Extension (OTN download)
- Explore data—discover new insights
- Build and evaluate data mining models
- Apply predictive models
- Share analytical workflows
- Deploy SQL Apply code/scripts



The Forrester Wave™: Predictive Analytics And Data Mining Solutions, Q1 2010

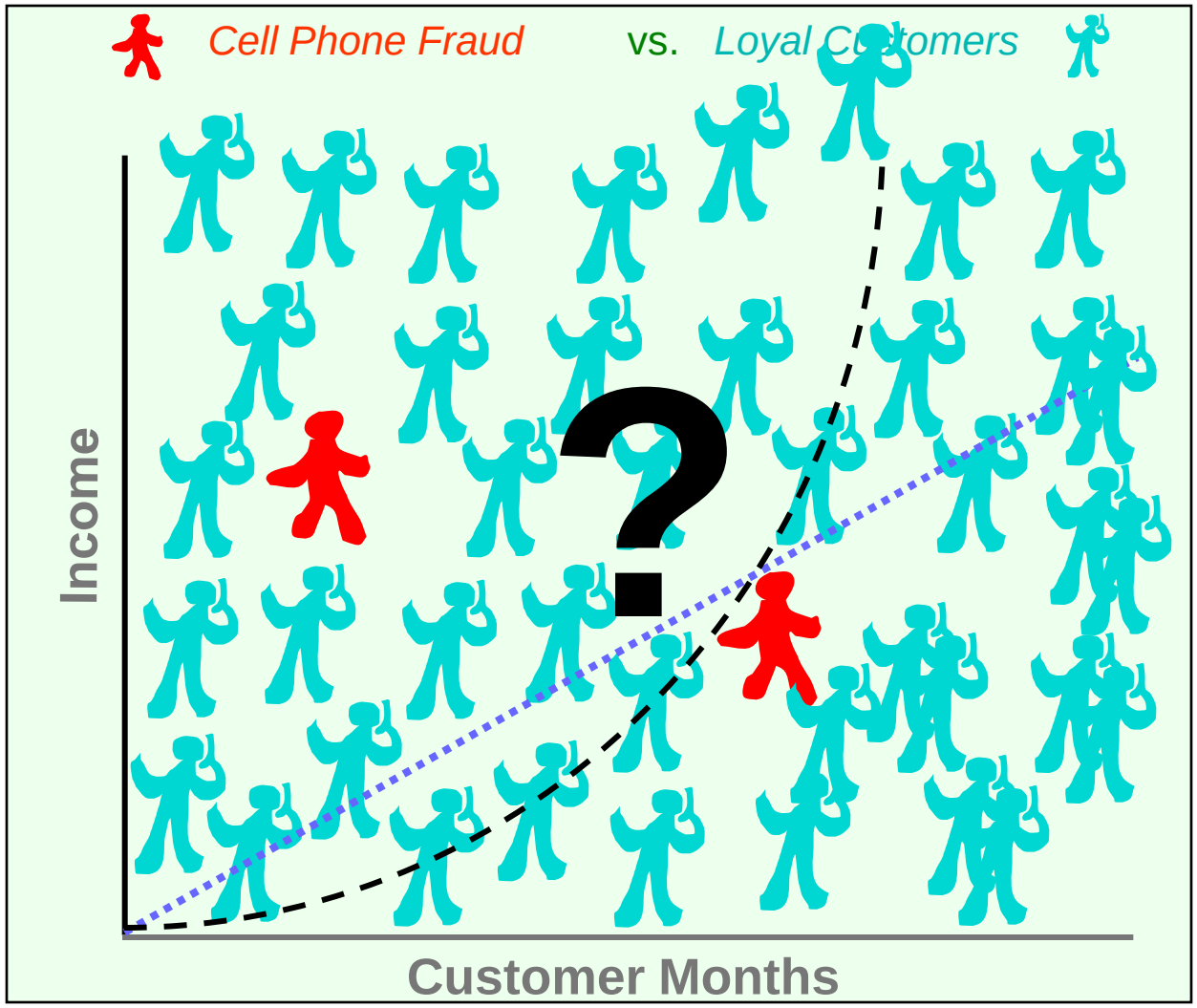
Oracle Data Mining Cited as a Leader; 2nd place in Current Offering

- Ranks 2nd place in Current Offering
- "Oracle focuses on in-database mining in the Oracle Database, on integration of Oracle Data Mining into the kernel of that database, and on leveraging that technology in Oracle's branded applications."



Data Mining Provides

Better Information, Valuable Insights and Predictions



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

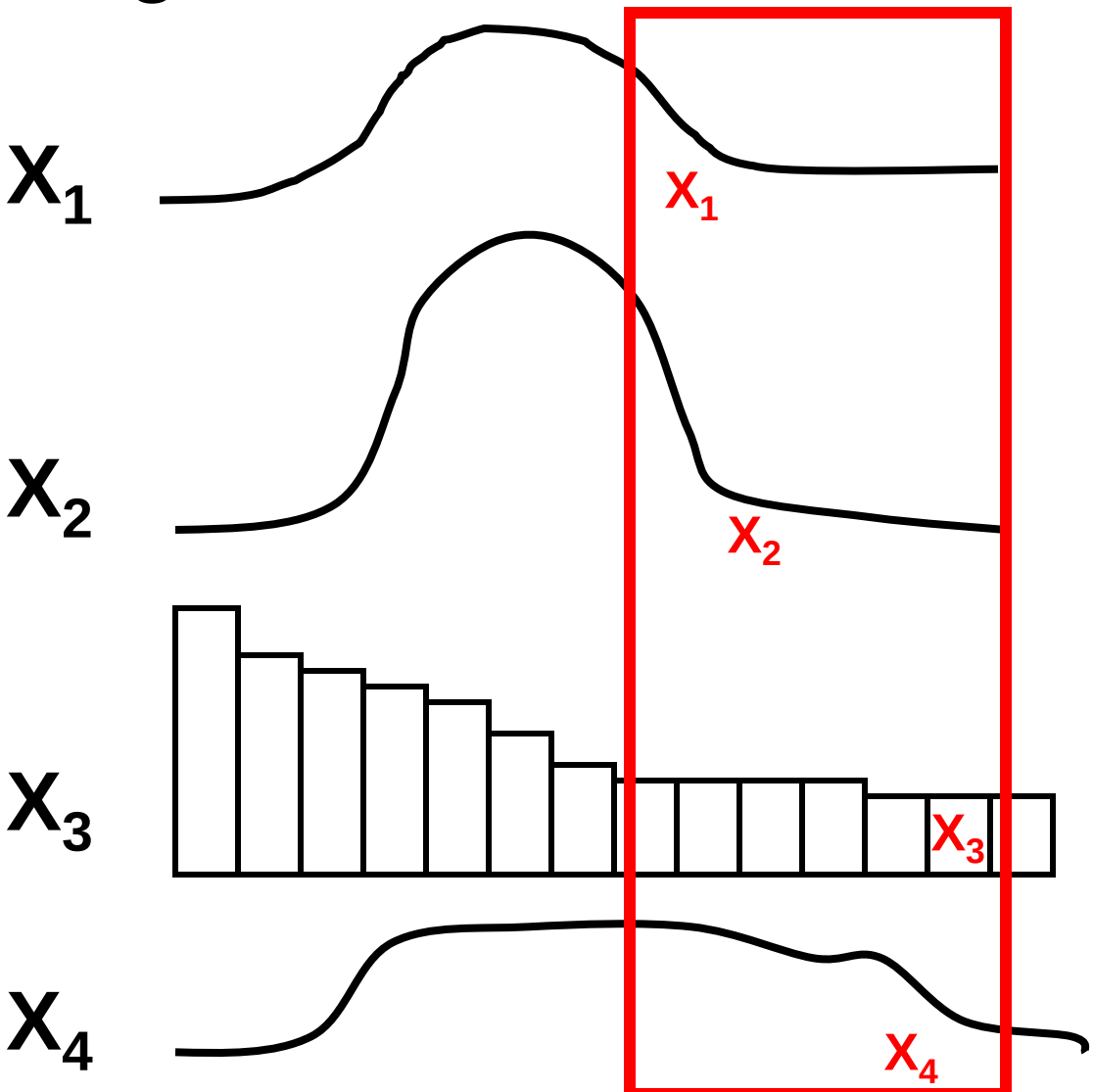
Finding Needles in Haystacks

- Haystacks are usually BIG
- Needles are typically **small** and rare

Look for What is "*Different*"

Challenge: Finding Anomalous Records

- Considering multiple attributes, taken collectively, a record may appear anomalous



In-Database Data Mining

Traditional Analytics

- Data Import
- Data Mining Model "Scoring"
- Data Preparation and Transformation
- Data Mining Model Building
- Data Prep & Transformation
- Data Extraction

Oracle Data Mining

Savings

- Model "Scoring"
- Embedded Data Prep
- Model Building
- Data Preparation

Results

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

Model "Scoring"
Data remains in the Database

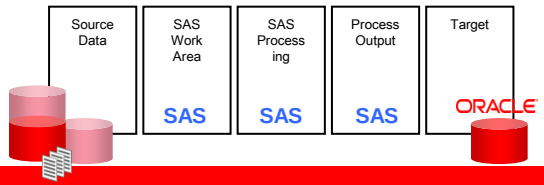
Embedded data preparation

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

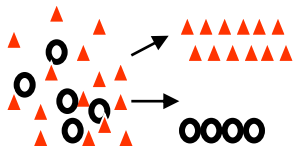
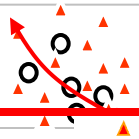
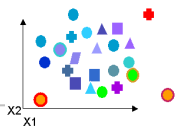
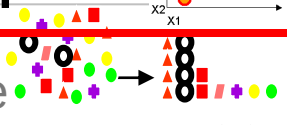
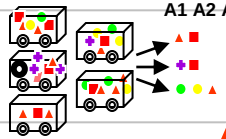
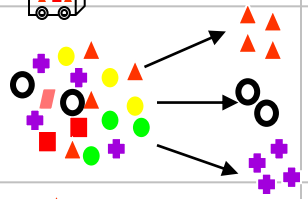
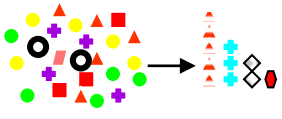
Hours, Days or Weeks



Secs. Mins or Hours

Oracle Data Mining

Algorithm Summary 11g Release 2

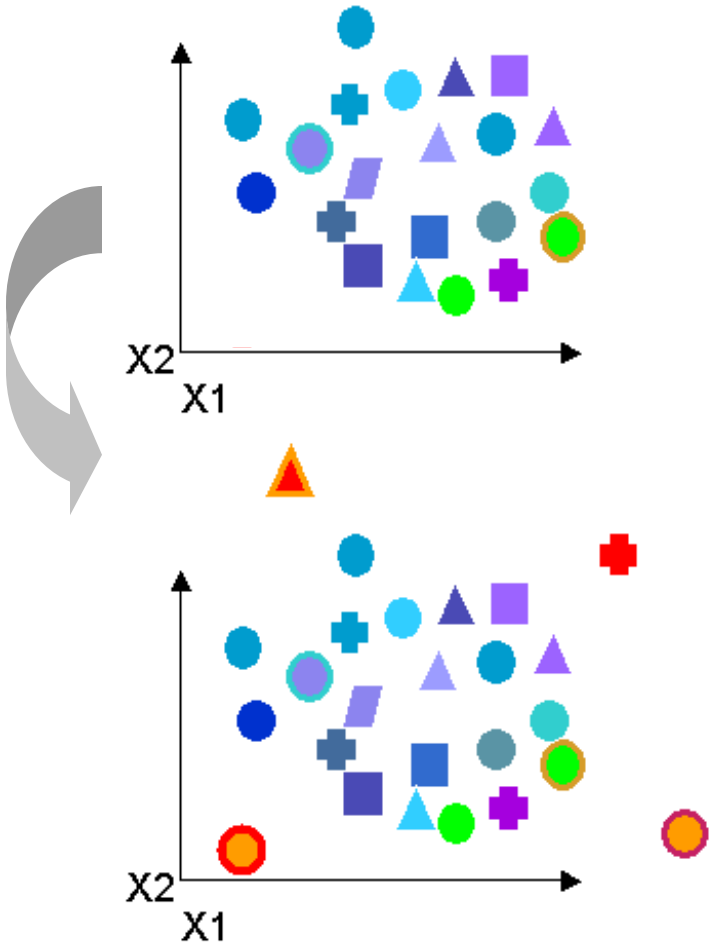
Problem	Algorithm	Applicability
Classification 	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression 	Multiple Regression (GLM) Support Vector Machine	Classical statistical technique Wide / narrow data / text
Anomaly Detection 	**One Class SVM**	**Lack examples**
Attribute Importance 	Minimum Description Length (MDL)	Attribute reduction Identify useful data Reduce data noise
Association Rules 	Apriori	Market basket analysis Link analysis
Clustering 	Hierarchical K-Means Hierarchical O-Cluster	Product grouping Text mining Gene and protein analysis
Feature Extraction 	NMF	Text analysis Feature reduction

