Bridging the Gap between Data and Value
Machine Learning in Oil & Gas
Energy Conference Network

Bill Robertson
Senior Manager, Enterprise Analytics
April 19, 2017
Enterprise Analytics

- Multi-disciplinary team of data analysts, domain experts, and application developers
- Over 120 years of industry experience
- Based in Houston, Texas
- Successful record delivering solutions from exploration to production optimization
Data is everywhere

“90% of the data in the world today has been created in the last two years”

IBM
Data is an asset

“Analytics will define the difference between the losers and winners”

“The average initial increase in profits from big data investments was 6%…that increased to 9% for investments spanning five years.”

McKinsey&Company
Leveraging data is hard

- Challenges
  - Which data should you use?
  - How should you handle analytics?
  - How can you use the insights you’ve gained to transform your operations?

Data analysis breakdown

“Data preparation accounts for almost 80% of the work of data scientists”

Forbes

What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%
Today’s enterprise

- Functional Groups
  - HR
  - Finance
  - HSE
  - Quality
  - Legal
  - Supply Chain
  - Enterprise Analytics
  - Data Governance

- Product Lines

- Operations
Data analysis evolution
## Analytic goals

<table>
<thead>
<tr>
<th>Goal</th>
<th>2011</th>
<th>2013</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve understanding of customers</td>
<td>33%</td>
<td>45%</td>
<td>46%</td>
</tr>
<tr>
<td>Retaining customers</td>
<td>30%</td>
<td>36%</td>
<td>37%</td>
</tr>
<tr>
<td>Improving customer experiences</td>
<td>22%</td>
<td>36%</td>
<td>36%</td>
</tr>
<tr>
<td>Selling products / services to existing customers</td>
<td>23%</td>
<td>33%</td>
<td>35%</td>
</tr>
<tr>
<td>Market research / survey analysis</td>
<td>29%</td>
<td>36%</td>
<td>34%</td>
</tr>
<tr>
<td>Acquiring customers</td>
<td>23%</td>
<td>32%</td>
<td>32%</td>
</tr>
<tr>
<td>Improving direct marketing programs</td>
<td>22%</td>
<td>27%</td>
<td>30%</td>
</tr>
<tr>
<td>Sales forecasting</td>
<td>19%</td>
<td>27%</td>
<td>27%</td>
</tr>
<tr>
<td>Fraud detection or prevention</td>
<td>21%</td>
<td>23%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Challenge the business case

**Stakeholder**

Do you know who our top customers are in region X?

I need to know what our revenue distribution looks like.

Last quarter’s revenues are lower than expected.

I don’t know, but I think we can increase our share of wallet by 5%.

**Analyst**

Why?

Why?

Why?

???
Predictive maintenance for Artificial Lift

- When will the electronic submersible pump (ESP) fail?
  - If we can predict this, where’s the potential value?

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Land</th>
<th>Offshore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost production</td>
<td>$$$</td>
<td>$$$</td>
</tr>
<tr>
<td>Rig availability</td>
<td>$</td>
<td>$$$</td>
</tr>
<tr>
<td>Materials inventory</td>
<td>$</td>
<td>$$$</td>
</tr>
<tr>
<td>Alternate suppliers</td>
<td>$</td>
<td>$$$</td>
</tr>
<tr>
<td>Potential value</td>
<td>$$</td>
<td>$$$$$$$</td>
</tr>
</tbody>
</table>
Barriers to success

- What is considered a failure?
- How far in advance can we make an accurate prediction?
- Do we have historical data for the intended use case?
- What’s the quality of the data?
- Is there management support for implementing a change?
Change management

Virginia Satir Change Process Model

- LATE STATUS QUO
- FOREIGN ELEMENT
- CHAOS
- NEW STATUS QUO
- PRACTICE AND INTEGRATION
- TRANSFORMING IDEA
Product selection for Flow Assurance

Given a fluid sample, which products will mitigate the production risks?

Before
- Compatibility tests
- Static bottle tests
- Kinetic turbidity tests
- Dynamic tube-blocking tests

After (i.e., use a prediction)

<table>
<thead>
<tr>
<th>Fluid sample ID</th>
<th>Product</th>
<th>Conc. (ppm)</th>
<th>Prediction</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>100</td>
<td>Not clear</td>
<td>61%</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>1,000</td>
<td>Not clear</td>
<td>61%</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>10,000</td>
<td>Not clear</td>
<td>65%</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>100</td>
<td>Clear</td>
<td>86%</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>1,000</td>
<td>Clear</td>
<td>88%</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>10,000</td>
<td>Clear</td>
<td>88%</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>100</td>
<td>Clear</td>
<td>71%</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>1,000</td>
<td>Clear</td>
<td>72%</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>10,000</td>
<td>Clear</td>
<td>70%</td>
</tr>
</tbody>
</table>
All data are not created equal

Structured

Cost of Ownership

Unstructured

Low Cost of Ownership High
Evaluating classifier performance

- True Positive
- True Negative
- False Positive
- False Negative

Clear

Probability of Passing

Not Clear

Actual

Not Clear

Threshold?

Lab Test

- 95%
- 73%
- 61%
- 49%
- 37%
- 25%
- 13%
- 1%

© 2016 Baker Hughes Incorporated. All Rights Reserved.
Data management

- Manage your data well so that good analyses are possible…
Takeaways

- Ensure your business cases are well-defined
- Manage your data
- Empower yourself

Bill Robertson
Senior Manager, Enterprise Analytics
Baker Hughes
bill.robertson@bakerhughes.com
+1-713-553-4451