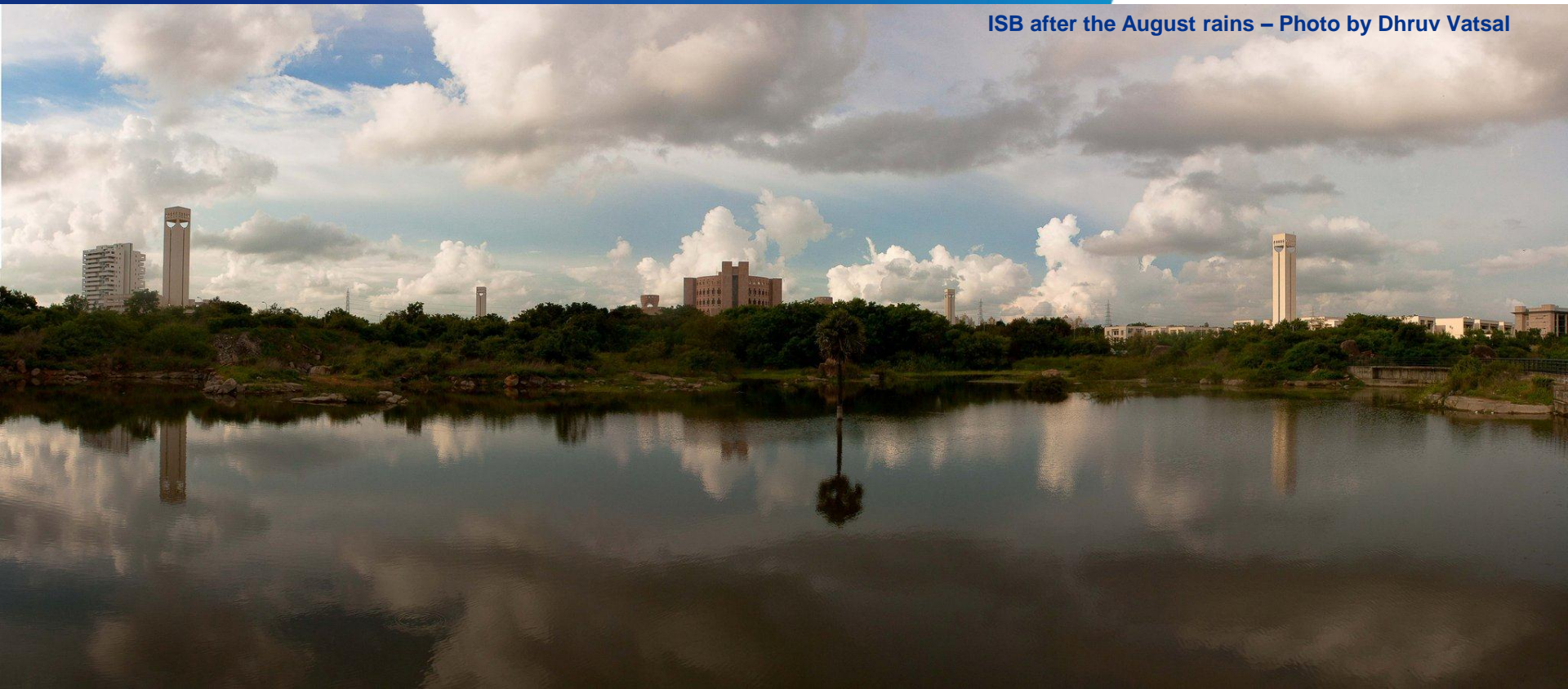


Predicting the pricing methodology for outsourcing contracts



28th December 2011

ISB after the August rains – Photo by Dhruv Vatsal



Agenda

1 Introduction – Problem

2 Methodology

3 Performance

4 Take-aways

Outsourcing contracts – themes and trends

From outsourcing for cost-arbitrage to strategic outsourcing

More deals each year

New work. New vendors/capabilities.

Pre-contract activities/spends increasing

Business Imperative – reduce pre-contract spends

Work and results – in brief

Database

- 22040 records 😊
- 50 fields 😊
- Good quality 😊
- Good info 😊
- Definitions document 😊 😊 😊
- A lot of info 😞

Preprocessing

- Removal of sparse data-columns.
- Remove textual, univariate, list data.
- Mostly ordinal data (~ 30 fields)
- Logarithmic Scale and Binning on continuous data.

Data Preparation

- Select data (remove unknown price-methodology – 25 % rows)
- Select attributes (remove attributes have low information or are collinear)
- Decide on use of 'confidence value' attribute
- Use Pricing methodology for classification – binary variable.

Sampling and Classification Model

- Sample into Train and Hold-Out
- *Cross-validation* runs on various classification methods
- ROC/Confusion matrix.