Oracle Data Mining 11g Release 2

Anomaly Detection

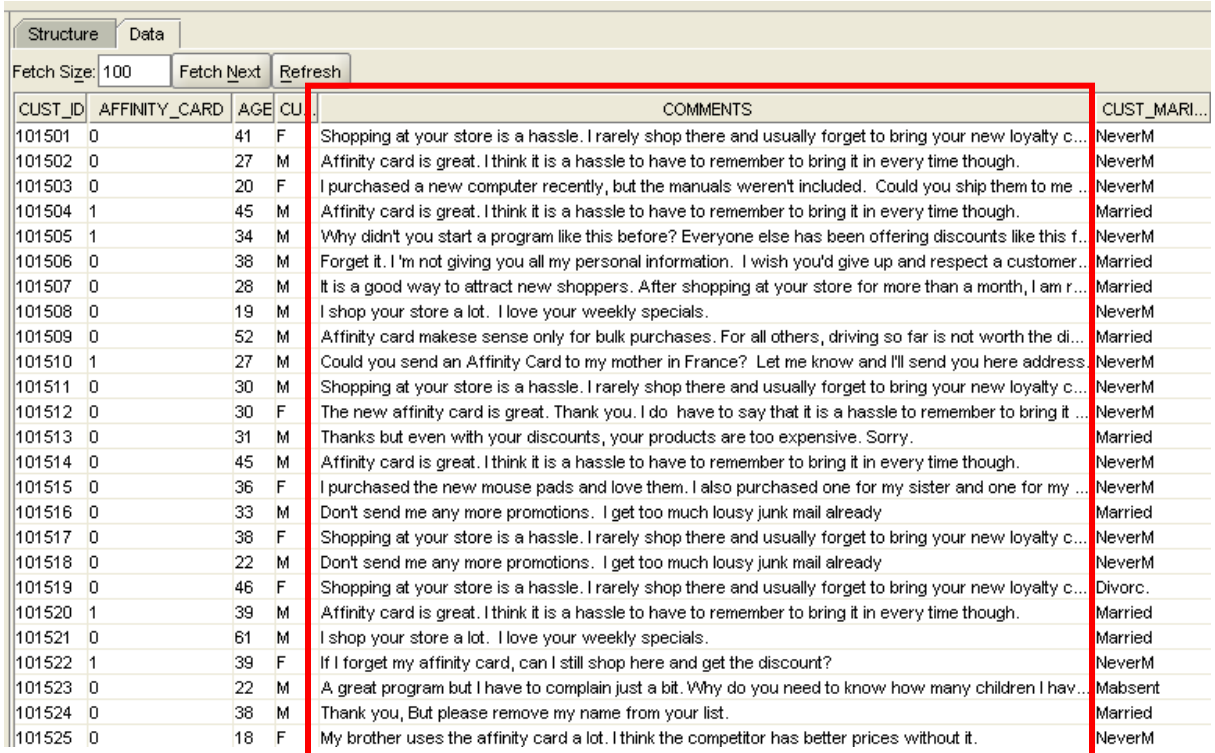
Problem: Detect rare cases

- Rare events, true novelty
- Fraud, noncompliance
- Disease outbreaks
- Outlier detection
- Network intrusion detection
- Oracle Data Mining
  - "One-Class" SVM Models



Oracle Data Mining and Unstructured Data

- Oracle Data Mining mines unstructured i.e. “text” data
- Include free text and comments in ODM models
- Cluster and Classify documents
- Oracle Text used to preprocess unstructured text

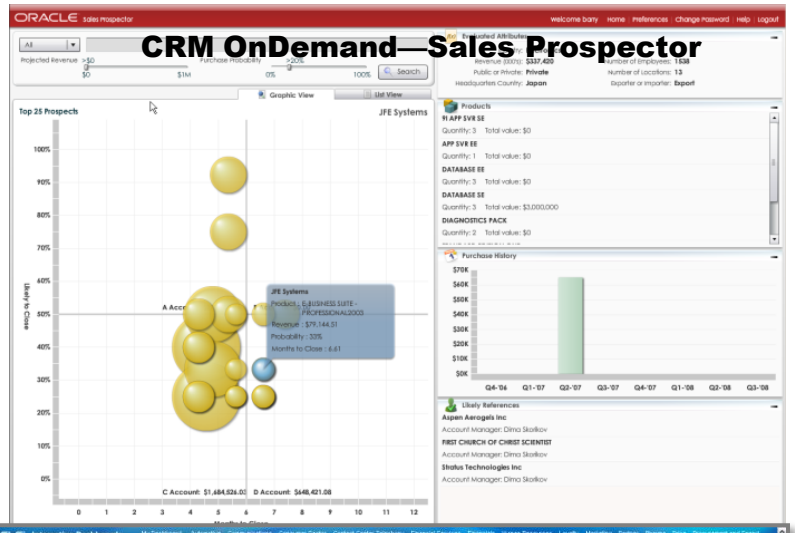

Structure | Data

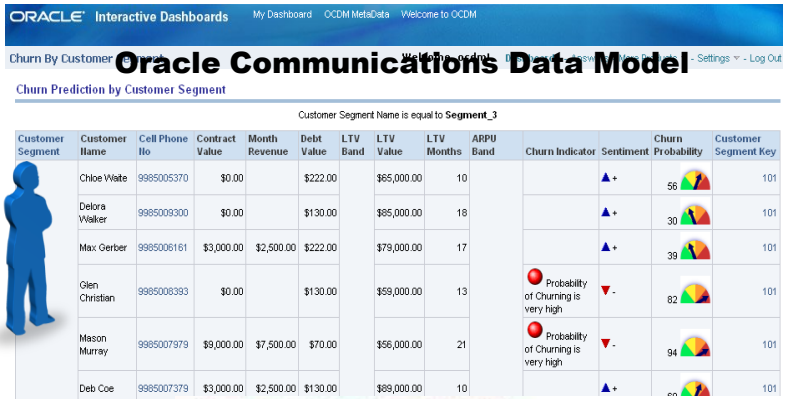
Fetch Size: 100 | Fetch Next | Refresh

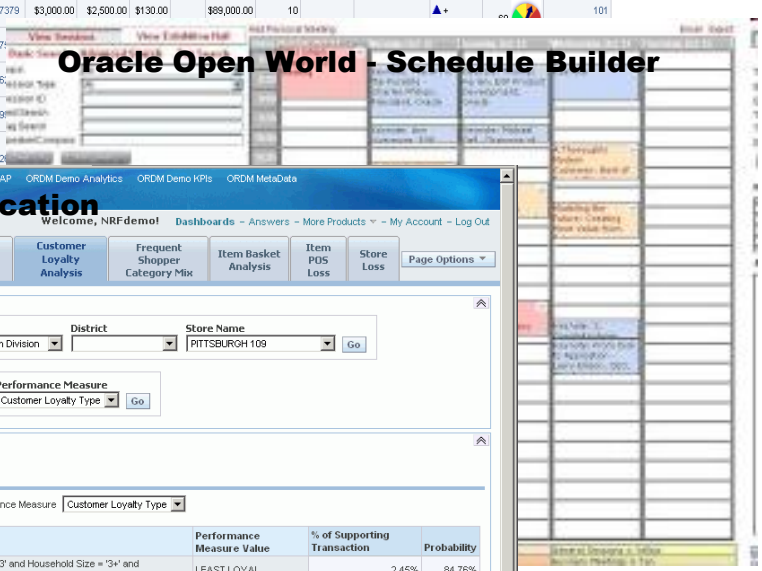
CUST_ID	AFFINITY_CARD	AGE	CU...	COMMENTS	CUST_MARI...
101501	0	41	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101502	0	27	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	NeverM
101503	0	20	F	I purchased a new computer recently, but the manuals weren't included. Could you ship them to me ...	NeverM
101504	1	45	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	Married
101505	1	34	M	Why didn't you start a program like this before? Everyone else has been offering discounts like this f...	NeverM
101506	0	38	M	Forget it. I 'm not giving you all my personal information. I wish you'd give up and respect a customer...	Married
101507	0	28	M	It is a good way to attract new shoppers. After shopping at your store for more than a month, I am r...	Married
101508	0	19	M	I shop your store a lot. I love your weekly specials.	NeverM
101509	0	52	M	Affinity card makese sense only for bulk purchases. For all others, driving so far is not worth the di...	Married
101510	1	27	M	Could you send an Affinity Card to my mother in France? Let me know and I'll send you here address.	NeverM
101511	0	30	M	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101512	0	30	F	The new affinity card is great. Thank you. I do have to say that it is a hassle to remember to bring it ...	NeverM
101513	0	31	M	Thanks but even with your discounts, your products are too expensive. Sorry.	Married
101514	0	45	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	NeverM
101515	0	36	F	I purchased the new mouse pads and love them. I also purchased one for my sister and one for my ...	NeverM
101516	0	33	M	Don't send me any more promotions. I get too much lousy junk mail already	Married
101517	0	38	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101518	0	22	M	Don't send me any more promotions. I get too much lousy junk mail already	NeverM
101519	0	46	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	Divorc.
101520	1	39	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	Married
101521	0	61	M	I shop your store a lot. I love your weekly specials.	Married
101522	1	39	F	If I forget my affinity card, can I still shop here and get the discount?	NeverM
101523	0	22	M	A great program but I have to complain just a bit. Why do you need to know how many children I hav...	Mabsent
101524	0	38	M	Thank you, But please remove my name from your list.	Married
101525	0	18	F	My brother uses the affinity card a lot. I think the competitor has better prices without it.	NeverM

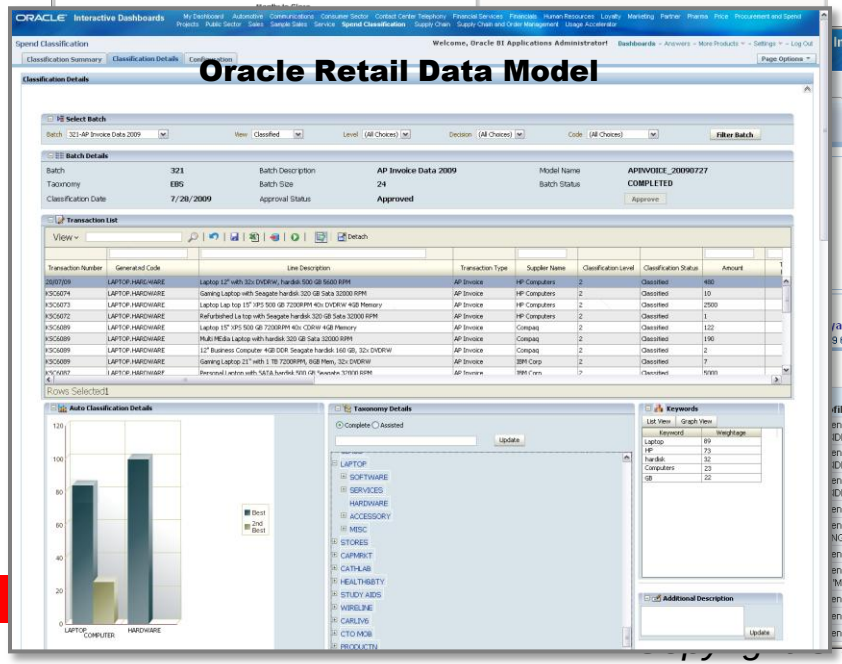
Predictive Analytics Applications

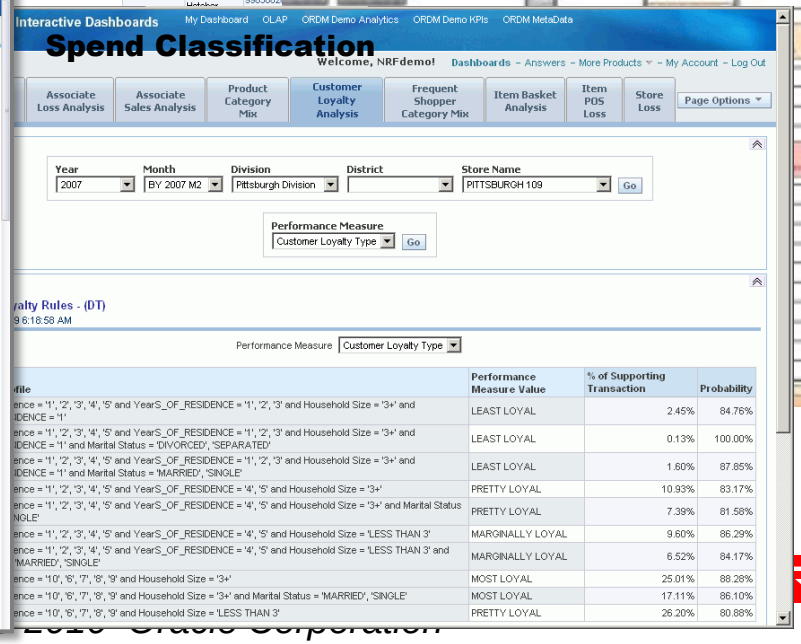
Powered by Oracle Data Mining (Partial List as of March 2010)









