Anna Nyulund
Past Projects Overview

Estimation of well production

Develop and demonstrate predictive model to estimate the production of wells within a given filed based on validation wells datasets and use this model to optimize well placement to maximize field production value.

- Targets: 6 Mo Cum Prod Gas (mmcf)
- 6 Mo Cum Prod Oil (mbo)
- 6 Mo Cum Prod Con (mbc)

Estimation of well production

- Data In: Unidentified well logs plus

IP Gas (mcf/d)	12 Mo Cum Prod Oil (mbo)	6 Mo Oil Rate per Stage (bopd)
IP Oil (bopd)	12 Mo Cum Prod Water (mbw)	6 Mo Cal Oil Rate (mcf/d)
IP Water (bwpd)	12 Mo Cum Prod Cond (mbc)	6 Mo Oil Rate (bopd)
IP Cond(bcpd)	18 Mo Cum Prod Oil (mbo)	12 Mo Cal Water Rate (bwpd)
Pres Prod Water (bwpd)	18 Mo Cum Prod Water (mbw)	12 Mo Cal Fluid Rate (bpd)
Pres Prod Cond (bcpd)	18 Mo Cum Prod Cond (mbc)	12 Mo Gas Rate per Stage (mcf/d)
Cum Prod Gas (mmcf)	6 Mo Cal Gas Rate (mcf/d)	12 Mo Cal Cond Rate (bcpd)
Cum Prod Cond (mbc)	6 Mo Cal Oil Rate (bopd)	12 Mo Oil Rate per Stage (bopd)

Possible models to use

- Random forest
- XG-boost
- Linear regression
- Feed forward neural net

NLP: keyword matching

- Resume classification based on skill

resume_key_words

Statistics	Machine Learning	Deep Learning	R Language	Python Language	NLP	Data Engineering
statistical models	linear regression	neural network	r	python	nlp	aws
statistical modeling	logstic regression	keras	ggplot	flask	natural language processing	ec2
probability	K means	pytorch	shiny	django	topic modeling	amazon redshift
normal distribution	random forest	theano	cran	pandas	lda	s3
poisson distribution	xgboost	face detection	dplyr	numpy	named entity recognition	docker
survival models	svm	neural networks	tidyr	scikitlearn	pos tagging	kubernetes
hypothesis testing	naive bayes	convolutional neural network (CNN)	lubridate	sklearn	word2vec	scala
bayesian inference	pca	recurrent neural network (RNN)	knitr	matplotlib	word embedding	teradata
factor analysis	decision trees	object detection		scipy	lsi	google big query
forecasting	svd	yolo		bokeh	spacy	aws lambda
markov chain	ensemble models	gpu		statsmodel	gensim	aws emr
monte carlo	boltzman machine	cuda		seaborn	nltk	hive
		tensorflow			nmf	hadoop
		lstm			doc2vec	sql
		gan			cbow	
		opencv			bag of words	
					skip gram	
					bert	
					sentiment analysis	
					chat bot	

NLP: Named Entity Recognition

NAMED ENTITY RECOGNITION (NER)

- **Conditional Random Fields (CRF)**
- **Spacy, Flair (Torch), NLTK**
- **Train custom NER on top of Spacy**
- **Human labeling**
  - **Dataturks**
  - **Amazon Mechanical Turk**

```
Name: [ANNA NYULUND]
Organizaton: []
Location: [Austin, Texas]
Phone or email: [redacted]
```

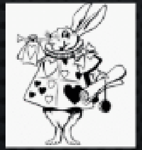
```
[All Names:          0          1
0         ANNA     NYULUND
1     Spotfire    Engineer
2         Iron      Python
3     OKLAHOMA        CITY
4         Data   Scientist
5    Reservoir    Engineer
6       Planit        None
7     Spotfire        None
[8      Fekete         RTA
9       Gohfer        None
10      Austin        None
```

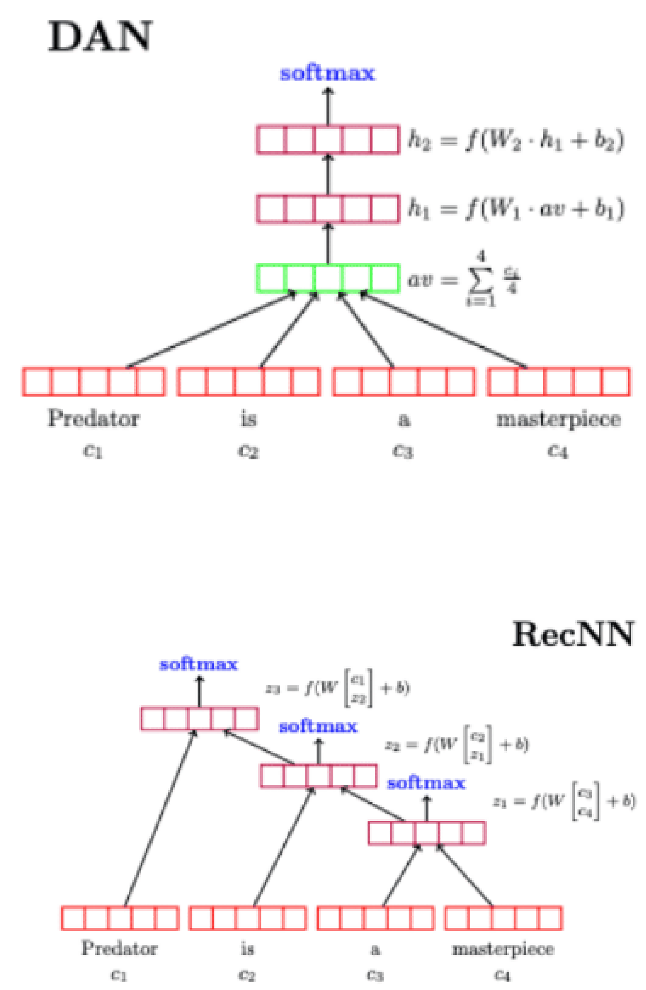
NLP: Semantic Analysis

SENTIMENT ANALYSIS DEEP AVERAGING NETWORK (DAN)

