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

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

Outline

- Fraud as an Industry Problems
- Predictive Analytics and Oracle Data Mining Overview
- Demonstration(s)
- Oracle Data Miner 11gR2 New GUI Preview
- Presentation of Results and Integration with Applications

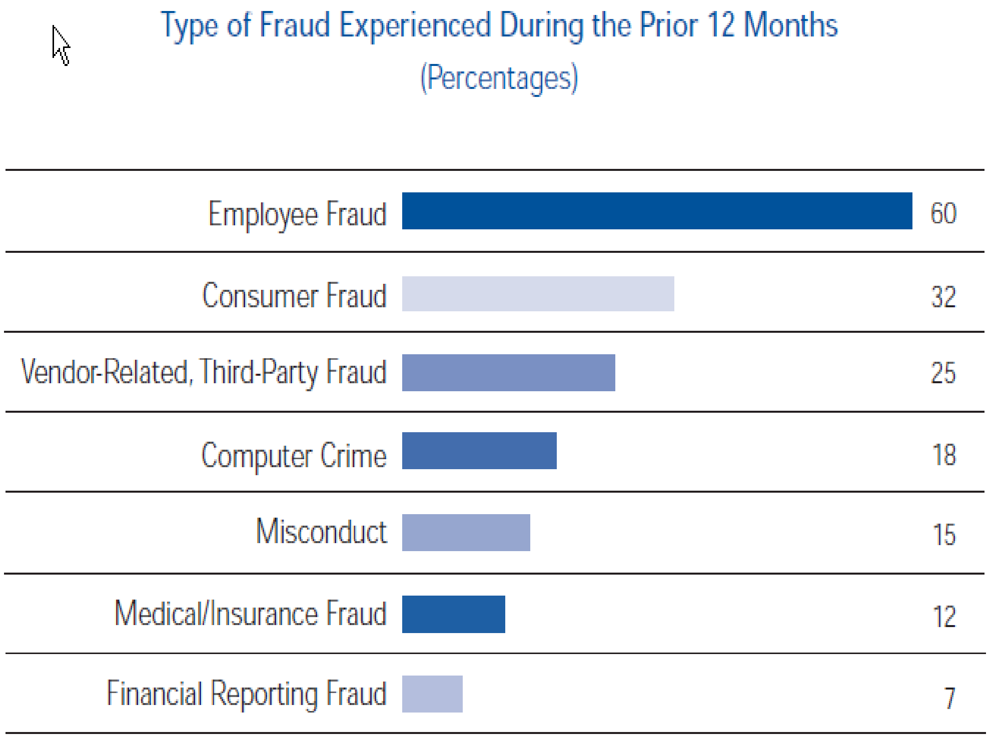
Insurance Fraud a Significant Problem

- Insurance Fraud a $31B industry in the US in 2003
- About 10-15 percent of all claims are fraudulent
- Over 25% of all claims have some element of fraud/abuse
- More than 1 of every 3 bodily-injury claims from car crashes involve fraud
- 17-20 cents of every dollar paid for bodily injury claims from auto policies involves fraud or claim buildup
- Property & Casualty Insurers Survey (2004)
  - About 11-30 cents — or more — of every claim dollar is lost to "soft" fraud (smalltime cheating by normally honest people)
- Fraud adds $5.2-$6.3 billion to the auto premiums that policyholders pay each year

Sources: *Coalition Against Insurance Fraud, Research Council-Insurance Services Office; (1996-2004 Statistics)*

Reports of Fraud Are on the Rise

- Expense account abuse and theft of assets showed the greatest percentage point increases since 1998
- 75% percent of companies surveyed report they experienced an instance of fraud (62% in 1998)
- Employee fraud is the most prevalent type of fraud
- Medical/insurance fraud & financial reporting fraud are the most costly



Types of Fraud

- Computer Crime
  - Hacking and other cybertheft
- Consumer Fraud
  - ATM theft
  - Check fraud
  - Credit card fraud
  - Fraudulent classification of merchandise for customers
  - Fraudulent merchandise returns
  - Identity theft
- Employee Fraud
  - Check fraud
  - Expense account abuse
  - Payroll fraud
  - Pension theft
  - Theft or misappropriation of assets
- Financial Reporting Fraud
  - Asset revenue misstatement
  - Concealed liabilities and expenses
  - Improper revenue recognition
  - Inadequate omissions and disclosures
- Medical/Insurance Fraud
  - Medical/insurance claims fraud
  - Policy churning
  - Workers' compensation fraud
- Misconduct
  - Conflicts of interest
  - Insider trading
- Vendor and 3rd-Party Fraud
  - Bid rigging and price fixing
  - Bribery
  - Diversion of sales
  - Duplicate billings
  - Extortion
  - False invoices and phantom vendors
  - Inventory theft
  - Kickbacks and conflicts of interest
  - Loan fraud
  - Theft of intellectual property

http://www.kpmg.com/aci/docs/surveys/Fraud%20Survey_040855_R5.pdf

People's Attitudes About Fraud—Consumers

- Nearly one of four Americans say it's ok to defraud insurers
  - Some 8 percent say it's "quite acceptable" to bilk insurers, while 16 percent say it's "somewhat acceptable."
  - About one in 10 people agree it's ok to submit claims for items that aren't lost or damaged, or for personal injuries that didn't occur.
  - Two of five people are "not very likely" or "not likely at all" to report someone who ripped of an insurer. Accenture Ltd.(2003)
- Nearly one of 10 Americans would commit insurance fraud if they knew they could get away with it.
- Nearly three of 10 Americans (29 percent) wouldn't report insurance scams committed by someone they know. Progressive Insurance (2001)

Coalition Against Insurance Fraud

ORACLE

People's Attitudes About Fraud—Consumers

- More than one of three Americans say it's ok to exaggerate insurance claims to make up for the deductible (40 percent in 1997).
  Insurance Research Council (2000)
- One of four Americans says it's ok to pad a claim to make up for premiums they've already paid.
  Insurance Research Council (2000)
- One of three Americans says it's ok for employees to stay off work and receive workers compensation benefits because they feel pain, even though their doctor says it's ok to return to work.
  Insurance Research Council (1999)

Coalition Against Insurance Fraud

People's Attitudes About Fraud—Physicians

- Nearly one of three physicians say it's necessary to *"game"* the health care system to provide high quality medical care.
- More than one of three physicians says patients have asked physicians to deceive third-party payers to help the patients obtain coverage for medical services in the last year.
- One of 10 physicians has reported medical signs or symptoms a patient didn't have in order to help the patient secure coverage for needed treatment or services in the last year.

Coalition Against Insurance Fraud

Journal of the American Medical Association (2000)

Typical Committer of Fraud

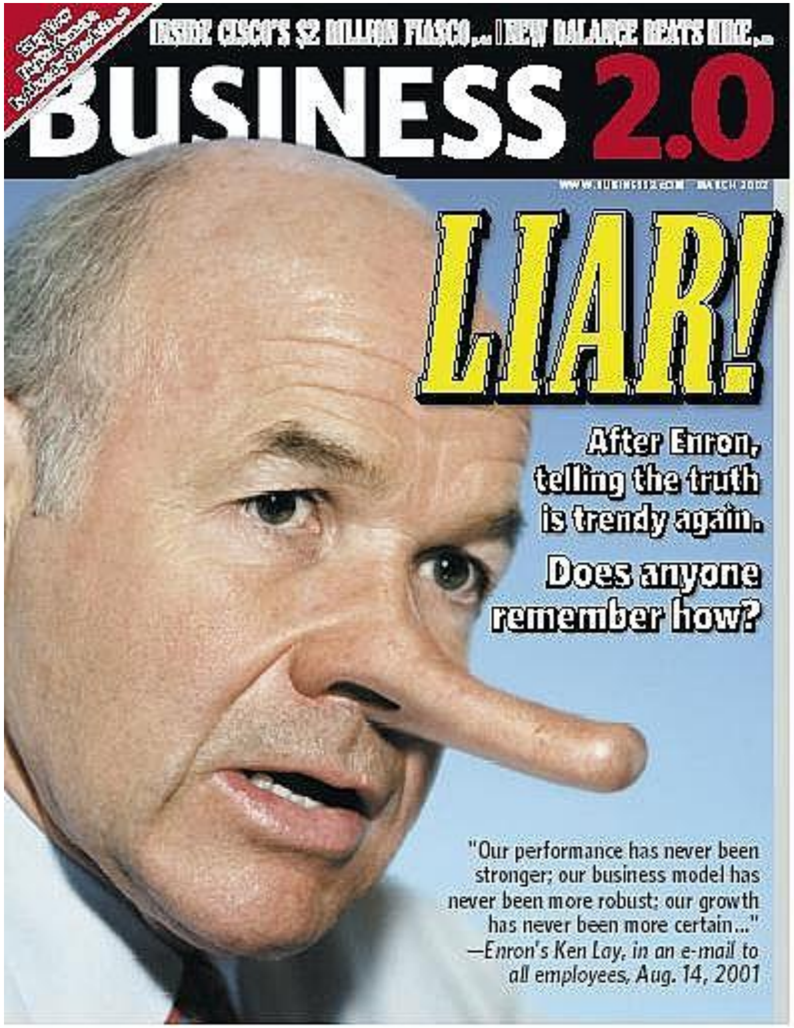
- The typical fraudster is male, between 36 and 55 years old
- By the time he starts profiting from his illegal means, he has usually been employed by the company for six or more years
- He typically works in the financial department, and commits the deed alone
- He is driven to crime by a desire for money and by opportunity

Typical Committer of Fraud

- A rogue trader has cost French bank Société Générale €4.9bn (£3.7bn) in the biggest fraud in financial history
- News of the fraud, which will virtually wipe out 2007 profits at France's second-largest bank, sent shockwaves through European markets, already battered by the escalating credit crisis.

Enron

- At the end of 2001 it was revealed that its reported financial condition was sustained substantially by institutionalized, systematic, and creatively planned accounting fraud, sometimes called the "Enron scandal".
- Enron has since become a popular symbol of willful corporate fraud and corruption.



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.

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.

20 Ways to Detect Fraud

7. Common Names and Addresses for Refunds

Sales employees frequently make bogus refunds to customers for merchandise. The address shown for the refund is then made to the employee's address, or to the address of a friend or co-worker.

8. Increasing Reconciling Items

Stolen deposits, or bogus checks written, are frequently not removed, or covered, from the reconciliation. Hence, over a period of time, the reconciling items tend to increase.

9. General Ledger Out-of-Balance

When funds, merchandise, or assets are stolen and not covered by a fictitious entry, the general ledger will be out of balance. An inventory of the merchandise or cash is needed to confirm the existence of the missing assets.

10. Adjustments to Receivables or Payables

In cases where customer payments are misappropriated, adjustments to receivables can be made to cover the shortage. Where payables are adjusted, the perpetrator can use a phony billing scheme to convert cash to his or her own use.

11. Excess Purchases

Excess purchases can be used to cover fraud in two ways:

Fictitious payees are used to convert funds

Excessive purchases may indicate a possible payoff of purchasing agent

12. Duplicate Payments

Duplicate payments are sometimes converted to the use of an employee. The employee may notice the duplicate payment, then he or she may prepare a phony endorsement of the check.

13. Ghost Employees

Ghost employee schemes are frequently uncovered when an auditor, fraud examiner, or other individual distributes paychecks to employees. Missing or otherwise unaccounted for employees could indicate the existence of a ghost employee scheme.

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

ORACLE

20 Ways to Detect Fraud

14. Employee Expense Accounts

Employees frequently conceal fraud in their individual expense account reimbursements. These reimbursements should be scrutinized for reasonableness and trends, especially in the area of cash transactions on the expense account.

15. Inventory Shortages

Normal shrinkage over a period of time can be computed through historical analysis. Excessive shrinkage could explain a host of fraudulent activity, from embezzlement to theft of inventory.

16. Increased Scrap

In the manufacturing process, an increased amount of scrap could indicate a scheme to steal and resell this material. Scrap is a favorite target of embezzlers because it is usually subject to less scrutiny than regular inventory.

17. Large Payments to Individuals

Excessively large payments to individuals may indicate instances of fraudulent disbursements.

18. Employee Overtime

Employees being paid for overtime hours not worked by altering time sheets before or after management approval.

19. Write-off of Accounts Receivable

Comparing the write-off of receivables by customers may lead to information indicating that the employee has absconded with customer payments.

20. Post Office Boxes as Shipping Addresses

In instances where merchandise is shipped to a post office box, this may indicate that an employee is shipping to a bogus purchaser.

Pretty Easy? Huh?

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

My Personal Experience

- Purchases were made in pairs of $75.00 purchases

Total purchases exceeds time period average

May 22	1:14 PM	FOOD	Monaco Café	$127.00
May 22	7:32 PM	WINE	Wine Bistro	$28.00
… Gas Station?				
June 14	2:05 PM	MISC	Mobil Mart	$75.00
June 14	2:06 PM	MISC	Mobil Mart	$75.00
June 15	11:48 AM	MISC	Mobil Mart	$75.00
June 15	11:49 AM	MISC	Mobil Mart	$75.00
May 22	7:32	WINE	Wine Bistro	$28.00
May 22	7:32	WINE	Wine Bistro	$28.00
June 16	11:48 AM	MISC	Mobil Mart	$75.00
June 16	11:49 AM	MISC	Mobil Mart	$75.00

France

Pairs of $75?

All same $75 amount?

ORACLE

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',
  '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 ('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 Beta

Mix_beta_logo_v2

Oracle OpenWorld 2009 Vote-a-Session

Login to Mix

Go to Mix and login »

Main | Streams | Tracks | FAQ

Home » Industries » Financial Services » Proposal: Building a Check Fraud Dete...

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

Type	Conference Session	Presenter	Jiang Zhou (Customer Speaker)
Track	Financial Services	Stream	Industries

Abstract

It is estimated that the nation's banks experience over $10 billion per year in attempted check fraud. In order to minimize this type of fraud loss, a data driven predictive model is assembled, tested and deployed within a single platform using Oracle 11g and Oracle Data Mining (ODM). Since all of the processes can be executed within a database, it provides a level of security crucial to banks. Oracle 11g and ODM allow users to perform a multitude of analysis using large volumes of transactional data. The features of Oracle 11g and ODM (e.g. analytical functions, various predictive models, ease of model deployment, materialized views, partitioning, etc.) result in substantially increased productivity, manageability and scalability.

About the presenter

Jiang Zhou

Voting is closed

Share this: | More...

25 people voted for this

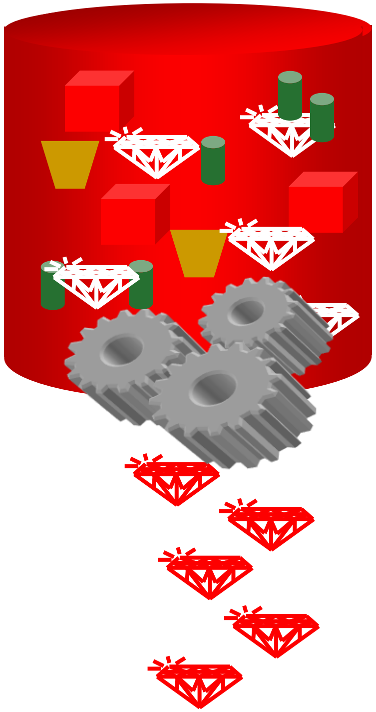
More in this stream and track

Managing Change in Banking: Integrated Performance Management Several forces impact retail, commercial, and investment bankers today: - Mor...

The next generation User Experience Everybody is talking about Web2.0 but how does it really enable your business...

Oracle Data Mining Option

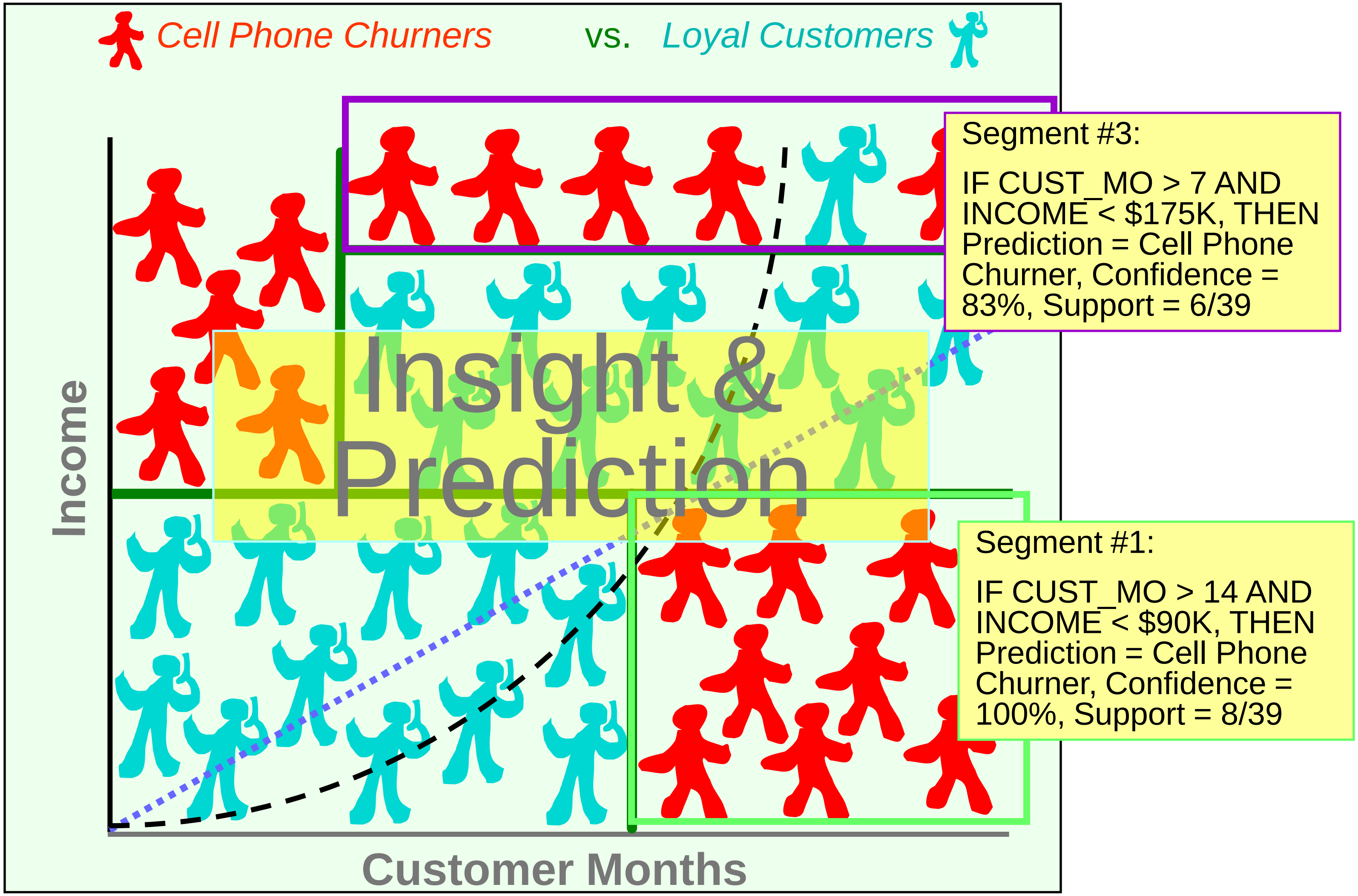
What is Data Mining?



