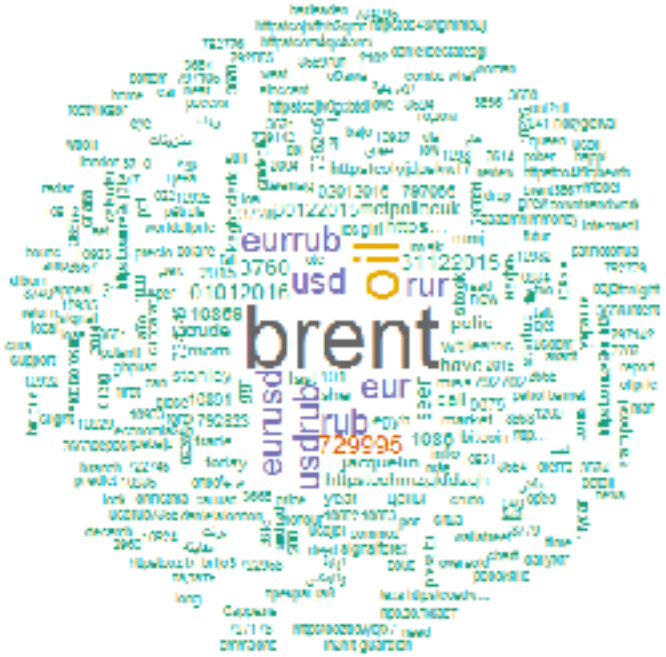

SATOSHI
SYSTEMS

SATOSHI SYSTEMS

Analysis of Things (AoT)



Big Data & Machine Learning

Applied to Brent Crude

Risk Edge Solutions

Executive Summary

SATOSHI SYSTEMS

Data

Selecting & Visualising Data

- We select historical, monthly, fundamental data
- We check for correlations between variables using multiple techniques
- Check for linear relationships within variables

Relations

Understanding Important Parameters & their Relations

- Reduced number of parameters to achieve optimal model – for both performance and accuracy
- We deploy recursive algorithms to guide us towards variable selection

Prediction

The Holy Grail

- We plot decision trees to understand the inner workings of the model
- We do some basic-level prediction for Brent Prices through 2016 based on the model

