Clustering Considerations for Machine Learning
With examples from exploration data

Philip Lesslar

Digital Energy Journal Forum 2019
3rd October 2019
ADAX Center, Bangsar South
Kuala Lumpur, Malaysia
Key messages

• Focus is only on clustering

• Understand internals to maximise ML effectiveness

• Classification is a big field

• Data analysis is not for the faint-hearted

• Usage with some example exploration data
Classification:
Creating meaningful groups out of a collection of objects

Build the Model:
Feature extraction to enable effective identification of new objects

Identification:
Use the model to identify new objects to one of the groups

Unsupervised learning

Training (Model building)

Testing

Supervised learning
The Machine Learning Workflow

https://towardsdatascience.com
Multivariate methods for classification and dimensionality reduction

• Cluster analysis
  • Finding “natural” or pre-determined groups in datasets

• Principal components analysis
  • Reducing the dimensionality of a data set by finding a smaller set of variables that still represents it

• Factor analysis
  • For data sets where a large number of observed variables are thought to reflect a smaller number of unobserved/latent variables.

• Multi dimensional scaling
  • Technique for visualising the level of similarity of samples transformed onto a 2D plane

• Linear & Multiple Regression
  • One or more independent variables are used to predict the value of a dependent variable

Some approaches to Clustering

• K-Means
  • Iterative computing of distances between points and group means. Requires specification of number of groups.

• Mean Shift Clustering
  • Sliding iterative method to find point groups of higher mean density.

• Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
  • Similar to Mean Shift but will identify noise and outliers.

• Expectation–Maximization (EM) Clustering using Gaussian Mixture Models (GMM)
  • Uses Gaussian approach to define clusters and uses both mean and std deviation unlike K-Means which only uses means. Detects elliptical clusters

• Agglomerative Hierarchical Clustering
  • Progressive pairwise clustering until all are merge into one tree in a dendrogram. Not too sensitive to choice of coefficient.

https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68
Cluster Analysis – Separating variables in n-dimensions

Visualization

2 dimensions

3 dimensions

4, 5, ......, n dimensions?

Cluster analysis requires:
1. Measure of pairwise proximities between points
2. Grouping method
Proximity measures

Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>a</td>
<td>d</td>
<td>c</td>
<td>a</td>
</tr>
<tr>
<td>j</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proximity measures

Measures of Similarity / Dissimilarity (Distance)

Matching coefficient

\[ S_{ij} = \frac{(a + d)}{(a + b + c + d)} \]

Jaccard coefficient (1908)

\[ S_{ij} = \frac{a}{(a + b + c)} \]

Rogers & Tanimoto (1960)

\[ S_{ij} = \frac{(a + d)}{[a + 2(b + c) + d]} \]

Sneath & Sokal (1973)

\[ S_{ij} = \frac{a}{[a + 2(b + c)]} \]

Gower & Legendre (1986)

\[ S_{ij} = \frac{(a + d)}{[a + \frac{1}{2}(b + c) + d]} \]

Euclidean Distance

Distance between vectors \( x \) & \( y \)

\[ d(x, y) = \sqrt{\sum_{i} (x_i - y_i)^2} \]

Canberra Distance

Distance between vectors \( u \) & \( v \)

\[ d(u, v) = \sum_{i} \frac{|u_i - v_i|}{|u_i| + |v_i|} \]

Proximity measures - Euclidean Distance – Pythagoras’s Theorem

In a right angled triangle, the length of the hypothenuse is equal to the square root of the sum of squares of the other 2 sides

\[ C = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \]

The Euclidean Distance \[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} = C, \quad n = 2 \]
Examples from Exploration data

1. Prospect Appraisal – Expectation values
2. Well logs – Curve values
3. Micropaleontology – Foraminiferal assemblages
Exploration Prospect Appraisal

DATA
Seismic interpretation
Geological picks & zones
Paleontology (incl. palyn, nanno etc)
Lithology & Lithofacies
Environments of deposition
Temperature
etc

