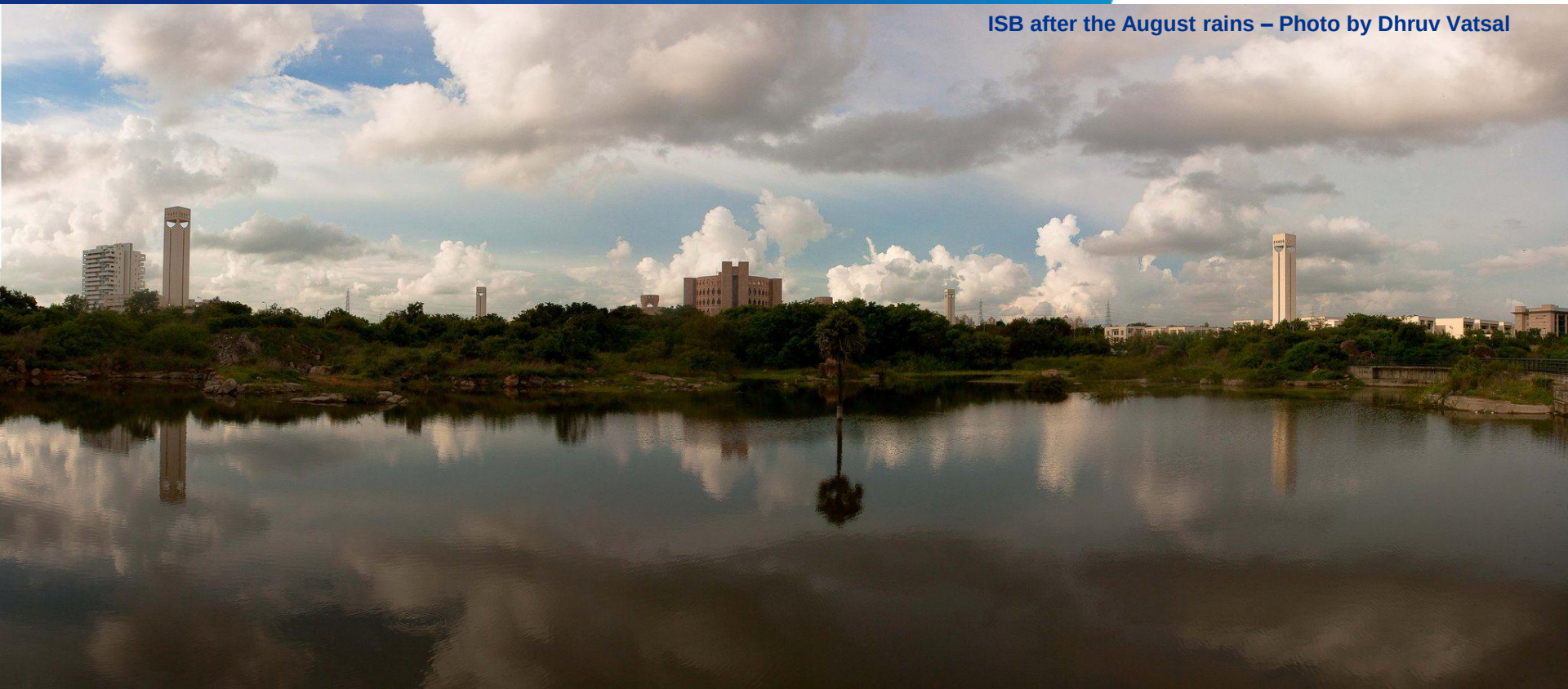
Predicting the pricing methodology for outsourcing contracts

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.

.

