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

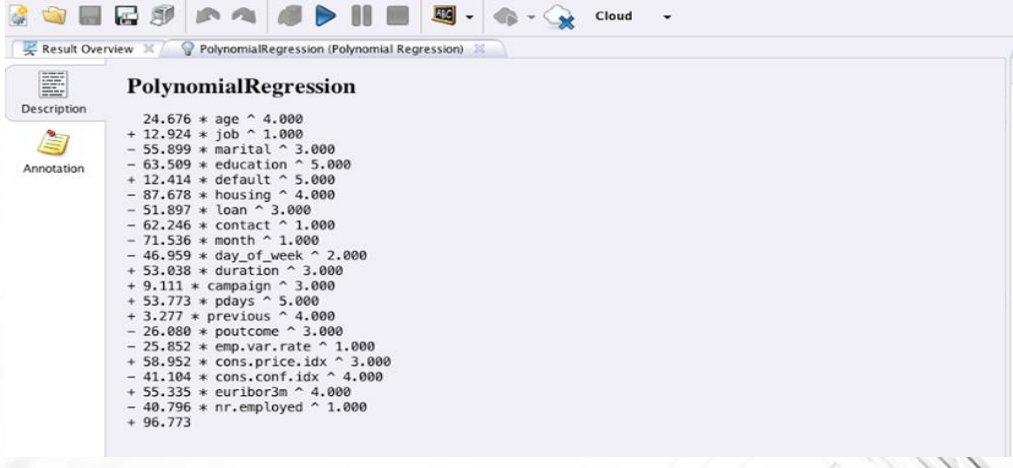
- Output variable: Has the client subscribed a Term deposit? Yes, No

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.employ	y
56	housemai	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	professio	no	no	no	telephone	may	mon	139	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
41	blue-colla	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexiste	1.1	93.994	-36.4	4.857	5191	no
24	technicia	single	professio	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-colla	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-colla	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

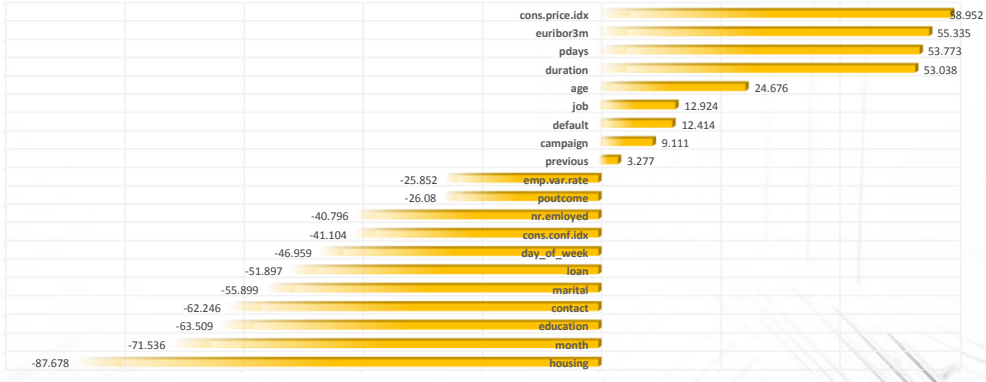
1. POLYNOMIAL REGRESSION



DATA ANALYSIS

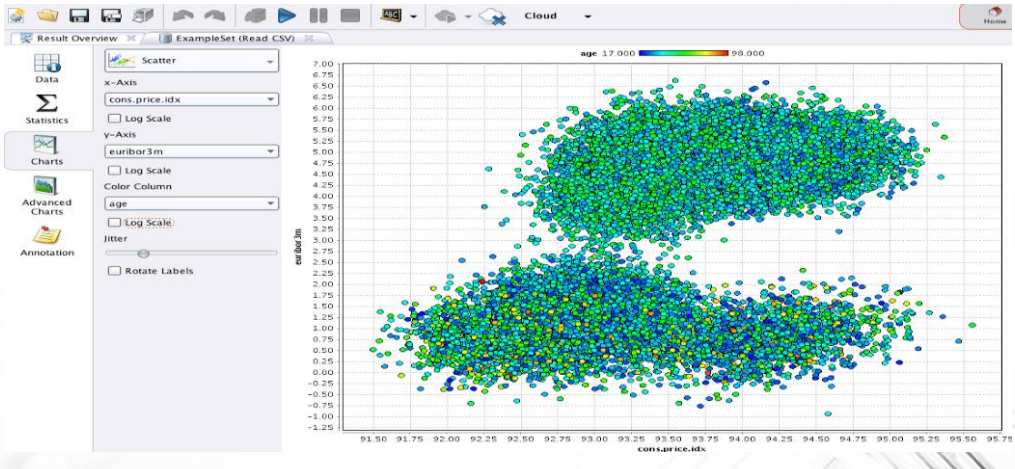
1-1. COEFFICIENT ANALYSIS

INDEPENDENT VARIABLES vs DEPENDENT VARIABLES (TERM DEPOSIT)