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

Data Mining Provides

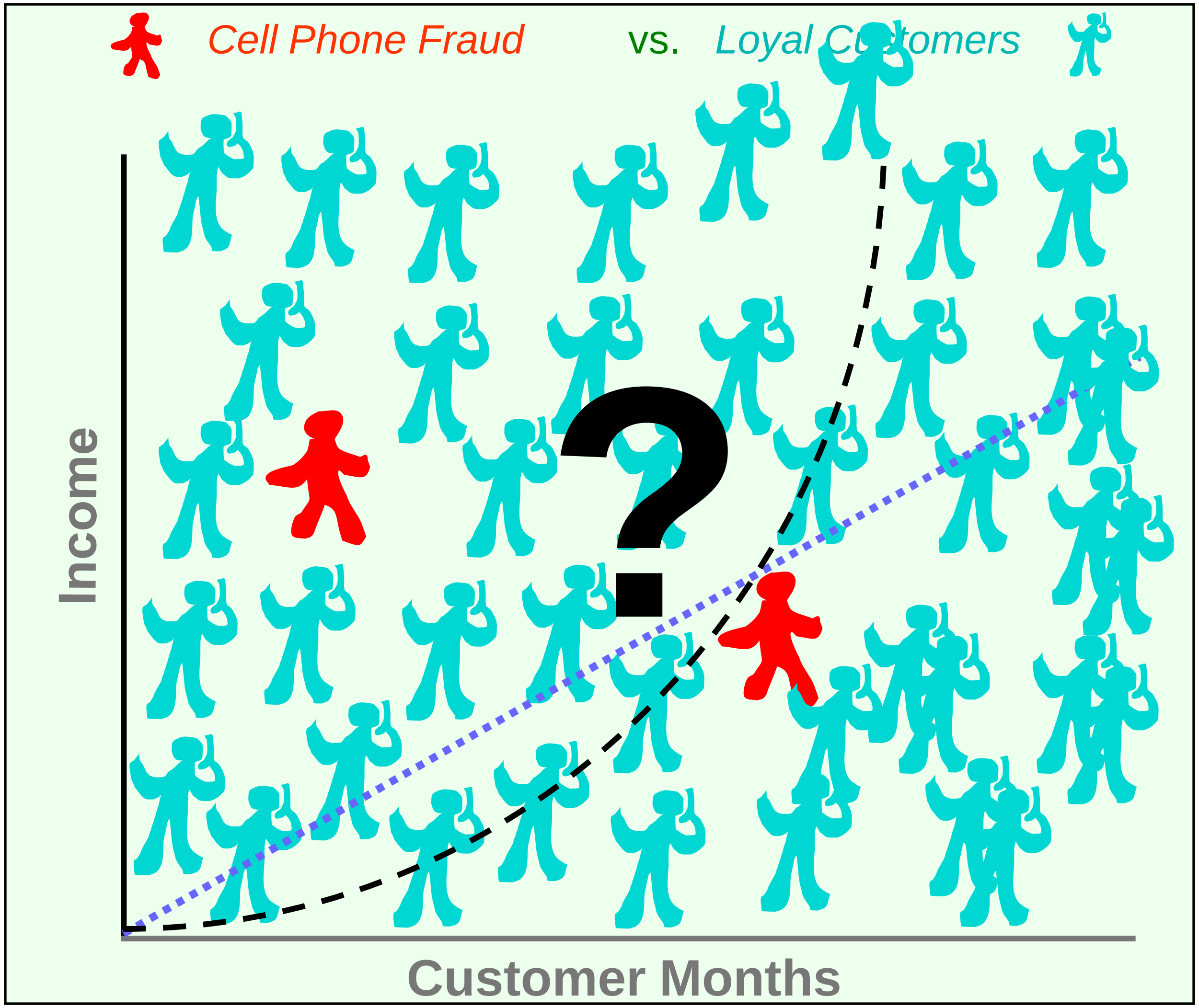
Better Information, Valuable Insights and Predictions



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

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

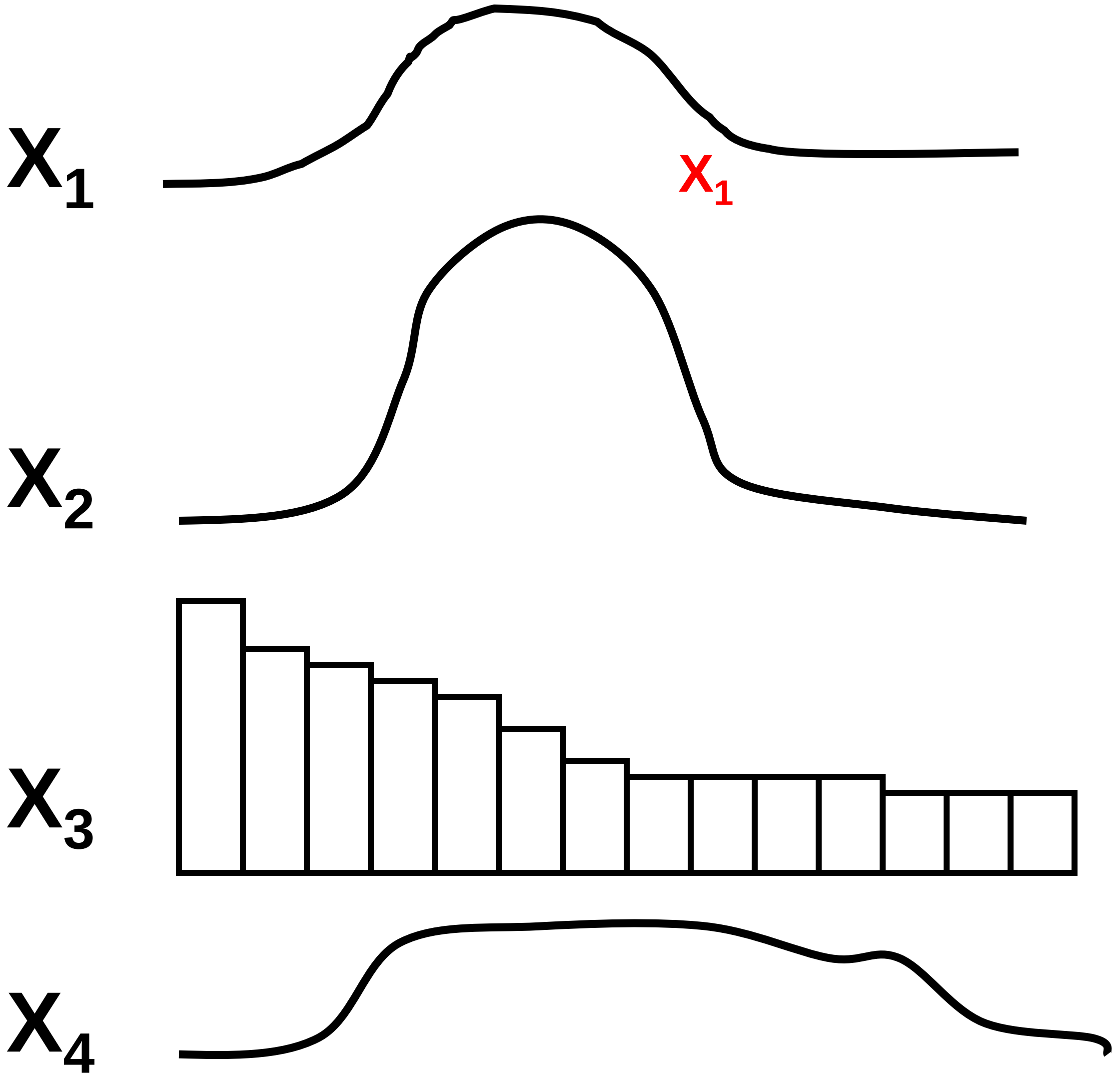
Finding Needles in Haystacks

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

Look for What is "*Different*"

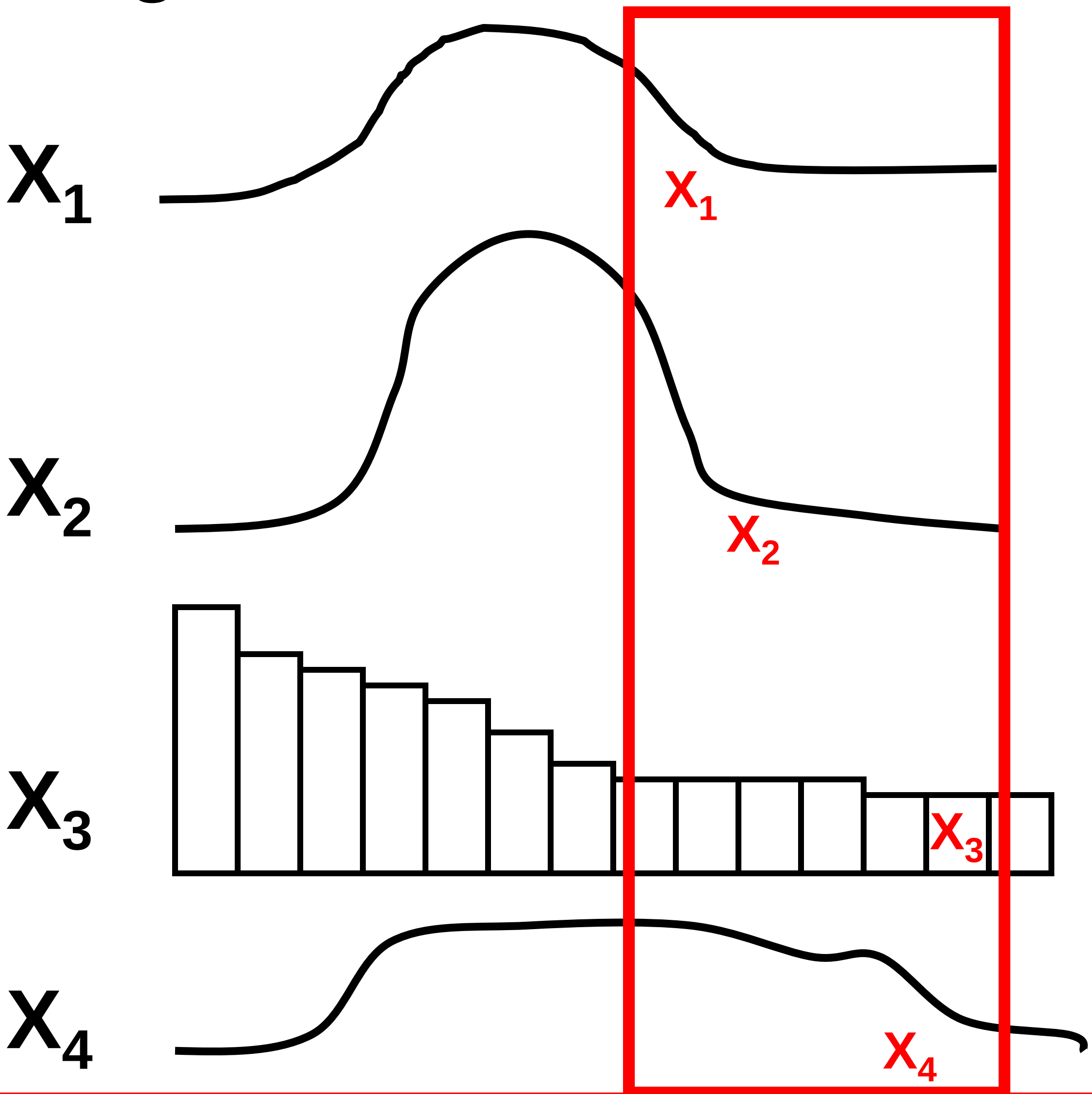
Challenge: Finding Anomalous Records

- Considering one attribute at a time, each record, may look “normal”



Challenge: Finding Anomalous Records

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



Oracle Data Mining

Overview (Classification)

Input Attributes

Target

Model

Functional Relationship:

$$Y = F(X_1, X_2, \ldots, X_m)$$

Cases

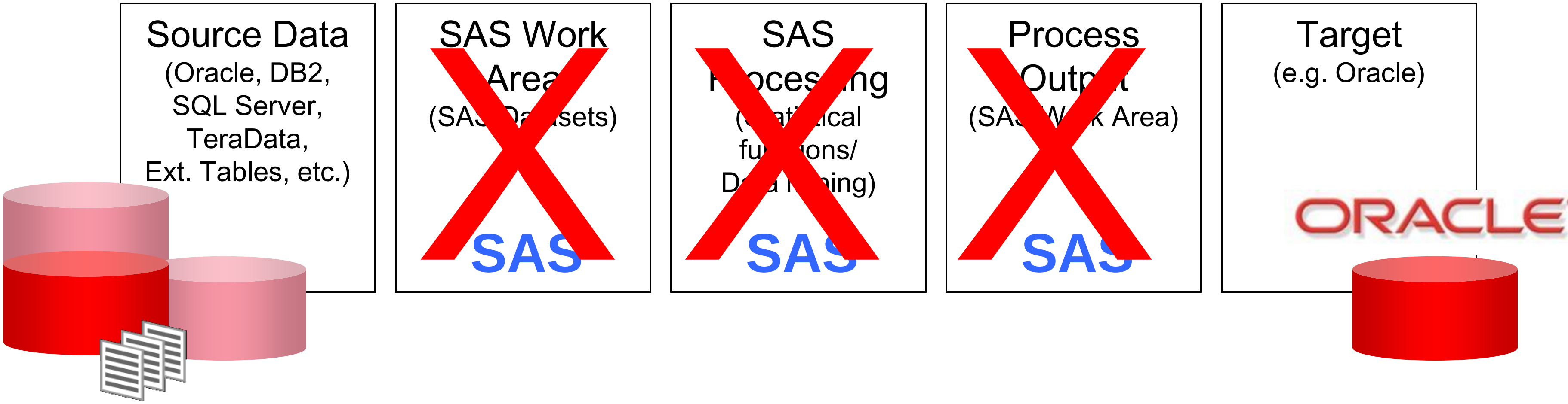
Name	Income	Age	. . . .		Fraud? 1 =Yes, 0 =No	Prediction	Confidence
Historic Data							
Jones	30,000	30			0		
Smith	55,000	67			1		
Lee	25,000	23			0		
Rogers	50,000	44			0		
New Data							
Campos	40,500	52			?	0	.85
Horn	37,000	73			?	0	.74
Habers	57,200	32			?	0	.93
Berger	95,600	34			?	1	.65

Prediction   Confidence

ORACLE DATABASE 11g

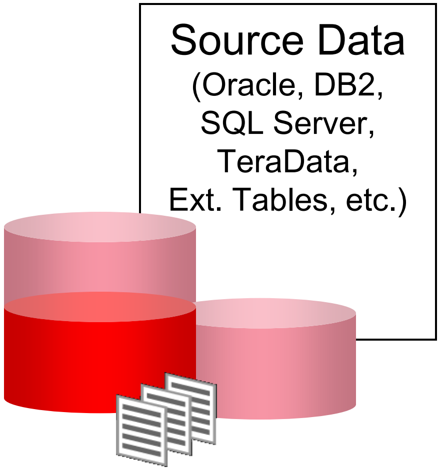
Fraud typically requires a **lot more data** to find examples of rare events

Traditional Analytics (SAS) Environment



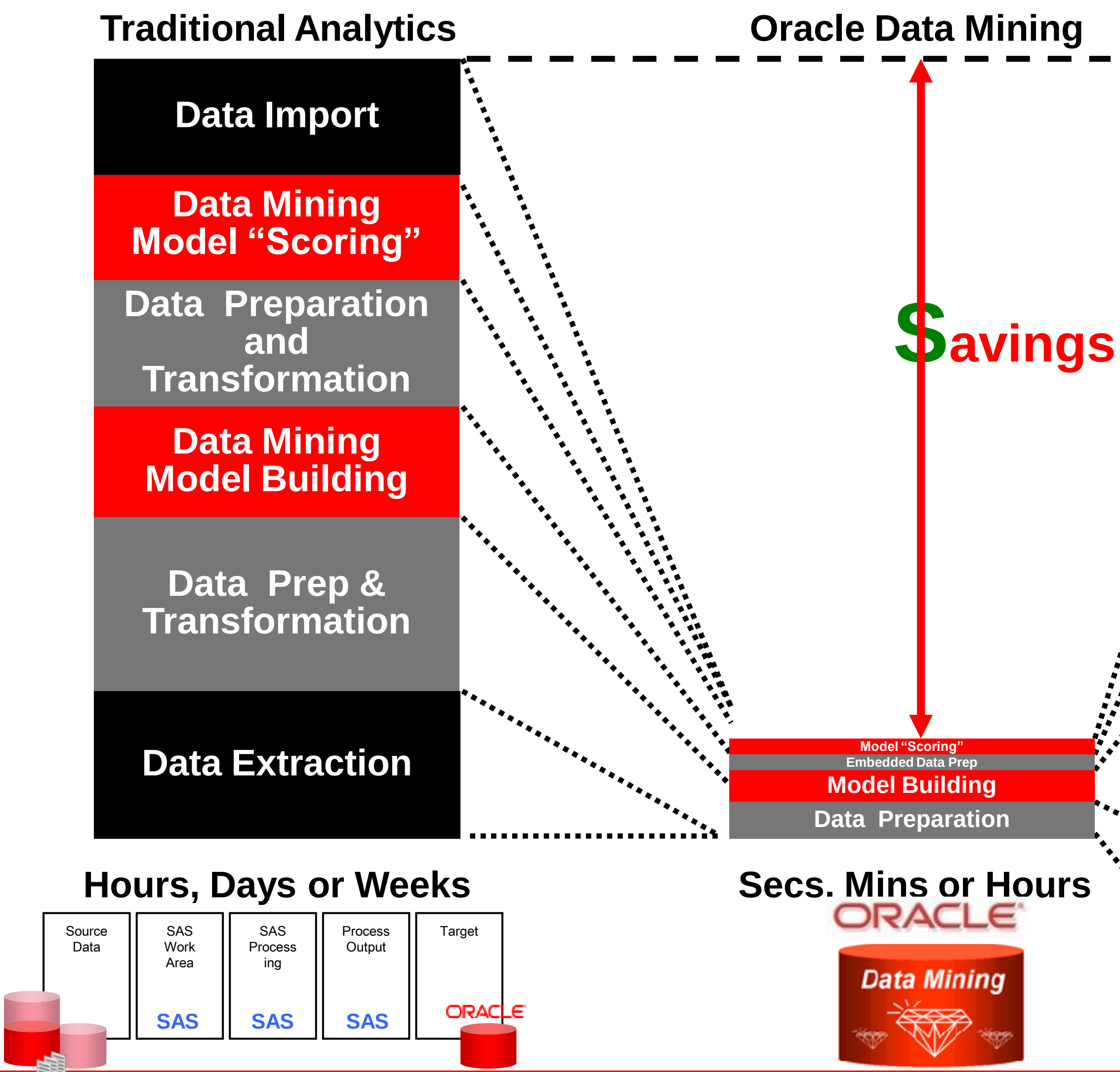
- SAS environment requires:
  - Data movement
  - Data duplication
  - Loss of security

Oracle Architecture



- Oracle environment:
  - Eliminates data movement
  - Eliminates data duplication
  - Preserves security

In-Database Data Mining



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

Oracle Database Machine & ODM

- Integrated data warehouse solution
- Extreme Performance
  - 10-100X faster than conventional DW systems
- Scalability to Petabytes
- Enterprise-Ready
  - Complete data warehouse functionality
  - Enterprise-level availability and security
- Scoring of Oracle Data Mining models
  - Blazingly fast performance
  - For example, find the US customers likely to churn:

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

Oracle Database Machine & ODM

- In 11gR2, SQL predicates and Oracle Data Mining models are pushed to storage level for execution

  For example, find the US customers likely to churn:

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

ODM 11gR2 Scoring: Offloaded to Exadata

- Data mining scoring executed in Exadata:

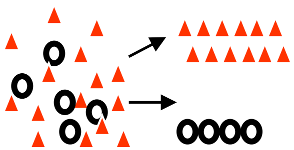
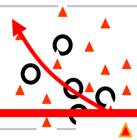
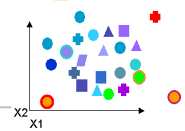
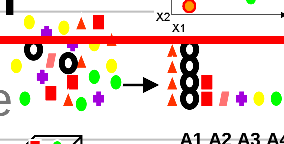
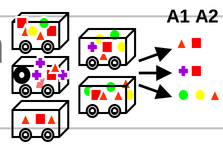
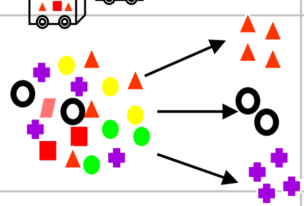
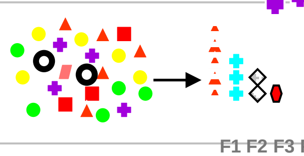
```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod, 'Y' using *) > 0.8;
```

Scoring function executed in Exadata

- All scoring functions offloaded to Exadata
- Benefits
  - Reduces data returned from Exadata to Database server
  - Reduces CPU utilization on Database Server
  - Up to 10x performance gains

Oracle Data Mining

Algorithm Summary 11g

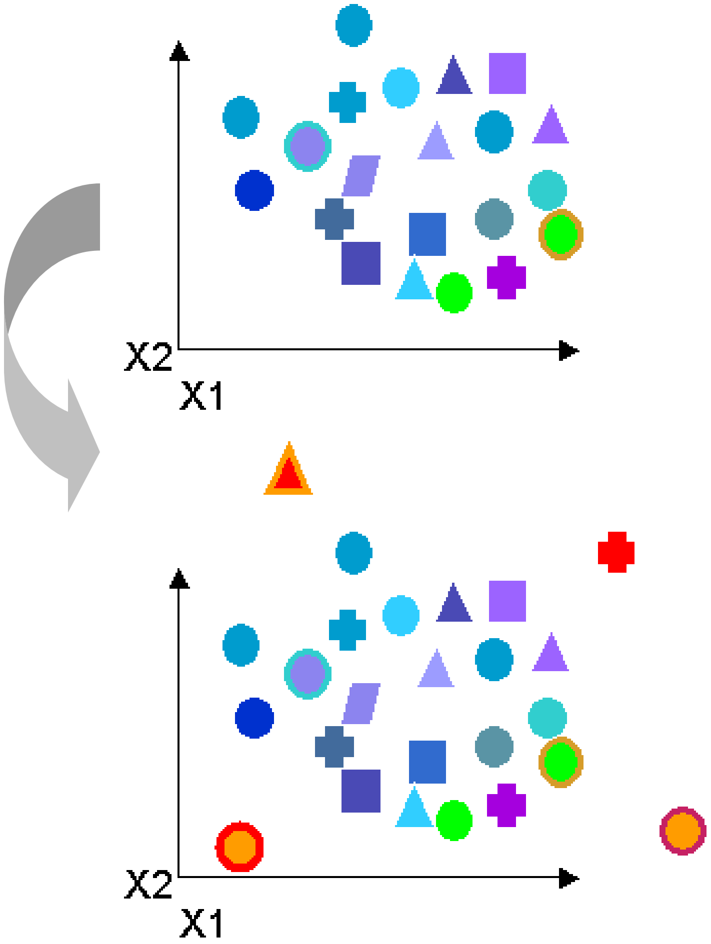
Problem	Algorithm	Applicability
Classification 	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression 	Multiple Regression (GLM) Support Vector Machine	Classical statistical 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

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

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

Brief Demonstration

1. Oracle Data Mining

Quick Demo: Oracle Data Mining

- Scenario: Auto Insurance Company
- Business problem(s):
  1. Better understand the business by looking at graphs
  2. Target Fraud
     a. Build a predictive model to discover “normal” records
     b. Apply the model to flag “most non-normal” records
     c. View results in an OBI EE Dashboard
        - (Entire process **can be automated** w/ PL/SQL and/or Java APIs)

Insurance Claims Fraud

Structure | Data

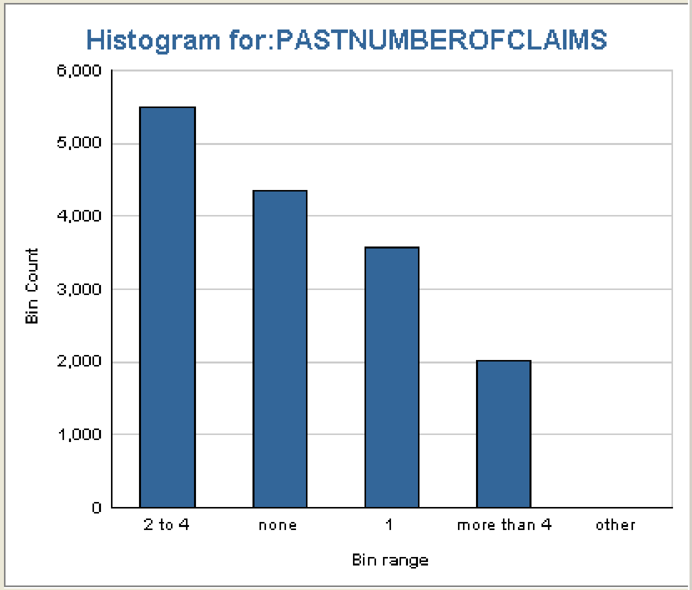
Fetch Size: 100 Fetch Next Refresh

MONTH	WEEKOFMO...	DAYOFWEEK	MAKE	ACCIDENTA...	DAYOFWEE...	MONTHCLAI...	WEEKOFMO...	SEX	MARITALST...	AGE	FAULT	POLICYTYPE
Dec	4	Sunday	Pontiac	Urban	Wednesday	Jan	2	Male	Married	35	Policy Holder	Sedan - Colli...
Nov	4	Thursday	Mazda	Urban	Friday	Nov	4	Male	Married	46	Third Party	Sedan - Liabi...
Feb	4	Tuesday	Accura	Urban	Friday	Apr	4	Male	Married	62	Policy Holder	Sedan - All P...
Feb	4	Tuesday	Dodge	Urban	Thursday	Mar	1	Male	Married	42	Third Party	Sedan - Colli...
Feb	4	Tuesday	Honda	Urban	Wednesday	Mar	1	Male	Married	40	Policy Holder	Sedan - All P...
Oct	2	Sunday	Mazda	Urban	Monday	Oct	2	Female	Single	33	Third Party	Sedan - All P...
Sep	3	Monday	Honda	Urban	Monday	Sep	3	Male	Married	44	Policy Holder	Sedan - Colli...
Sep	3	Saturday	Pontiac	Urban	Monday	Sep	4	Male	Single	35	Policy Holder	Sedan - Liabi...
Sep	1	Monday	Toyota	Urban	Wednesday	Sep	2	Male	Married	57	Policy Holder	Sedan - Liabi...
Sep	1	Monday	Pontiac	Urban	Thursday	Sep	2	Male	Married	38	Policy Holder	Sedan - All P...
Aug	4	Friday	Mazda	Urban	Wednesday	Aug	5	Male	Married	30	Policy Holder	Utility - All Pe...
Jul	3	Tuesday	Mazda	Urban	Thursday	Jul	4	Female	Married	38	Policy Holder	Sedan - Liabi...
Jul	3	Friday	Pontiac	Urban	Tuesday	Jul	3	Male	Married	54	Policy Holder	Sedan - Liabi...
Nov	2	Wednesday	Toyota	Urban	Monday	Nov	2	Male	Married	52	Third Party	Sedan - All P...
Aug	1	Tuesday	Toyota	Urban	Thursday	Aug	1	Female	Single	26	Policy Holder	Sedan - Colli...
Jul	3	Monday	Honda	Urban	Monday	Jul	3	Male	Married	39	Policy Holder	Sedan - Liabi...
Jun	3	Wednesday	Chevrolet	Urban	Tuesday	Jun	3	Male	Married	69	Policy Holder	Sedan - Colli...
May	3	Wednesday	Toyota	Urban	Thursday	May	3	Male	Single	35	Policy Holder	Sedan - Liabi...
May	2	Friday	Chevrolet	Urban	Friday	May	2	Male	Married	47	Policy Holder	Sedan - Liabi...
Apr	4	Friday	Pontiac	Urban	Monday	May	3	Male	Married	55	Policy Holder	Sedan - All P...
Mar	4	Saturday	Mazda	Urban	Monday	Mar	4	Male	Married	26	Policy Holder	Sedan - Liabi...
Oct	4	Thursday	Mazda	Urban	Friday	Nov	1	Male	Married	25	Policy Holder	Sedan - Colli...
Oct	4	Thursday	Chevrolet	Urban	Monday	Nov	2	Male	Married	30	Policy Holder	Sedan - Colli...
Oct	2	Thursday	Honda	Urban	Saturday	Oct	3	Male	Married	31	Policy Holder	Sedan - Liabi...
Oct	2	Monday	Pontiac	Urban	Thursday	Oct	4	Male	Married	28	Policy Holder	Sedan - Liabi...
Sep	2	Monday	Toyota	Urban	Tuesday	Sep	3	Male	Married	61	Third Party	Sedan - Liabi...
Aug	4	Thursday	Saab	Urban	Saturday	Sep	2	Male	Married	63	Third Party	Sedan - All P...
Aug	2	Friday	Mazda	Urban	Tuesday	Sep	1	Male	Married	43	Policy Holder	Sedan - Colli...
Jun	2	Tuesday	Toyota	Urban	Wednesday	Jun	3	Female	Married	37	Policy Holder	Sedan - Liabi...
May	4	Monday	Pontiac	Urban	Monday	May	5	Male	Married	44	Third Party	Sedan - Liabi...
Jun	3	Monday	Pontiac	Urban	Monday	Jun	3	Male	Single	30	Policy Holder	Sedan - Colli...

Insurance Claims Fraud

Histogram for selected attribute

Data Source:CBERGER.AAA_INSUR_CLAIMS Attribute:

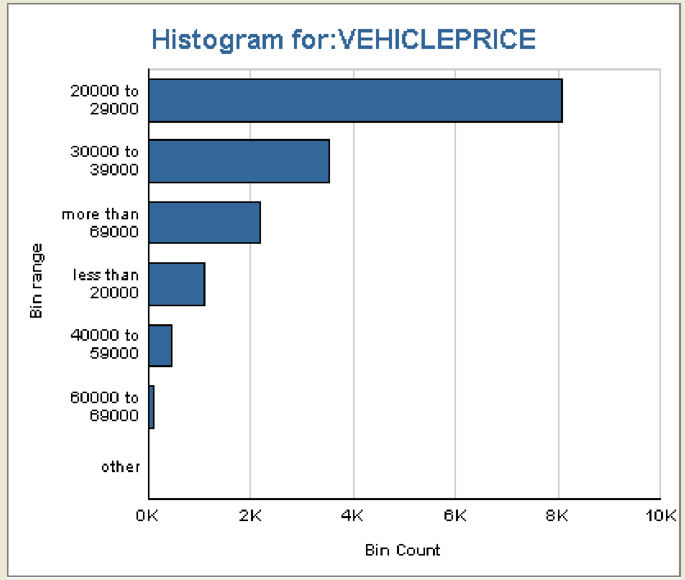


Group	Value(s)	Bin Count	% of Total
2 to 4	2 to 4	5485	35.57
none	none	4352	28.22
1	1	3573	23.17
more than 4	more than 4	2010	13.04
other	nulls	0	0.0

Help

Histogram for selected attribute

Data Source:CBERGER.AAA_INSUR_CLAIMS Attribute: VEHICLEPRICE



Statistics:

Sample count: 15420

Mode: 20000 to

Group	Value(s)	Bin Count	% of Total
20000 to 29000	20000 to 29000	8079	52.39
30000 to 39000	30000 to 39000	3533	22.91
more than 69000	more than 69000	2164	14.03
less than 20000	less than 20000	1096	7.11
40000 to 59000	40000 to 59000	461	2.99
60000 to 69000	60000 to 69000	87	0.56
other	nulls	0	0.0

Binning Strategy:

Top N...

Graph orientation:

Vertical

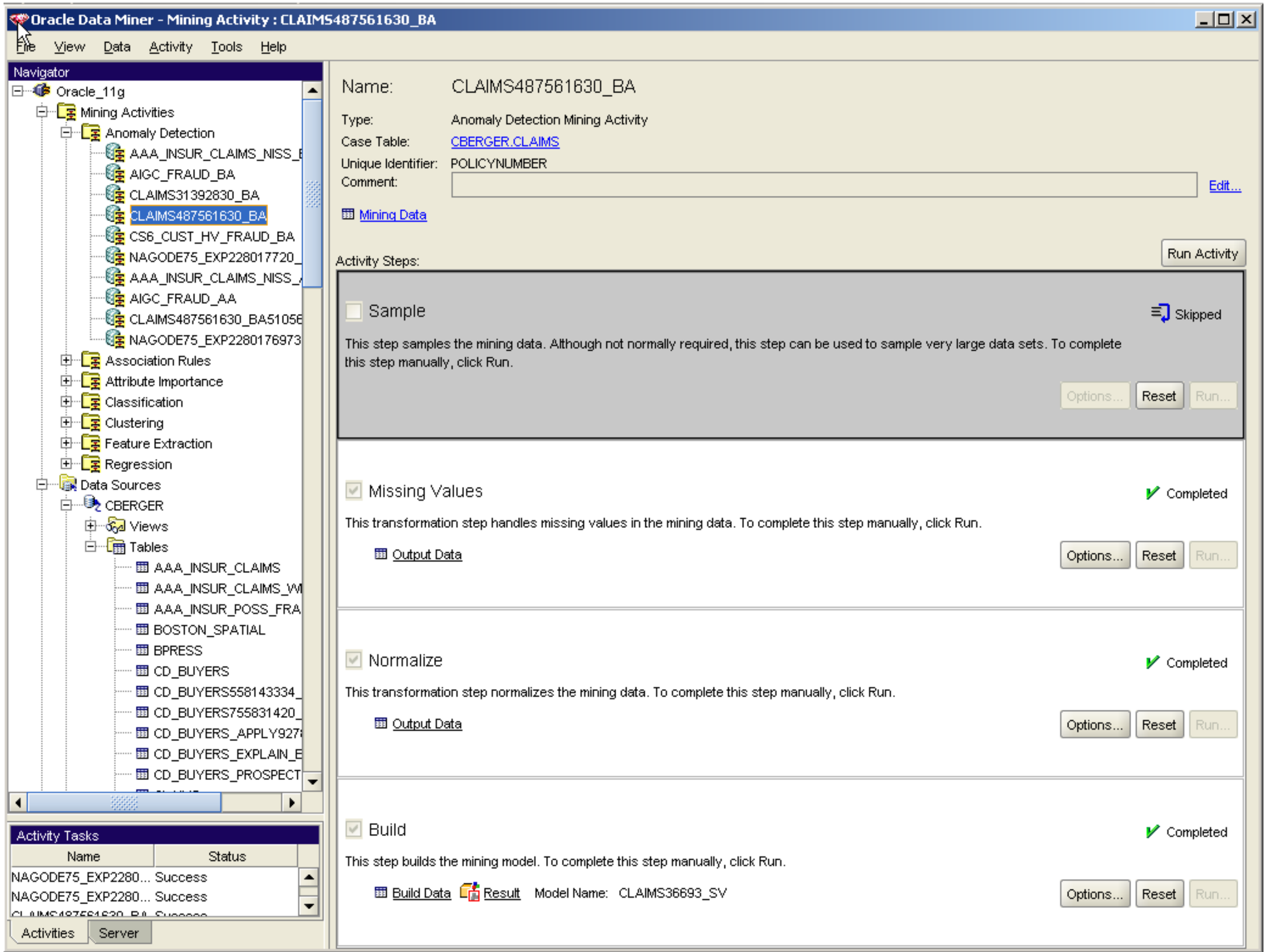
Horizontal

Help

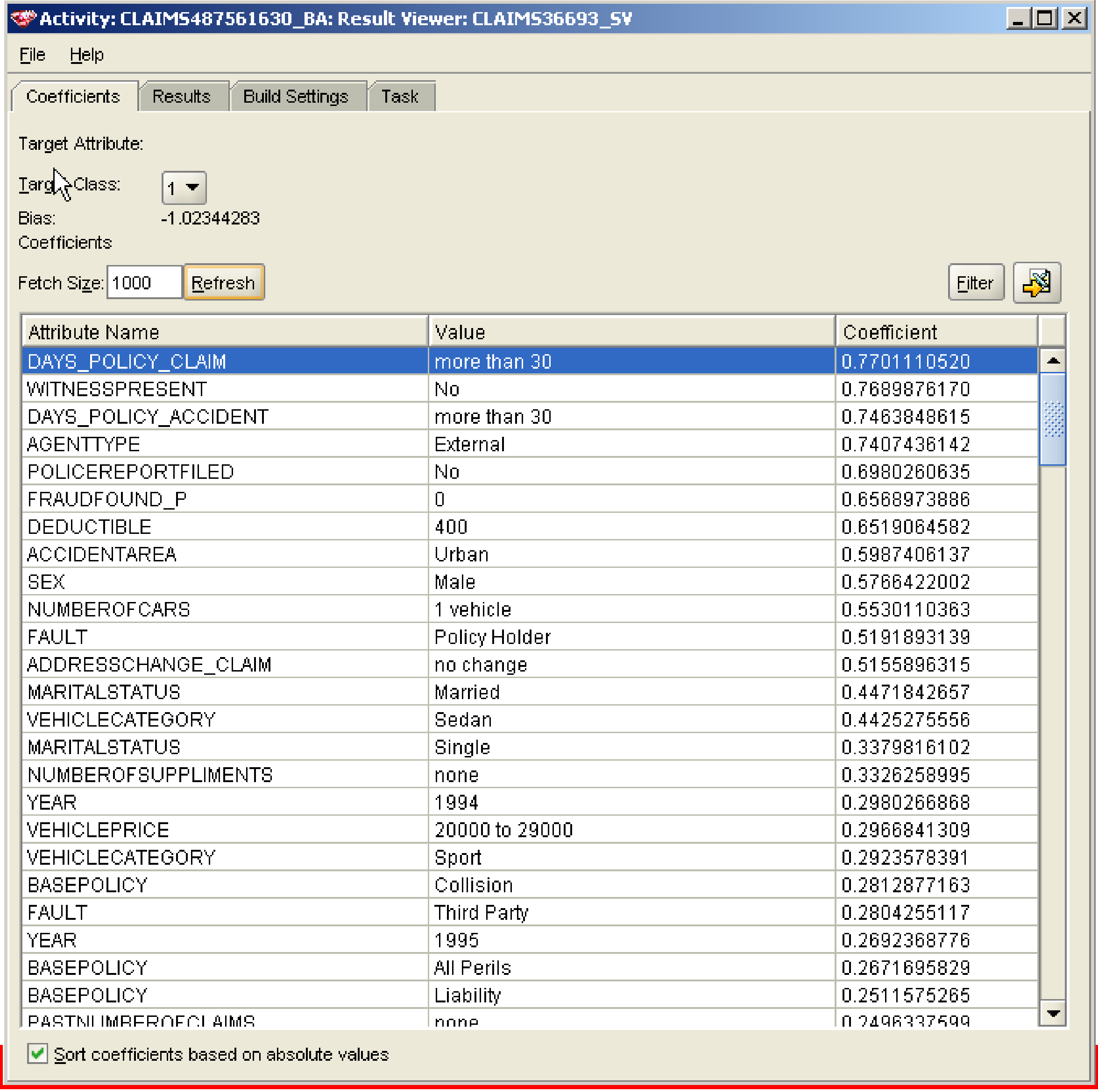
OK

ORACLE

Insurance Claims Fraud



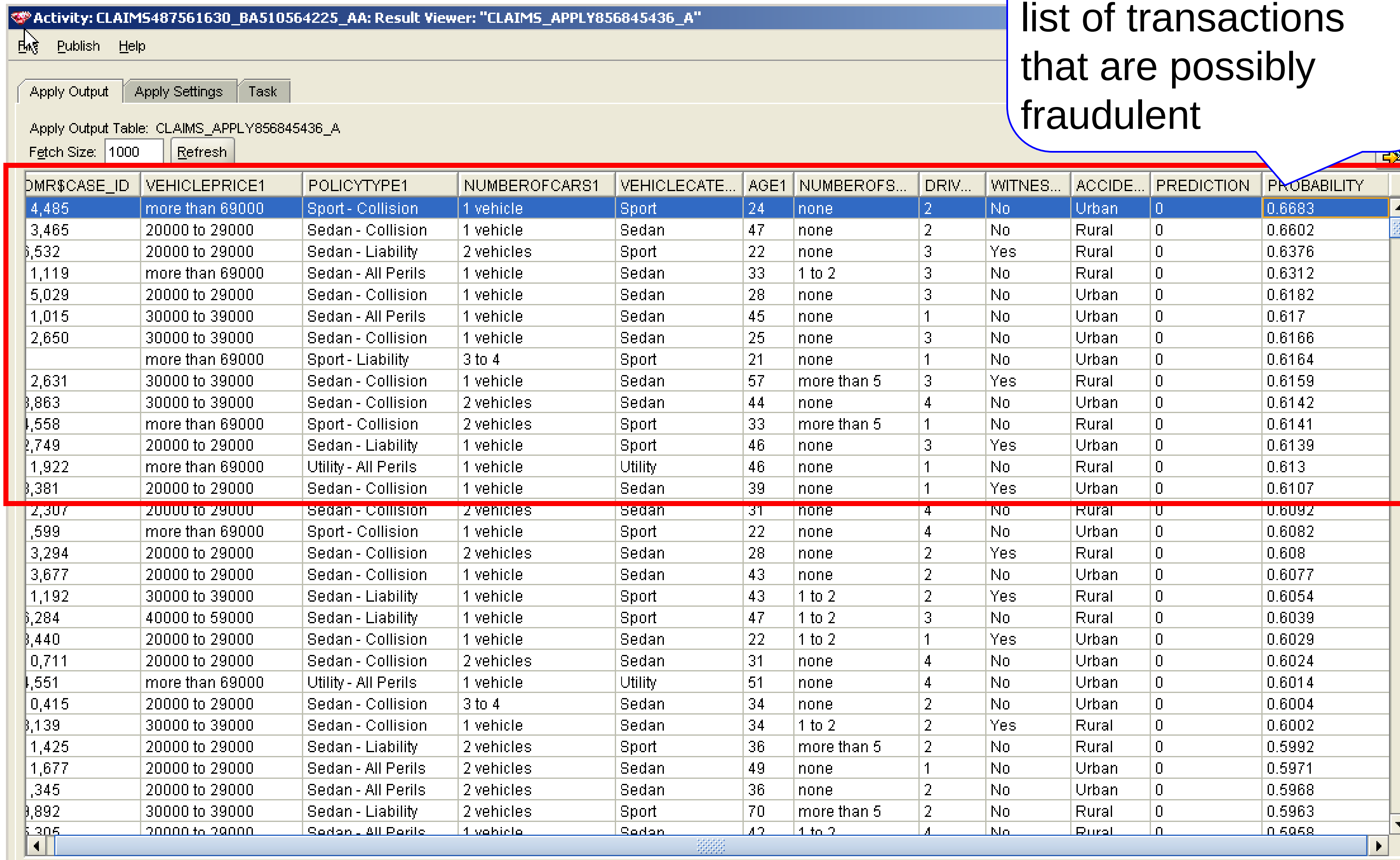
Insurance Claims Fraud



Attribute Name	Value	Coefficient
DAYS_POLICY_CLAIM	more than 30	0.7701110520
WITNESSPRESENT	No	0.7689876170
DAYS_POLICY_ACCIDENT	more than 30	0.7463848615
AGENTTYPE	External	0.7407436142
POLICEREPORTFILED	No	0.6980260635
FRAUDFOUND_P	0	0.6568973886
DEDUCTIBLE	400	0.6519064582
ACCIDENTAREA	Urban	0.5987406137
SEX	Male	0.5766422002
NUMBEROFCARS	1 vehicle	0.5530110363
FAULT	Policy Holder	0.5191893139
ADDRESSCHANGE_CLAIM	no change	0.5155896315
MARITALSTATUS	Married	0.4471842657
VEHICLECATEGORY	Sedan	0.4425275556
MARITALSTATUS	Single	0.3379816102
NUMBEROFSUPPLIMENTS	none	0.3326258995
YEAR	1994	0.2980266868
VEHICLEPRICE	20000 to 29000	0.2966841309
VEHICLECATEGORY	Sport	0.2923578391
BASEPOLICY	Collision	0.2812877163
FAULT	Third Party	0.2804255117
YEAR	1995	0.2692368776
BASEPOLICY	All Perils	0.2671695829
BASEPOLICY	Liability	0.2511575265
PASTNUMBEROFCLAIMS	none	0.2496337599

Insurance Claims Fraud

Oracle Data Mining creates a prioritized list of transactions that are possibly fraudulent

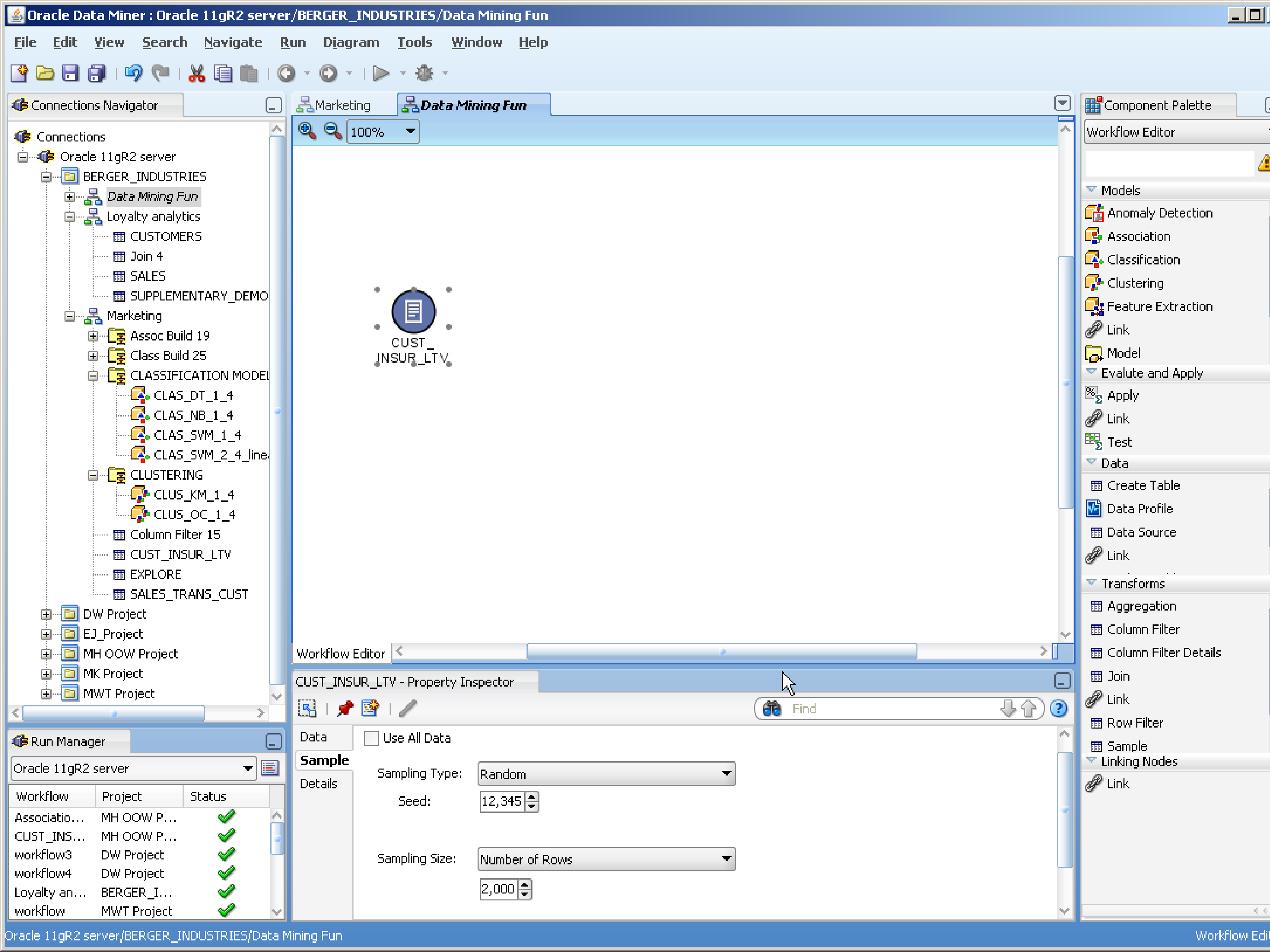


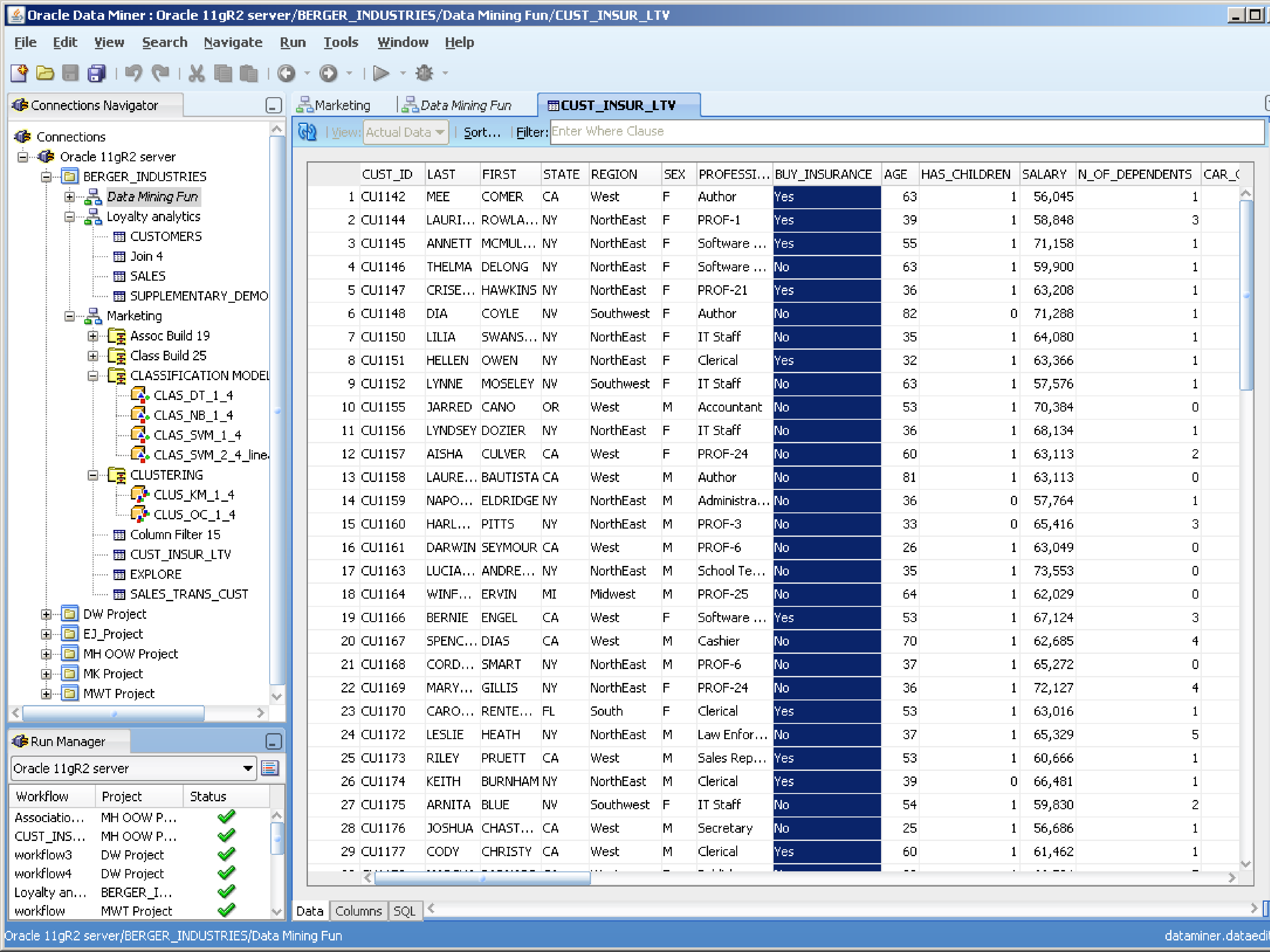
DMR$CASE_ID	VEHICLEPRICE1	POLICYTYPE1	NUMBEROFCARS1	VEHICLECATE...	AGE1	NUMBEROFS...	DRIV...	WITNES...	ACCIDE...	PREDICTION	PROBABILITY
4,485	more than 69000	Sport - Collision	1 vehicle	Sport	24	none	2	No	Urban	0	0.6683
3,465	20000 to 29000	Sedan - Collision	1 vehicle	Sedan	47	none	2	No	Rural	0	0.6602
,532	20000 to 29000	Sedan - Liability	2 vehicles	Sport	22	none	3	Yes	Rural	0	0.6376
1,119	more than 69000	Sedan - All Perils	1 vehicle	Sedan	33	1 to 2	3	No	Rural	0	0.6312
5,029	20000 to 29000	Sedan - Collision	1 vehicle	Sedan	28	none	3	No	Urban	0	0.6182
1,015	30000 to 39000	Sedan - All Perils	1 vehicle	Sedan	45	none	1	No	Urban	0	0.617
2,650	30000 to 39000	Sedan - Collision	1 vehicle	Sedan	25	none	3	No	Urban	0	0.6166
	more than 69000	Sport - Liability	3 to 4	Sport	21	none	1	No	Urban	0	0.6164
2,631	30000 to 39000	Sedan - Collision	1 vehicle	Sedan	57	more than 5	3	Yes	Rural	0	0.6159
,863	30000 to 39000	Sedan - Collision	2 vehicles	Sedan	44	none	4	No	Urban	0	0.6142
,558	more than 69000	Sport - Collision	2 vehicles	Sport	33	more than 5	1	No	Rural	0	0.6141
,749	20000 to 29000	Sedan - Liability	1 vehicle	Sport	46	none	3	Yes	Urban	0	0.6139
1,922	more than 69000	Utility - All Perils	1 vehicle	Utility	46	none	1	No	Rural	0	0.613
,381	20000 to 29000	Sedan - Collision	1 vehicle	Sedan	39	none	1	Yes	Urban	0	0.6107
2,307	20000 to 29000	Sedan - Collision	2 vehicles	Sedan	31	none	4	No	Rural	0	0.6092
,599	more than 69000	Sport - Collision	1 vehicle	Sport	22	none	4	No	Urban	0	0.6082
3,294	20000 to 29000	Sedan - Collision	2 vehicles	Sedan	28	none	2	Yes	Rural	0	0.608
3,677	20000 to 29000	Sedan - Collision	1 vehicle	Sedan	43	none	2	No	Urban	0	0.6077
1,192	30000 to 39000	Sedan - Liability	1 vehicle	Sport	43	1 to 2	2	Yes	Rural	0	0.6054
,284	40000 to 59000	Sedan - Liability	1 vehicle	Sport	47	1 to 2	3	No	Rural	0	0.6039
,440	20000 to 29000	Sedan - Collision	1 vehicle	Sedan	22	1 to 2	1	Yes	Urban	0	0.6029
0,711	20000 to 29000	Sedan - Collision	2 vehicles	Sedan	31	none	4	No	Urban	0	0.6024
,551	more than 69000	Utility - All Perils	1 vehicle	Utility	51	none	4	No	Urban	0	0.6014
0,415	20000 to 29000	Sedan - Collision	3 to 4	Sedan	34	none	2	No	Urban	0	0.6004
,139	30000 to 39000	Sedan - Collision	1 vehicle	Sedan	34	1 to 2	2	Yes	Rural	0	0.6002
1,425	20000 to 29000	Sedan - Liability	2 vehicles	Sport	36	more than 5	2	No	Rural	0	0.5992
1,677	20000 to 29000	Sedan - All Perils	2 vehicles	Sedan	49	none	1	No	Urban	0	0.5971
,345	20000 to 29000	Sedan - All Perils	2 vehicles	Sedan	36	none	2	No	Urban	0	0.5968
,892	30000 to 39000	Sedan - Liability	2 vehicles	Sport	70	more than 5	2	No	Rural	0	0.5963
,305	20000 to 29000	Sedan - All Perils	1 vehicle	Sedan	42	1 to 2	4	No	Rural	0	0.5958

Oracle Data Miner 11gR2 (GUI) Preview

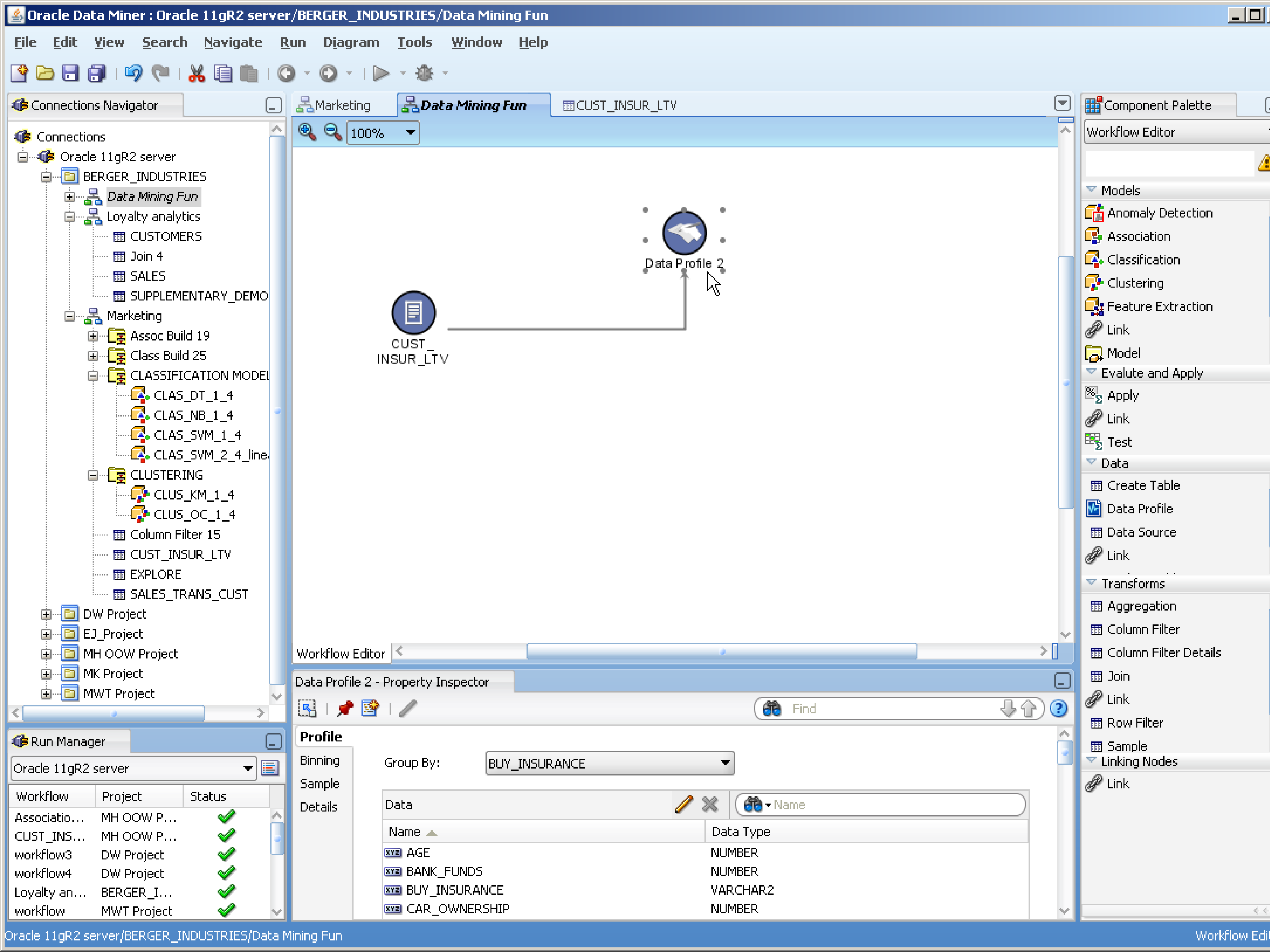
[ODM’r “New”]