SATOSHI SYSTEMS

Data Selection

Visualising and Normalising data

Data : Fundamental, Monthly, Historical

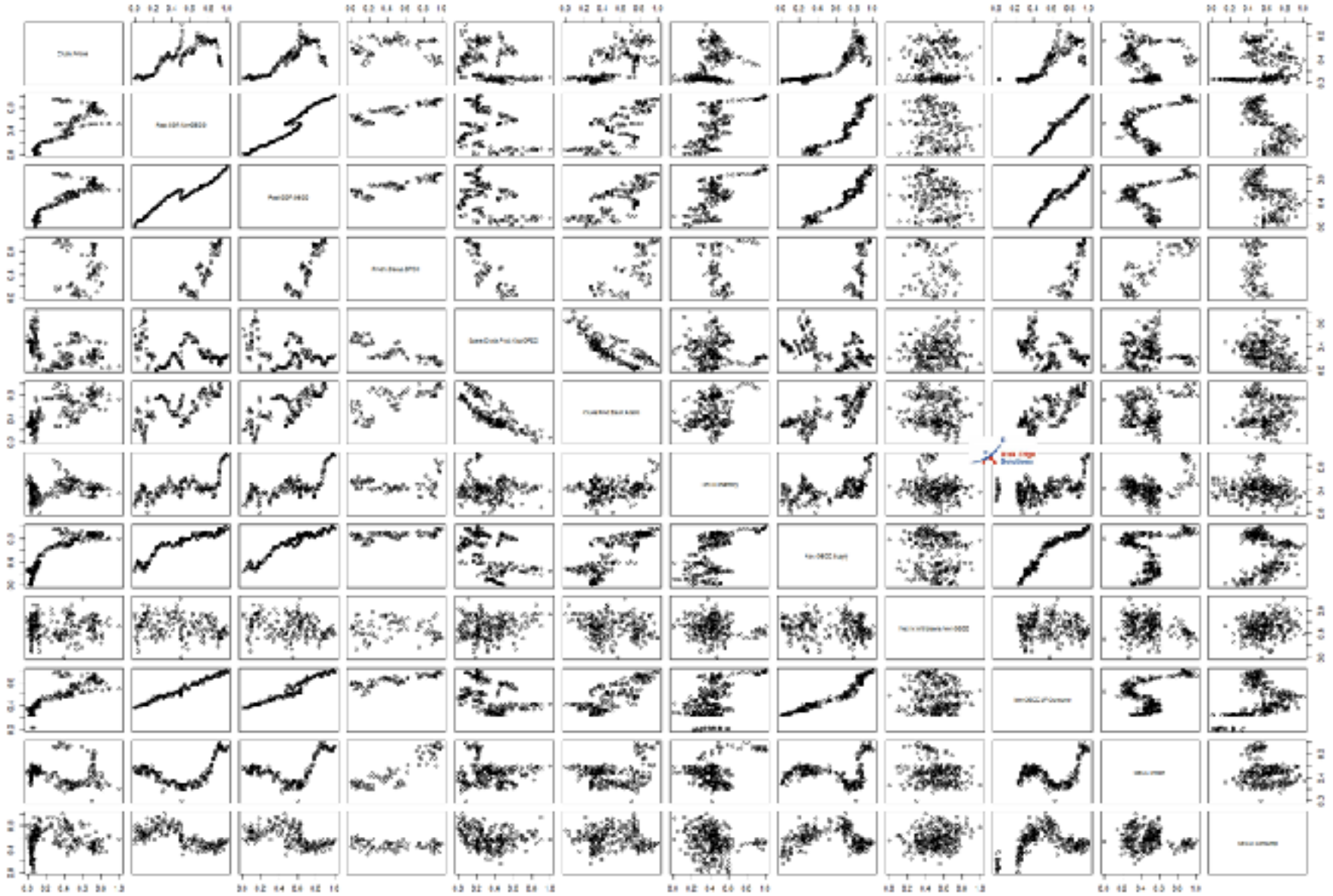
SATOSHI SYSTEMS

1. OECD Real Gross Domestic Product,
2. Non-OECD Real Gross Domestic Product,
3. OPEC Total Spare Crude Oil Production Capacity,
4. Crude Oil Production, Saudi Arabia,
5. Unplanned crude oil production disruptions, OPEC,
6. Unplanned liquid-fuel production disruptions, non-OPEC,
7. OECD End-of-period Commercial Crude Oil and Other Liquids Inventory,
8. Crude Oil and Liquid Fuels Supply, Total Non-OECD,
9. Net Inventory Withdrawals, Total Non-OECD Crude Oil and Other Liquids,
10. OECD Petroleum Production
11. Non-OECD Liquid Fuels consumption
12. OECD Liquid Fuels consumption
13. Crude Prices, End of Month