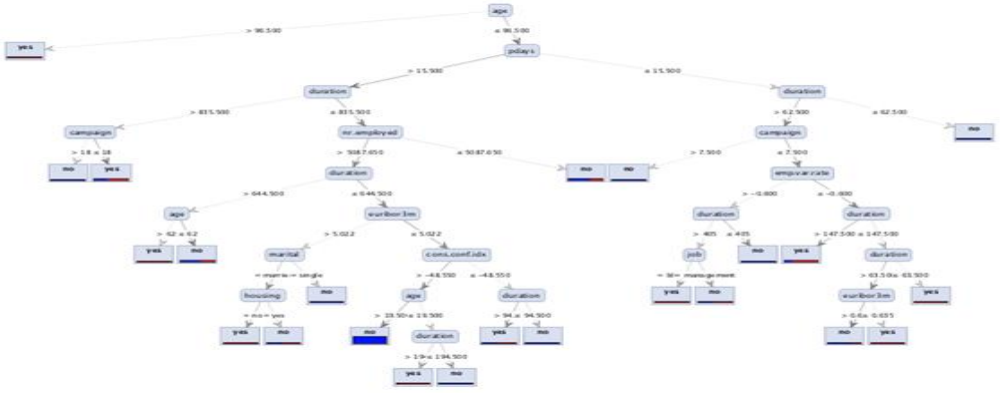
DATA ANALYSIS

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



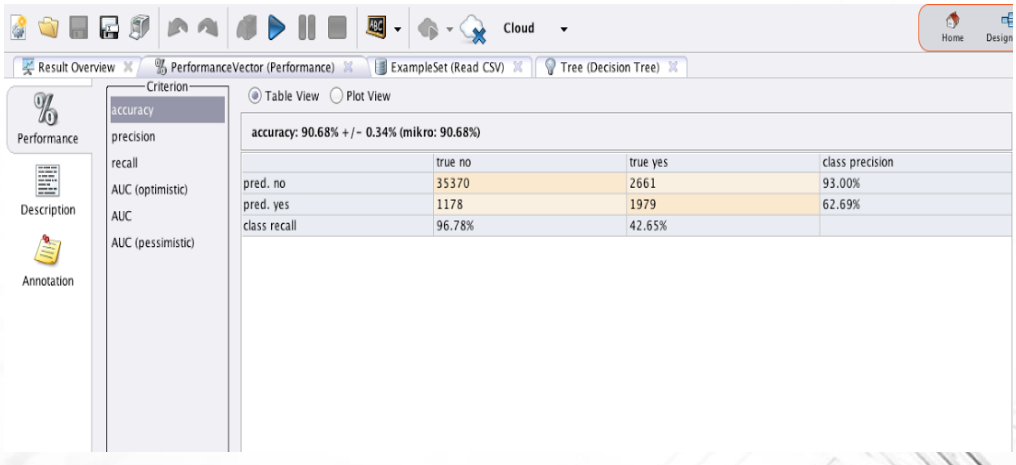
DATA ANALYSIS

2. DECISION TREE



DATA ANALYSIS

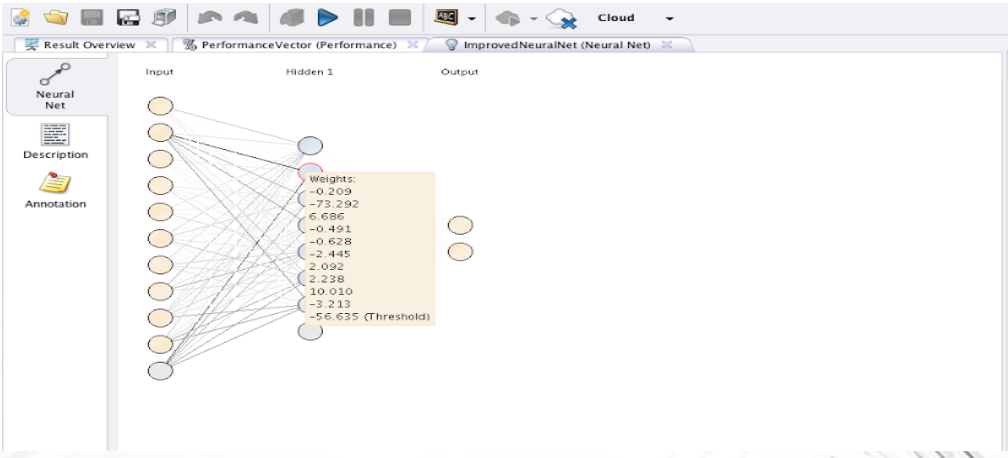
2-1. A Cross-Validation Evaluating Decision Tree Model (Accuracy : 90.68%)



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

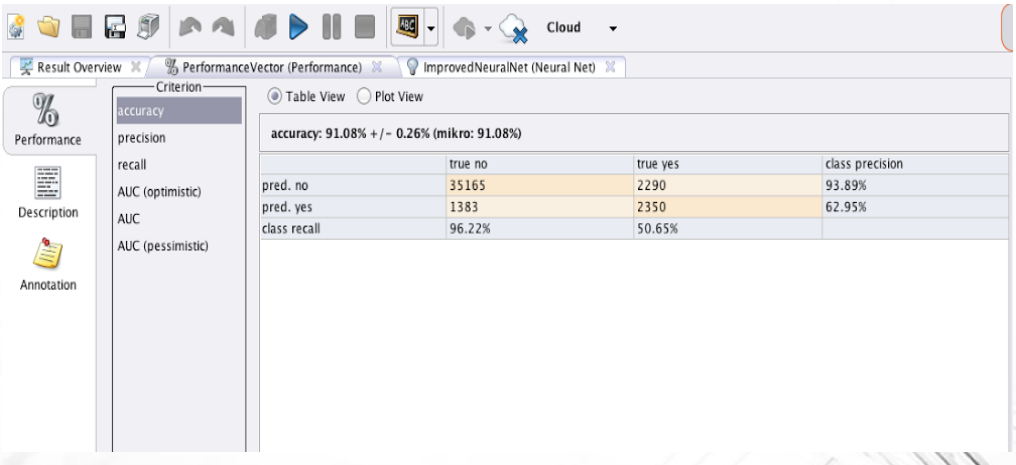
DATA ANALYSIS

3. NEURAL NET



DATA ANALYSIS

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



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. ☺