ORACLE

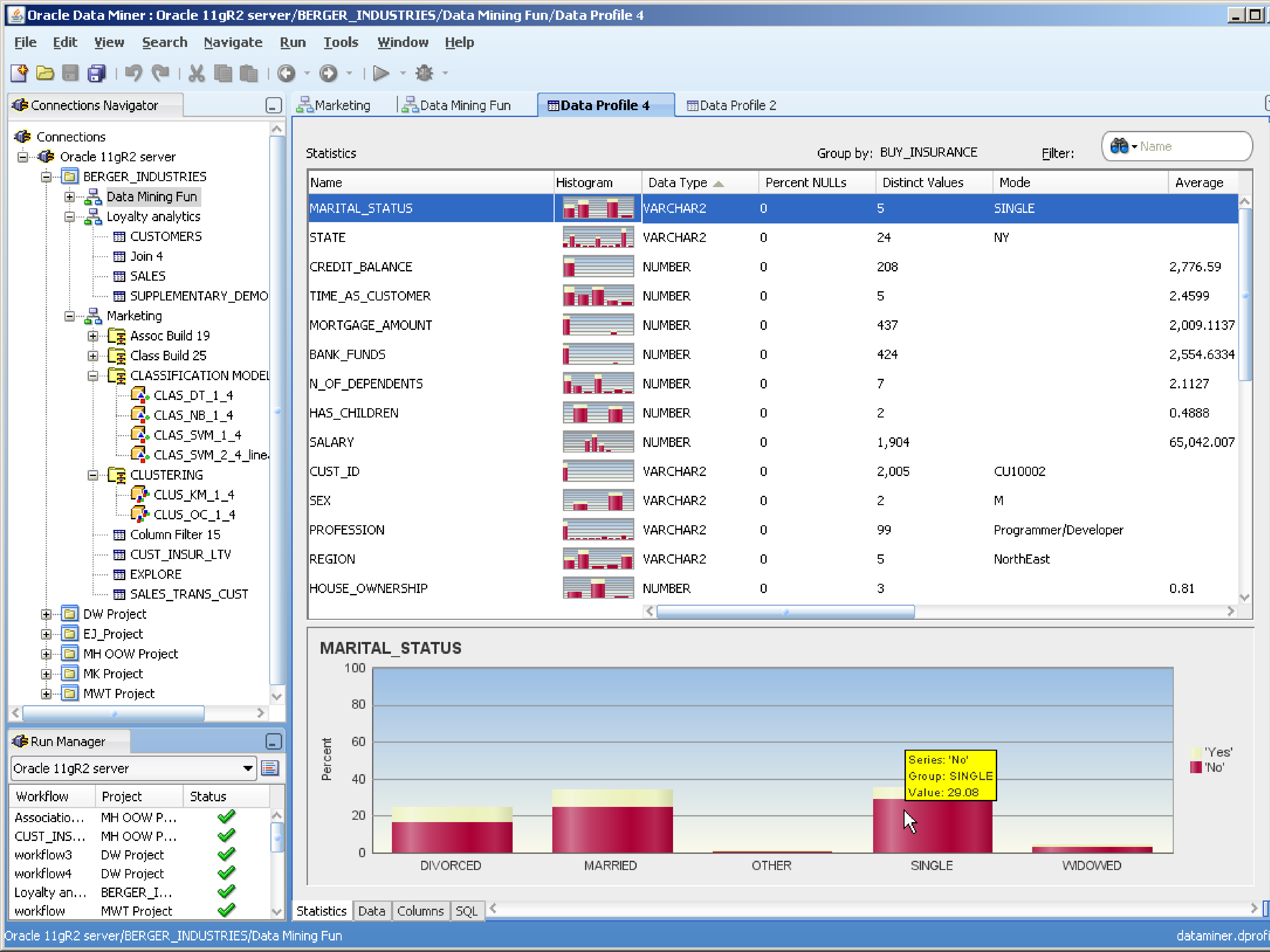

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Run Manager
Oracle 11gR2 server
Workflow
Project
Status
Associatio...
MH OOW P...
CUST_INS...
MH OOW P...
workflow3
DW Project
workflow4
DW Project
Loyalty an...
BERGER_I...
workflow
MWT Project
Marketing
Data Mining Fun
100%
CUST_
INSUR_LTV
Workflow Editor
CUST_INSUR_LTV - Property Inspector
Find
Data
Sample
Details
Use All Data
Sampling Type:
Random
Seed:
12,345
Sampling Size:
Number of Rows
2,000
Component Palette
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Evalute and Apply
Apply
Link
Test
Data
Create Table
Data Profile
Data Source
Link
Transforms
Aggregation
Column Filter
Column Filter Details
Join
Link
Row Filter
Sample
Linking Nodes
Link
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
Workflow Edi

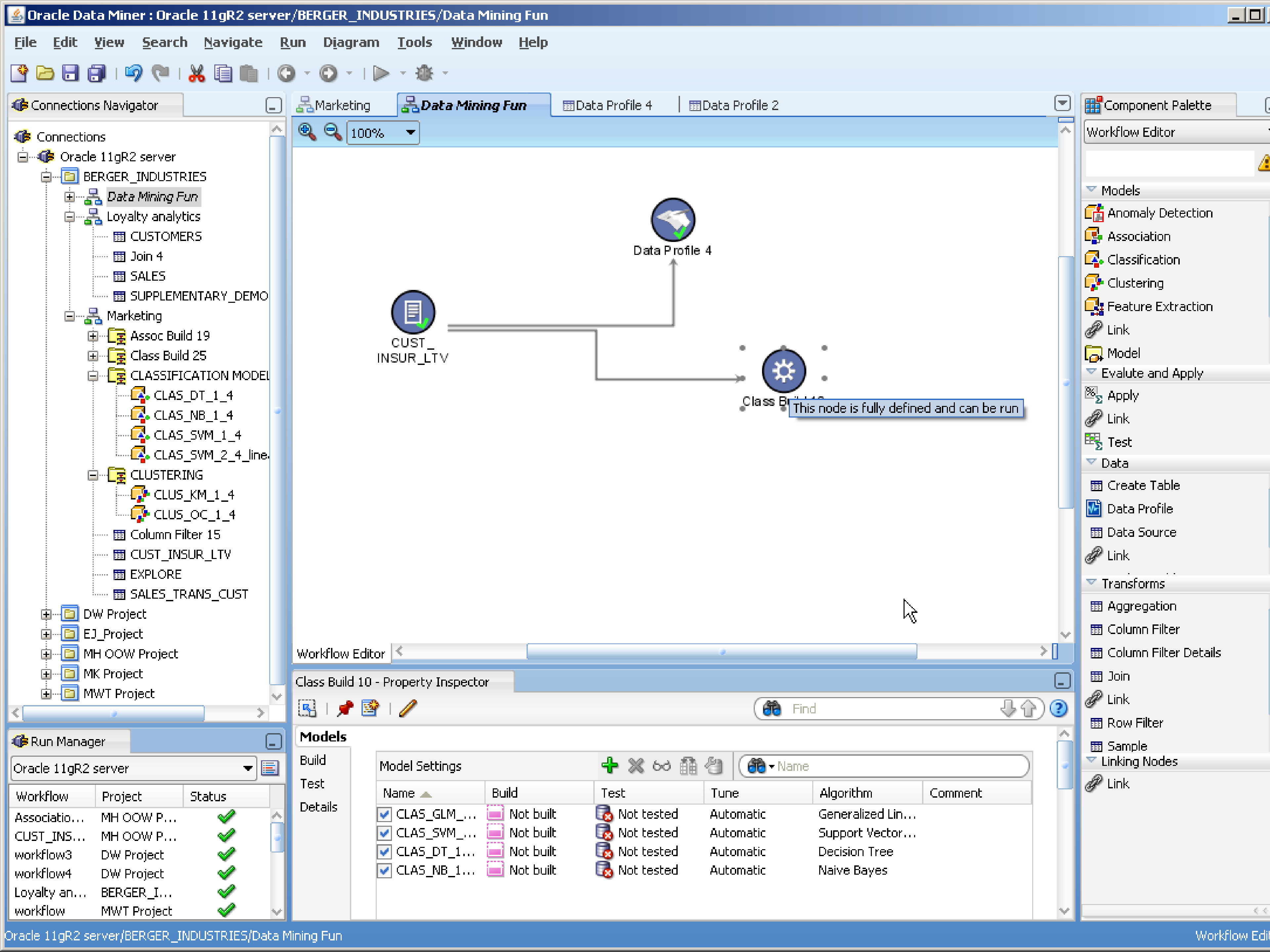


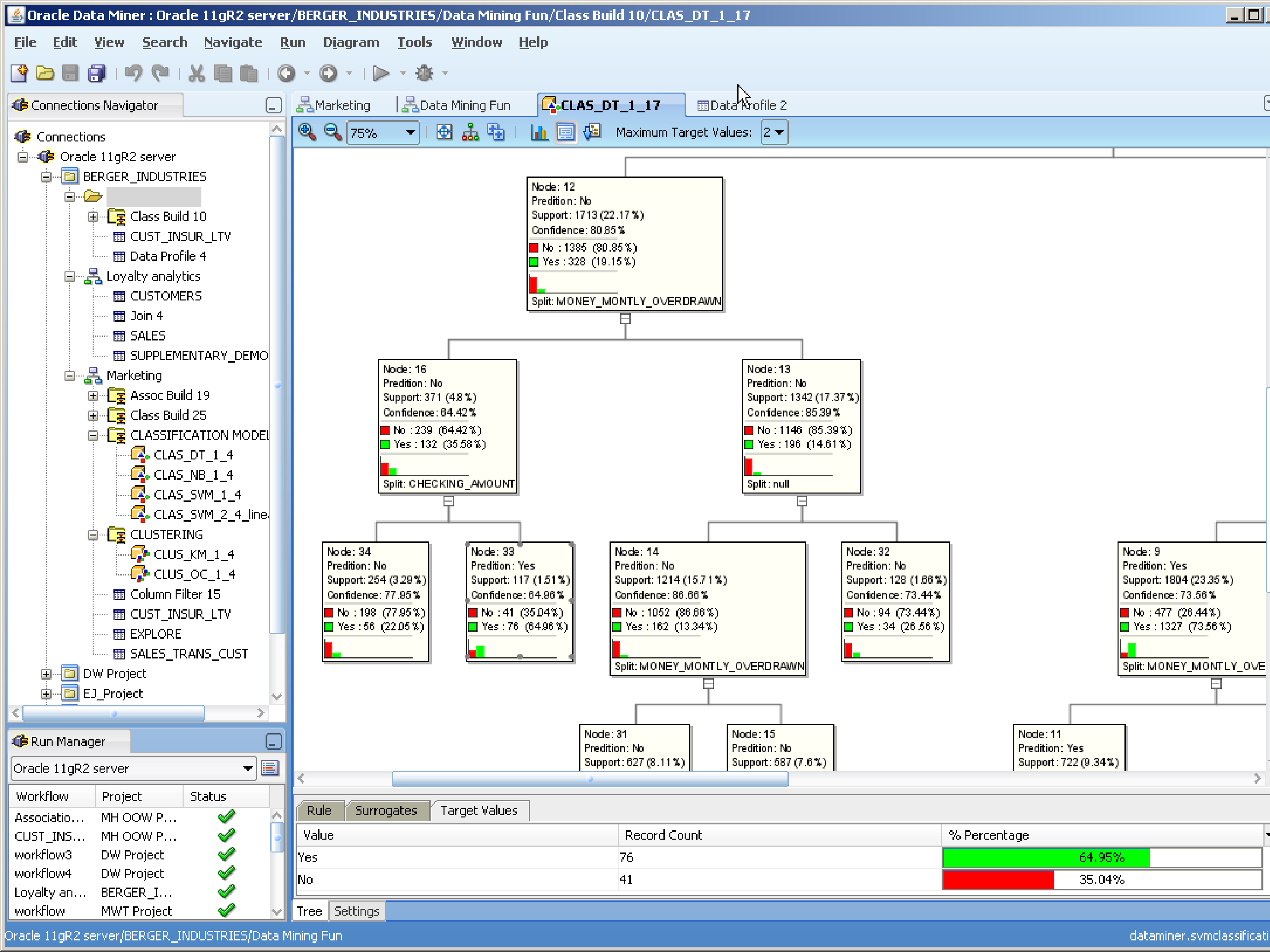
	CUST_ID	LAST	FIRST	STATE	REGION	SEX	PROFESSI...	BUY_INSURANCE	AGE	HAS_CHILDREN	SALARY	N_OF_DEPENDENTS	CAR_(
1	CU1142	MEE	COMER	CA	West	F	Author	Yes	63	1	56,045	1	
2	CU1144	LAURI...	ROWLA...	NY	NorthEast	F	PROF-1	Yes	39	1	58,848	3	
3	CU1145	ANNETT	MCMUL...	NY	NorthEast	F	Software ...	Yes	55	1	71,158	1	
4	CU1146	THELMA	DELONG	NY	NorthEast	F	Software ...	No	63	1	59,900	1	
5	CU1147	CRISE...	HAWKINS	NY	NorthEast	F	PROF-21	Yes	36	1	63,208	1	
6	CU1148	DIA	COYLE	NV	Southwest	F	Author	No	82	0	71,288	1	
7	CU1150	LILIA	SWANS...	NY	NorthEast	F	IT Staff	No	35	1	64,080	1	
8	CU1151	HELLEN	OWEN	NY	NorthEast	F	Clerical	Yes	32	1	63,366	1	
9	CU1152	LYNNE	MOSELEY	NV	Southwest	F	IT Staff	No	63	1	57,576	1	
10	CU1155	JARRED	CANO	OR	West	M	Accountant	No	53	1	70,384	0	
11	CU1156	LYNDSEY	DOZIER	NY	NorthEast	F	IT Staff	No	36	1	68,134	1	
12	CU1157	AISHA	CULVER	CA	West	F	PROF-24	No	60	1	63,113	2	
13	CU1158	LAURE...	BAUTISTA	CA	West	M	Author	No	81	1	63,113	0	
14	CU1159	NAPO...	ELDRIDGE	NY	NorthEast	M	Administra...	No	36	0	57,764	1	
15	CU1160	HARL...	PITTS	NY	NorthEast	M	PROF-3	No	33	0	65,416	3	
16	CU1161	DARWIN	SEYMOUR	CA	West	M	PROF-6	No	26	1	63,049	0	
17	CU1163	LUCIA...	ANDRE...	NY	NorthEast	M	School Te...	No	35	1	73,553	0	
18	CU1164	WINF...	ERVIN	MI	Midwest	M	PROF-25	No	64	1	62,029	0	
19	CU1166	BERNIE	ENGEL	CA	West	F	Software ...	Yes	53	1	67,124	3	
20	CU1167	SPENC...	DIAS	CA	West	M	Cashier	No	70	1	62,685	4	
21	CU1168	CORD...	SMART	NY	NorthEast	M	PROF-6	No	37	1	65,272	0	
22	CU1169	MARY...	GILLIS	NY	NorthEast	F	PROF-24	No	36	1	72,127	4	
23	CU1170	CARO...	RENTE...	FL	South	F	Clerical	Yes	53	1	63,016	1	
24	CU1172	LESLIE	HEATH	NY	NorthEast	M	Law Enfor...	No	37	1	65,329	5	
25	CU1173	RILEY	PRUETT	CA	West	M	Sales Rep...	Yes	53	1	60,666	1	
26	CU1174	KEITH	BURNHAM	NY	NorthEast	M	Clerical	Yes	39	0	66,481	1	
27	CU1175	ARNITA	BLUE	NV	Southwest	F	IT Staff	No	54	1	59,830	2	
28	CU1176	JOSHUA	CHAST...	CA	West	M	Secretary	No	25	1	56,686	1	
29	CU1177	CODY	CHRISTY	CA	West	M	Clerical	Yes	60	1	61,462	1	