Source: *https://www.quandl.com/data/EIA*

Visualising Pair-wise Relationships

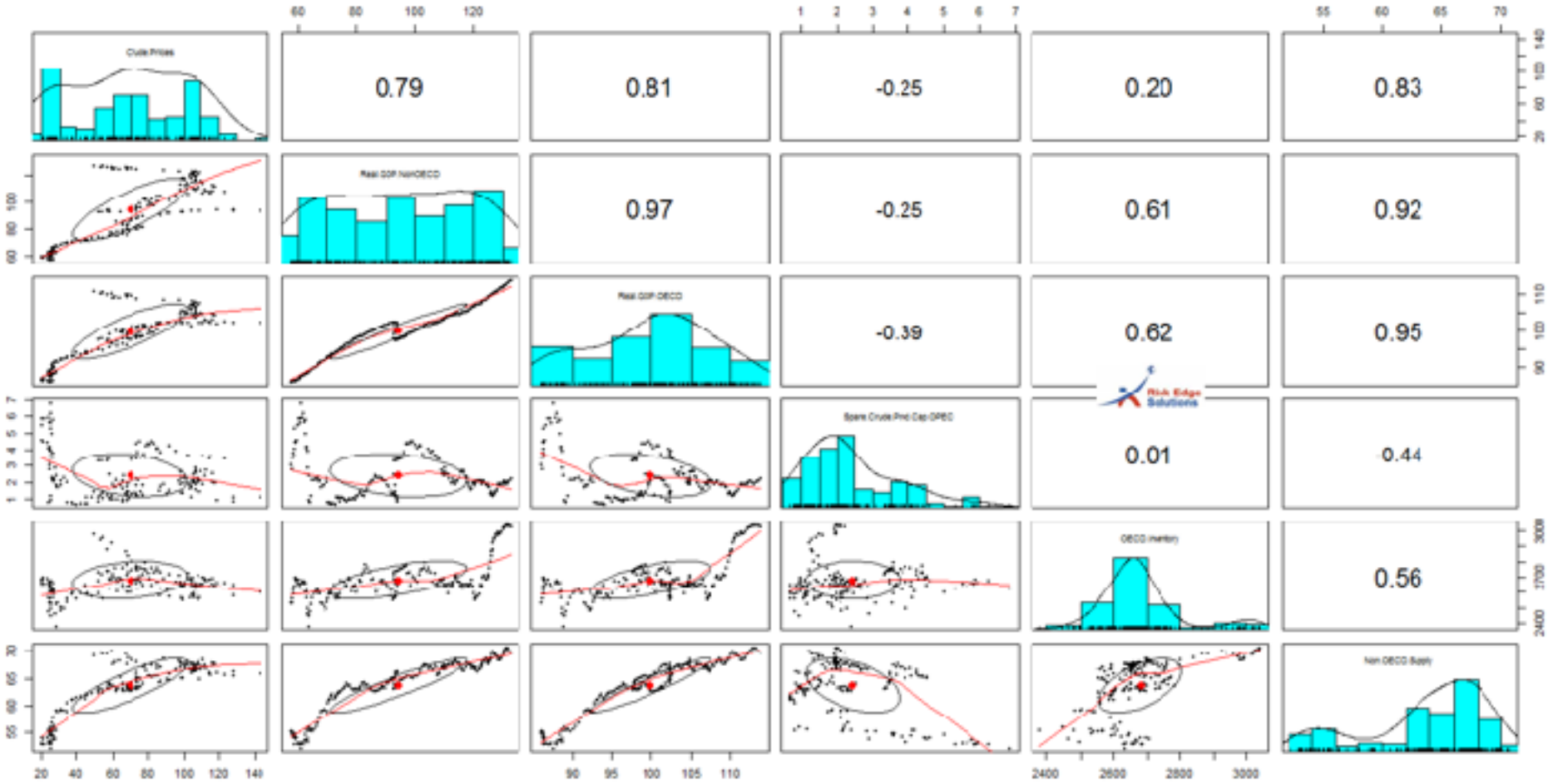
SATOSHI SYSTEMS



Correlations: Sample pairs

SATOSHI SYSTEMS

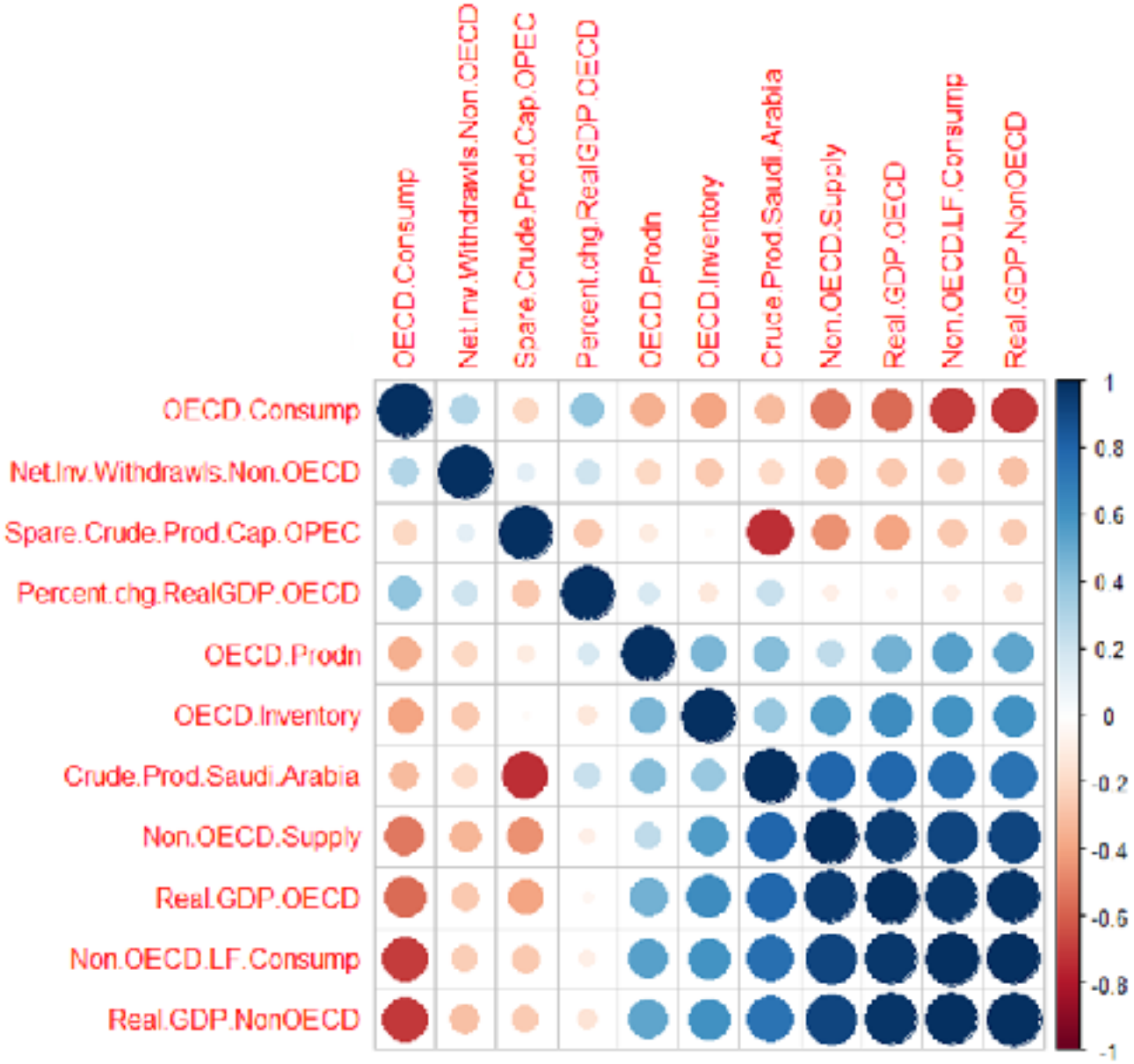
Distribution, Correlation Ellipses, Means and Loess Smoothing