Prospect Appraisal System

Expectations
- POS
- MSV
- HSV
- REC
- STOIIP
- GIIP

Cutoffs
- 0 mbbls
- 30 mbbls
- 0 bcf/tcf

Probabilistic
- Bootstrap
- Monte Carlo

Copyright©2019 Precision-DM
## Exploration Prospect Appraisal – The DATA

<table>
<thead>
<tr>
<th>POS</th>
<th>MSV</th>
<th>HSV</th>
<th>Expectation</th>
<th>STOIIP</th>
<th>OIL (0 mmbbls cutoff)</th>
<th>POS</th>
<th>MSV</th>
<th>HSV</th>
<th>Expectation</th>
<th>STOIIP</th>
<th>GAS (0 bscf cutoff)</th>
<th>POS</th>
<th>MSV</th>
<th>HSV</th>
<th>Expectation</th>
<th>STOIIP</th>
<th>(values/POS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>24</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>96</td>
<td>79</td>
<td>133</td>
<td>76</td>
<td>122</td>
<td>30</td>
<td>127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>11</td>
<td>26</td>
<td>7</td>
<td>23</td>
<td>10</td>
<td>38</td>
<td>60</td>
<td>64</td>
<td>25</td>
<td>57</td>
<td>16</td>
<td>27</td>
<td>36</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>11</td>
<td>23</td>
<td>8</td>
<td>31</td>
<td>15</td>
<td>29</td>
<td>38</td>
<td>80</td>
<td>41</td>
<td>90</td>
<td>33</td>
<td>55</td>
<td>46</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>5</td>
<td>9</td>
<td>4</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>85</td>
<td>15</td>
<td>32</td>
<td>13</td>
<td>25</td>
<td>29</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>7</td>
<td>16</td>
<td>5</td>
<td>22</td>
<td>6</td>
<td>29</td>
<td>40</td>
<td>80</td>
<td>27</td>
<td>64</td>
<td>22</td>
<td>36</td>
<td>31</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87</td>
<td>13</td>
<td>30</td>
<td>11</td>
<td>18</td>
<td>14</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>29</td>
<td>49</td>
<td>29</td>
<td>49</td>
<td>14</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>11</td>
<td>22</td>
<td>9</td>
<td>43</td>
<td>18</td>
<td>28</td>
<td>36</td>
<td>90</td>
<td>55</td>
<td>114</td>
<td>50</td>
<td>82</td>
<td>53</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>8</td>
<td>19</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>29</td>
<td>36</td>
<td>29</td>
<td>35</td>
<td>75</td>
<td>10</td>
<td>16</td>
<td>38</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>72</td>
<td>34</td>
<td>59</td>
<td>24</td>
<td>34</td>
<td>18</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>22</td>
<td>41</td>
<td>18</td>
<td>73</td>
<td>40</td>
<td>36</td>
<td>52</td>
<td>95</td>
<td>113</td>
<td>219</td>
<td>107</td>
<td>184</td>
<td>86</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>18</td>
<td>33</td>
<td>14</td>
<td>29</td>
<td>10</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>18</td>
<td>33</td>
<td>14</td>
<td>29</td>
<td>10</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>18</td>
<td>37</td>
<td>16</td>
<td>76</td>
<td>29</td>
<td>37</td>
<td>56</td>
<td>99</td>
<td>53</td>
<td>109</td>
<td>52</td>
<td>88</td>
<td>84</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>20</td>
<td>48</td>
<td>17</td>
<td>81</td>
<td>29</td>
<td>47</td>
<td>75</td>
<td>94</td>
<td>57</td>
<td>135</td>
<td>54</td>
<td>92</td>
<td>96</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>11</td>
<td>21</td>
<td>9</td>
<td>37</td>
<td>12</td>
<td>26</td>
<td>31</td>
<td>83</td>
<td>61</td>
<td>110</td>
<td>51</td>
<td>91</td>
<td>46</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>11</td>
<td>21</td>
<td>9</td>
<td>37</td>
<td>12</td>
<td>26</td>
<td>31</td>
<td>83</td>
<td>61</td>
<td>110</td>
<td>51</td>
<td>91</td>
<td>46</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>12</td>
<td>24</td>
<td>9</td>
<td>46</td>
<td>16</td>
<td>28</td>
<td>37</td>
<td>90</td>
<td>61</td>
<td>125</td>
<td>55</td>
<td>92</td>
<td>58</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>12</td>
<td>24</td>
<td>9</td>
<td>46</td>
<td>16</td>
<td>28</td>
<td>37</td>
<td>90</td>
<td>61</td>
<td>125</td>
<td>55</td>
<td>92</td>
<td>58</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>6</td>
<td>11</td>
<td>4</td>
<td>17</td>
<td>1</td>
<td>27</td>
<td>34</td>
<td>80</td>
<td>29</td>
<td>61</td>
<td>23</td>
<td>36</td>
<td>25</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The purpose:** Exploring ‘natural’ groups of prospects may trigger ideas
Exploration Prospect Appraisal - Clustering

Clustering method: Ward
Coefficient: Squared Euclidean Distance

Clustering method: Centroid
Coefficient: Squared Euclidean Distance

Clustering method: Average Linkage
Coefficient: Squared Euclidean Distance

Clustering method: Complete Linkage
Coefficient: Squared Euclidean Distance

1. Not very distinct clusters
2. Review data to remove non-discriminatory data
3. Rerun and review

Cluster analysis using Spyder / Anaconda
Scipy.cluster.hierarchy.dendrogram
## Well Curves - The DATA