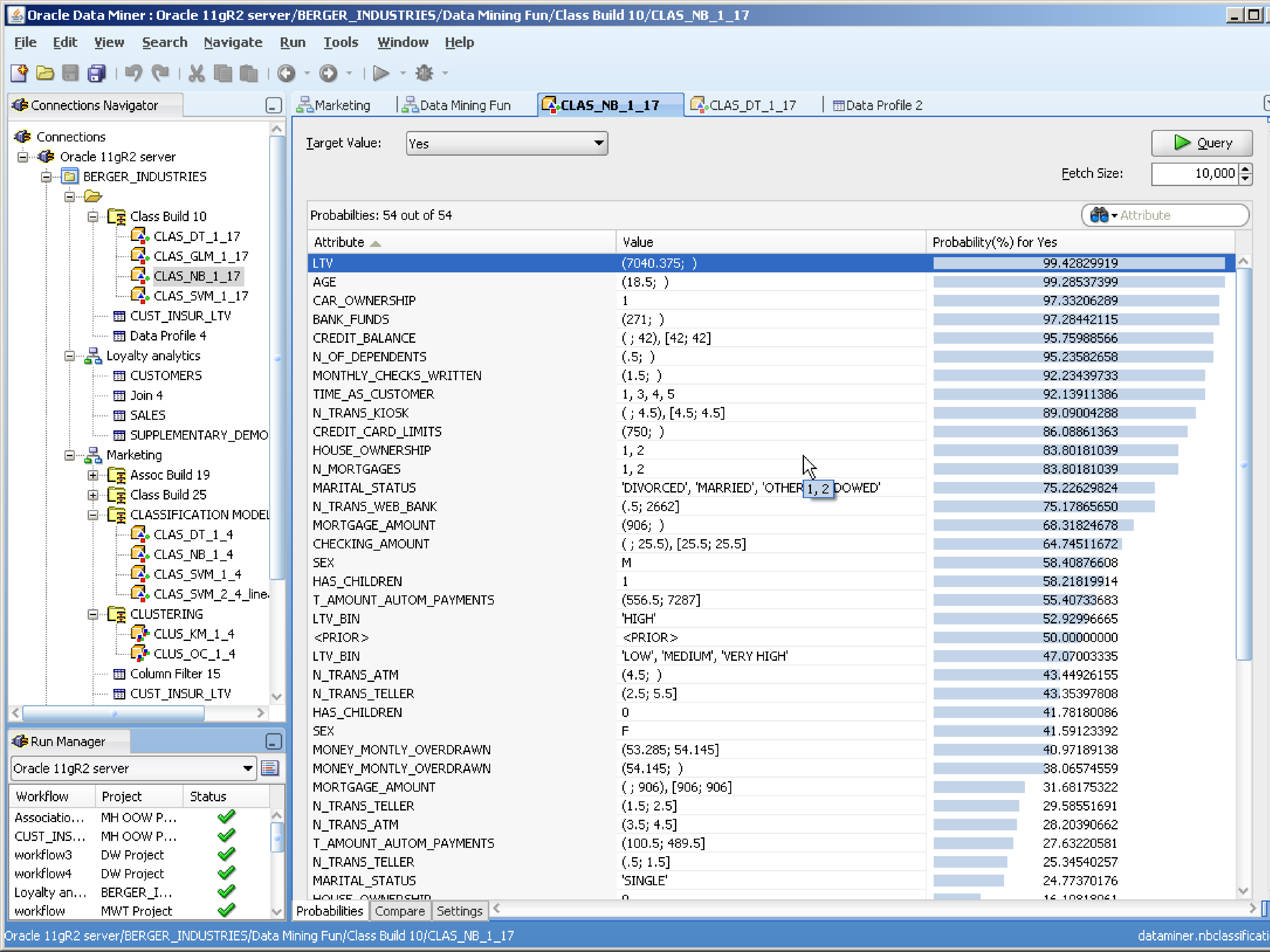


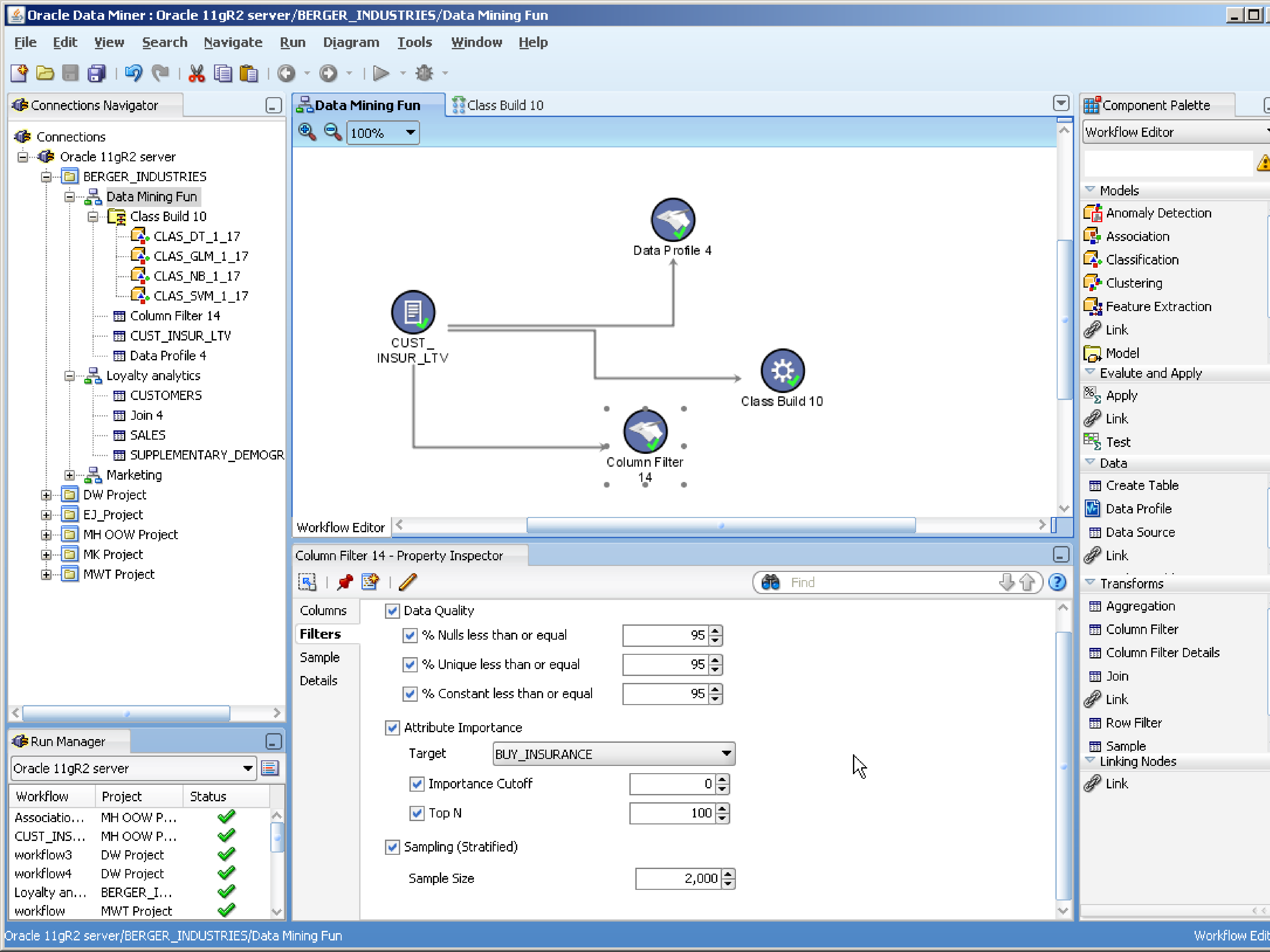
Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Marketing
Data Mining Fun
CUST_INSUR_LTV
100%
Data Profile 2
CUST_
INSUR_LTV
Workflow Editor
Component Palette
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Evalute and Apply
Apply
Link
Test
Data
Create Table
Data Profile
Data Source
Link
Transforms
Aggregation
Column Filter
Column Filter Details
Join
Link
Row Filter
Sample
Linking Nodes
Link
Run Manager
Oracle 11gR2 server
Workflow
Project
Status
Associatio...
MH OOW P...
CUST_INS...
MH OOW P...
workflow3
DW Project
workflow4
DW Project
Loyalty an...
BERGER_I...
workflow
MWT Project
Data Profile 2 - Property Inspector
Find
Profile
Binning
Sample
Details
Group By:
BUY_INSURANCE
Data
Name
Name
Data Type
AGE
NUMBER
BANK_FUNDS
NUMBER
BUY_INSURANCE
VARCHAR2
CAR_OWNERSHIP
NUMBER
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
Workflow Edit

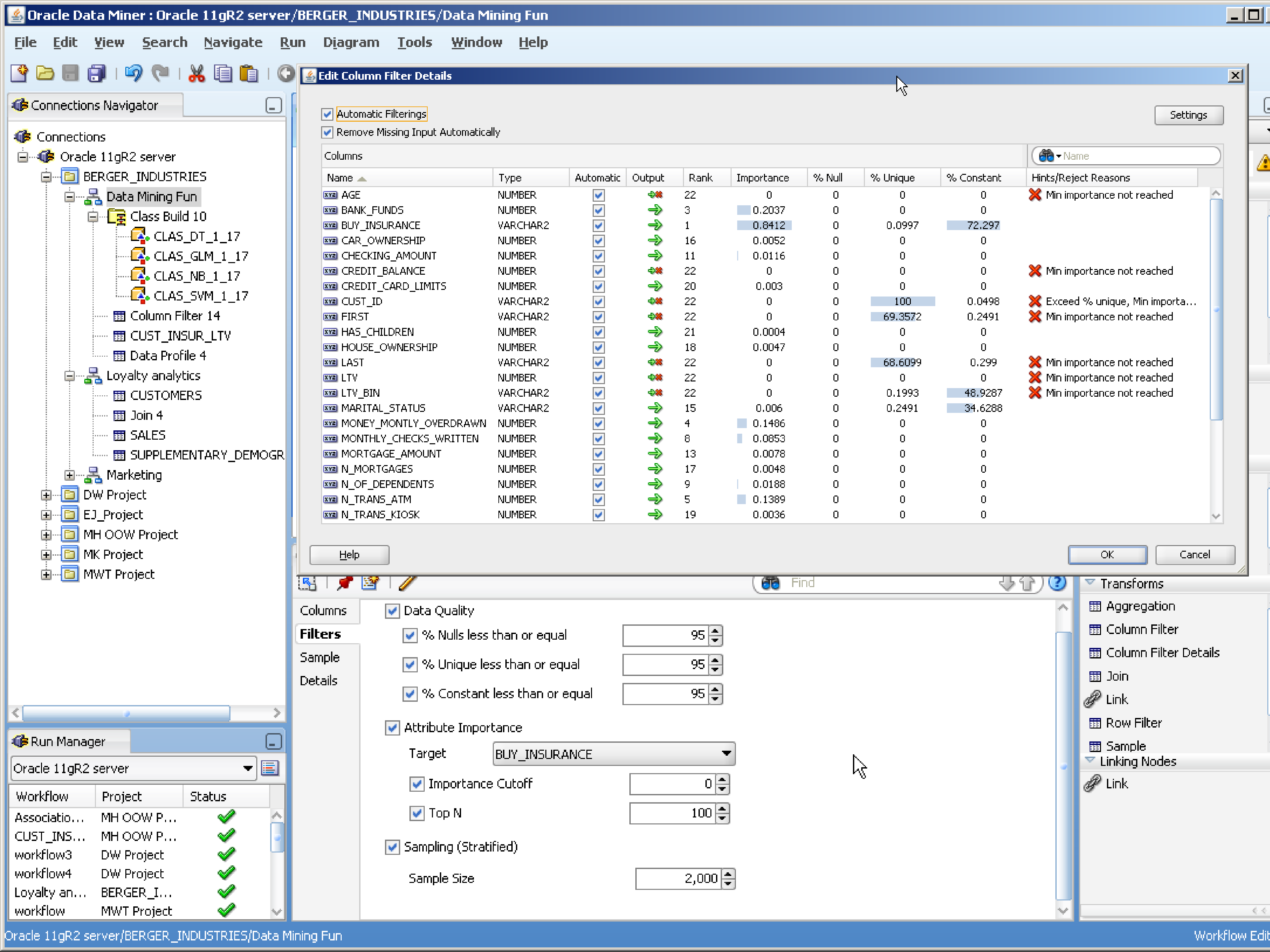

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Data Profile 4
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Run Manager
Oracle 11gR2 server
Workflow | Project | Status
Associatio... | MH OOW P...
CUST_INS... | MH OOW P...
workflow3 | DW Project
workflow4 | DW Project
Loyalty an... | BERGER_I...
workflow | MWT Project
Marketing
Data Mining Fun
Data Profile 4
Data Profile 2
Statistics
Group by: BUY_INSURANCE
Filter:
Name
Name | Histogram | Data Type | Percent NULLs | Distinct Values | Mode | Average
MARITAL_STATUS | VARCHAR2 | 0 | 5 | SINGLE
STATE | VARCHAR2 | 0 | 24 | NY
CREDIT_BALANCE | NUMBER | 0 | 208 | 2,776.59
TIME_AS_CUSTOMER | NUMBER | 0 | 5 | 2.4599
MORTGAGE_AMOUNT | NUMBER | 0 | 437 | 2,009.1137
BANK_FUNDS | NUMBER | 0 | 424 | 2,554.6334
N_OF_DEPENDENTS | NUMBER | 0 | 7 | 2.1127
HAS_CHILDREN | NUMBER | 0 | 2 | 0.4888
SALARY | NUMBER | 0 | 1,904 | 65,042.007
CUST_ID | VARCHAR2 | 0 | 2,005 | CU10002
SEX | VARCHAR2 | 0 | 2 | M
PROFESSION | VARCHAR2 | 0 | 99 | Programmer/Developer
REGION | VARCHAR2 | 0 | 5 | NorthEast
HOUSE_OWNERSHIP | NUMBER | 0 | 3 | 0.81
MARITAL_STATUS
100
80
60
40
20
0
Percent
Series: 'No'
Group: SINGLE
Value: 29.08
'Yes'
'No'
DIVORCED
MARRIED
OTHER
SINGLE
WIDOWED
Statistics
Data
Columns
SQL
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
dataminer.dprofi

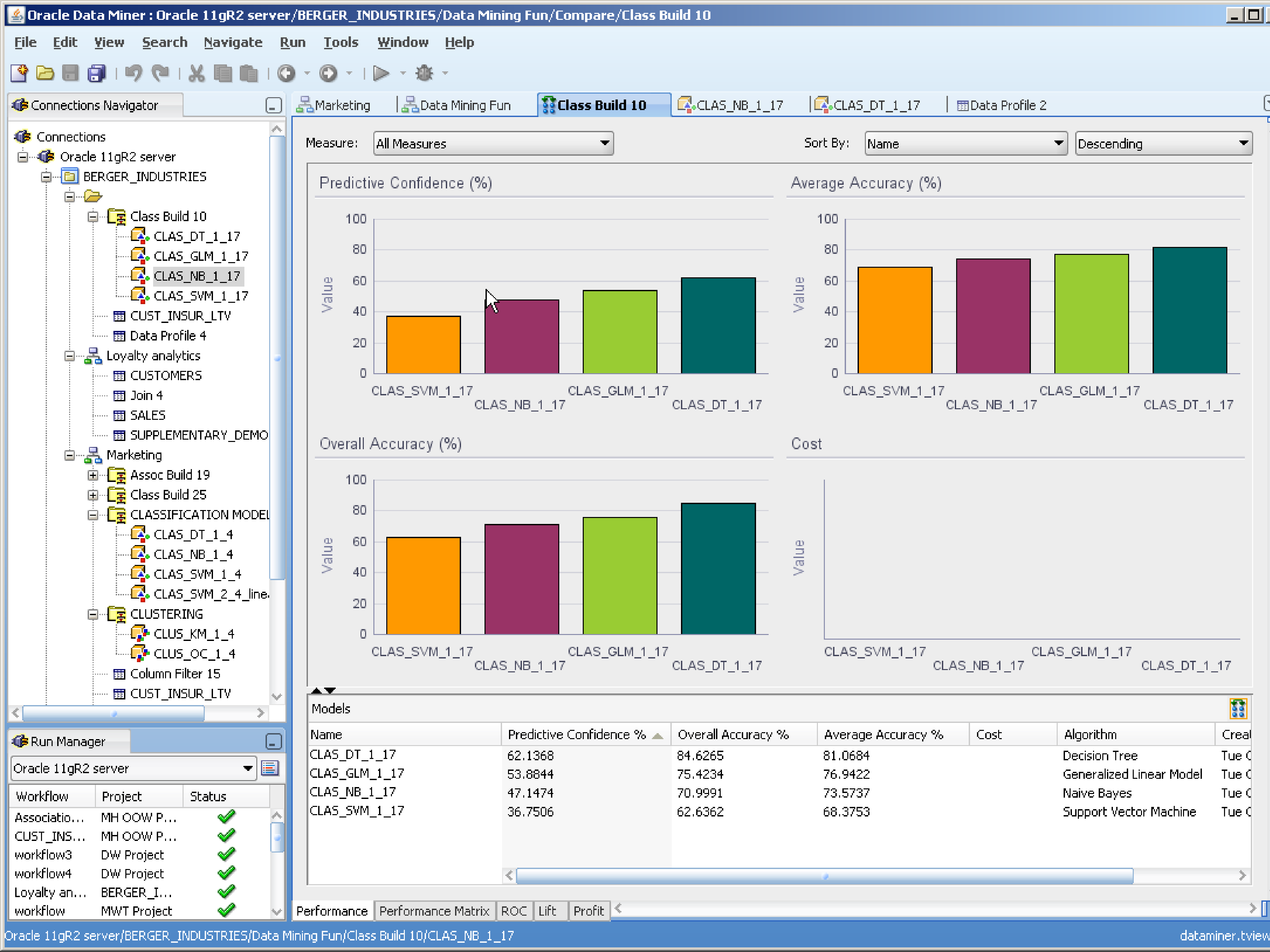

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Marketing
Data Mining Fun
Data Profile 4
Data Profile 2
100%
Data Profile 4
CUST_
INSUR_LTV
This node is fully defined and can be run
Workflow Editor
Component Palette
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Evalute and Apply
Apply
Link
Test
Data
Create Table
Data Profile
Data Source
Link
Transforms
Aggregation
Column Filter
Column Filter Details
Join
Link
Row Filter
Sample
Linking Nodes
Link
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Class Build 10 - Property Inspector
Find
Models
Build
Test
Details
Model Settings
Name
Name Build Test Tune Algorithm Comment
CLAS_GLM_... Not built Not tested Automatic Generalized Lin...
CLAS_SVM_... Not built Not tested Automatic Support Vector...
CLAS_DT_1... Not built Not tested Automatic Decision Tree
CLAS_NB_1... Not built Not tested Automatic Naive Bayes
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun

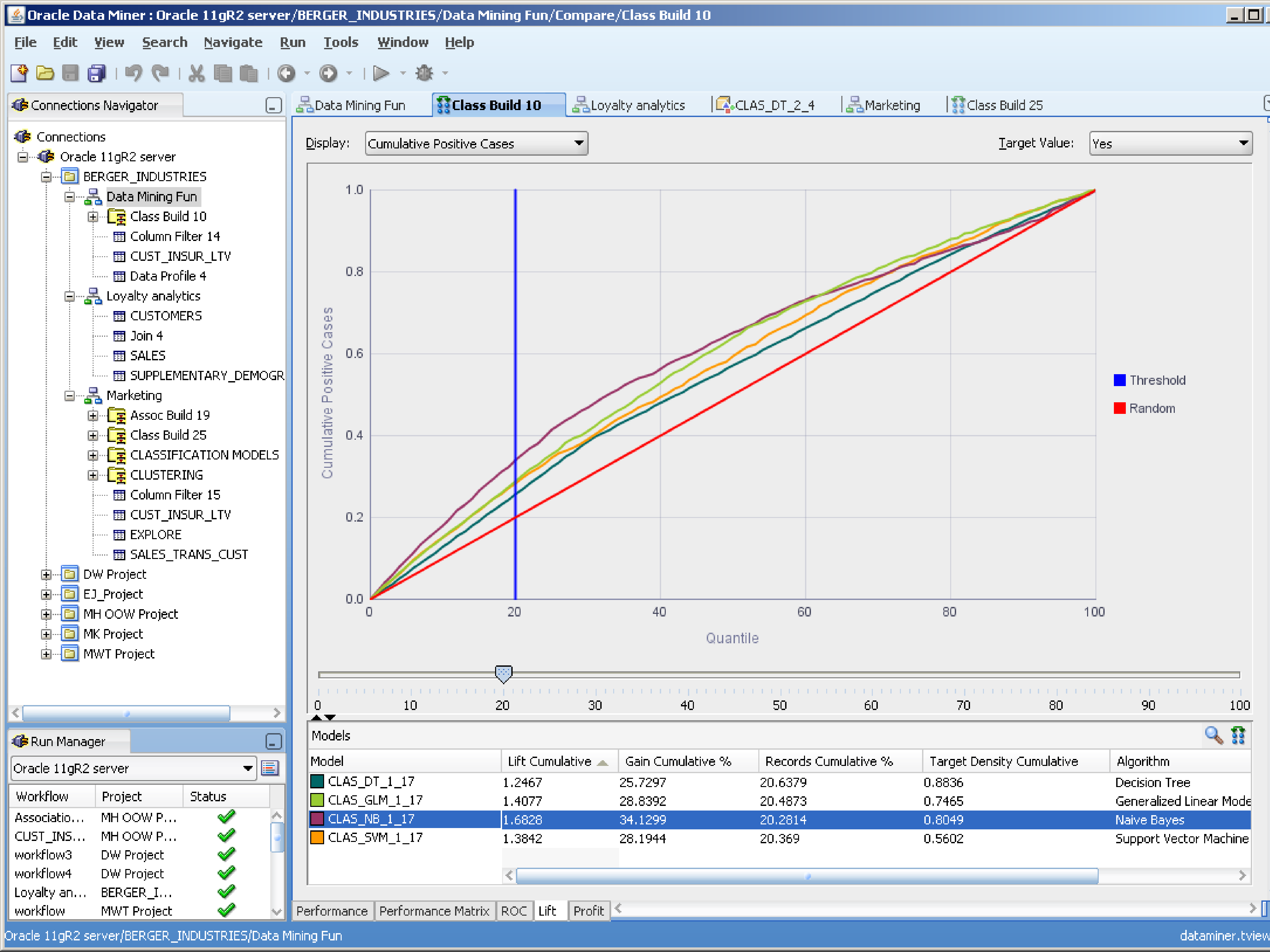

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Class Build 10/CLAS_DT_1_17
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Class Build 10
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Marketing
Data Mining Fun
CLAS_DT_1_17
Data Profile 2
75%
Maximum Target Values: 2
Node: 12
Predition: No
Support: 1713 (22.17%)
Confidence: 80.85%
No : 1385 (80.85%)
Yes : 328 (19.15%)
Split: MONEY_MONTLY_OVERDRAWN
Node: 16
Predition: No
Support: 371 (4.8%)
Confidence: 64.42%
No : 239 (64.42%)
Yes : 132 (35.58%)
Split: CHECKING_AMOUNT
Node: 13
Predition: No
Support: 1342 (17.37%)
Confidence: 85.39%
No : 1146 (85.39%)
Yes : 196 (14.61%)
Split: null
Node: 34
Predition: No
Support: 254 (3.29%)
Confidence: 77.95%
No : 198 (77.95%)
Yes : 56 (22.05%)
Node: 33
Predition: Yes
Support: 117 (1.51%)
Confidence: 64.96%
No : 41 (35.04%)
Yes : 76 (64.96%)
Node: 14
Predition: No
Support: 1214 (15.71%)
Confidence: 86.66%
No : 1052 (86.66%)
Yes : 162 (13.34%)
Split: MONEY_MONTLY_OVERDRAWN
Node: 32
Predition: No
Support: 128 (1.66%)
Confidence: 73.44%
No : 94 (73.44%)
Yes : 34 (26.56%)
Node: 9
Predition: Yes
Support: 1804 (23.35%)
Confidence: 73.56%
No : 477 (26.44%)
Yes : 1327 (73.56%)
Split: MONEY_MONTLY_OVE
Node: 31
Predition: No
Support: 627 (8.11%)
Node: 15
Predition: No
Support: 587 (7.6%)
Node: 11
Predition: Yes
Support: 722 (9.34%)
Rule Surrogates Target Values
Value Record Count % Percentage
Yes 76 64.95%
No 41 35.04%
Tree Settings
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
dataminer.svmclassificati

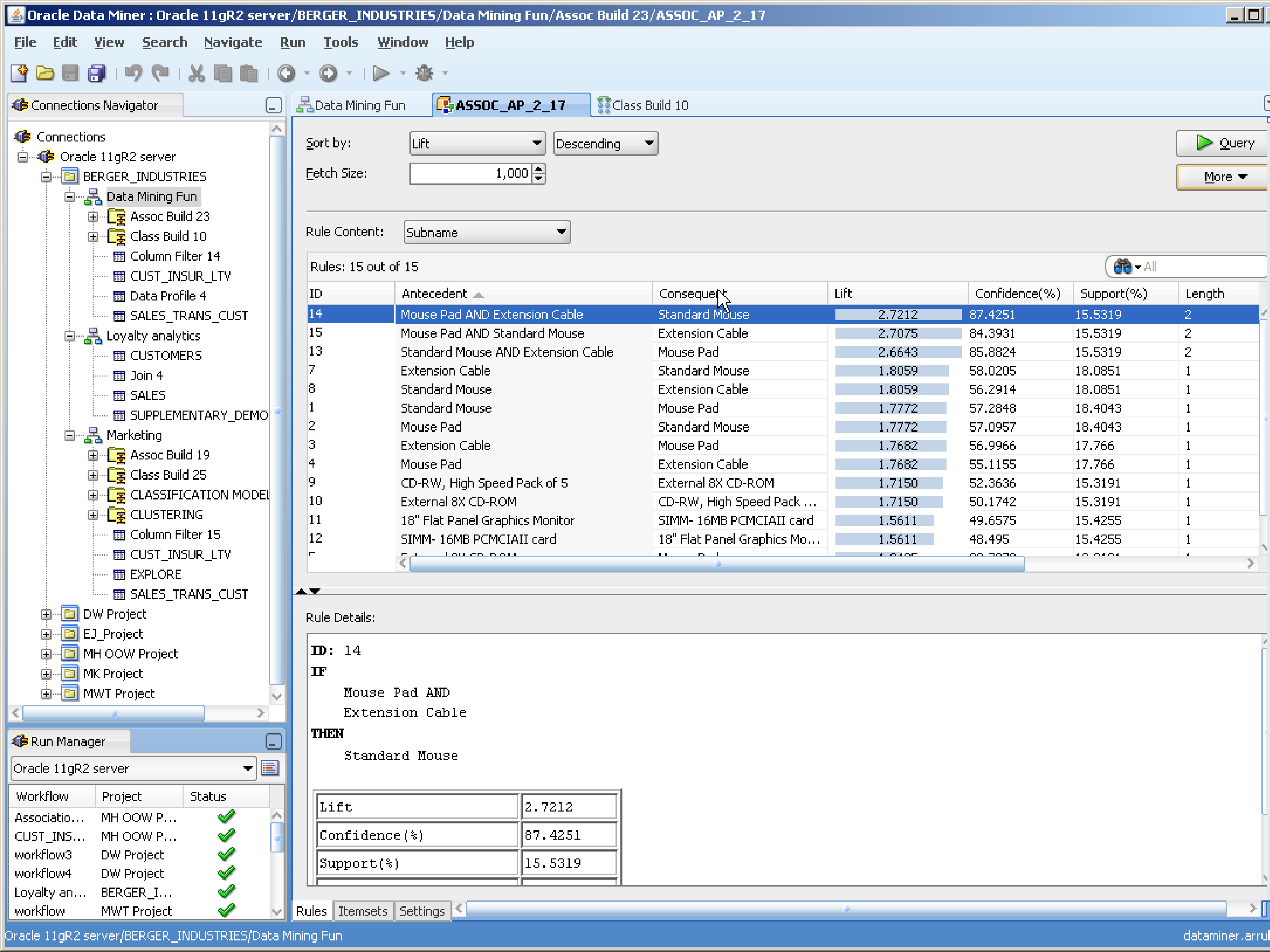

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Class Build 10/CLAS_NB_1_17
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Class Build 10
CLAS_DT_1_17
CLAS_GLM_1_17
CLAS_NB_1_17
CLAS_SVM_1_17
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Marketing
Data Mining Fun
CLAS_NB_1_17
CLAS_DT_1_17
Data Profile 2
Target Value: Yes
Query
Fetch Size: 10,000
Probabilties: 54 out of 54
Attribute
Attribute | Value | Probability(%) for Yes
LTV | (7040.375;) | 99.42829919
AGE | (18.5;) | 99.28537399
CAR_OWNERSHIP | 1 | 97.33206289
BANK_FUNDS | (271;) | 97.28442115
CREDIT_BALANCE | (; 42), [42; 42] | 95.75988566
N_OF_DEPENDENTS | (.5;) | 95.23582658
MONTHLY_CHECKS_WRITTEN | (1.5;) | 92.23439733
TIME_AS_CUSTOMER | 1, 3, 4, 5 | 92.13911386
N_TRANS_KIOSK | (; 4.5), [4.5; 4.5] | 89.09004288
CREDIT_CARD_LIMITS | (750;) | 86.08861363
HOUSE_OWNERSHIP | 1, 2 | 83.80181039
N_MORTGAGES | 1, 2 | 83.80181039
MARITAL_STATUS | 'DIVORCED', 'MARRIED', 'OTHER 1, 2 DOWED' | 75.22629824
N_TRANS_WEB_BANK | (.5; 2662] | 75.17865650
MORTGAGE_AMOUNT | (906;) | 68.31824678
CHECKING_AMOUNT | (; 25.5), [25.5; 25.5] | 64.74511672
SEX | M | 58.40876608
HAS_CHILDREN | 1 | 58.21819914
T_AMOUNT_AUTOM_PAYMENTS | (556.5; 7287] | 55.40733683
LTV_BIN | 'HIGH' | 52.92996665
<PRIOR> | <PRIOR> | 50.00000000
LTV_BIN | 'LOW', 'MEDIUM', 'VERY HIGH' | 47.07003335
N_TRANS_ATM | (4.5;) | 43.44926155
N_TRANS_TELLER | (2.5; 5.5] | 43.35397808
HAS_CHILDREN | 0 | 41.78180086
SEX | F | 41.59123392
MONEY_MONTLY_OVERDRAWN | (53.285; 54.145] | 40.97189138
MONEY_MONTLY_OVERDRAWN | (54.145;) | 38.06574559
MORTGAGE_AMOUNT | (; 906), [906; 906] | 31.68175322
N_TRANS_TELLER | (1.5; 2.5] | 29.58551691
N_TRANS_ATM | (3.5; 4.5] | 28.20390662
T_AMOUNT_AUTOM_PAYMENTS | (100.5; 489.5] | 27.63220581
N_TRANS_TELLER | (.5; 1.5] | 25.34540257
MARITAL_STATUS | 'SINGLE' | 24.77370176
Probabilities Compare Settings
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Class Build 10/CLAS_NB_1_17
dataminer.nbclassificati

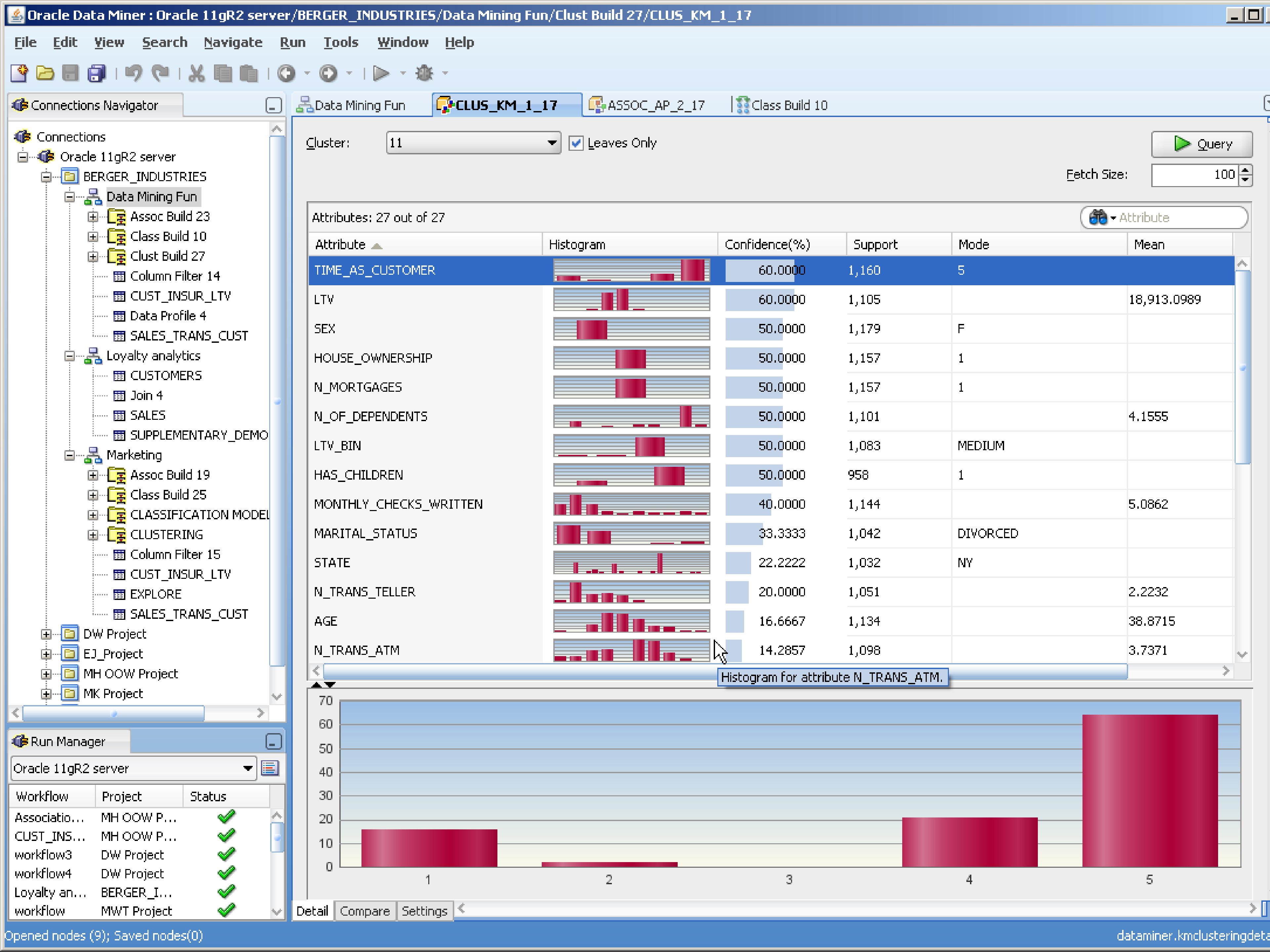

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Class Build 10
CLAS_DT_1_17
CLAS_GLM_1_17
CLAS_NB_1_17
CLAS_SVM_1_17
Column Filter 14
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMOGR
Marketing
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Data Mining Fun
Class Build 10
100%
Data Profile 4
CUST_INSUR_LTV
Class Build 10
Column Filter 14
Workflow Editor
Column Filter 14 - Property Inspector
Find
Columns
Filters
Sample
Details
Data Quality
% Nulls less than or equal 95
% Unique less than or equal 95
% Constant less than or equal 95
Attribute Importance
Target BUY_INSURANCE
Importance Cutoff 0
Top N 100
Sampling (Stratified)
Sample Size 2,000
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Component Palette
Workflow Editor
Models
Anomaly Detection
Association
Classification
Clustering
Feature Extraction
Link
Model
Evalute and Apply
Apply
Link
Test
Data
Create Table
Data Profile
Data Source
Link
Transforms
Aggregation
Column Filter
Column Filter Details
Join
Link
Row Filter
Sample
Linking Nodes
Link
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
Workflow Edit


Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
File Edit View Search Navigate Run Diagram Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Class Build 10
CLAS_DT_1_17
CLAS_GLM_1_17
CLAS_NB_1_17
CLAS_SVM_1_17
Column Filter 14
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMOGR
Marketing
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Edit Column Filter Details
Automatic Filterings
Remove Missing Input Automatically
Settings
Columns
Name
Name Type Automatic Output Rank Importance % Null % Unique % Constant Hints/Reject Reasons
AGE NUMBER 22 0 0 0 0 Min importance not reached
BANK_FUNDS NUMBER 3 0.2037 0 0 0
BUY_INSURANCE VARCHAR2 1 0.8412 0 0.0997 72.297
CAR_OWNERSHIP NUMBER 16 0.0052 0 0 0
CHECKING_AMOUNT NUMBER 11 0.0116 0 0 0
CREDIT_BALANCE NUMBER 22 0 0 0 0 Min importance not reached
CREDIT_CARD_LIMITS NUMBER 20 0.003 0 0 0
CUST_ID VARCHAR2 22 0 0 100 0.0498 Exceed % unique, Min importa...
FIRST VARCHAR2 22 0 0 69.3572 0.2491 Min importance not reached
HAS_CHILDREN NUMBER 21 0.0004 0 0 0
HOUSE_OWNERSHIP NUMBER 18 0.0047 0 0 0
LAST VARCHAR2 22 0 0 68.6099 0.299 Min importance not reached
LTV NUMBER 22 0 0 0 0 Min importance not reached
LTV_BIN VARCHAR2 22 0 0 0.1993 48.9287 Min importance not reached
MARITAL_STATUS VARCHAR2 15 0.006 0 0.2491 34.6288
MONEY_MONTLY_OVERDRAWN NUMBER 4 0.1486 0 0 0
MONTHLY_CHECKS_WRITTEN NUMBER 8 0.0853 0 0 0
MORTGAGE_AMOUNT NUMBER 13 0.0078 0 0 0
N_MORTGAGES NUMBER 17 0.0048 0 0 0
N_OF_DEPENDENTS NUMBER 9 0.0188 0 0 0
N_TRANS_ATM NUMBER 5 0.1389 0 0 0
N_TRANS_KIOSK NUMBER 19 0.0036 0 0 0
Help
OK
Cancel
Find
Columns
Filters
Sample
Details
Data Quality
% Nulls less than or equal 95
% Unique less than or equal 95
% Constant less than or equal 95
Attribute Importance
Target BUY_INSURANCE
Importance Cutoff 0
Top N 100
Sampling (Stratified)
Sample Size 2,000
Transforms
Aggregation
Column Filter
Column Filter Details
Join
Link
Row Filter
Sample
Linking Nodes
Link
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
Workflow Edit


Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Compare/Class Build 10
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Class Build 10
CLAS_DT_1_17
CLAS_GLM_1_17
CLAS_NB_1_17
CLAS_SVM_1_17
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLAS_DT_1_4
CLAS_NB_1_4
CLAS_SVM_1_4
CLAS_SVM_2_4_line
CLUSTERING
CLUS_KM_1_4
CLUS_OC_1_4
Column Filter 15
CUST_INSUR_LTV
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Marketing
Data Mining Fun
Class Build 10
CLAS_NB_1_17
CLAS_DT_1_17
Data Profile 2
Measure: All Measures
Sort By: Name
Descending
Predictive Confidence (%)
Average Accuracy (%)
Overall Accuracy (%)
Cost
Value
100
80
60
40
20
0
CLAS_SVM_1_17
CLAS_NB_1_17
CLAS_GLM_1_17
CLAS_DT_1_17
Models
Name Predictive Confidence % Overall Accuracy % Average Accuracy % Cost Algorithm Creat
CLAS_DT_1_17 62.1368 84.6265 81.0684 Decision Tree Tue
CLAS_GLM_1_17 53.8844 75.4234 76.9422 Generalized Linear Model Tue
CLAS_NB_1_17 47.1474 70.9991 73.5737 Naive Bayes Tue
CLAS_SVM_1_17 36.7506 62.6362 68.3753 Support Vector Machine Tue
Performance
Performance Matrix
ROC
Lift
Profit
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Class Build 10/CLAS_NB_1_17
dataminer.tview