Loess → Local Polynomial Regression; Perfect Ellipse → Weak Correlation

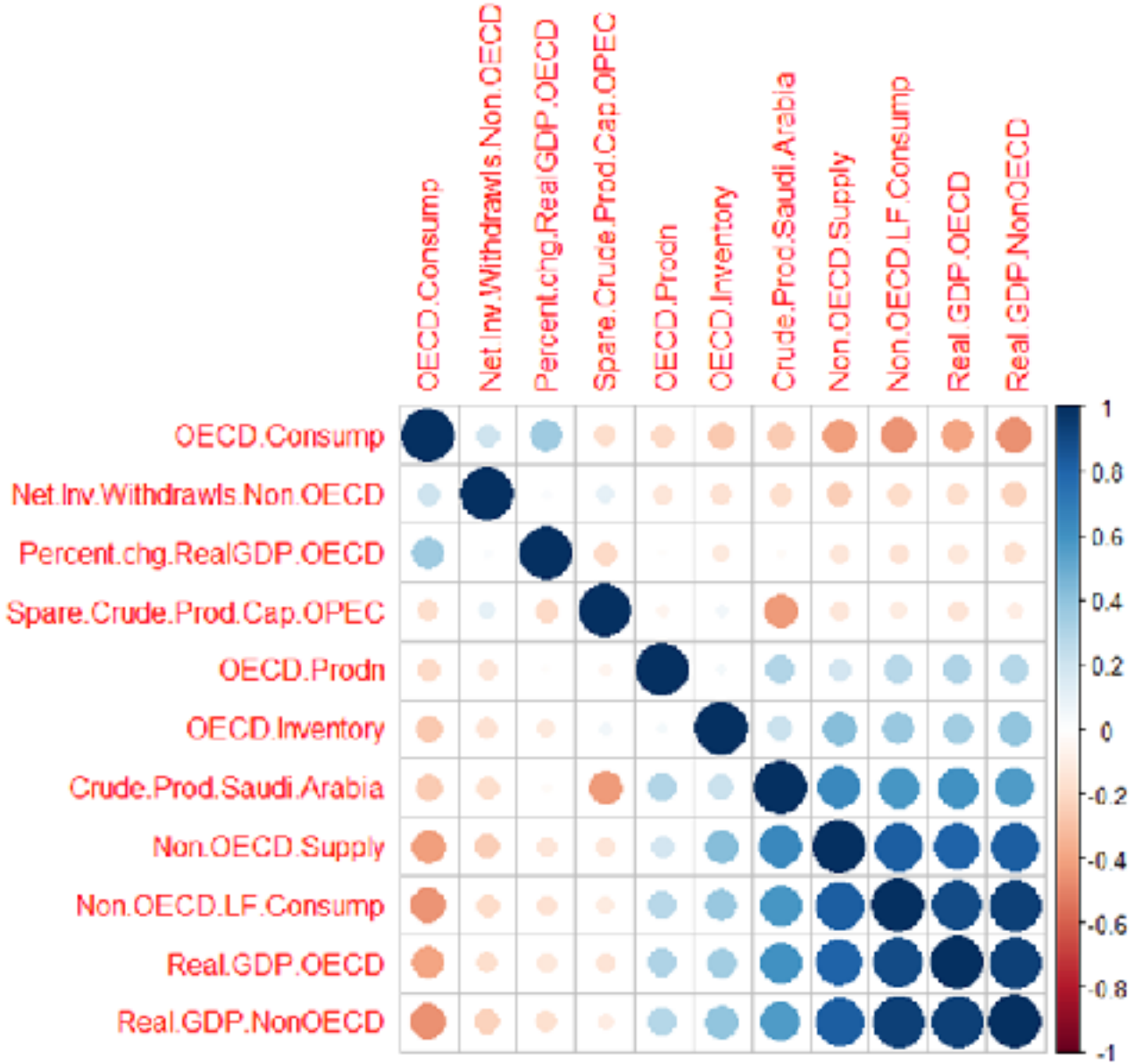
Correlations : Pearson Coefficient

SATOSHI SYSTEMS



Viewing Correlations between various variables gives us a general sense of what kind of model we can expect