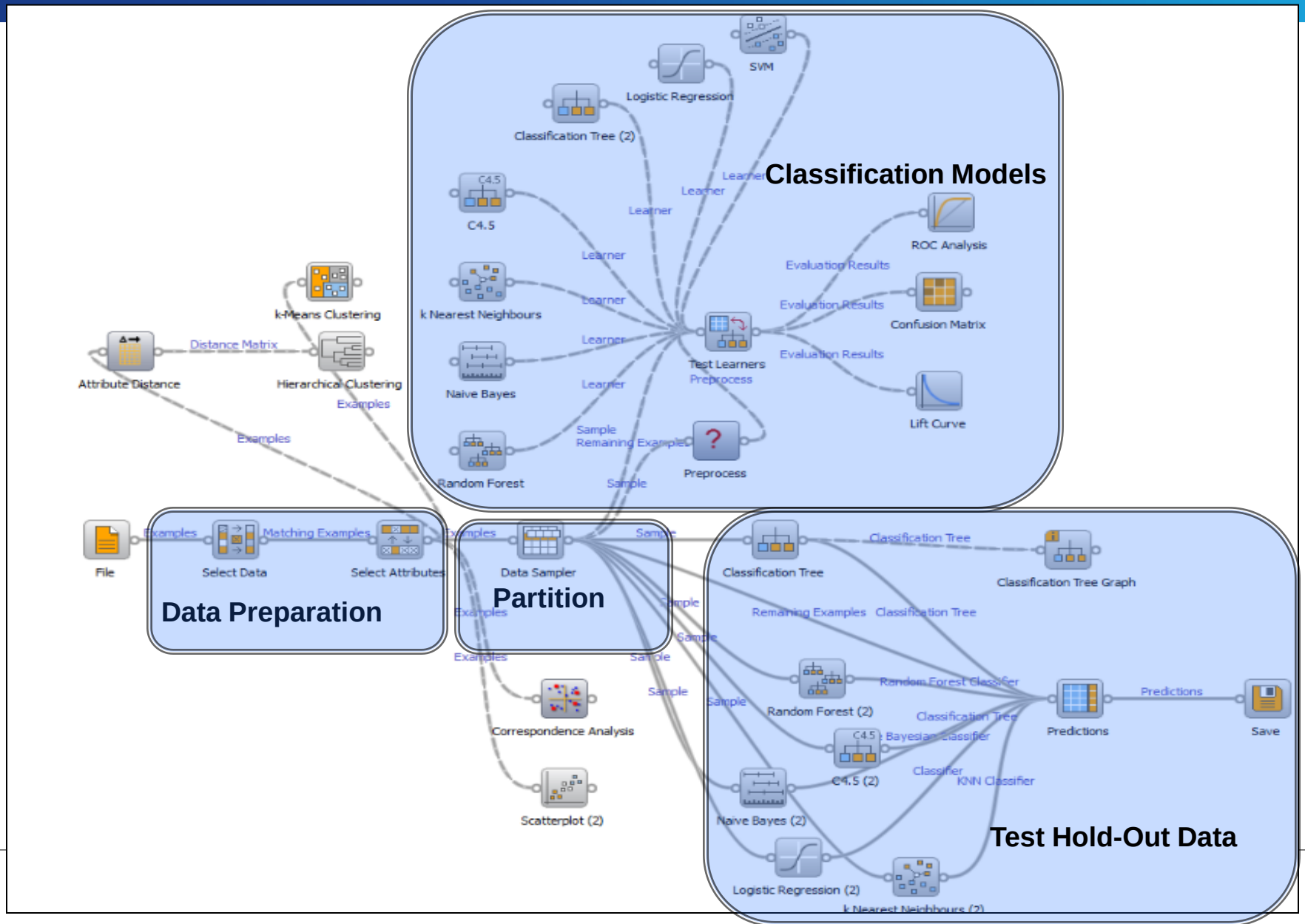
Results

- 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

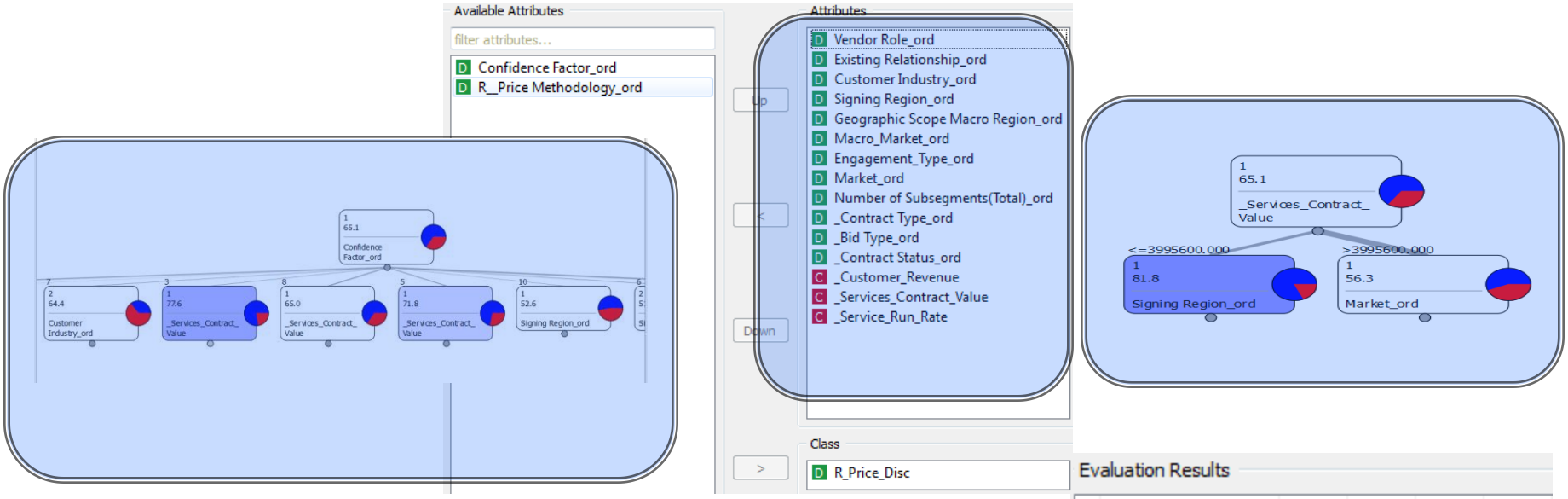
- Models have ability to classify fixed-price contracts.

Methodology – Process flow



Comparison of classification results of different methods

Selected Attributes



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

With 'Confidence Value'

Evaluation Results