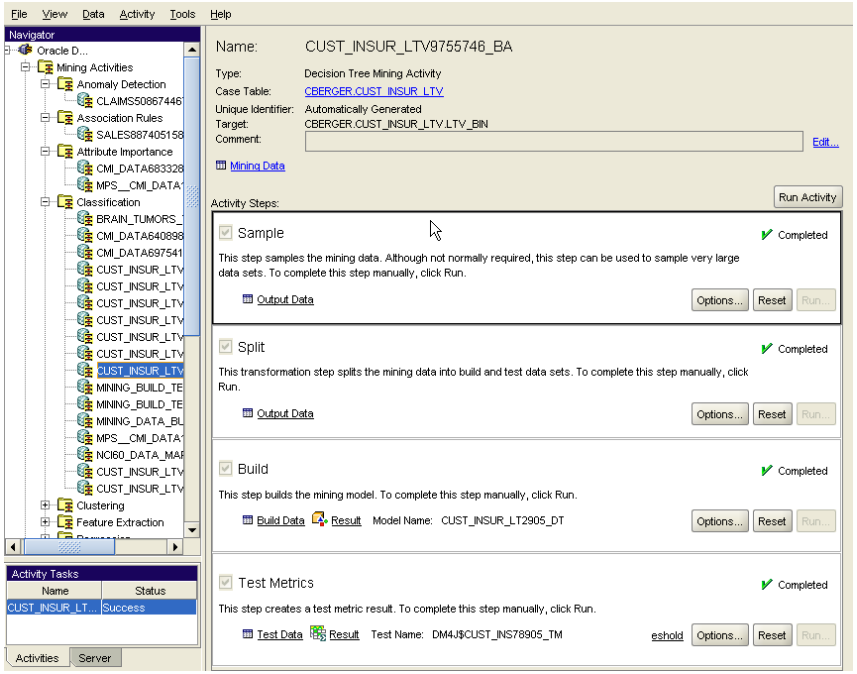


Oracle Data Miner 11g Release 2 GUI

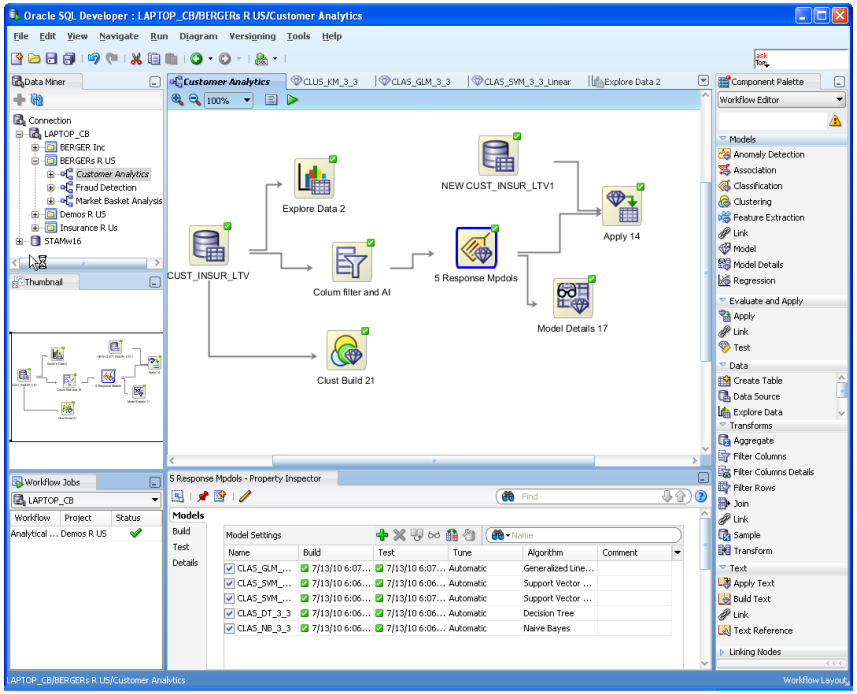
Oracle Data Miner GUIs

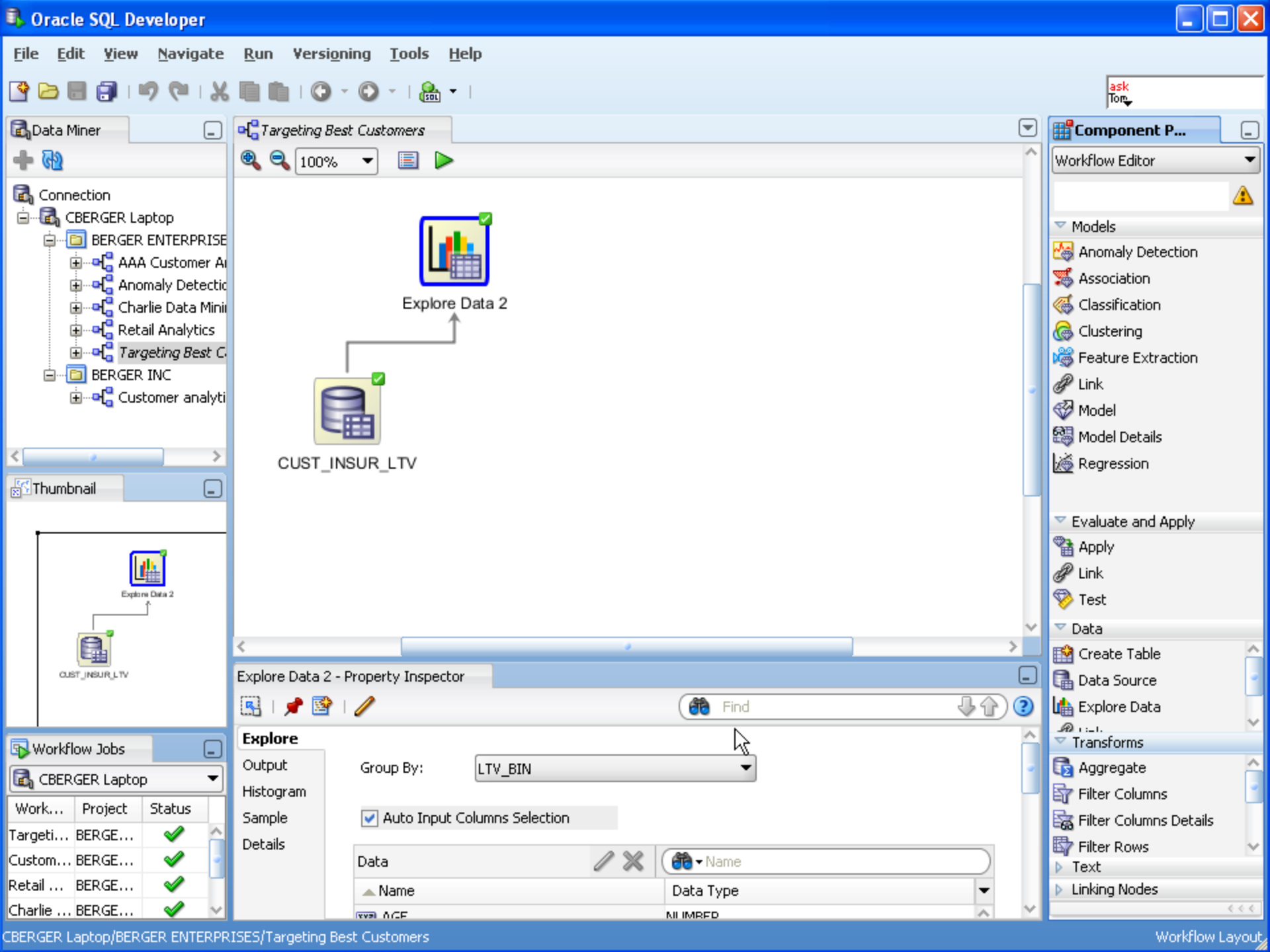
Free Products on OTN

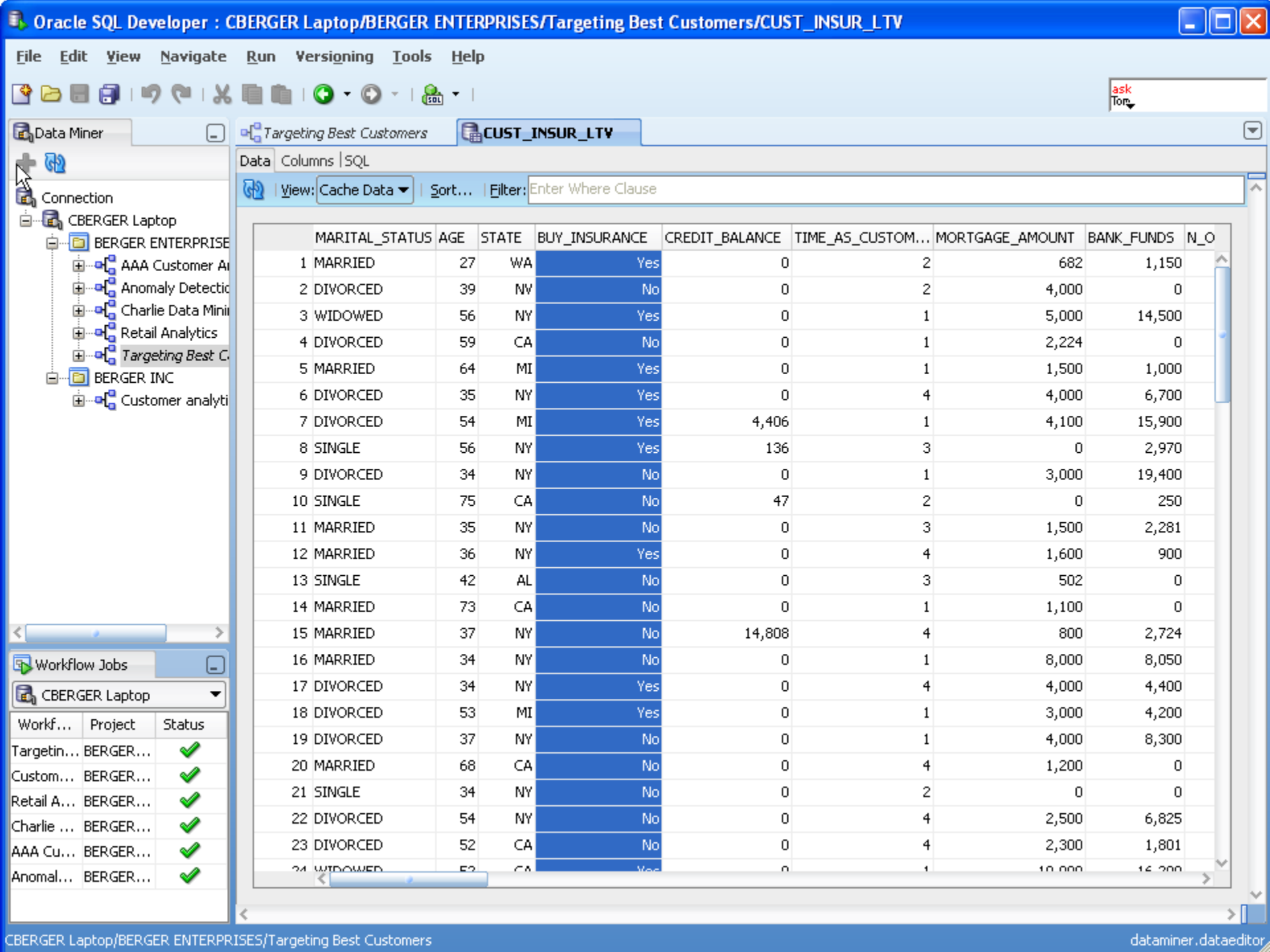
11g Release 2 and previous releases—ODM'r "Classic"



11g Release 2 and future—ODM'r "New" or just ODM'r




Oracle SQL Developer
File Edit View Navigate Run Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer A
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
Explore Data 2
CUST_INSUR_LTV
Workflow Jobs
CBERGER Laptop
Work...
Project
Status
Targeti...
BERGE...
Custom...
BERGE...
Retail ...
BERGE...
Charlie ...
BERGE...
Targeting Best Customers
100%
Explore Data 2
CUST_INSUR_LTV
Explore Data 2 - Property Inspector
Find
Explore
Output
Histogram
Sample
Details
Group By:
LTV_BIN
Auto Input Columns Selection
Data
Name
Name
Data Type
AGE
NUMBER
Component P...
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Model Details
Regression
Evaluate and Apply
Apply
Link
Test
Data
Create Table
Data Source
Explore Data
Transforms
Aggregate
Filter Columns
Filter Columns Details
Filter Rows
Text
Linking Nodes
CBERGER Laptop/BERGER ENTERPRISES/Targeting Best Customers
Workflow Layout


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/Targeting Best Customers/CUST_INSUR_LTV
File Edit View Navigate Run Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Workflow Jobs
CBERGER Laptop
Workf... Project Status
Targetin... BERGER...
Custom... BERGER...
Retail A... BERGER...
Charlie ... BERGER...
AAA Cu... BERGER...
Anomal... BERGER...
Targeting Best Customers
CUST_INSUR_LTV
Data Columns SQL
View: Cache Data
Sort...
Filter: Enter Where Clause
MARITAL_STATUS AGE STATE BUY_INSURANCE CREDIT_BALANCE TIME_AS_CUSTOM... MORTGAGE_AMOUNT BANK_FUNDS N_O
1 MARRIED 27 WA Yes 0 2 682 1,150
2 DIVORCED 39 NV No 0 2 4,000 0
3 WIDOWED 56 NY Yes 0 1 5,000 14,500
4 DIVORCED 59 CA No 0 1 2,224 0
5 MARRIED 64 MI Yes 0 1 1,500 1,000
6 DIVORCED 35 NY Yes 0 4 4,000 6,700
7 DIVORCED 54 MI Yes 4,406 1 4,100 15,900
8 SINGLE 56 NY Yes 136 3 0 2,970
9 DIVORCED 34 NY No 0 1 3,000 19,400
10 SINGLE 75 CA No 47 2 0 250
11 MARRIED 35 NY No 0 3 1,500 2,281
12 MARRIED 36 NY Yes 0 4 1,600 900
13 SINGLE 42 AL No 0 3 502 0
14 MARRIED 73 CA No 0 1 1,100 0
15 MARRIED 37 NY No 14,808 4 800 2,724
16 MARRIED 34 NY No 0 1 8,000 8,050
17 DIVORCED 34 NY Yes 0 4 4,000 4,400
18 DIVORCED 53 MI Yes 0 1 3,000 4,200
19 DIVORCED 37 NY No 0 1 4,000 8,300
20 MARRIED 68 CA No 0 4 1,200 0
21 SINGLE 34 NY No 0 2 0 0
22 DIVORCED 54 NY No 0 4 2,500 6,825
23 DIVORCED 52 CA No 0 4 2,300 1,801
CBERGER Laptop/BERGER ENTERPRISES/Targeting Best Customers
dataminer.dataeditor

Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/Targeting Best Customers/Explore Data 2

File Edit View Navigate Run Versioning Tools Help

Data Miner

Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti

Workflow Jobs

CBERGER Laptop

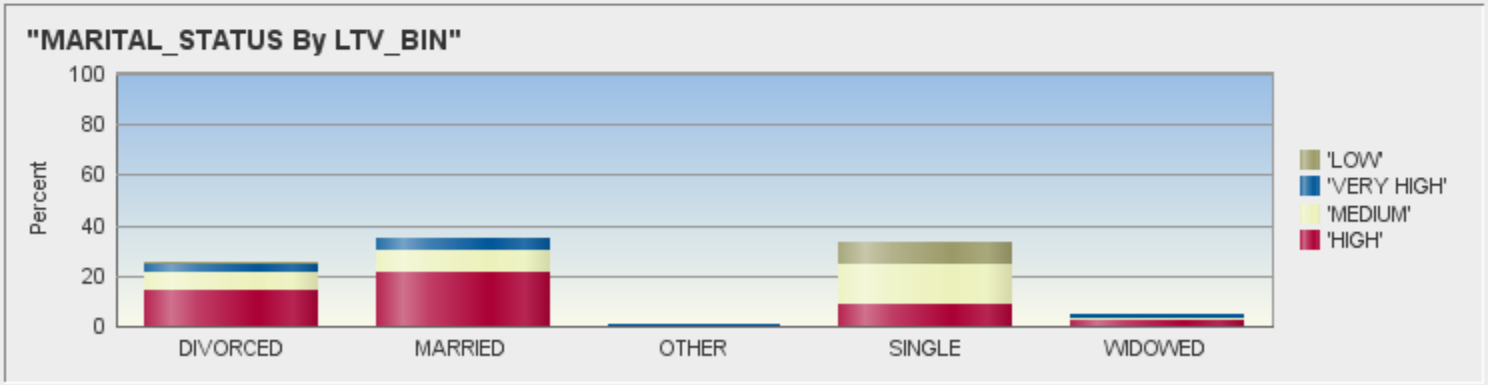
Workf...	Project	Status
Targetin...	BERGER...	✓
Custom...	BERGER...	✓
Retail A...	BERGER...	✓
Charlie ...	BERGER...	✓
AAA Cu...	BERGER...	✓
Anomal...	BERGER...	✓

Targeting Best Customers | Explore Data 2 | CUST_INSUR_LTV

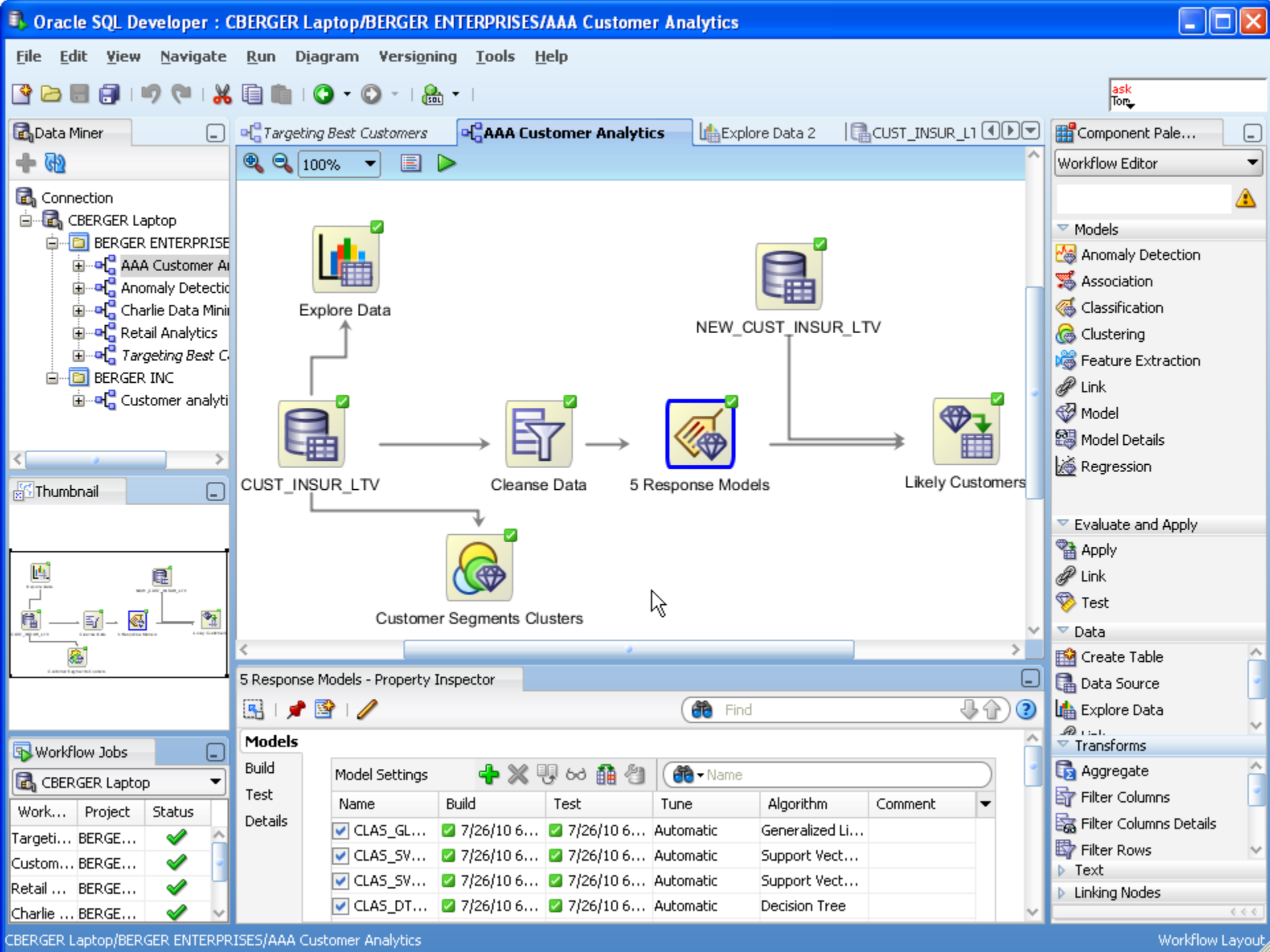
Statistics | Data | Columns | SQL

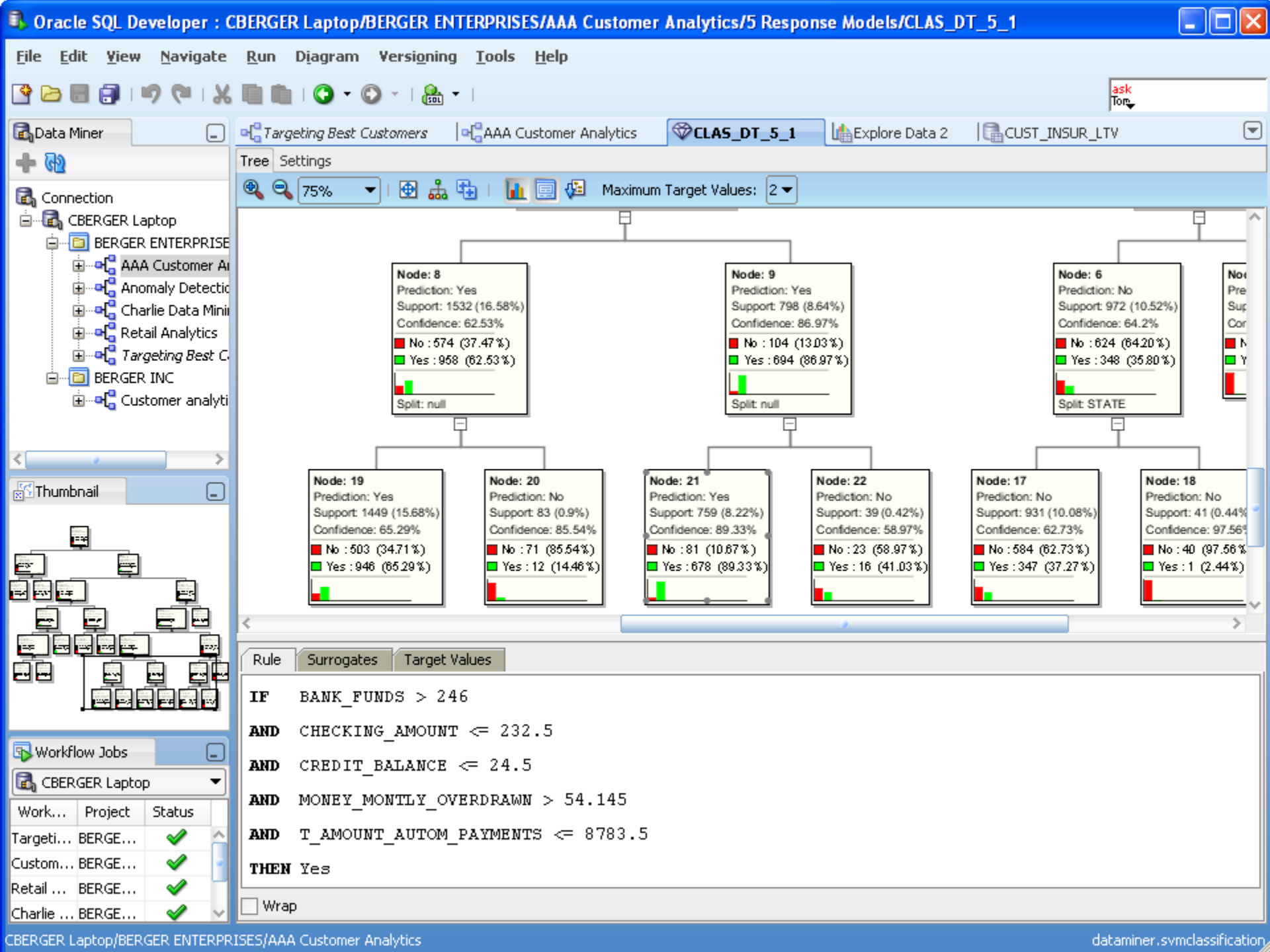
Statistics Group by: "LTV_BIN" Filter: Name

Name	Histogram	Data Type	Percent NULLs	Distinct Values	Mode
"CREDIT_CARD_LIMITS"		NUMBER	0	28	
"CHECKING_AMOUNT"		NUMBER	0	636	
"MARITAL_STATUS"		VARCHAR2	0	5	MARRIED
"LTV"		NUMBER	0	1,930	
"HAS_CHILDREN"		NUMBER	0	2	
"LAST"		VARCHAR2	0	1,396	EMERY
"MONEY_MONTLY_OVERDRAWN"		NUMBER	0	371	
"STATE"		VARCHAR2	0	24	NY
"SALARY"		NUMBER	0	1,906	
"BANK_FUNDS"		NUMBER	0	446	
"MORTGAGE_AMOUNT"		NUMBER	0	413	



CBERGER Laptop/BERGER ENTERPRISES/Targeting Best Customers dataminer.dprofiler


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics
File Edit View Navigate Run Diagram Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer An
Anomaly Detectio
Charlie Data Minin
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
Workflow Jobs
CBERGER Laptop
Work... Project Status
Targeti... BERGE...
Custom... BERGE...
Retail ... BERGE...
Charlie ... BERGE...
Targeting Best Customers
AAA Customer Analytics
Explore Data 2
CUST_INSUR_L1
100%
Explore Data
NEW_CUST_INSUR_LTV
CUST_INSUR_LTV
Cleanse Data
5 Response Models
Likely Customers
Customer Segments Clusters
5 Response Models - Property Inspector
Find
Models
Build
Test
Details
Model Settings
Name
Name Build Test Tune Algorithm Comment
CLAS_GL... 7/26/10 6... 7/26/10 6... Automatic Generalized Li...
CLAS_SV... 7/26/10 6... 7/26/10 6... Automatic Support Vect...
CLAS_SV... 7/26/10 6... 7/26/10 6... Automatic Support Vect...
CLAS_DT... 7/26/10 6... 7/26/10 6... Automatic Decision Tree
Component Pale...
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Model Details
Regression
Evaluate and Apply
Apply
Link
Test
Data
Create Table
Data Source
Explore Data
Transforms
Aggregate
Filter Columns
Filter Columns Details
Filter Rows
Text
Linking Nodes
CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics
Workflow Layout