Correlations : Kendall Coefficient



The same Correlation chart using another method – Kendall's

SATOSHI SYSTEMS

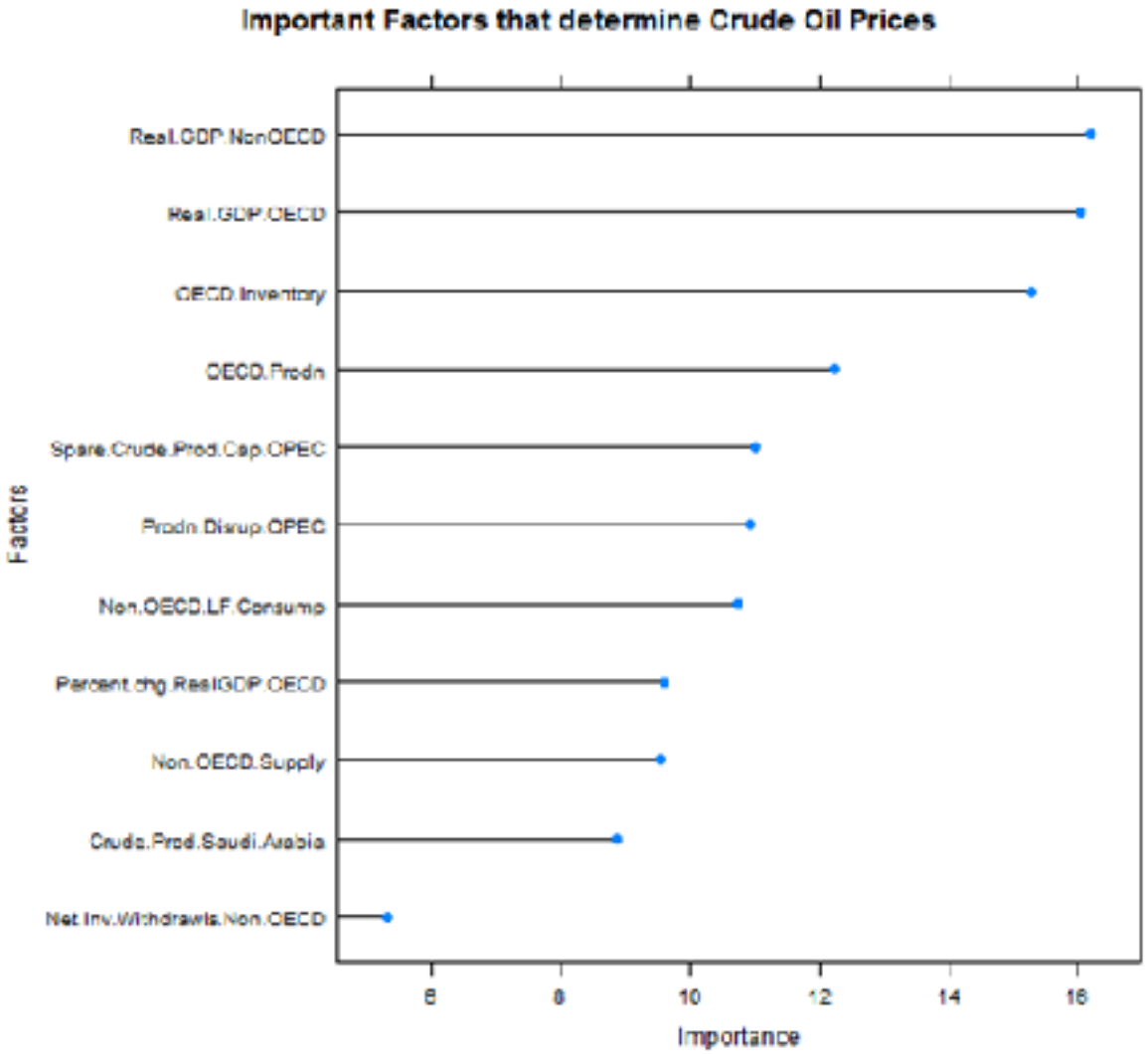
Variable Selection

Which Variables give the most Optimal Model

Prioritising Parameters

SATOSHI SYSTEMS

Which variables to keep and discard



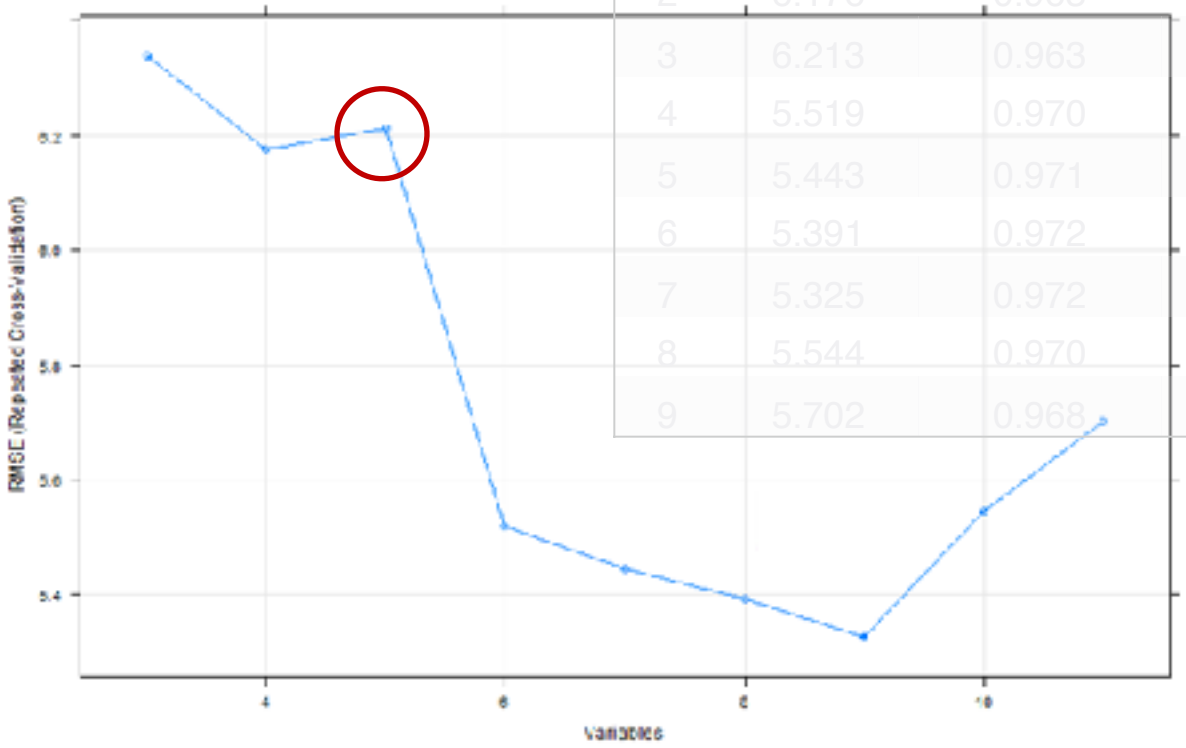
Analysing R-squared

- Analysing each set of features to see which ones should make the cut without increasing the **model complexity** and without compromising on the **model quality**.
- **Recursive elimination algorithm** used for selecting the best possible combination of parameters.

Recursive Elimination

Automated Algo tells us to use 5 variables for optimal model

Var	RMSE	R squared	RMSE SD	R squared SD
1	6.339	0.961	1.747	0.0198
2	6.176	0.963	1.847	0.0207
3	6.213	0.963	1.574	0.0159
4	5.519	0.970	1.655	0.0163
5	5.443	0.971	1.697	0.0161
6	5.391	0.972	1.638	0.0156
7	5.325	0.972	1.638	0.0158
8	5.544	0.970	1.837	0.0194
9	5.702	0.968	1.774	0.0188



Top five variables suggested by automated algorithm

1. OECD Real GDP
2. Non-OECD Real GDP
3. OECD Inventory
4. Non-OECD Supply
5. Non-OECD LF consumption

Optimizing Selection

SATOSHI SYSTEMS

All 9 variable Considered

#	Description / # of Variables	9	8	7	6	5
1	Real.GDP.NonOECD					
2	Real.GDP.OECD	***	***	***	***	***
3	OECD.Inventory	***	***	***	***	***
4	Non.OECD.Supply		.	.		
5	Non.OECD.LF.Consump	**	*	**	*	
6	Spare.Crude.Prod.Cap.OPEC	***	***	***	***	***
7	OECD.Prodn	***	***	***	***	***
8	Percent.chg.RealGDP.OECD	.				
9	Crude.Prod.Saudi.Arabia	***	***	***	***	***
	Multiple R^2	**0.9006**	**0.8994**	**0.8988**	**0.8968**	**0.8935**
	Adjusted R^2	**0.8954**	**0.8947**	**0.8947**	**0.8932**	**0.8904**