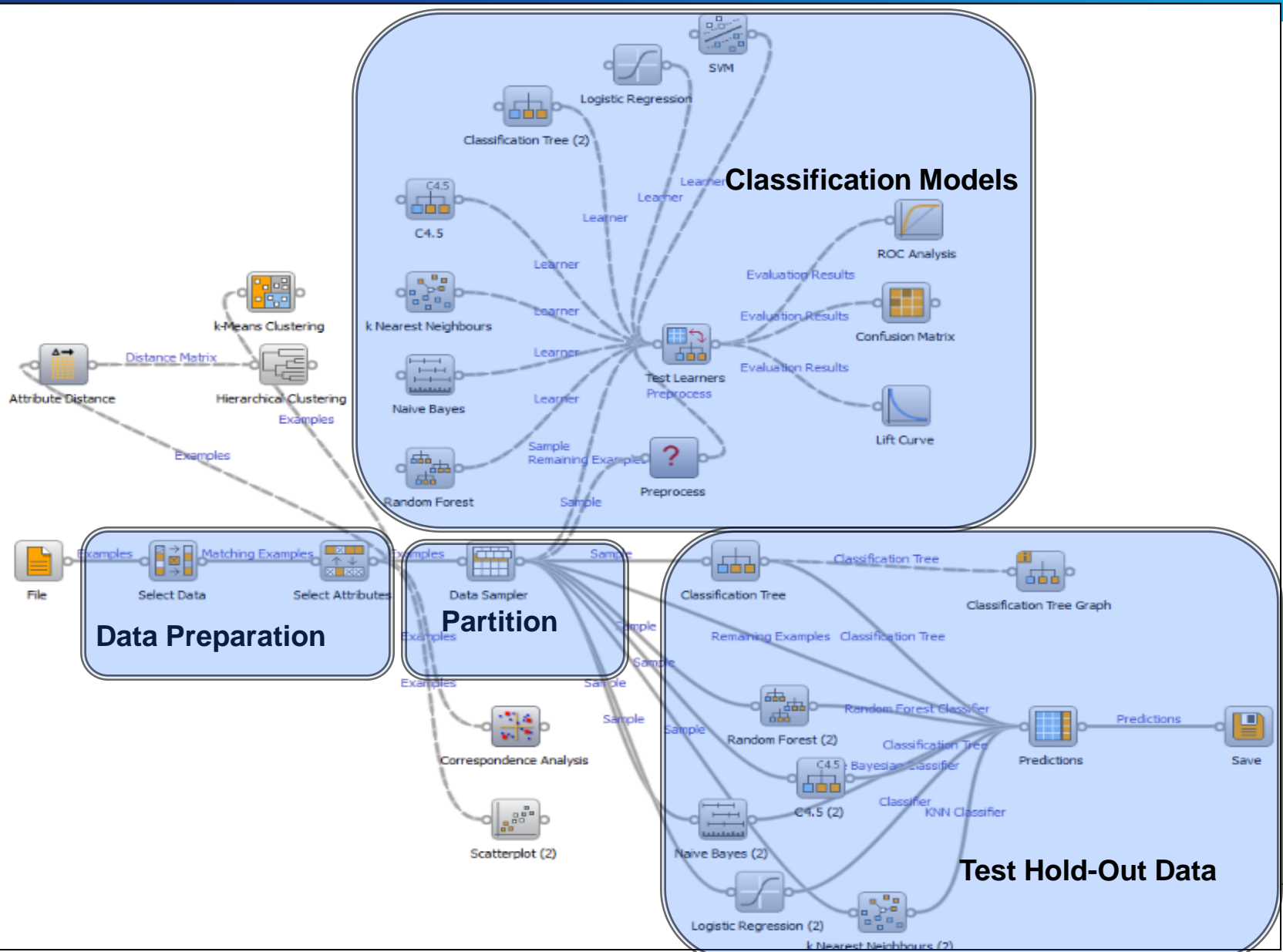
Results

- Prediction on holdout set
- Classification accuracy similar across various methods.
- RF slightly better – use of Ensemble
- High Sensitivity -> Ability to predict fixed-price contract.