### Depth, SGRC, SGRA, SGRB, SEXP, SEP, SEMP, SEDP, SEXC, SESC, SEMC, SEDC, STEM, SDDE, SPLF, SNA, SNFA, SBD, SCOR, SBD2, SCO2, SNEB, SFBD, SNPE, SHSI (ft), (api), (api), (ohmm), (ohmm), (ohmm), (ohmm), (ohmm), (ohmm), (ohmm), (deqf), (ptpf), (in), (c), (g/cc), (in)

<table>
<thead>
<tr>
<th>Depth</th>
<th>SGRC</th>
<th>SGRA</th>
<th>SGRB</th>
<th>SEXP</th>
<th>SEP</th>
<th>SEMP</th>
<th>SEDP</th>
<th>SEXC</th>
<th>SESC</th>
<th>SEMC</th>
<th>SEDC</th>
<th>STEM</th>
<th>SDDE</th>
<th>SPLF</th>
<th>SNA</th>
<th>SNFA</th>
<th>SBD</th>
<th>SCOR</th>
<th>SBD2</th>
<th>SCO2</th>
<th>SNEB</th>
<th>SFBD</th>
<th>SNPE</th>
<th>SHSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ft]</td>
<td>[api]</td>
<td>[api]</td>
<td>[api]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
<td>[ohmm]</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10291</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>120.34</td>
<td>975.0</td>
<td>0.09</td>
<td>0.09</td>
<td>36.32</td>
<td>194.42</td>
<td>142.46</td>
<td>-999.25</td>
<td>0.5</td>
<td>2440</td>
<td>475</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
</tr>
<tr>
<td>10292</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>105.39</td>
<td>981.11</td>
<td>0.09</td>
<td>0.09</td>
<td>34.45</td>
<td>193.68</td>
<td>136.26</td>
<td>-999.25</td>
<td>0.5</td>
<td>2445</td>
<td>475</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>10292.5</td>
</tr>
<tr>
<td>10293</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>52.02</td>
<td>952.05</td>
<td>0.09</td>
<td>0.09</td>
<td>28.88</td>
<td>188.31</td>
<td>142.95</td>
<td>-999.25</td>
<td>0.49</td>
<td>2467</td>
<td>472</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>10293.5</td>
</tr>
<tr>
<td>10294</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>64.08</td>
<td>1005.84</td>
<td>0.09</td>
<td>0.09</td>
<td>17.79</td>
<td>190.26</td>
<td>144.8</td>
<td>-999.25</td>
<td>0.49</td>
<td>2475</td>
<td>482</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>10294.5</td>
</tr>
<tr>
<td>10295</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>125.01</td>
<td>886.83</td>
<td>0.09</td>
<td>0.09</td>
<td>13.62</td>
<td>192.59</td>
<td>159.28</td>
<td>5.66</td>
<td>0.48</td>
<td>2471</td>
<td>487</td>
<td>2.24</td>
<td>-0.29</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>2.52</td>
<td>-999.25</td>
<td>8.5</td>
</tr>
<tr>
<td>10296</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>214.79</td>
<td>848.52</td>
<td>0.09</td>
<td>0.09</td>
<td>10.03</td>
<td>194.46</td>
<td>145.93</td>
<td>32.03</td>
<td>0.48</td>
<td>2466</td>
<td>490</td>
<td>2.23</td>
<td>-0.29</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>2.52</td>
<td>-999.25</td>
<td>8.5</td>
</tr>
<tr>
<td>10297</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>382.27</td>
<td>976.65</td>
<td>0.09</td>
<td>0.09</td>
<td>5.39</td>
<td>266.43</td>
<td>51.08</td>
<td>0.48</td>
<td>2491</td>
<td>491</td>
<td>2.23</td>
<td>-0.29</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>2.52</td>
<td>-999.25</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>10298</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>0.06</td>
<td>0.06</td>
<td>46.32</td>
<td>1565.17</td>
<td>0.09</td>
<td>0.09</td>
<td>0.29</td>
<td>2440</td>
<td>507.64</td>
<td>26.48</td>
<td>0.47</td>
<td>2467</td>
<td>497</td>
<td>2.23</td>
<td>-0.29</td>
<td>-999.25</td>
<td>-999.25</td>
<td>-999.25</td>
<td>2.52</td>
<td>-999.25</td>
<td>8.5</td>
</tr>
</tbody>
</table>