*Significance codes: very significant ****

Selected Parameters

SATOSHI SYSTEMS

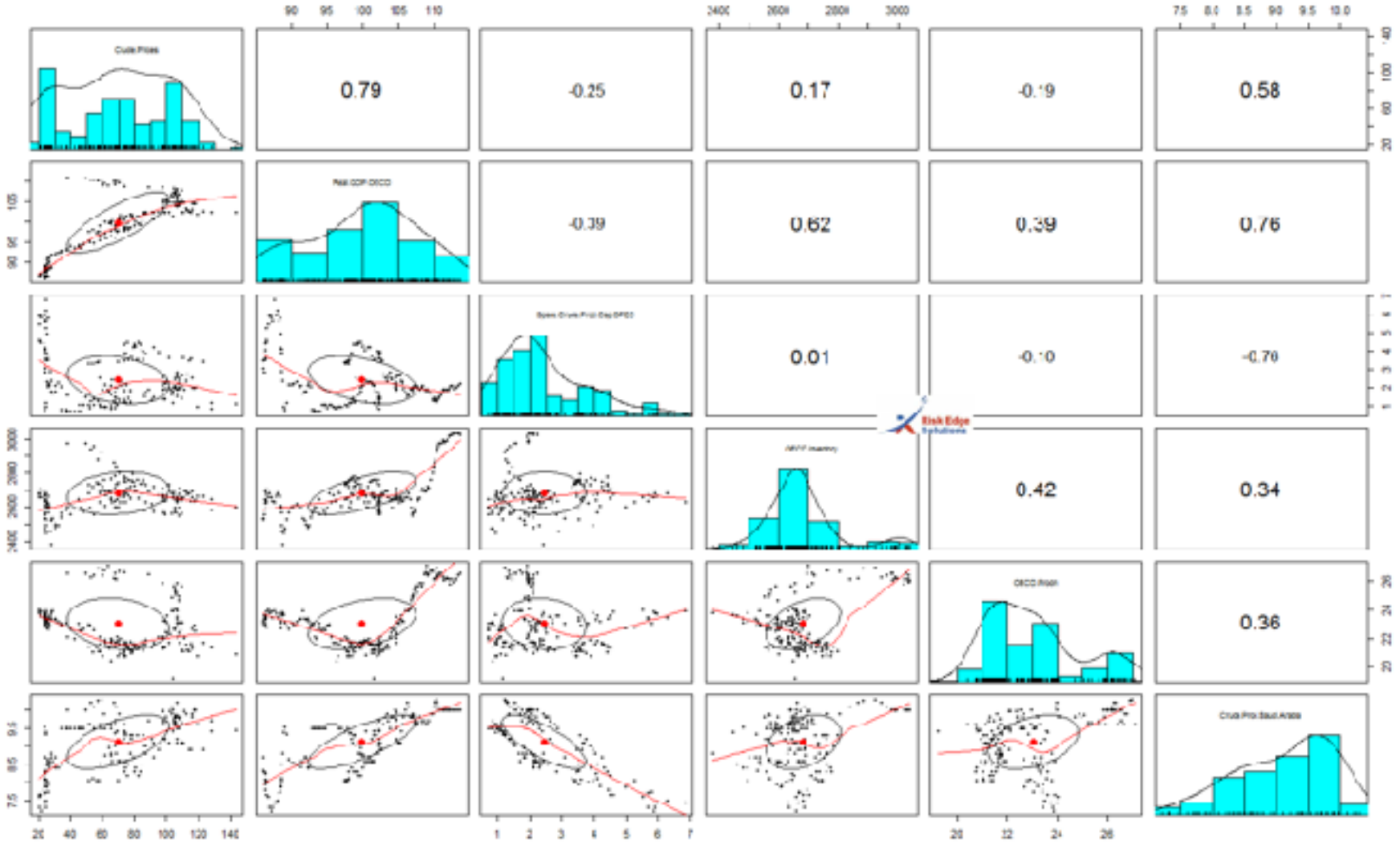
After Elimination / Pruning

1. **OECD Real Gross Domestic Product,**
2. **OPEC Total Spare Crude Oil Production Capacity,**
3. **Crude Oil Production, Saudi Arabia,**
4. **OECD End-of-period Commercial Crude Oil and Other Liquids Inventory,**
5. **OECD Petroleum Production**
6. ~~Non-OECD Real Gross Domestic Product,~~
7. ~~Unplanned crude oil production disruptions, OPEC,~~
8. ~~Unplanned liquid-fuel production disruptions, non-OPEC,~~
9. ~~Percentage change in Real GDP of OECD Countries~~
10. ~~Crude Oil and Liquid Fuels Supply, Total Non-OECD,~~
11. ~~Net Inventory Withdrawals, Total Non-OECD Crude Oil and Other Liquids,~~
12. ~~Non-OECD Liquid Fuels consumption~~
13. ~~OECD Liquid Fuels consumption~~