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics/5 Response Models/CLAS_DT_5_1
File Edit View Navigate Run Diagram Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer An
Anomaly Detectio
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
Workflow Jobs
CBERGER Laptop
Work... Project Status
Targeti... BERGE...
Custom... BERGE...
Retail ... BERGE...
Charlie ... BERGE...
Targeting Best Customers
AAA Customer Analytics
CLAS_DT_5_1
Explore Data 2
CUST_INSUR_LTV
Tree Settings
75%
Maximum Target Values: 2
Node: 8
Prediction: Yes
Support: 1532 (16.58%)
Confidence: 62.53%
No : 574 (37.47%)
Yes : 958 (62.53%)
Split: null
Node: 9
Prediction: Yes
Support: 798 (8.64%)
Confidence: 86.97%
No : 104 (13.03%)
Yes : 694 (86.97%)
Split: null
Node: 6
Prediction: No
Support: 972 (10.52%)
Confidence: 64.2%
No : 624 (64.20%)
Yes : 348 (35.80%)
Split: STATE
Node: 19
Prediction: Yes
Support: 1449 (15.68%)
Confidence: 65.29%
No : 503 (34.71%)
Yes : 946 (65.29%)
Node: 20
Prediction: No
Support: 83 (0.9%)
Confidence: 85.54%
No : 71 (85.54%)
Yes : 12 (14.46%)
Node: 21
Prediction: Yes
Support: 759 (8.22%)
Confidence: 89.33%
No : 81 (10.67%)
Yes : 678 (89.33%)
Node: 22
Prediction: No
Support: 39 (0.42%)
Confidence: 58.97%
No : 23 (58.97%)
Yes : 16 (41.03%)
Node: 17
Prediction: No
Support: 931 (10.08%)
Confidence: 62.73%
No : 584 (62.73%)
Yes : 347 (37.27%)
Node: 18
Prediction: No
Support: 41 (0.44%)
Confidence: 97.56%
No : 40 (97.56%)
Yes : 1 (2.44%)
Rule Surrogates Target Values
IF BANK_FUNDS > 246
AND CHECKING_AMOUNT <= 232.5
AND CREDIT_BALANCE <= 24.5
AND MONEY_MONTLY_OVERDRAWN > 54.145
AND T_AMOUNT_AUTOM_PAYMENTS <= 8783.5
THEN Yes
Wrap
CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics
dataminer.svmclassification

Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics/5 Response Models/CLAS_SVM_4_1_Linear

File Edit View Navigate Run Versioning Tools Help

Data Miner

Connection
CBERGER Laptop
BERGER ENTERPRISES
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti

Thumbnail

Workflow Jobs

CBERGER Laptop

Work...	Project	Status
Targeti...	BERGE...	✓
Custom...	BERGE...	✓
Retail ...	BERGE...	✓
Charlie ...	BERGE...	✓

Targeting Best Customers | AAA Customer Analytics | CLAS_SVM_4_1_Linear | CLAS_DT_5_1 | Explore Data 2

Coefficients | Compare | Settings

Target Value: Yes

Query

Sort by absolute value

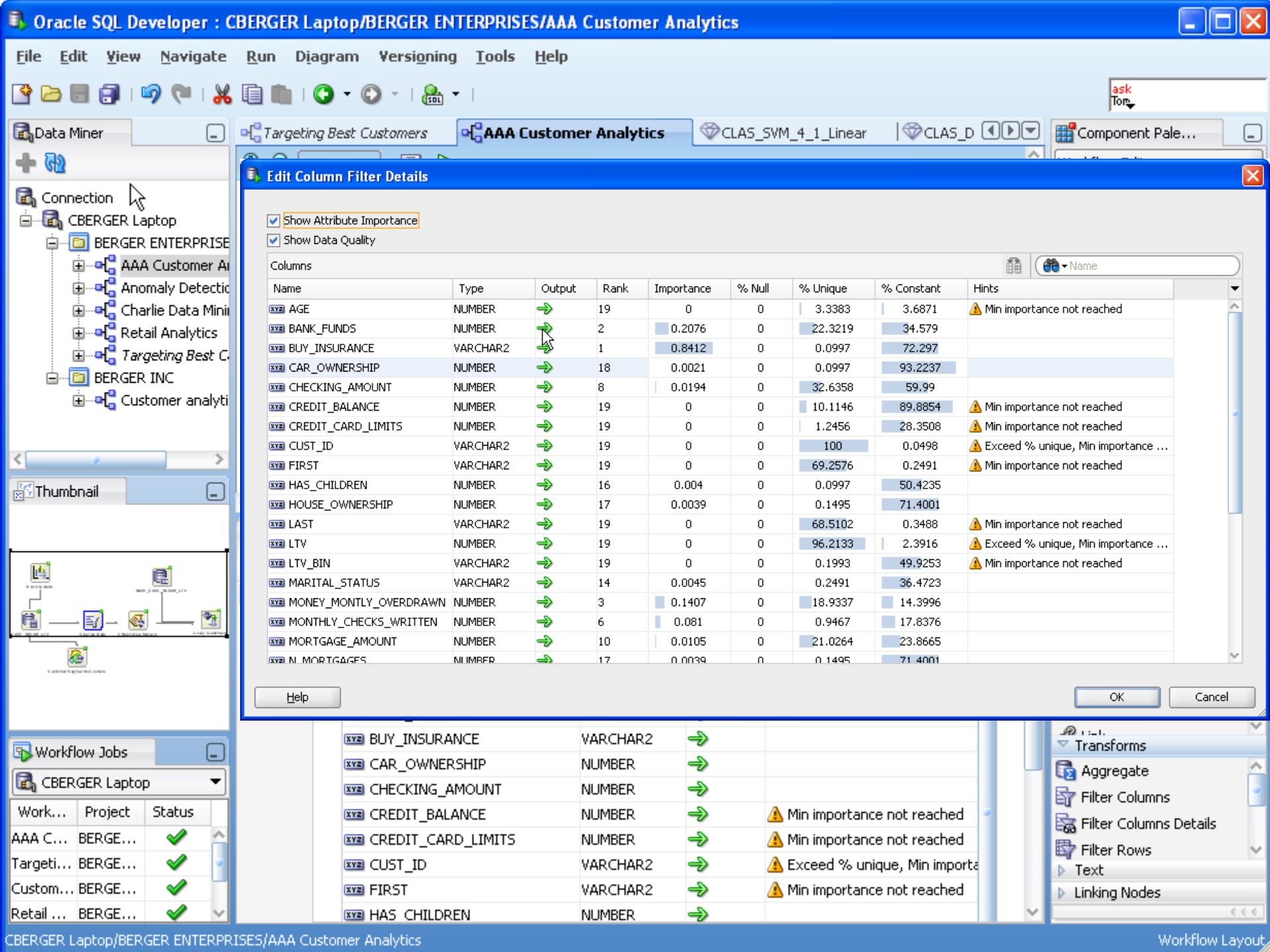
Fetch Size: 10,000

Coefficients: 72 out of 72

Attribute	Value	Coefficient
N_TRANS_ATM	NULL	4.61320837
T_AMOUNT_AUTOM_PAYMENTS	NULL	-4.25073236
CREDIT_BALANCE	NULL	-3.93007442
CHECKING_AMOUNT	NULL	-3.73001909
N_TRANS_WEB_BANK	NULL	-3.63538508
<Intercept>	NULL	-2.76534902
N_TRANS_TELLER	NULL	2.20391478
LTV	NULL	2.14265801
AGE	NULL	-1.66463722
N_OF_DEPENDENTS	NULL	0.86557898
MORTGAGE_AMOUNT	NULL	-0.85111609
STATE	UT	-0.76673662
STATE	TX	0.68542101
SALARY	NULL	-0.55453918
STATE	AZ	-0.48900215
HOUSE_OWNERSHIP	2	-0.34647433
STATE	MO	-0.34306028
N_TRANS_KIOSK	NULL	0.32794196
STATE	NV	0.32315484
STATE	WI	0.30801241
HOUSE_OWNERSHIP	0	0.28906774
TIME_AS_CUSTOMER	5	-0.27872530
TIME_AS_CUSTOMER	1	0.24943188

CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics

dataminer.svmclassification


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics
File Edit View Navigate Run Diagram Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
Workflow Jobs
CBERGER Laptop
Work... Project Status
AAA C... BERGE...
Targeti... BERGE...
Custom... BERGE...
Retail ... BERGE...
Targeting Best Customers
AAA Customer Analytics
CLAS_SVM_4_1_Linear
CLAS_D
Component Pale...
Edit Column Filter Details
Show Attribute Importance
Show Data Quality
Columns
Name
Name Type Output Rank Importance % Null % Unique % Constant Hints
AGE NUMBER 19 0 0 3.3383 3.6871 Min importance not reached
BANK_FUNDS NUMBER 2 0.2076 0 22.3219 34.579
BUY_INSURANCE VARCHAR2 1 0.8412 0 0.0997 72.297
CAR_OWNERSHIP NUMBER 18 0.0021 0 0.0997 93.2237
CHECKING_AMOUNT NUMBER 8 0.0194 0 32.6358 59.99
CREDIT_BALANCE NUMBER 19 0 0 10.1146 89.8854 Min importance not reached
CREDIT_CARD_LIMITS NUMBER 19 0 0 1.2456 28.3508 Min importance not reached
CUST_ID VARCHAR2 19 0 0 100 0.0498 Exceed % unique, Min importance ...
FIRST VARCHAR2 19 0 0 69.2576 0.2491 Min importance not reached
HAS_CHILDREN NUMBER 16 0.004 0 0.0997 50.4235
HOUSE_OWNERSHIP NUMBER 17 0.0039 0 0.1495 71.4001
LAST VARCHAR2 19 0 0 68.5102 0.3488 Min importance not reached
LTV NUMBER 19 0 0 96.2133 2.3916 Exceed % unique, Min importance ...
LTV_BIN VARCHAR2 19 0 0 0.1993 49.9253 Min importance not reached
MARITAL_STATUS VARCHAR2 14 0.0045 0 0.2491 36.4723
MONEY_MONTLY_OVERDRAWN NUMBER 3 0.1407 0 18.9337 14.3996
MONTHLY_CHECKS_WRITTEN NUMBER 6 0.081 0 0.9467 17.8376
MORTGAGE_AMOUNT NUMBER 10 0.0105 0 21.0264 23.8665
Help
OK
Cancel
BUY_INSURANCE VARCHAR2
CAR_OWNERSHIP NUMBER
CHECKING_AMOUNT NUMBER
CREDIT_BALANCE NUMBER Min importance not reached
CREDIT_CARD_LIMITS NUMBER Min importance not reached
CUST_ID VARCHAR2 Exceed % unique, Min importa
FIRST VARCHAR2 Min importance not reached
HAS_CHILDREN NUMBER
Transforms
Aggregate
Filter Columns
Filter Columns Details
Filter Rows
Text
Linking Nodes
CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics
Workflow Layout

Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics/Compare/5 Response Models

File Edit View Navigate Run Versioning Tools Help

Data Miner

Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer An
Anomaly Detectio
Charlie Data Minin
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti

Workflow Jobs

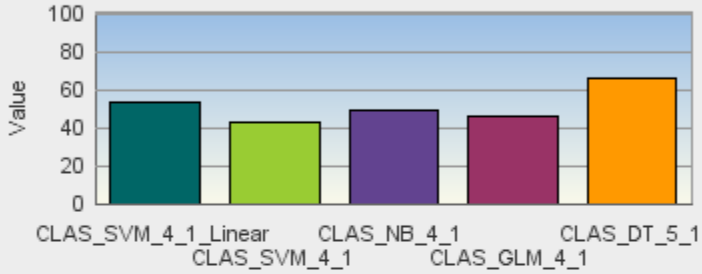
CBERGER Laptop

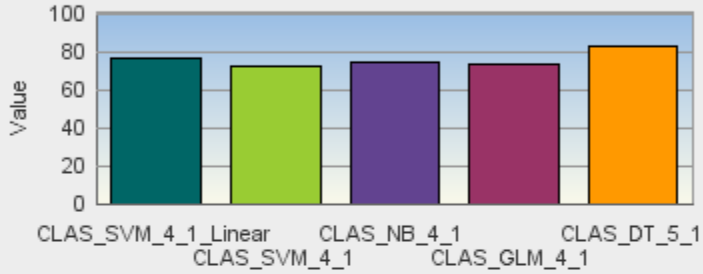
Workf...	Project	Status
AAA Cu...	BERGER...	✓
Targetin...	BERGER...	✓
Custom...	BERGER...	✓
Retail A...	BERGER...	✓
Charlie ...	BERGER...	✓
Anomal...	BERGER...	✓

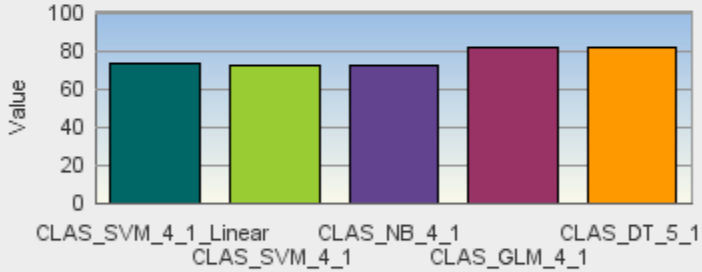
Targeting Best Customers | AAA Customer Analytics | 5 Response Models | CLAS_SVM_4_1_Linear | CLAS_DT_5_1

Performance | Performance Matrix | ROC | Lift | Profit

Measure: All Measures Sort By: Name Descending







Models

Name	Predictive Confidence %	Overall Accuracy %	Average Accuracy %	Algorithm	Cre
CLAS_DT_5_1	65.96	81.3268	82.98	Decision Tree	7/2
CLAS_GLM_4_1	45.775	81.4087	72.8875	Generalized Linear Model	7/2
CLAS_NB_4_1	49.4425	72.2031	74.7212	Naive Bayes	7/2
CLAS_SVM_4_1	43.3568	71.7936	71.6784	Support Vector Machine	7/2
CLAS_SVM_4_1_Linear	53.4227	73.317	76.7114	Support Vector Machine	7/2

CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics dataminer.tviewer

Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics/Compare/5 Response Models

File Edit View Navigate Run Versioning Tools Help

Data Miner

Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer An
Anomaly Detectio
Charlie Data Minin
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti

Workflow Jobs

CBERGER Laptop

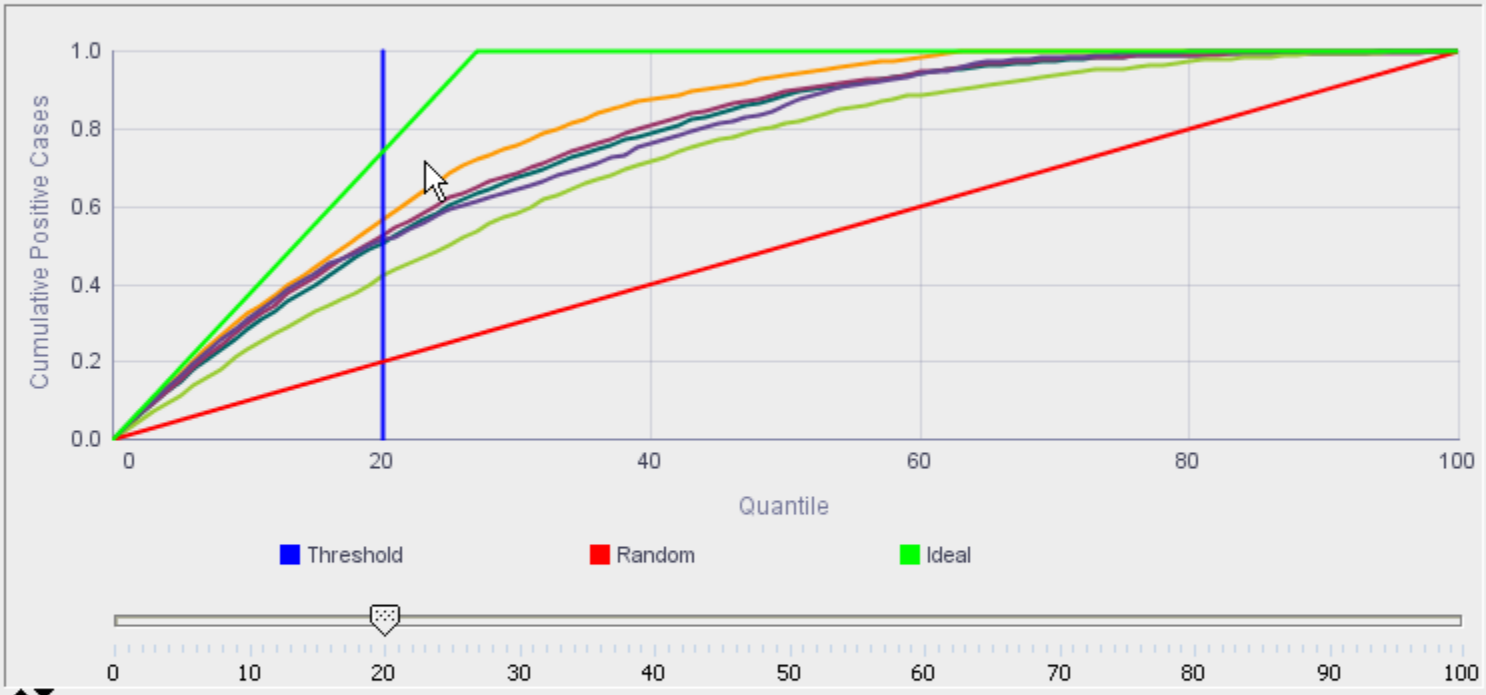
Workf...	Project	Status
AAA Cu...	BERGER...	✓
Targetin...	BERGER...	✓
Custom...	BERGER...	✓
Retail A...	BERGER...	✓
Charlie ...	BERGER...	✓
Anomal...	BERGER...	✓

Targeting Best Customers | AAA Customer Analytics | 5 Response Models | CLAS_SVM_4_1_Linear | CLAS_DT_5_1

Performance | Performance Matrix | ROC | Lift | Profit

Display: Cumulative Positive Cases

Target Value: Yes

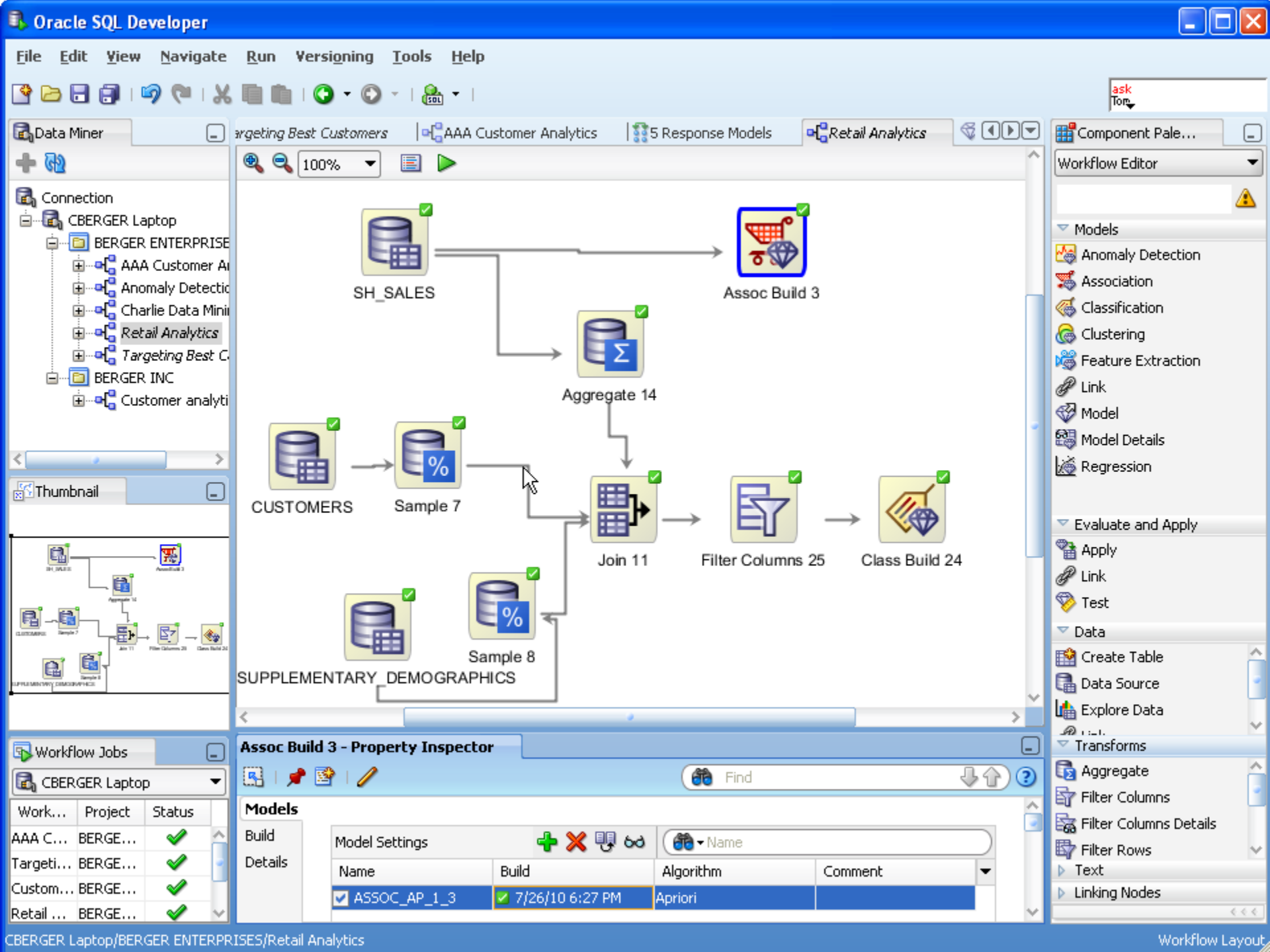


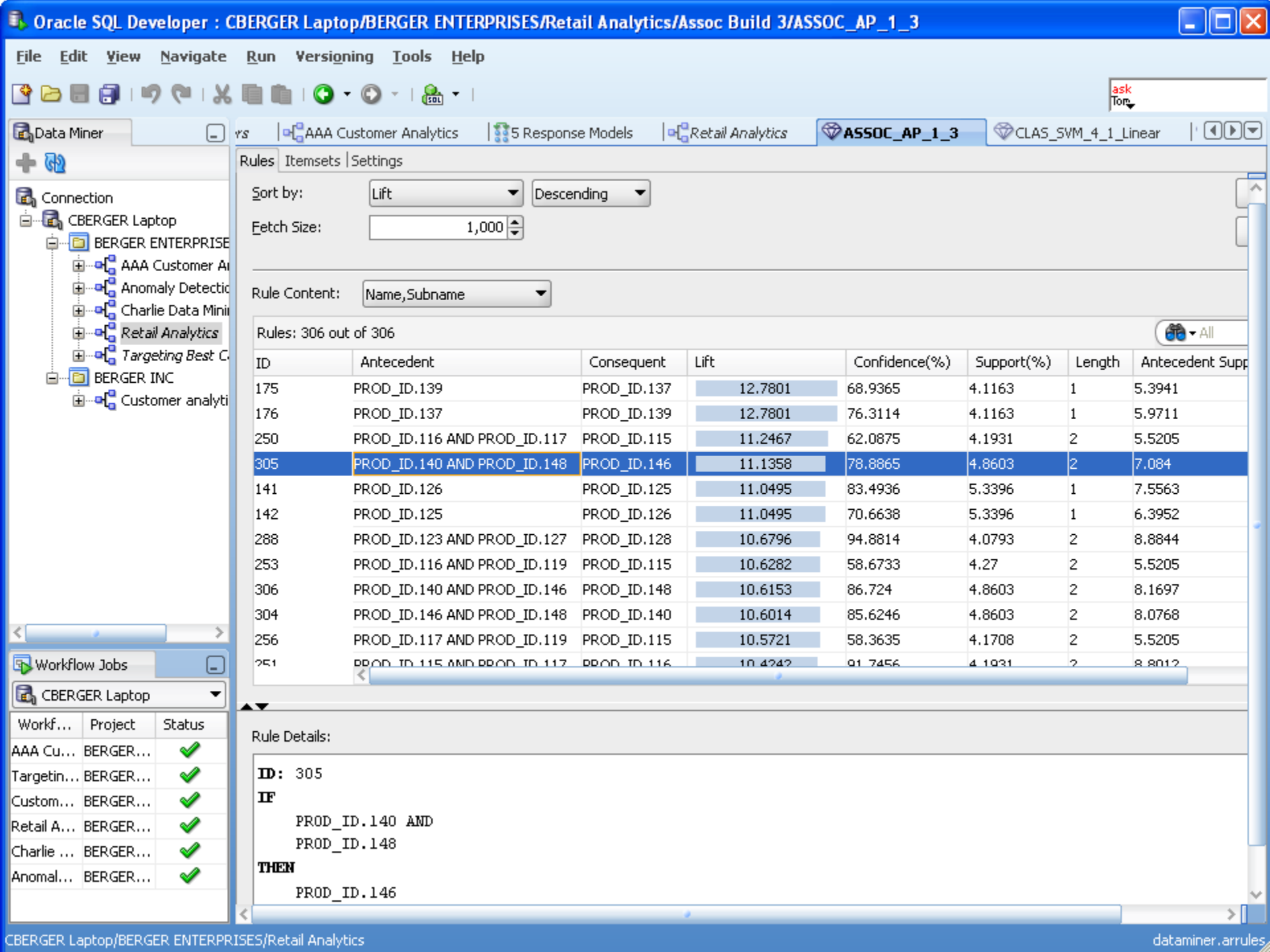
Models

Model	Lift Cumulative	Gain Cumulative %	Record...	Target Den...	Algorithm	Creation Date
CLAS_DT_5_1	2.8172	56.5285	20.0655	0.7591	Decision Tree	7/26/10 6:13 PM
CLAS_GLM_4_1	2.6267	52.7052	20.0655	0.7078	Generalized Linear Model	7/26/10 6:13 PM
CLAS_NB_4_1	2.56	51.3678	20.0655	0.6898	Naive Bayes	7/26/10 6:13 PM
CLAS_SVM_4_1	2.1056	42.2492	20.0655	0.5673	Support Vector Machine	7/26/10 6:13 PM
CLAS_SVM_4_1_Linear	2.5297	50.7599	20.0655	0.6816	Support Vector Machine	7/26/10 6:13 PM

CBERGER Laptop/BERGER ENTERPRISES/AAA Customer Analytics

dataminer.tviewer


Oracle SQL Developer
File Edit View Navigate Run Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
rgeting Best Customers
AAA Customer Analytics
5 Response Models
Retail Analytics
100%
SH_SALES
Assoc Build 3
Aggregate 14
CUSTOMERS
Sample 7
Join 11
Filter Columns 25
Class Build 24
Sample 8
SUPPLEMENTARY_DEMOGRAPHICS
Component Pale...
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Model Details
Regression
Evaluate and Apply
Apply
Link
Test
Data
Create Table
Data Source
Explore Data
Transforms
Aggregate
Filter Columns
Filter Columns Details
Filter Rows
Text
Linking Nodes
Workflow Jobs
CBERGER Laptop
Work...
Project
Status
AAA C...
BERGE...
Targeti...
BERGE...
Custom...
BERGE...
Retail ...
BERGE...
Assoc Build 3 - Property Inspector
Find
Models
Build
Details
Model Settings
Name
Build
Algorithm
Comment
ASSOC_AP_1_3
7/26/10 6:27 PM
Apriori
CBERGER Laptop/BERGER ENTERPRISES/Retail Analytics
Workflow Layout


