

BANK MARKETING DATA ANALYSIS

Instructor:
Professor Soon Ae Chun

Subject Name: BDA761
Big Data Management in a Supercomputing
Environment

Date : 10th Dec 2015

Student Name: Eun Jin Kwak

BANK DATA ?

- Basic & useful information for various business field
- To predict future client with high possibility
- Prioritizing and selecting the next customers to be contacted for future marketing
- Minimize the cost, and time saving for the business perspective
- Maximize the profit from the marketing result



DATA SUMMARY

- Data Source: UCI Machine Learning Repository
<http://archive.ics.uci.edu/ml/>
- Data Period: From May 2008 to June 2013, in a total of 52,944 phone contracts from Portuguese banking institutions
- Data Characteristic: Classification
- Data Management & Visualization Tools: R, RapidMiner
- Data Modeling: Decision Tree , Neural Net

DATA INFORMATION

- No of Observations: 41,188
- Input Variable: 20 variables with 3 categories
- 1) Bank client data_7 variables: Age, Job, Marital Status, Education, Default, Housing Loan, Personal Loan
- 2) Related with the last contact to the current campaign Contact_8 variables: Contact Type, Contacted Month, Contacted Day of Week, Campaign Duration, No of Contacted, Passed days after the last contact, No of Previous contact, Outcome from previous campaign
- 3) Social and economic context attributes_5 variables: Employment Variation Rate, Consumer Price Index, Consumer Confidence Index, Euribor 3 Month, Number of Employees
- Output variable: Has the client subscribed a Term deposit? Yes, No

DATA FORMAT

age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp.var.r	cons.price	cons.conf	euribor3m	nr.employed	y
56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
57	services	married	high.scho	unknown	no	no	telephone	may	mon	149	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
37	services	married	high.scho	no	yes	no	telephone	may	mon	226	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
56	services	married	high.scho	no	no	yes	telephone	may	mon	307	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
45	services	married	basic.9y	unknown	no	no	telephone	may	mon	198	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
59	admin.	married	profession	no	no	no	telephone	may	mon	139	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
24	technician	single	profession	no	yes	no	telephone	may	mon	380	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
25	services	single	high.scho	no	yes	no	telephone	may	mon	50	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	55	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
25	services	single	high.scho	no	yes	no	telephone	may	mon	222	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
29	blue-collar	single	high.scho	no	no	yes	telephone	may	mon	137	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no

DATA ANALYSIS

1. POLYNOMIAL REGRESSION

Result Overview PolynomialRegression (Polynomial Regression)

PolynomialRegression

Description

Annotation

```

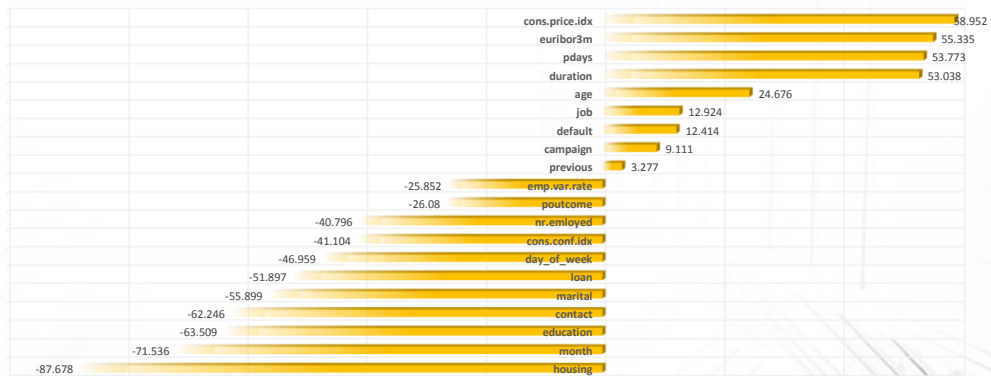
24.676 * age ^ 4.000
+ 12.924 * job ^ 1.000
- 55.899 * marital ^ 3.000
- 63.509 * education ^ 5.000
+ 12.414 * default ^ 5.000
- 87.678 * housing ^ 4.000
- 51.897 * loan ^ 3.000
- 62.246 * contact ^ 1.000
- 71.536 * month ^ 1.000
- 46.959 * day_of_week ^ 2.000
+ 53.038 * duration ^ 3.000
+ 9.111 * campaign ^ 3.000
+ 53.773 * pdays ^ 5.000
+ 3.277 * previous ^ 4.000
- 26.080 * poutcome ^ 3.000
- 25.852 * emp.var.rate ^ 1.000
+ 58.952 * cons.price.idx ^ 3.000
- 41.104 * cons.conf.idx ^ 4.000
+ 55.335 * euribor3m ^ 4.000
- 40.796 * nr.employed ^ 1.000
+ 96.773