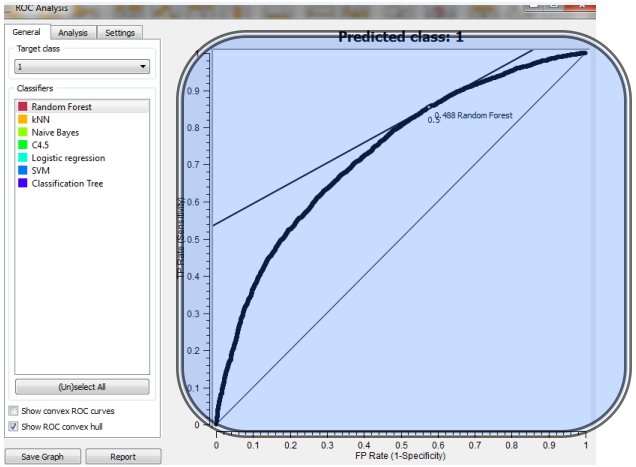
	Method	CA	Sens	Spec	AUC
1	Random Forest	0.7047	0.8973	0.3455	0.7311
2	Naive Bayes	0.6587	0.7296	0.5263	0.6973
3	kNN	0.6742	0.7985	0.4424	0.6705
4	C4.5	0.6934	0.8412	0.4179	0.6730
5	Classification Tree	0.6818	0.7972	0.4667	0.6101
6	Logistic regression	0.6769	0.9189	0.2256	0.6900
7	SVM	0.6617	0.9510	0.1222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

Without 'Confidence Value'