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/Retail Analytics/Assoc Build 3/ASSOC_AP_1_3
File Edit View Navigate Run Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer Ar
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Workflow Jobs
CBERGER Laptop
Workf... Project Status
AAA Cu... BERGER...
Targetin... BERGER...
Custom... BERGER...
Retail A... BERGER...
Charlie ... BERGER...
Anomal... BERGER...
AAA Customer Analytics
5 Response Models
Retail Analytics
ASSOC_AP_1_3
CLAS_SVM_4_1_Linear
Rules Itemsets Settings
Sort by: Lift Descending
Fetch Size: 1,000
Rule Content: Name,Subname
Rules: 306 out of 306
ID Antecedent Consequent Lift Confidence(%) Support(%) Length Antecedent Supp
175 PROD_ID.139 PROD_ID.137 12.7801 68.9365 4.1163 1 5.3941
176 PROD_ID.137 PROD_ID.139 12.7801 76.3114 4.1163 1 5.9711
250 PROD_ID.116 AND PROD_ID.117 PROD_ID.115 11.2467 62.0875 4.1931 2 5.5205
305 PROD_ID.140 AND PROD_ID.148 PROD_ID.146 11.1358 78.8865 4.8603 2 7.084
141 PROD_ID.126 PROD_ID.125 11.0495 83.4936 5.3396 1 7.5563
142 PROD_ID.125 PROD_ID.126 11.0495 70.6638 5.3396 1 6.3952
288 PROD_ID.123 AND PROD_ID.127 PROD_ID.128 10.6796 94.8814 4.0793 2 8.8844
253 PROD_ID.116 AND PROD_ID.119 PROD_ID.115 10.6282 58.6733 4.27 2 5.5205
306 PROD_ID.140 AND PROD_ID.146 PROD_ID.148 10.6153 86.724 4.8603 2 8.1697
304 PROD_ID.146 AND PROD_ID.148 PROD_ID.140 10.6014 85.6246 4.8603 2 8.0768
256 PROD_ID.117 AND PROD_ID.119 PROD_ID.115 10.5721 58.3635 4.1708 2 5.5205
Rule Details:
ID: 305
IF
PROD_ID.140 AND
PROD_ID.148
THEN
PROD_ID.146
CBERGER Laptop/BERGER ENTERPRISES/Retail Analytics
dataminer.arrules

Oracle Data Miner Development(4) : STAMW16/BERGER INSURANCE/Targeting Best Customers/Clustering Models/CLUS_KM_1_8

File Edit View Search Navigate Tools Window Help

Connections Navigator

- Connections
  - STAMW16
    - Allegro
    - Allegro1
    - BERGER Bank & Loan
    - BERGER INSURANCE
      - Analyzing Customer Data
      - Targeting Best Customers
    - BM Project
    - BPottle
    - BRIAN BRAIN Company
    - Brian Macdonald
    - BWise
    - manish-project
    - MFH Project
    - MK Project
    - SLAREMAR

Run Manager

STAMW16

Workflow	Project	Status
Targeting B...	BPottle	✓
WF-1	MFH Project	✓
Association...	BM Project	✓
SPE-Cluste...	manish-pro...	✓
Targeting ...	BRIAN BRA...	✓
myworkflow	BWise	✓
TEST-VGUA...	MK Project	✗
Analyzing ...	BERGER IN...	✓

Targeting Best Customers | CLUS_KM_1_8

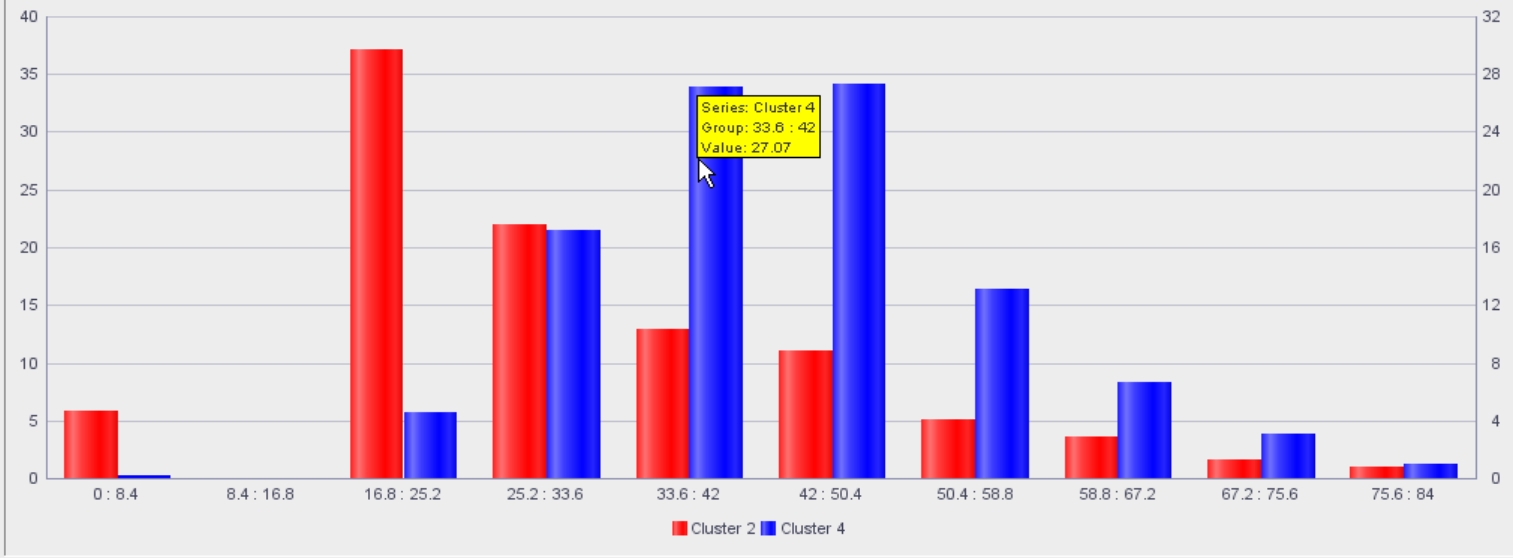
Cluster 1: 2 Edit Cluster 2: 4 Edit Query

Leaves Only Fetch Size: 2,000

Attributes: 27 out of 27 Attribute

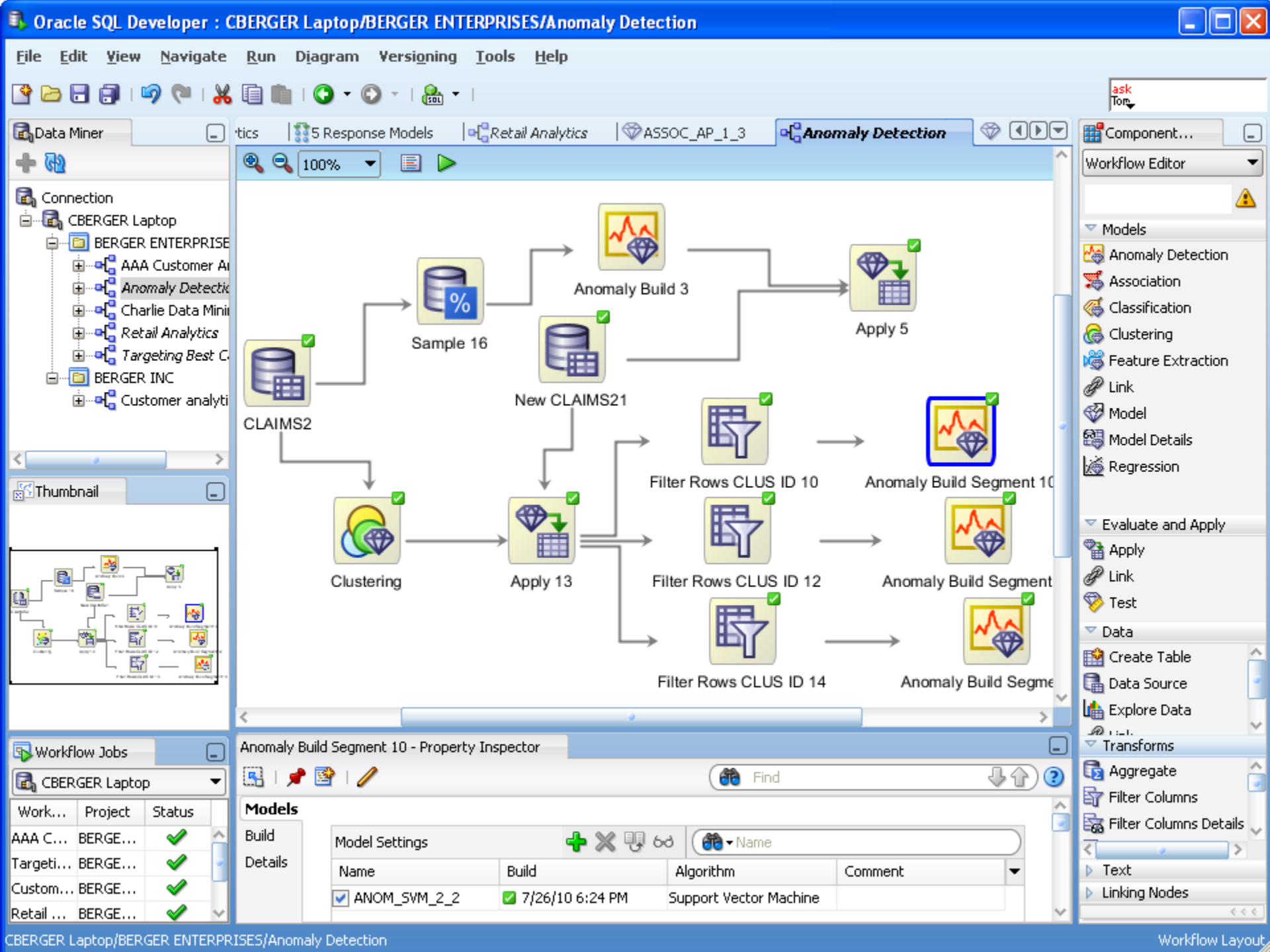
Attribute	Rank	Divergence	Distribution	2(Centroid)	4(Centroid)
MONTHLY_CHECKS_WRITTEN	16	0.0289		3.5687573	4.77708638
BANK_FUNDS	17	0.0221		763.20023374	3,650.86987555
BUY_INSURANCE	18	0.0205		No	No
SALARY	19	0.0164		64,642.89988313	66,417.14915813
HOUSE_OWNERSHIP	2	0.3577		0	1
N_MORTGAGES	2	0.3577		0	1
MORTGAGE_AMOUNT	20	0.0145		144.79333853	3,115.02379209
STATE	21	0.0083		NY	NY
CHECKING_AMOUNT	22	0.0065		614.84047526	1,238.01610542
N_TRANS_KIOSK	23	0.0048		1.88878068	1.95241581
REGION	24	0.0028		West	NorthEast