Conclusion

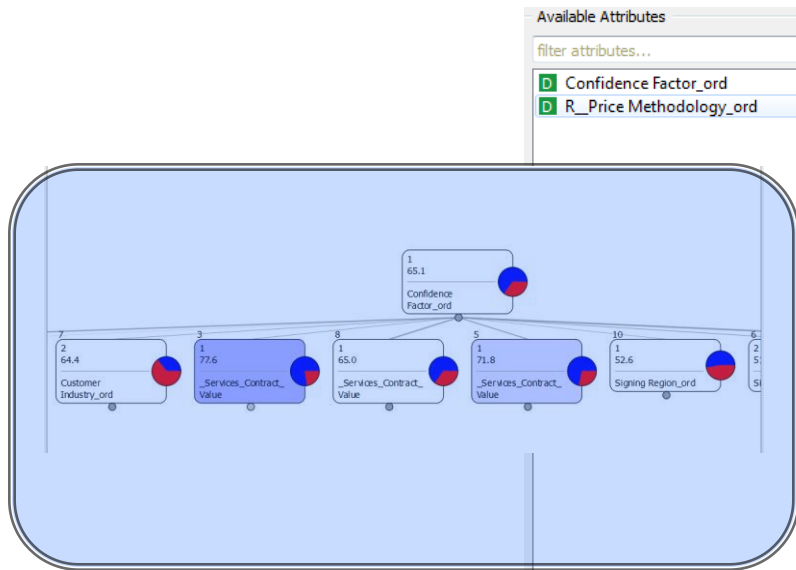
- Models have ability to classify fixed-price contracts.

Methodology – Process flow



Comparison of classification results of different methods

Selected Attributes



Available Attributes

filter attributes...

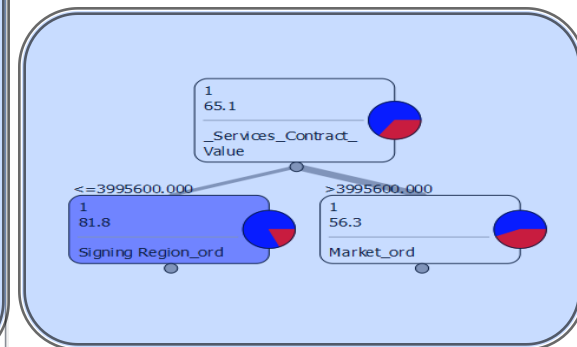
- Confidence Factor_ord
- R_Price Methodology_ord

Attributes

- Vendor Role_ord
- Existing Relationship_ord
- Customer Industry_ord
- Signing Region_ord
- Geographic Scope Macro Region_ord
- Macro_Market_ord
- Engagement_Type_ord
- Market_ord
- Number of Subsegments(Total)_ord
- _Contract Type_ord
- _Bid Type_ord
- _Contract Status_ord
- _Customer Revenue
- _Services_Contract_Value
- _Service_Run_Rate

Class

- R_Price_Disc



Evaluation Results

Method	CA	Sens	Spec	AUC
1 Naive Bayes	0.6653	0.7300	0.5446	0.7205
2 kNN	0.6984	0.8202	0.4712	0.7087
3 C4.5	0.7183	0.8581	0.4577	0.7196
4 Classification Tree	0.6941	0.8056	0.4862	0.6243
5 Logistic regression	0.7115	0.8894	0.3798	0.7355

Evaluation Results

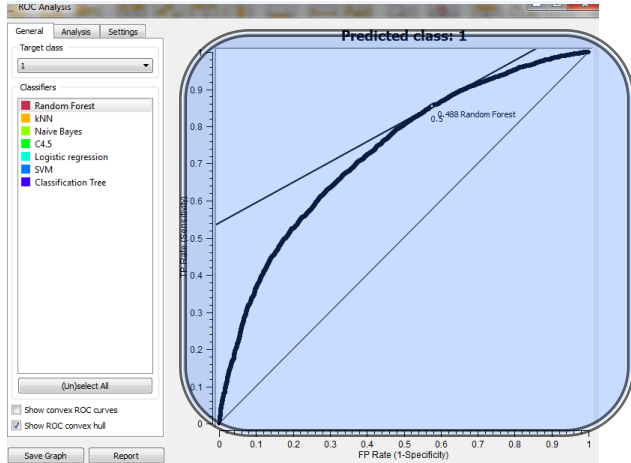
Method	CA	Sens	Spec	AUC
1 Random Forest	0.7047	0.8973	0.3455	0.7311
2 Naive Bayes	0.6587	0.7296	0.5263	0.6973
3 kNN	0.6742	0.7985	0.4424	0.6705
4 C4.5	0.6934	0.8412	0.4179	0.6730
5 Classification Tree	0.6818	0.7972	0.4667	0.6101
6 Logistic regression	0.6769	0.9189	0.2256	0.6900
7 SVM	0.6617	0.9510	0.1222	0.6956

With 'Confidence Value'