### Copyright © 2019 Precision-DM
Well Curves – Clustering

1. Some distinct clusters, majority of points are mixed
2. Review data to remove non-discriminatory data
3. Investigate end points. Rerun and review

Clustering method: Ward
Coefficient: Squared Euclidean Distance

Clustering method: Centroid
Coefficient: Squared Euclidean Distance

Clustering method: Average Linkage
Coefficient: Squared Euclidean Distance

Clustering method: Complete Linkage
Coefficient: Squared Euclidean Distance

Cluster analysis using Spyder / Anaconda
Scipy.cluster.hierarchy.dendrogram
Well Curves – Change of coefficient

Cluster analysis using Spyder / Anaconda
.Scipy.cluster.hierarchy.dendrogram

1. More distinct clusters, easier to differentiate
2. Investigate groups for significance
3. Review data for noise
Micropaleontology

Benthonic Foraminifera – Protozoa. Live(d) on the sea bottom. Size ~ 200-2000 microns
Best viewed with binocular microscope at 25x – 80x magnification

North West Borneo
Environmental Scheme
(Shell, 1970s)

Holomarine Inner Neritic
0 – 40m water depth

Holomarine Middle Neritic
40 – 100m water depth

Holomarine Middle Neritic
100 – 200m water depth

Fluviomarine realm
The purpose: Group samples belonging to the same environment of deposition based on species content
Micropaleontology – Well foraminiferal samples

Cluster analysis using Spyder / Anaconda
Scipy.cluster.hierarchy.dendrogram

1. Some distinct clusters, mostly mixed
2. Investigate groups for significance
3. Review data for noise
Data Science opportunities – Paleoenvironmental reconstruction

Stratigraphy
- Litho, bio, chrono
- Sea level changes
- Flooding surfaces

Structural
- Faults
- Uplifts
- Eustatic
- Erosion
- Uplifts
- Missing sections

Sedimentary facies
- Types
- Characteristics
- Bedding, dips etc
- Log shape interpretation

Seismic
- Seismic features (seismostrat)
- Traces
- Checkshots
- Time-depth curve
- Vertical seismic profiling (VSP)

Well Logs
- Gamma ray
- Sonic
- Density
- Neutron
- Resistivities
- Caliper

Minerals
- Glauculite
- Siderite
- Pyrite
- Mica

Paleontology
- Benthics
- Planktonics
- Larger forams
- Nannofossils
- Palynology
- Ostracods
- Trace fossils
Data Science opportunities—Source Rocks

**Pressure**
- Spot readings
- Trends

**Temperature**
- Sample readings
- Gradients

**Surrounding wells**
- Well data
- Source rock distribution patterns
- Maps & trends

**Burial History**
- Sedimentation rates
- Sediment types
- Missing sections
- Palinspastic reconstruction

**Well Logs**
- Gamma ray
- Sonic
- Density
- Resistivities
- Caliper

**Sedimentary facies**
- Types
- Characteristics
- Bedding, dips etc
- Log shape interpretation

**Rock properties**
- Porosity
- Permeability
- Diagenesis

**Macerals**
- Organic type (Lip. vs Vit.)
- Kitchen area
- Migration paths
- Maturity levels (DOM, VR/E)

**Computer simulation**
- Methods (eg Migration Models)
- Probabilistic vs deterministic

**Paleontology**
- Benthics
- Planktonics
- Larger forams
- Nannofossils
- Palynology
- Ostracods

Copyright © 2019 Precision-DM
Data Science opportunities – Prospect appraisal

Temperature
- Sample readings
- Gradients

Pressure
- Spot readings
- Trends

Surrounding wells
- Well data
- Correlation
- Maps & trends

Rock properties
- Porosity
- Permeability
- Diagenesis

Analogues
- Local comparators
- Regional
- Global

Sedimentary facies
- Sediment types
- Characteristics
- Bedding, dips etc
- Log shape interpretation

Structural
- Faults
- Closures
- Seals

Burial History
- Sedimentation rates
- Sediment types
- Missing sections
- Palinspastic reconstruction

Well Logs
- Gamma ray
- Sonic
- Density
- Neutron
- Resistivities
- Caliper

Computer simulation
- Methods (eg Monte Carlo)
- Probabilistic vs deterministic

Paleontology
- Benthics
- Planktonics
- Larger forams
- Nannofossils
- Palynology
- Ostracods

Source Rocks
- Type (lip. vs vit.)
- Kitchen area
- Maturity

Copyright © 2019 Precision-DM
Conclusions

• Machine learning is not a black box

• Understand the ML workflow components, behaviors and limitations

• Look at the DATA

• Give importance to feature selection & feature extraction

• Look at the results

• Look at the DATA again
Questions