```

DATA ANALYSIS

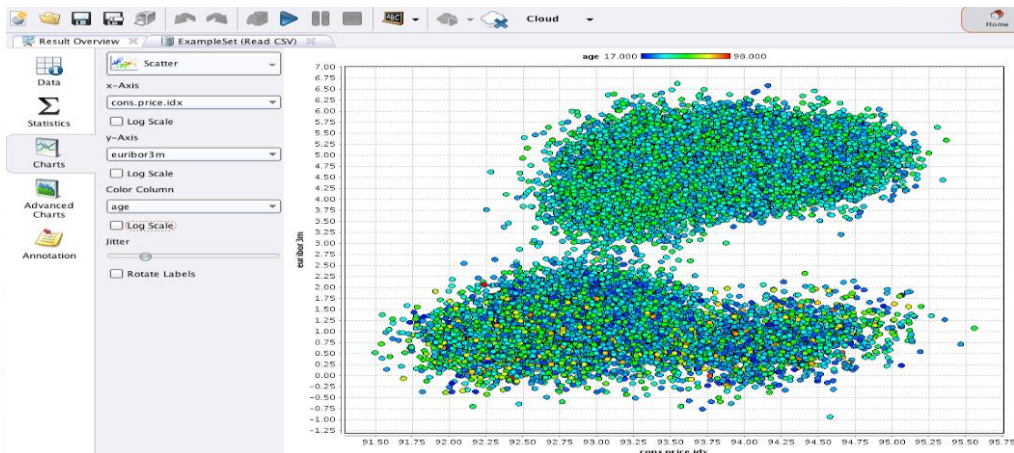
1-1. COEFFICIENT ANALYSIS

INDEPENDENT VARIABLES vs DEPENDENT VARIABLES (TERM DEPOSIT)

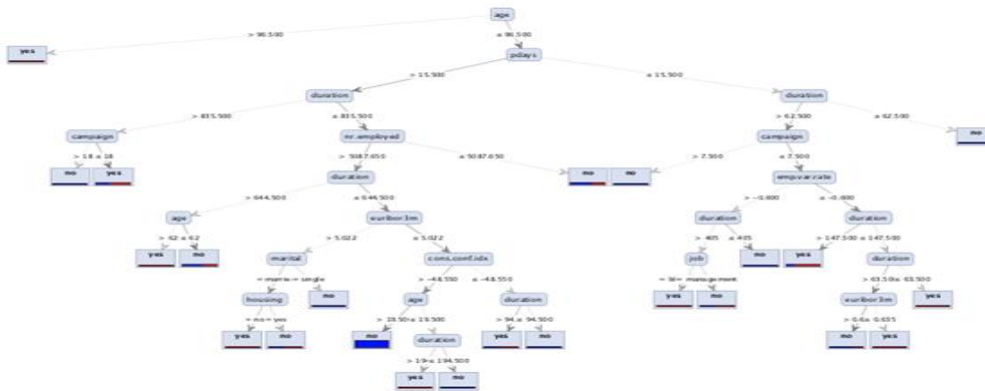


DATA ANALYSIS

1-2. CLUSTER ANALYSIS with 3 variables on Positive-relation

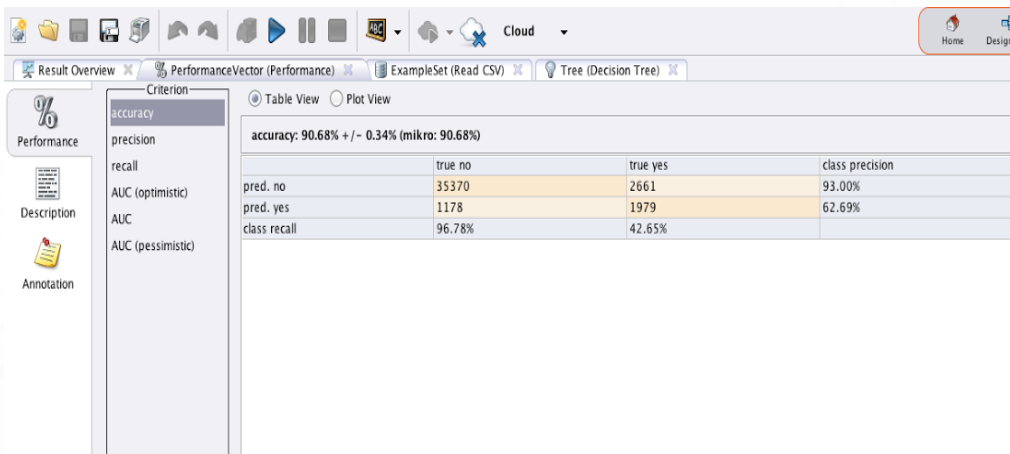


2. DECISION TREE



DATA ANALYSIS

2-1. A Cross-Validation Evaluating Decision Tree Model (Accuracy : 90.68%)