Correlations: Selected Variables

SATOSHI SYSTEMS

Distribution, Correlation Ellipses, Means and Loess Smoothing



SATOSHI SYSTEMS

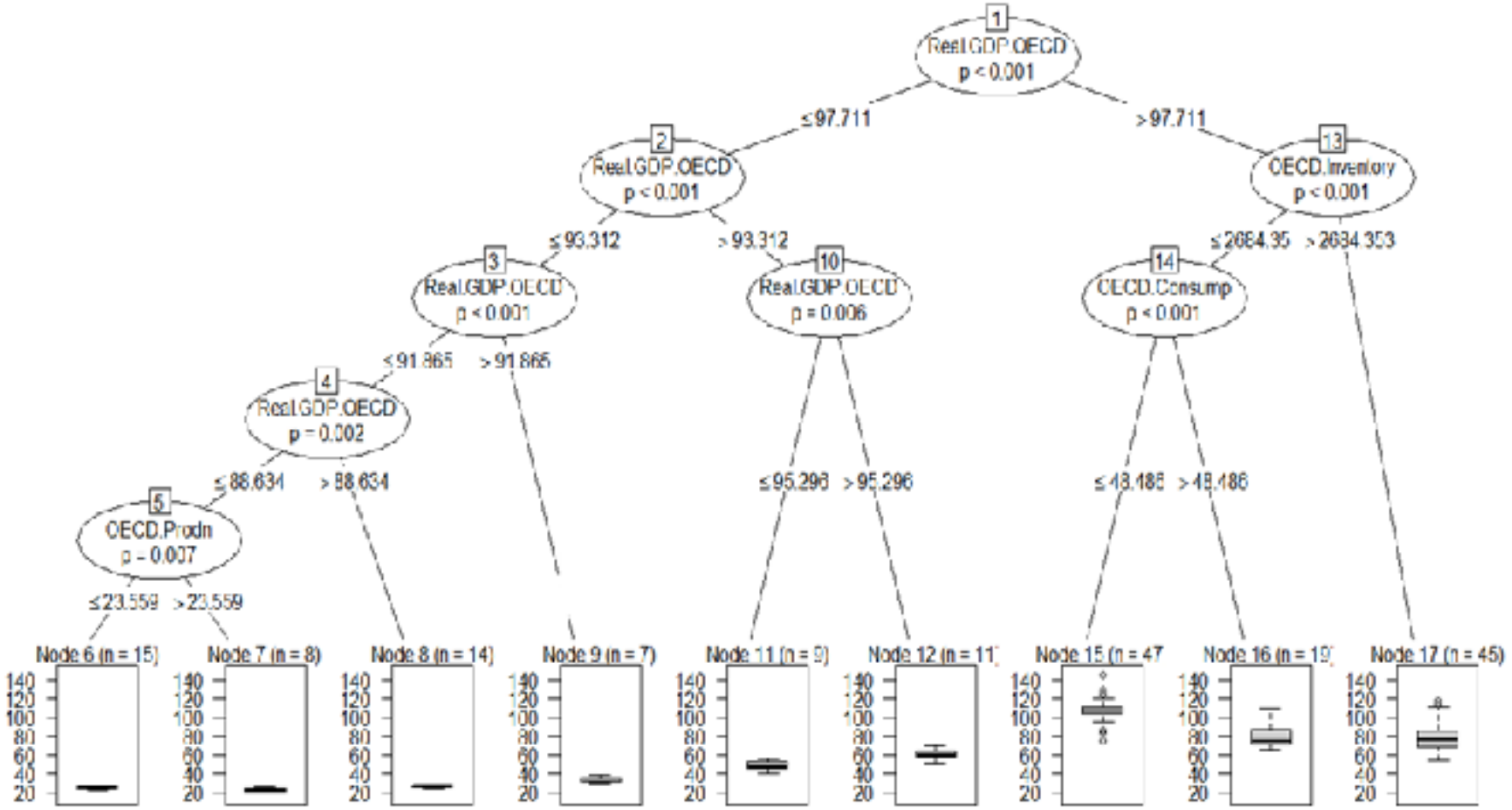
End Result

Decision Trees and Predictive Analytics

Decision Tree: Final Model

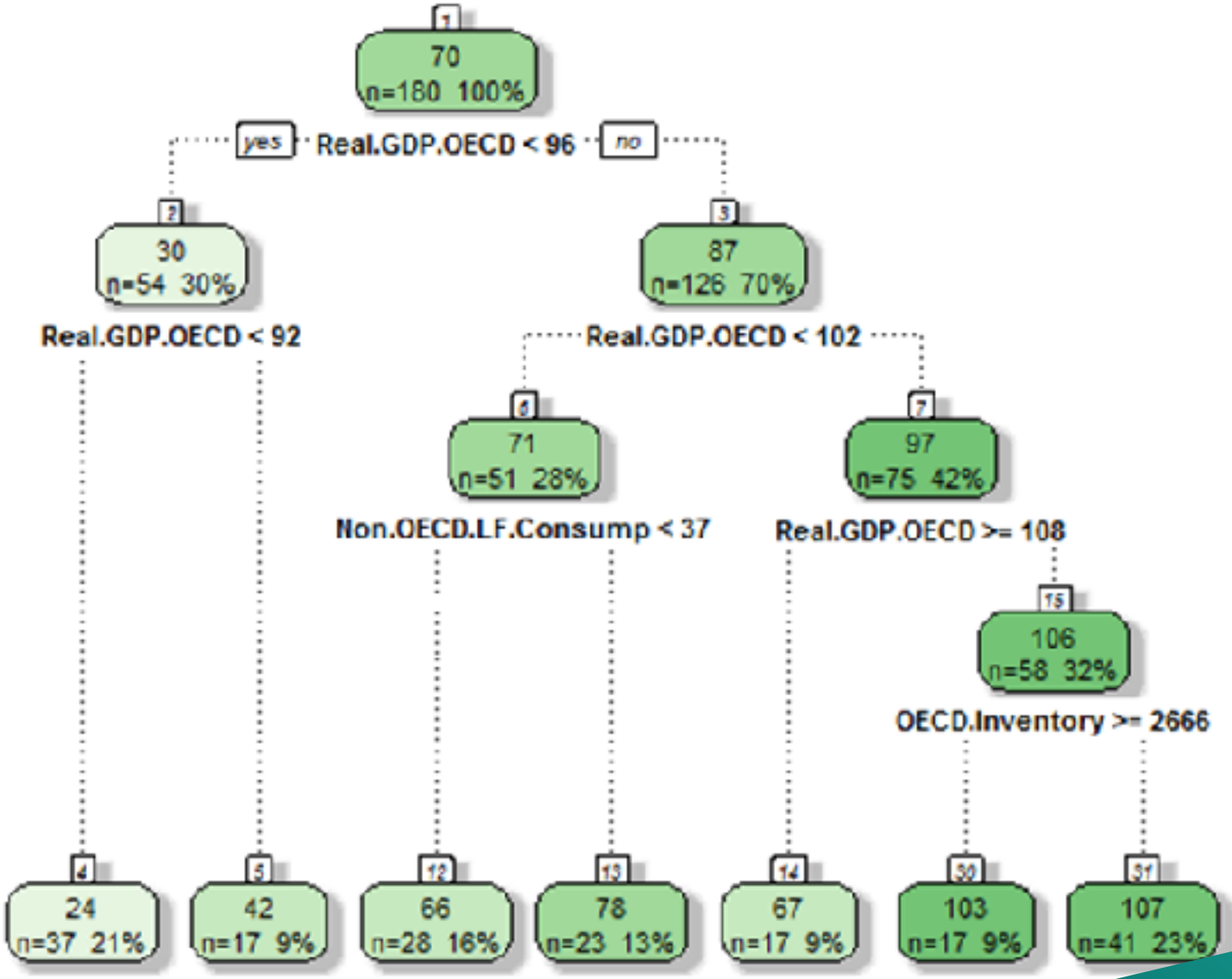
SATOSHI SYSTEMS

Taking the most relevant trees only



Decision Tree: Simplified Final Model

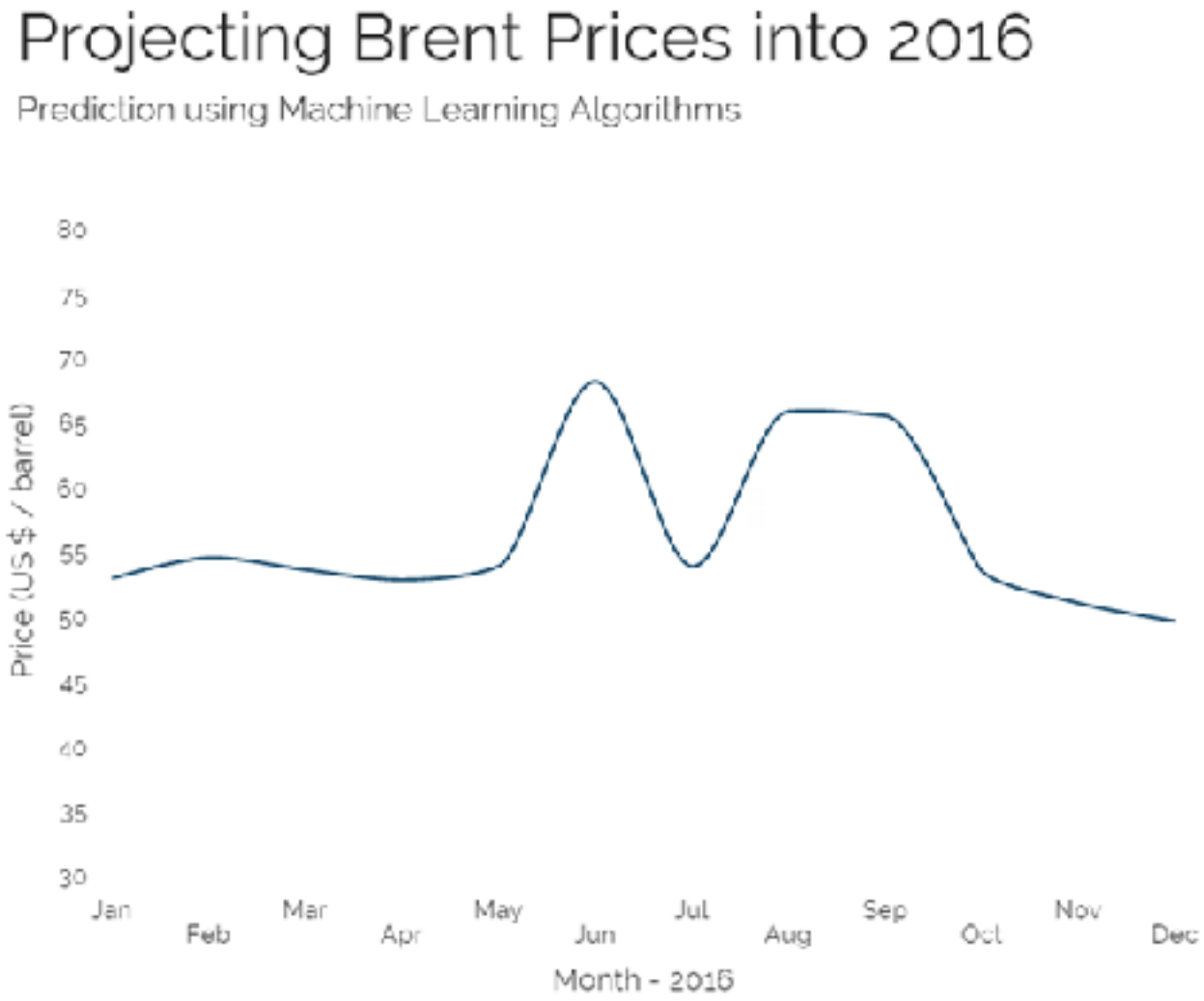
SATOSHI SYSTEMS



Result: Predicting Brent in 2016

SATOSHI SYSTEMS

Applying Big Data Analytics & Machine Learning



- *The range is predicted to be between 50-68 $/barrel (only month-end prices were projected.)*
- *Based on 1 year forward estimates of fundamental factors as given by EIA.*

End Notes: Disclaimers & Conclusions

SATOSHI SYSTEMS

So, the point of it all…?

- To show that Machine Learning, a part of Big Data Analytics, can be applied to Commodities
- To understand the relations between various fundamental factors and how they affect prices

How accurate is the prediction?

- We've used Decision Forest and such models which show higher accuracy than other Machine Learning Algos, but still the prediction should be looked at as a range, rather than exact numbers.
- Such models applied to fundamental data like Demand / Supply might provide better accuracy.

Could it have been better?

- We've used month-end data for our analysis, from 2001 (180 data points). So yes, more frequent data, with different variables could have given better results.
- Assumptions were made on future Crude Production in Saudi Arabia.

Can there be Other Applications of this model?

- Similar models can be built to predict Supply & Demand, Detect Fraud in Trading or Operations; Counter-party Credit Risk, Sales Analytics, Predict Prices and many other Custom Solutions

SATOSHI
SYSTEMS

THANK YOU