AGE



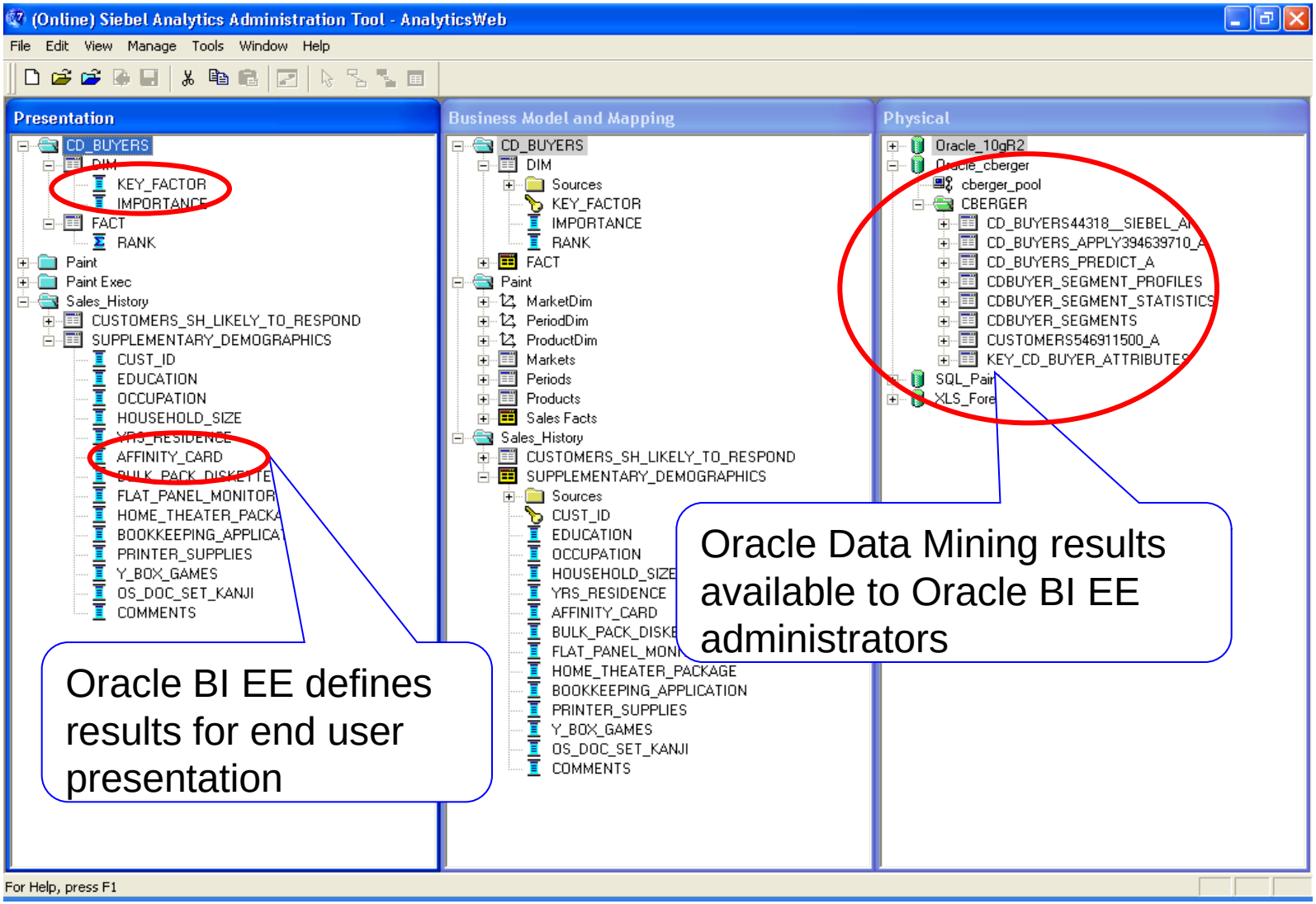
Tree Detail Compare Settings

STAMW16/BERGER INSURANCE/Targeting Best Customers dataminer.kmclusteringdetails


Oracle SQL Developer : CBERGER Laptop/BERGER ENTERPRISES/Anomaly Detection
File Edit View Navigate Run Diagram Versioning Tools Help
Data Miner
Connection
CBERGER Laptop
BERGER ENTERPRISE
AAA Customer A
Anomaly Detectic
Charlie Data Mini
Retail Analytics
Targeting Best C
BERGER INC
Customer analyti
Thumbnail
5 Response Models
Retail Analytics
ASSOC_AP_1_3
Anomaly Detection
100%
Anomaly Build 3
Sample 16
Apply 5
New CLAIMS21
CLAIMS2
Filter Rows CLUS ID 10
Anomaly Build Segment 10
Clustering
Apply 13
Filter Rows CLUS ID 12
Anomaly Build Segment
Filter Rows CLUS ID 14
Anomaly Build Segme
Component...
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Model Details
Regression
Evaluate and Apply
Apply
Link
Test
Data
Create Table
Data Source
Explore Data
Transforms
Aggregate
Filter Columns
Filter Columns Details
Text
Linking Nodes
Workflow Jobs
CBERGER Laptop
Work...
Project
Status
AAA C...
BERGE...
Targeti...
Custom...
Retail ...
Anomaly Build Segment 10 - Property Inspector
Find
Models
Build
Details
Model Settings
Name
Build
Algorithm
Comment
ANOM_SVM_2_2
7/26/10 6:24 PM
Support Vector Machine
CBERGER Laptop/BERGER ENTERPRISES/Anomaly Detection
Workflow Layout

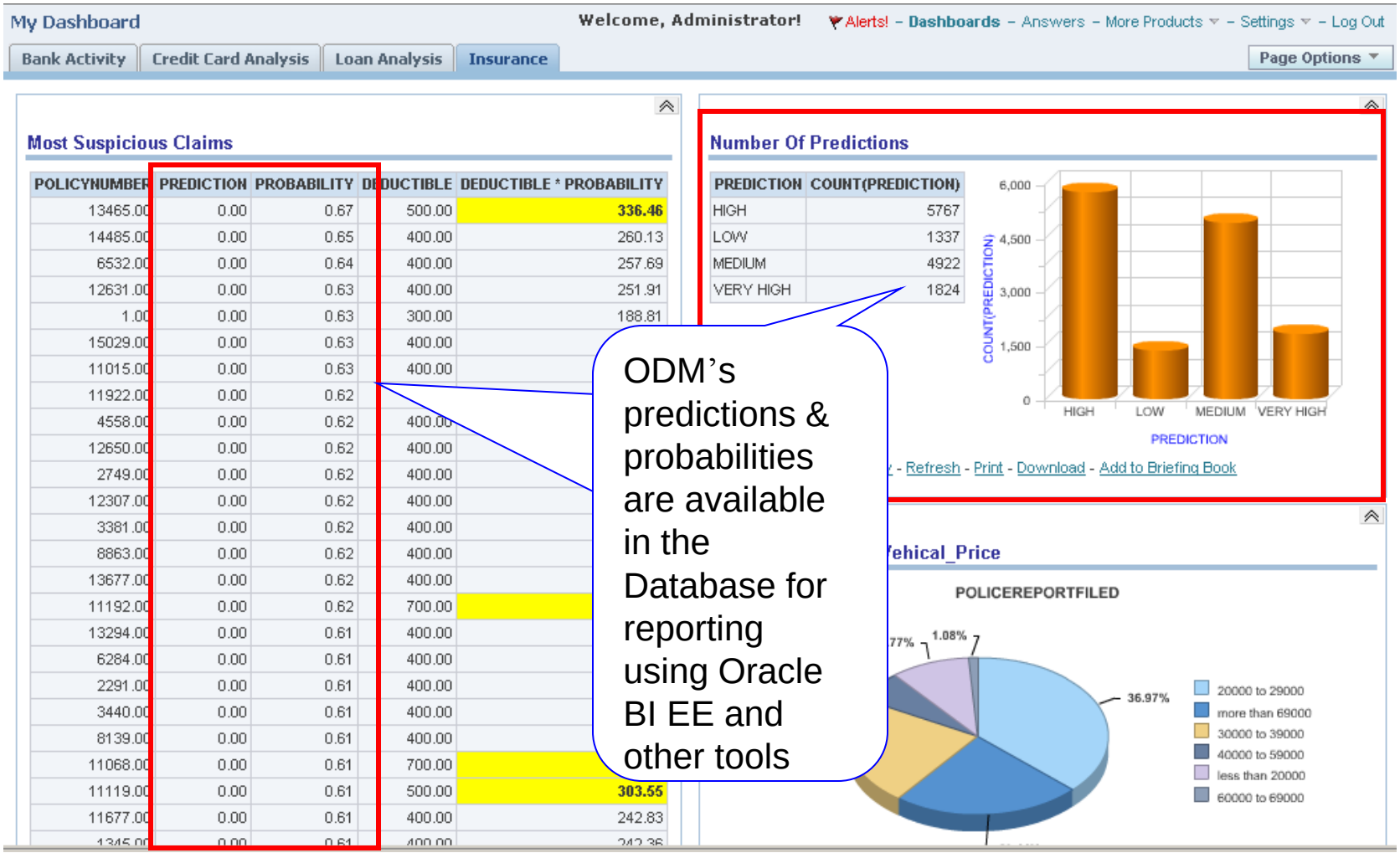
Presentation of Results and Integration with Applications

Integration with Oracle BI EE



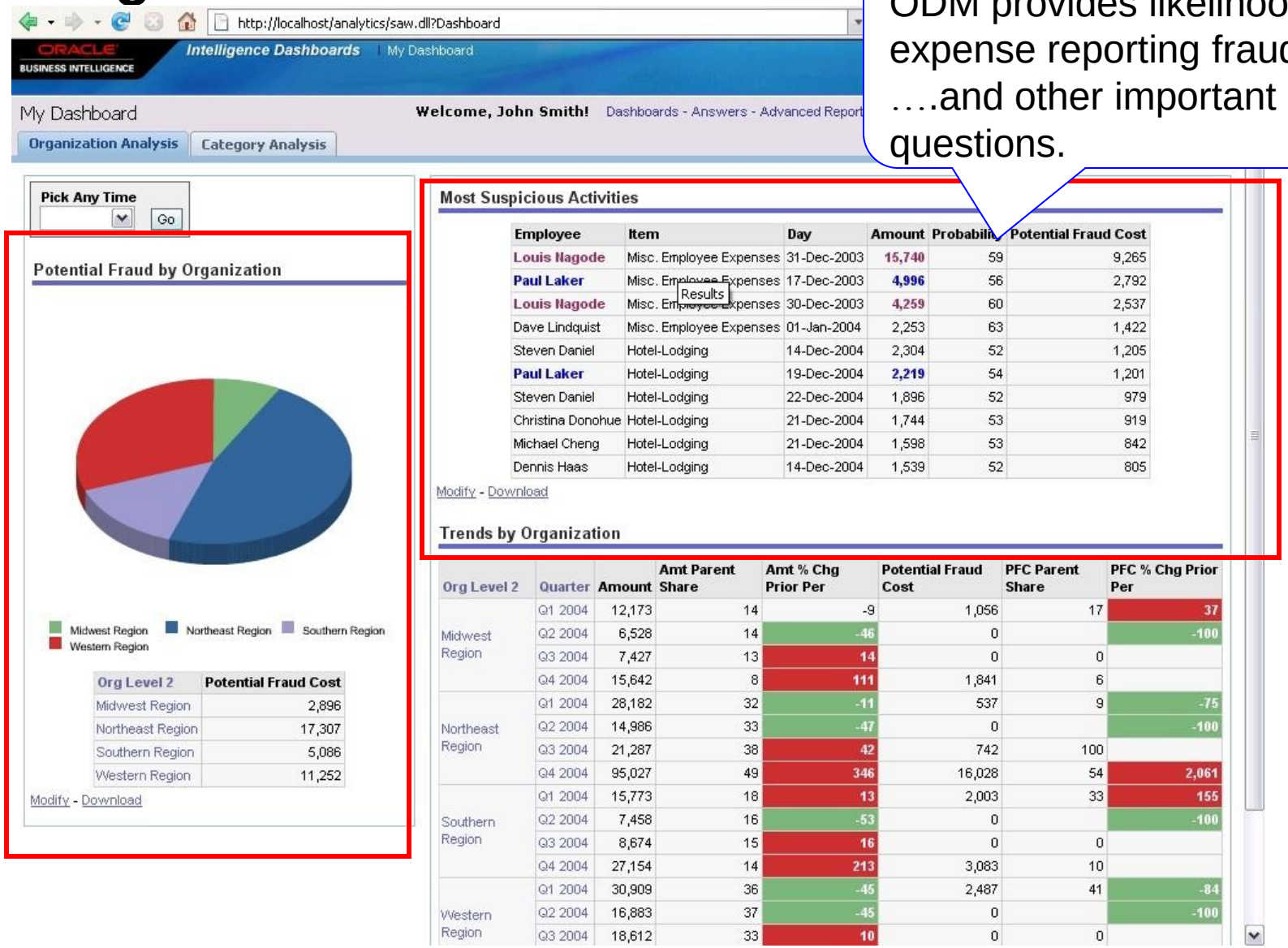
Example

Better Information for OBI EE Reports and Dashboards



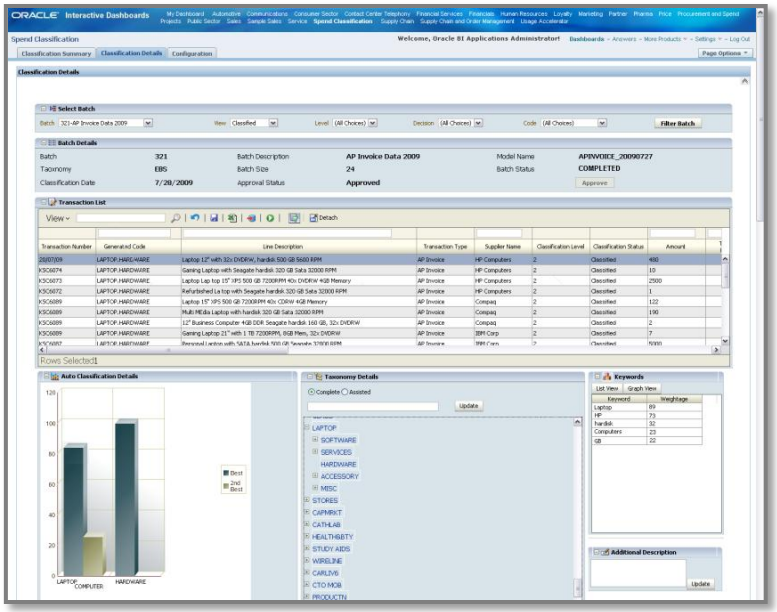
Integration with Oracle BI EE

ODM provides likelihood of expense reporting fraud ….and other important questions.



Oracle Spend Classification

Classify Spend into Purchasing Categories



- Features
  - Hierarchical classification and scoring
  - Auto Spend Classification – Inline and Batch
  - Integration to OBIA Procurement & Spend Analytics 7.9.6
- Benefits
  - Classifies spend data from various sources into procurement category hierarchies
  - Category normalization aids strategic sourcing and contract negotiation
  - In-line mode integrated with EBS iProcurement

Additional Information

- Preview of the new Oracle Data Miner 11g R2 “work flow” New GUI
- Oracle Data Mining 11gR2 presentation at Oracle Open World 2009
- Oracle Data Mining Blog
- Funny YouTube video that features Oracle Data Mining
- Oracle Data Mining on the Amazon Cloud
- Oracle Data Mining 11gR2 data sheet
- Oracle Data Mining 11gR2 white paper
- New TechCast (audio and video recording): ODM overview and several demos
- Fraud and Anomaly Detection using Oracle Data Mining 11g presentation
- Algorithm technical summary with links to Documentation
- Getting Started w/ ODM page w/ instructions to download
  - Oracle Data Miner graphical user interface (GUI),
  - ODM Step-by-Step Tutorial
  - Demo datasets
- ODM Discussion Forum on OTN (great for posting questions/answers)
- ODM 11g Sample Code (examples of ODM SQL and Java APIs applied in several use cases; great for developers)
- Oracle’s 50+ SQL based statistical functions (t-test, ANOVA, Pearson’s, etc.)

Google™ Oracle Data Mining