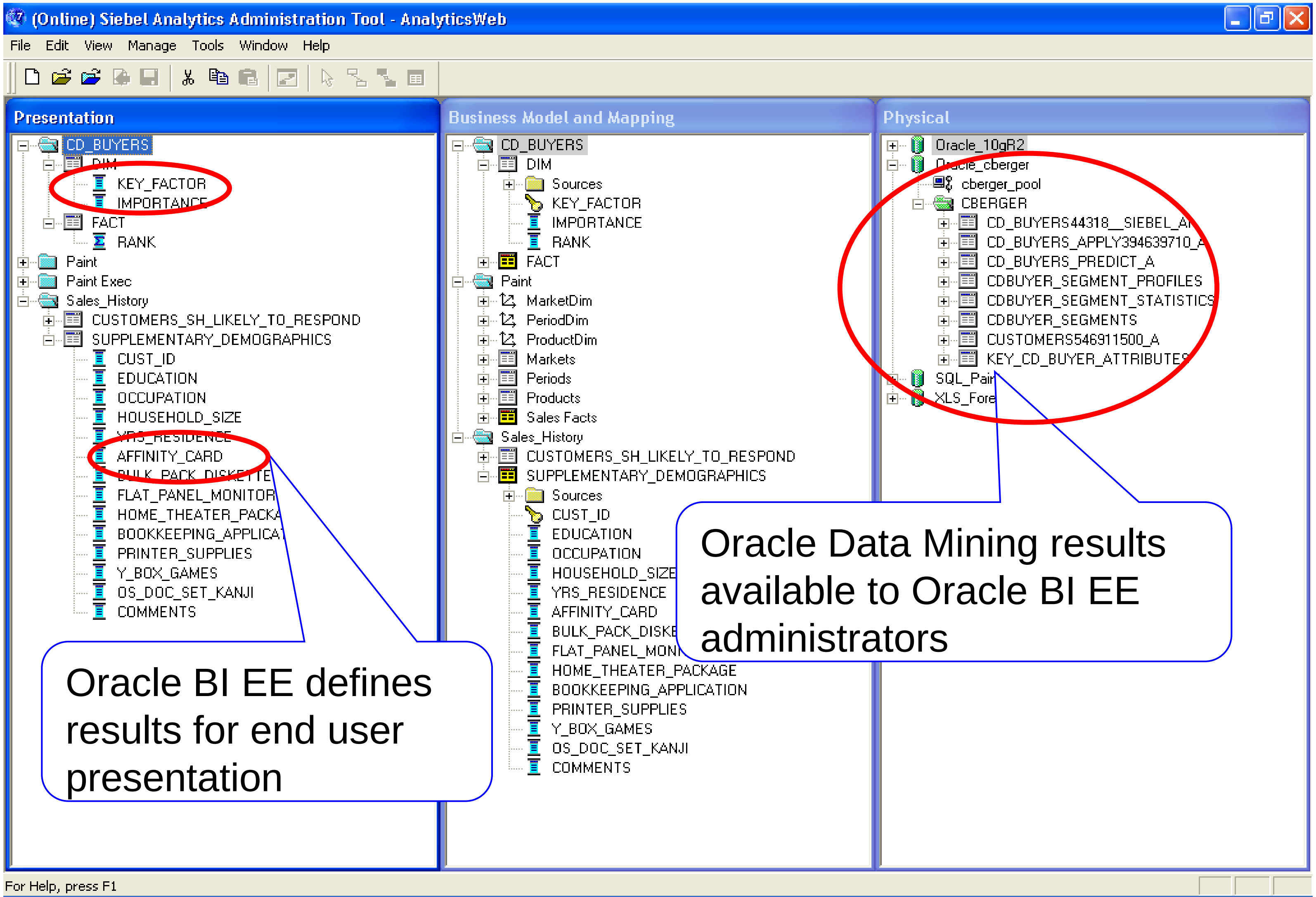

Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Compare/Class Build 10
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Class Build 10
Column Filter 14
CUST_INSUR_LTV
Data Profile 4
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMOGR
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODELS
CLUSTERING
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Data Mining Fun
Class Build 10
Loyalty analytics
CLAS_DT_2_4
Marketing
Class Build 25
Display: Cumulative Positive Cases
Target Value: Yes
Cumulative Positive Cases
Quantile
Threshold
Random
Models
Model Lift Cumulative Gain Cumulative % Records Cumulative % Target Density Cumulative Algorithm
CLAS_DT_1_17 1.2467 25.7297 20.6379 0.8836 Decision Tree
CLAS_GLM_1_17 1.4077 28.8392 20.4873 0.7465 Generalized Linear Mode
CLAS_NB_1_17 1.6828 34.1299 20.2814 0.8049 Naive Bayes
CLAS_SVM_1_17 1.3842 28.1944 20.369 0.5602 Support Vector Machine
Performance
Performance Matrix
ROC
Lift
Profit
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
dataminer.tview


Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Assoc Build 23/ASSOC_AP_2_17
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Assoc Build 23
Class Build 10
Column Filter 14
CUST_INSUR_LTV
Data Profile 4
SALES_TRANS_CUST
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLUSTERING
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
MWT Project
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Data Mining Fun
ASSOC_AP_2_17
Class Build 10
Sort by: Lift Descending
Query
Fetch Size: 1,000
More
Rule Content: Subname
Rules: 15 out of 15
All
ID Antecedent Consequent Lift Confidence(%) Support(%) Length
14 Mouse Pad AND Extension Cable Standard Mouse 2.7212 87.4251 15.5319 2
15 Mouse Pad AND Standard Mouse Extension Cable 2.7075 84.3931 15.5319 2
13 Standard Mouse AND Extension Cable Mouse Pad 2.6643 85.8824 15.5319 2
7 Extension Cable Standard Mouse 1.8059 58.0205 18.0851 1
8 Standard Mouse Extension Cable 1.8059 56.2914 18.0851 1
1 Standard Mouse Mouse Pad 1.7772 57.2848 18.4043 1
2 Mouse Pad Standard Mouse 1.7772 57.0957 18.4043 1
3 Extension Cable Mouse Pad 1.7682 56.9966 17.766 1
4 Mouse Pad Extension Cable 1.7682 55.1155 17.766 1
9 CD-RW, High Speed Pack of 5 External 8X CD-ROM 1.7150 52.3636 15.3191 1
10 External 8X CD-ROM CD-RW, High Speed Pack ... 1.7150 50.1742 15.3191 1
11 18" Flat Panel Graphics Monitor SIMM- 16MB PCMCIAII card 1.5611 49.6575 15.4255 1
12 SIMM- 16MB PCMCIAII card 18" Flat Panel Graphics Mo... 1.5611 48.495 15.4255 1
Rule Details:
ID: 14
IF
Mouse Pad AND
Extension Cable
THEN
Standard Mouse
Lift 2.7212
Confidence(%) 87.4251
Support(%) 15.5319
Rules Itemsets Settings
Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun
dataminer.arrul


Oracle Data Miner : Oracle 11gR2 server/BERGER_INDUSTRIES/Data Mining Fun/Clust Build 27/CLUS_KM_1_17
File Edit View Search Navigate Run Tools Window Help
Connections Navigator
Connections
Oracle 11gR2 server
BERGER_INDUSTRIES
Data Mining Fun
Assoc Build 23
Class Build 10
Clust Build 27
Column Filter 14
CUST_INSUR_LTV
Data Profile 4
SALES_TRANS_CUST
Loyalty analytics
CUSTOMERS
Join 4
SALES
SUPPLEMENTARY_DEMO
Marketing
Assoc Build 19
Class Build 25
CLASSIFICATION MODEL
CLUSTERING
Column Filter 15
CUST_INSUR_LTV
EXPLORE
SALES_TRANS_CUST
DW Project
EJ_Project
MH OOW Project
MK Project
Run Manager
Oracle 11gR2 server
Workflow Project Status
Associatio... MH OOW P...
CUST_INS... MH OOW P...
workflow3 DW Project
workflow4 DW Project
Loyalty an... BERGER_I...
workflow MWT Project
Data Mining Fun
CLUS_KM_1_17
ASSOC_AP_2_17
Class Build 10
Cluster: 11
Leaves Only
Query
Fetch Size: 100
Attributes: 27 out of 27
Attribute
Attribute Histogram Confidence(%) Support Mode Mean
TIME_AS_CUSTOMER 60.0000 1,160 5
LTV 60.0000 1,105 18,913.0989
SEX 50.0000 1,179 F
HOUSE_OWNERSHIP 50.0000 1,157 1
N_MORTGAGES 50.0000 1,157 1
N_OF_DEPENDENTS 50.0000 1,101 4.1555
LTV_BIN 50.0000 1,083 MEDIUM
HAS_CHILDREN 50.0000 958 1
MONTHLY_CHECKS_WRITTEN 40.0000 1,144 5.0862
MARITAL_STATUS 33.3333 1,042 DIVORCED
STATE 22.2222 1,032 NY
N_TRANS_TELLER 20.0000 1,051 2.2232
AGE 16.6667 1,134 38.8715
N_TRANS_ATM 14.2857 1,098 3.7371
Histogram for attribute N_TRANS_ATM.
70
60
50
40
30
20
10
0
1
2
3
4
5
Detail Compare Settings
Opened nodes (9); Saved nodes(0)
dataminer.kmclusteringdeta

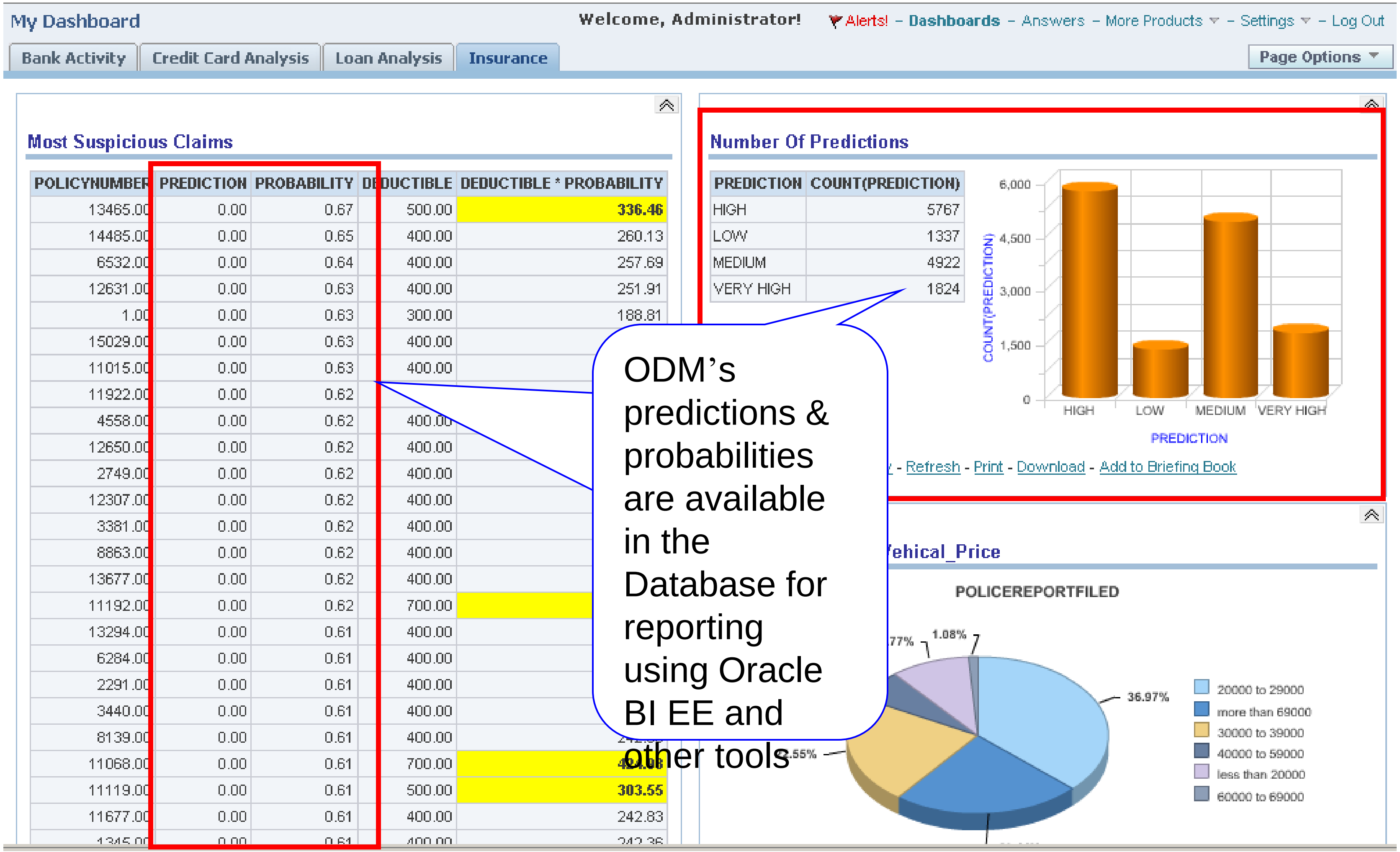
Presentation of Results and Integration with Applications

Integration with Oracle BI EE



Example

Better Information for OBI EE Reports and Dashboards



ODM's predictions & probabilities are available in the Database for reporting using Oracle BI EE and other tools

ORACLE

Integration with Oracle BI EE

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

http://localhost/analytics/saw.dll?Dashboard

ORACLE BUSINESS INTELLIGENCE | Intelligence Dashboards | My Dashboard

My Dashboard

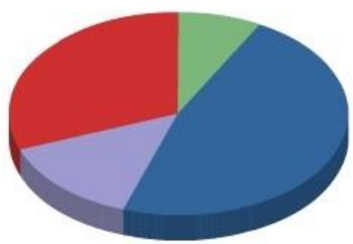
Welcome, John Smith! Dashboards - Answers - Advanced Report

Organization Analysis | Category Analysis

Pick Any Time

Go

Potential Fraud by Organization



Org Level 2	Potential Fraud Cost
Midwest Region	2,896
Northeast Region	17,307
Southern Region	5,086
Western Region	11,252

Modify - Download

Most Suspicious Activities

Employee	Item	Day	Amount	Probability	Potential Fraud Cost
Louis Nagode	Misc. Employee Expenses	31-Dec-2003	15,740	59	9,265
Paul Laker	Misc. Employee Expenses	17-Dec-2003	4,996	56	2,792
Louis Nagode	Misc. Employee Expenses	30-Dec-2003	4,259	60	2,537
Dave Lindquist	Misc. Employee Expenses	01-Jan-2004	2,253	63	1,422
Steven Daniel	Hotel-Lodging	14-Dec-2004	2,304	52	1,205
Paul Laker	Hotel-Lodging	19-Dec-2004	2,219	54	1,201
Steven Daniel	Hotel-Lodging	22-Dec-2004	1,896	52	979
Christina Donohue	Hotel-Lodging	21-Dec-2004	1,744	53	919
Michael Cheng	Hotel-Lodging	21-Dec-2004	1,598	53	842
Dennis Haas	Hotel-Lodging	14-Dec-2004	1,539	52	805

Results

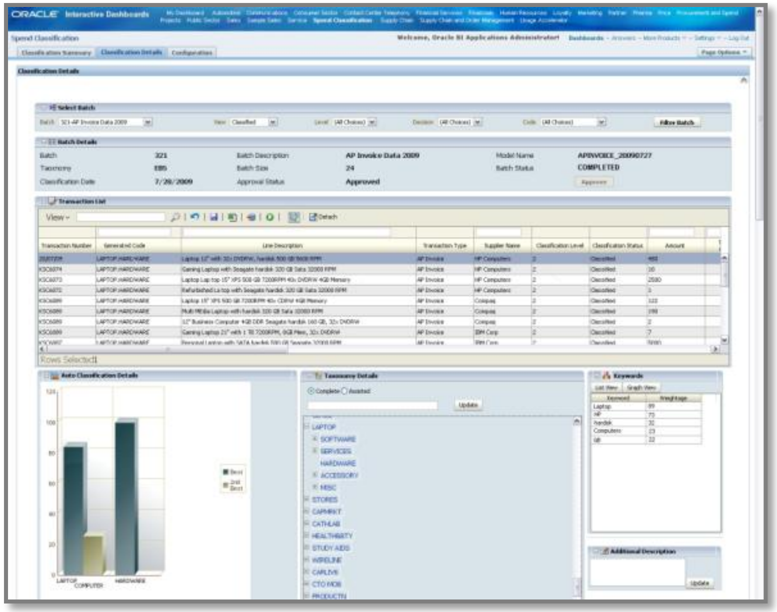
Modify - Download

Trends by Organization

Org Level 2	Quarter	Amount	Amt Parent Share	Amt % Chg Prior Per	Potential Fraud Cost	PFC Parent Share	PFC % Chg Prior Per
Midwest Region	Q1 2004	12,173	14	-9	1,056	17	37
	Q2 2004	6,528	14	-46	0		-100
	Q3 2004	7,427	13	14	0	0	
	Q4 2004	15,642	8	111	1,841	6	
Northeast Region	Q1 2004	28,182	32	-11	537	9	-75
	Q2 2004	14,986	33	-47	0		-100
	Q3 2004	21,287	38	42	742	100	
	Q4 2004	95,027	49	346	16,028	54	2,061
Southern Region	Q1 2004	15,773	18	13	2,003	33	155
	Q2 2004	7,458	16	-53	0		-100
	Q3 2004	8,674	15	16	0	0	
	Q4 2004	27,154	14	213	3,083	10	
Western Region	Q1 2004	30,909	36	-45	2,487	41	-84
	Q2 2004	16,883	37	-45	0		-100
	Q3 2004	18,612	33	10	0	0	

Oracle Spend Classification

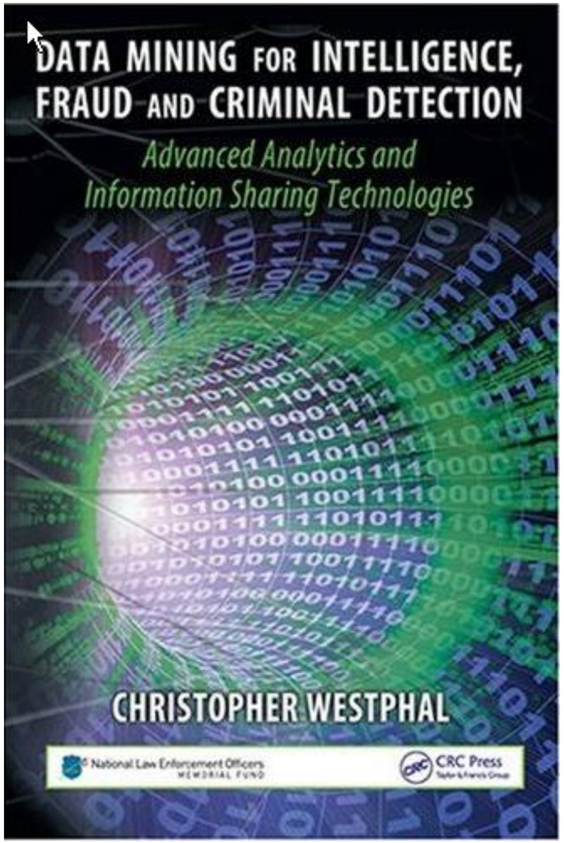
Classify Spend into Purchasing Categories



- 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

Further Reading

- Oracle Data Mining Documentation
- Oracle Data Mining Sample PL/SQL and Java Code Examples
- ODM Consultants
- OTN Discussion Forum
- Numerous Books and Papers e.g. ….



More Information:

Oracle Data Mining 11g

- oracle.com/technology/products/bi/odm/index.html

Oracle Statistical Functions

- http://www.oracle.com/technology/products/bi/stats_fns/index.html

Oracle Business Intelligence Solutions

- oracle.com/bi

http://search.oracle.com

oracle data mining

Contact Information: Email: Charlie.berger@oracle.com

ORACLE IS THE INFORMATION COMPANY

Q&A

QUESTIONS

ANSWERS