The screenshot shows the Orange3 data mining software interface. The 'PerformanceVector (Performance)' widget is active, displaying a table of performance metrics for a Decision Tree model. The table shows the following data:

	true no	true yes	class precision
pred. no	35370	2661	93.00%
pred. yes	1178	1979	62.69%
class recall	96.78%	42.65%	

The overall accuracy is 90.68% +/- 0.34% (micro: 90.68%). The 'Tree (Decision Tree)' widget is also visible in the background, showing a decision tree structure.

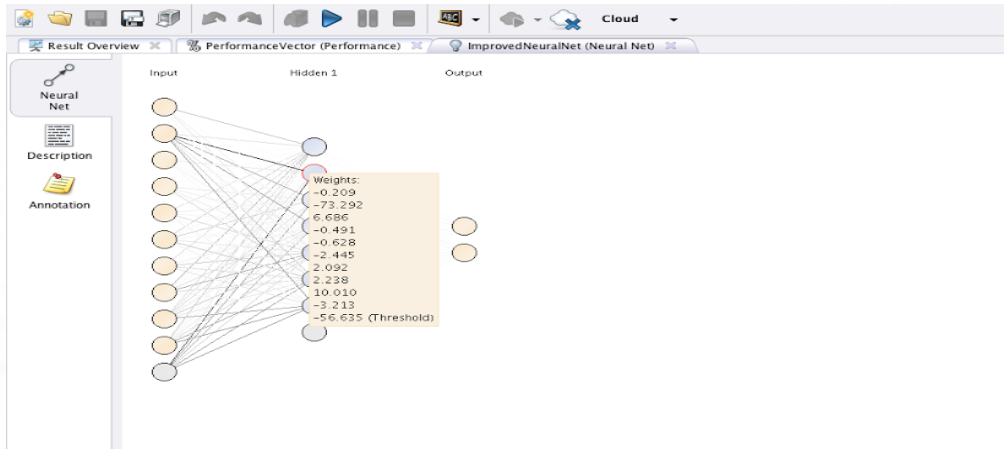
The screenshot shows the Orange3 software interface. The top toolbar includes icons for file operations, data loading, and model execution. The main window has several tabs: 'Result Overview', 'PerformanceVector (Performance)', 'ExampleSet (Read CSV)', and 'Tree (Decision Tree)'. The 'PerformanceVector (Performance)' tab is active, showing a table of performance metrics for a Decision Tree model. The 'Criterion' tab is selected, displaying the following data:

Criterion			
accuracy: 90.68% +/- 0.34% (mikro: 90.68%)			
recall	true no	true yes	class precision
AUC (optimistic)	35370	2661	93.00%
AUC	1178	1979	62.69%
AUC (pessimistic)	96.78%	42.65%	

accuracy: 90.68% +/- 0.34% (mikro: 90.68%)			
	true no	true yes	class precision
pred. no	35370	2661	93.00%
pred. yes	1178	1979	62.69%
class recall	96.78%	42.65%	

DATA ANALYSIS

3. NEURAL NET



DATA ANALYSIS

3-1. A Cross-Validation Evaluating Neural Net Model (Accuracy : 91.08%)

Criterion: ☒ Table View ☐ Plot View

accuracy: 91.08% +/- 0.26% (mikro: 91.08%)

	true no	true yes	class precision
pred. no	35165	2290	93.89%
pred. yes	1383	2350	62.95%
class recall	96.22%	50.65%	

FUTURE DIRECTION

- Comprehensive Analysis on various marketing methods; Internet, Banner, E-mail, Social Media, Text message, News Paper, Commercial, etc
- Detailed & Specified Data ; Contacted Time, Location of Banner, Length or Size of Commercial, Design Type of Commercials, etc
- Expended Attributes on Social Contexts and Economic Indicator; Foreign Exchange rate, Producer Price Index, Stock Market Index, etc.

Thanks. 😊