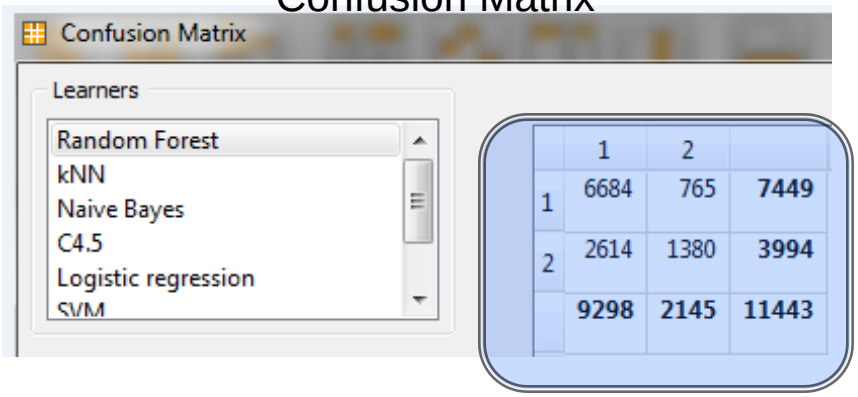
Results – Cross Validation and Hold-Out Performance

Cross Validation Results

ROC Curve

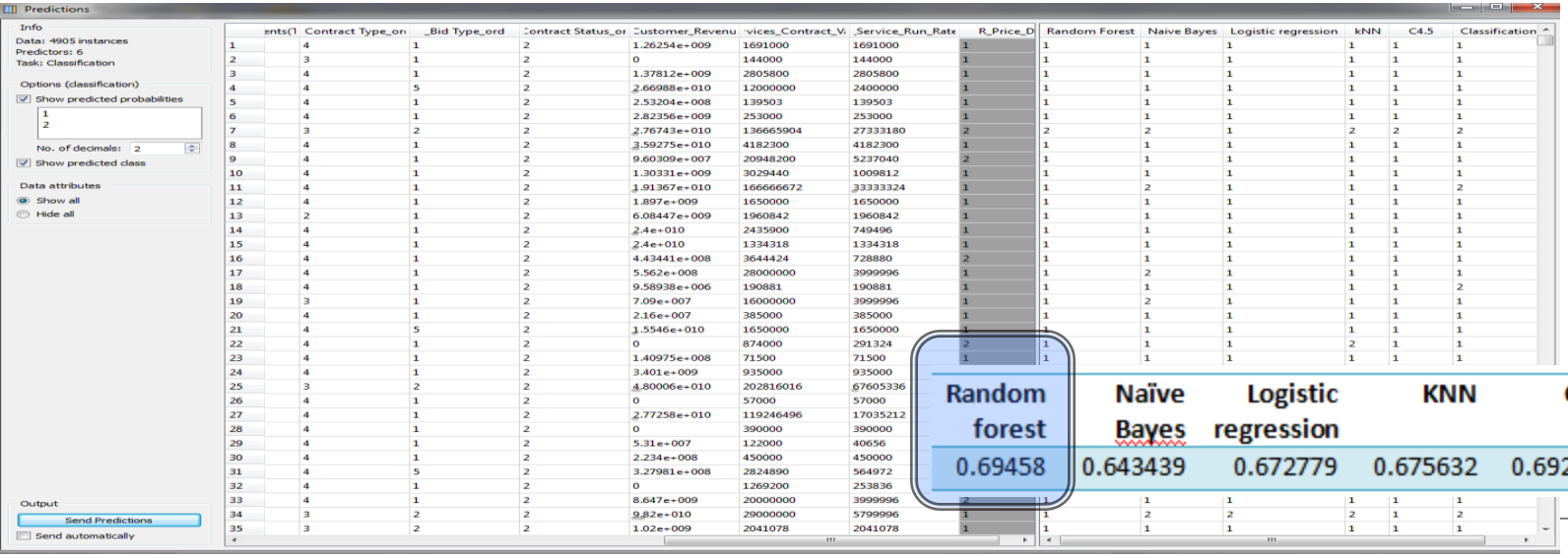


Confusion Matrix



	1	2	
1	6684	765	7449
2	2614	1380	3994
	9298	2145	11443

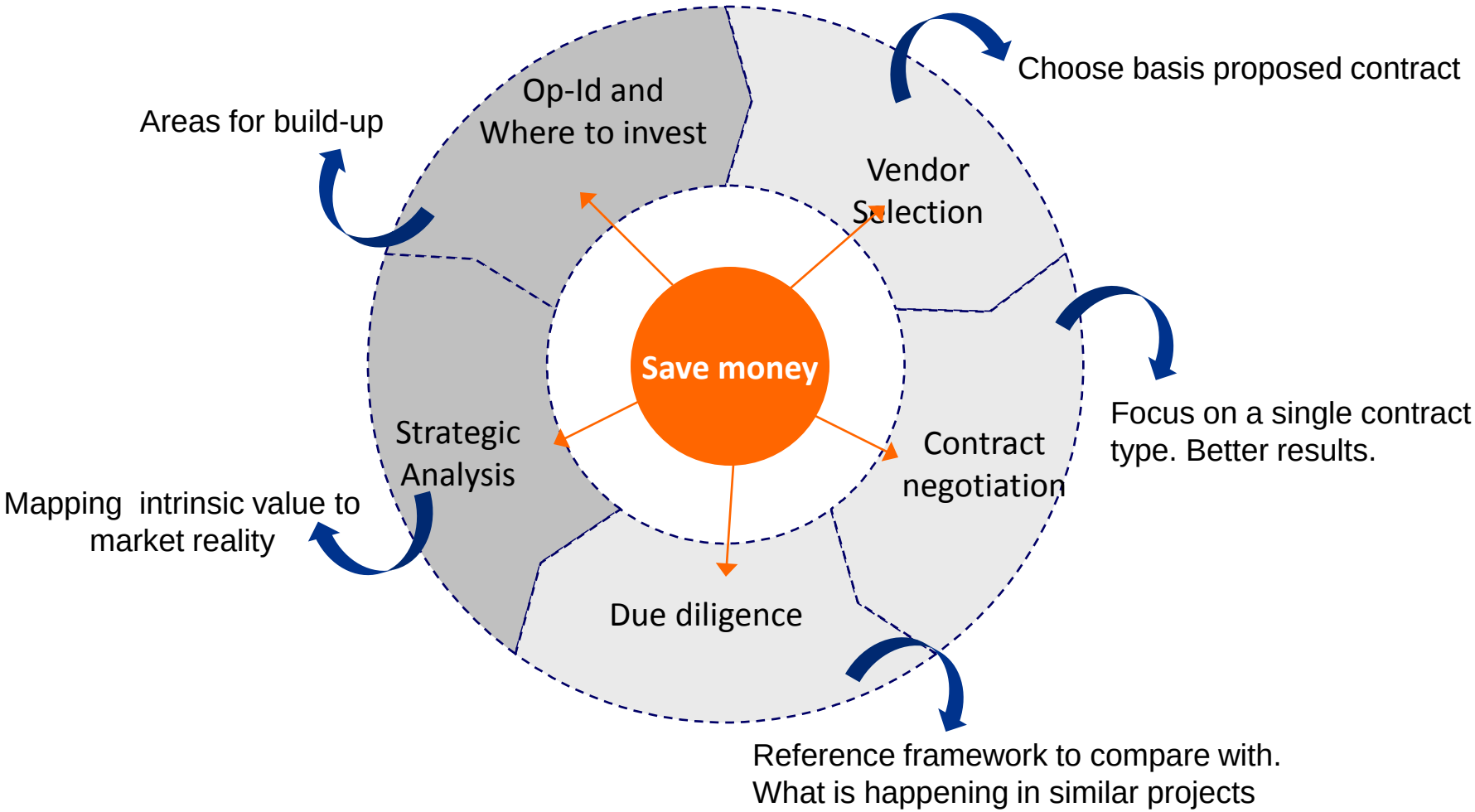
Results on Hold-Out DataSet



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


Areas for build-up
Op-Id and
Where to invest
Choose basis proposed contract
Vendor
Selection
Save money
Focus on a single contract
type. Better results.
Strategic
Analysis
Contract
negotiation
Mapping intrinsic value to
market reality
Due diligence
Reference framework to compare with.
What is happening in similar projects