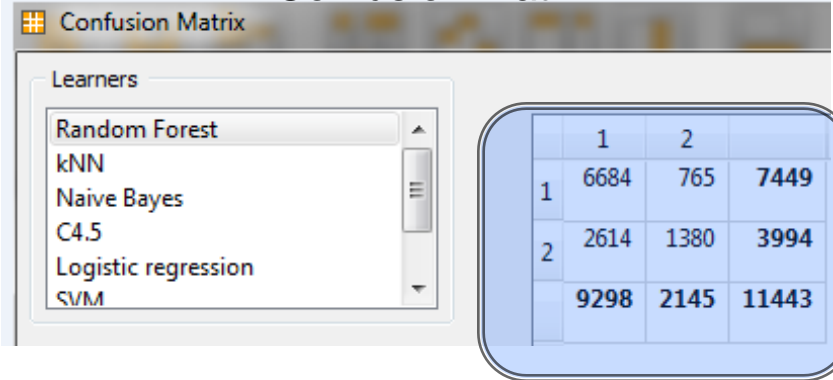
Without 'Confidence Value'

Results – Cross Validation and Hold-Out Performance

ROC Curve



Cross Validation Results



Results on Hold-Out DataSet

	snts(1)	Contract Type_ord	_Bid Type_ord	Contract Status_ord	Customer_Revenu	ices_Contract_V	Service_Run_Rate	R_Price_D	Random Forest	Naive Bayes	Logistic regression	kNN	C4.5	Classification
1	4	1	2	1.26254e+009	1691000	1691000	1	1	1	1	1	1	1	1
2	3	1	2	0	144000	144000	1	1	1	1	1	1	1	1
3	4	1	2	1.37812e+009	2805800	2805800	1	1	1	1	1	1	1	1
4	4	5	2	2.66988e+010	12000000	2400000	1	1	1	1	1	1	1	1
5	4	1	2	2.53204e+008	139503	139503	1	1	1	1	1	1	1	1
6	4	1	2	2.82356e+009	253000	253000	1	1	1	1	1	1	1	1
7	3	2	2	2.76743e+010	136665804	27333180	2	2	2	1	2	2	2	2
8	4	1	2	3.59275e+010	4182300	4182300	1	1	1	1	1	1	1	1
9	4	1	2	9.60309e+007	20948200	5237040	2	1	1	1	1	1	1	1
10	4	1	2	1.30331e+009	3029440	1009812	1	1	1	1	1	1	1	1
11	4	1	2	1.91367e+010	166666672	33333324	1	1	2	1	1	1	1	2
12	4	1	2	1.897e+009	1650000	1650000	1	1	1	1	1	1	1	1
13	2	1	2	6.08447e+009	1960842	1960842	1	1	1	1	1	1	1	1
14	4	1	2	2.4e+010	2435900	749496	1	1	1	1	1	1	1	1
15	4	1	2	2.4e+010	1334318	1334318	1	1	1	1	1	1	1	1
16	4	1	2	4.43444e+008	3644424	728880	2	1	1	1	1	1	1	1
17	4	1	2	5.562e+008	28000000	3999996	1	1	2	1	1	1	1	1
18	4	1	2	9.58938e+006	190881	190881	1	1	1	1	1	1	1	1
19	3	1	2	7.09e+007	16000000	3999996	1	1	2	1	1	1	1	1
20	4	1	2	2.16e+007	385000	385000	1	1	1	1	1	1	1	1
21	4	5	2	1.5546e+010	1650000	1650000	1	1	1	1	1	1	1	1
22	4	1	2	0	874000	291324	2	1	1	1	2	1	1	1
23	4	1	2	1.40975e+008	71500	71500	1	1	1	1	1	1	1	1
24	4	1	2	3.401e+009	935000	935000	1	1	1	1	1	1	1	1
25	3	2	2	4.80006e+010	202816016	67605336	1	1	1	1	1	1	1	1
26	4	1	2	0	57000	57000	1	1	1	1	1	1	1	1
27	4	1	2	2.77258e+010	119246496	17035212	1	1	1	1	1	1	1	1
28	4	1	2	0	390000	390000	1	1	1	1	1	1	1	1
29	4	1	2	5.31e+007	122000	40656	1	1	1	1	1	1	1	1
30	4	1	2	2.234e+008	450000	450000	1	1	1	1	1	1	1	1
31	4	5	2	3.27981e+008	2824890	564972	1	1	1	1	1	1	1	1
32	4	1	2	0	1269200	253836	1	1	1	1	1	1	1	1
33	4	1	2	8.647e+009	20000000	3999996	1	1	1	1	1	1	1	1
34	3	2	2	9.82e+010	29000000	5799996	1	1	1	1	1	1	1	1
35	3	2	2	1.02e+009	2041078	2041078	1	1	1	1	1	1	1	1

	Random forest	Naïve Bayes	Logistic regression	KNN	C4.5	Classification Tree
	0.69458	0.643439	0.672779	0.675632	0.692747	0.667278

Random Forest * (Sensitivity) = 0.89257

Take away