- PROVIDES STATE-OF-ART ACCURACIES ON A VARIETY OF SENTENCE AND DOCUMENT-LEVEL TASKS WITH JUST MINUTES OF TRAINING TIME AND AN AVERAGE LAPTOP COMPUTER
- TAKES THE VECTOR AVERAGE OF THE EMBEDDINGS ASSOCIATED WITH AN INPUT SEQUENCE OF TOKENS
- PASSES THE AVERAGE THROUGH ONE OR MORE FEED FORWARD LAYERS
- PERFORMS CLASSIFICATION ON THE FINAL LAYER'S REPRESENTATION
- REFERENCE: DEEP UNORDERED COMPOSITION RIVALS SYNTACTIC METHODS FOR TEXT CLASSIFICATION

- SOFTMAX: TAKES AN INPUT OF K REAL NUMBERS, AND NORMALIZES IT INTO A PROBABILITY DISTRIBUTION CONSISTING OF K PROBABILITIES.





NLP: Semantic Analysis

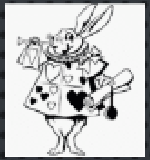
```
1795 / 224 / 225 train/dev/test examples
Read in 17615 vectors of size 300
=====Train Accuracy=====
Accuracy: 1465 / 1795 = 0.816156
Precision: 1083 / 1308 = 0.827982
Recall: 1083 / 1188 = 0.911616
F1: 0.867788
=====Dev Accuracy=====
Accuracy: 180 / 224 = 0.803571
Precision: 136 / 166 = 0.819277
Recall: 136 / 150 = 0.906667
F1: 0.860759
Time for training and evaluation: 9.98 seconds
```

DEEP AVERAGING NETWORK (DAN) EVALUATION

- **ACCURACY = TP + TN/TP+FP+FN+TN**
- **PRECISION = TP/TP + FP**
- **RECALL = TP/TP + FN**
- **F1 SCORE = 2*(RECALL*PRECISION)/(RECALL + PRECISION)**

NLP: Semantic Analysis

DEEP AVERAGING NETWORK (DAN) RESULTS



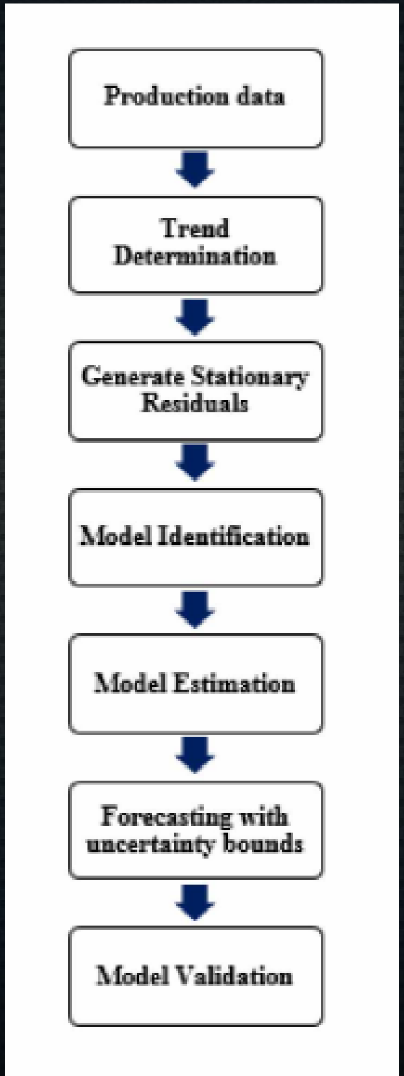
- Pretrained on Netflix Movie Reviews
- Used PyTorch
- Ran on oil-and-gas contractor reviews
- Challenge: oil-and-gas and work skill review language: hardworking, problem solver, consistent, driller, etc. and algorithmic bias.

- Label and retrain
- Regularize and redirect word embeddings
- Select different model (BERT – Bidirectional Encoder Representations from Transformers)
- Experiment with different word embeddings

```
1    Learns fast, great attention to detail. Performs well in high stress situations.
1    Very smart guy , Good on his feet hard worker, Great Employee , Good people , Always make it to work very dependable
0    Consistent performer
0    [redacted] is in my opinion, a go to guy for problems on workovers.
0    [redacted] is one of the hardest workers I've ever had the pleasure to work with.
```

Barnett Shale Gas production: ARMA

-



Barnett Shale Gas production: ARMA

- Detrend (subtract the deterministic component from the production data) using LGA (Logistic Growth Analysis – modified by Clark (2011)for monthly production per unit time. Represents hyperbolic decline of production from extremely low permeability oil and gas wells :

$$q(t) = \frac{dQ}{dt} = \frac{Knat^{n-1}}{(a+t^n)^2}$$

Barnett Shale Gas production: ARMA

- OLS (ordinary least squares) and WLS (weighted least squares) regressions to fit production data set for LGA)
- Autoregressive Moving average

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

Petroleum Producing Assets Portfolio

- Price forecast based on Sequential Gaussian Simulation

  Holmes et al. 2006
- Monte Carlo to simulate the distribution of reserves
- Estimation of after tax net cash flows
- Estimation of covariance between return distributions and combination into portfolios

Monte-Carlo to estimate the reserves

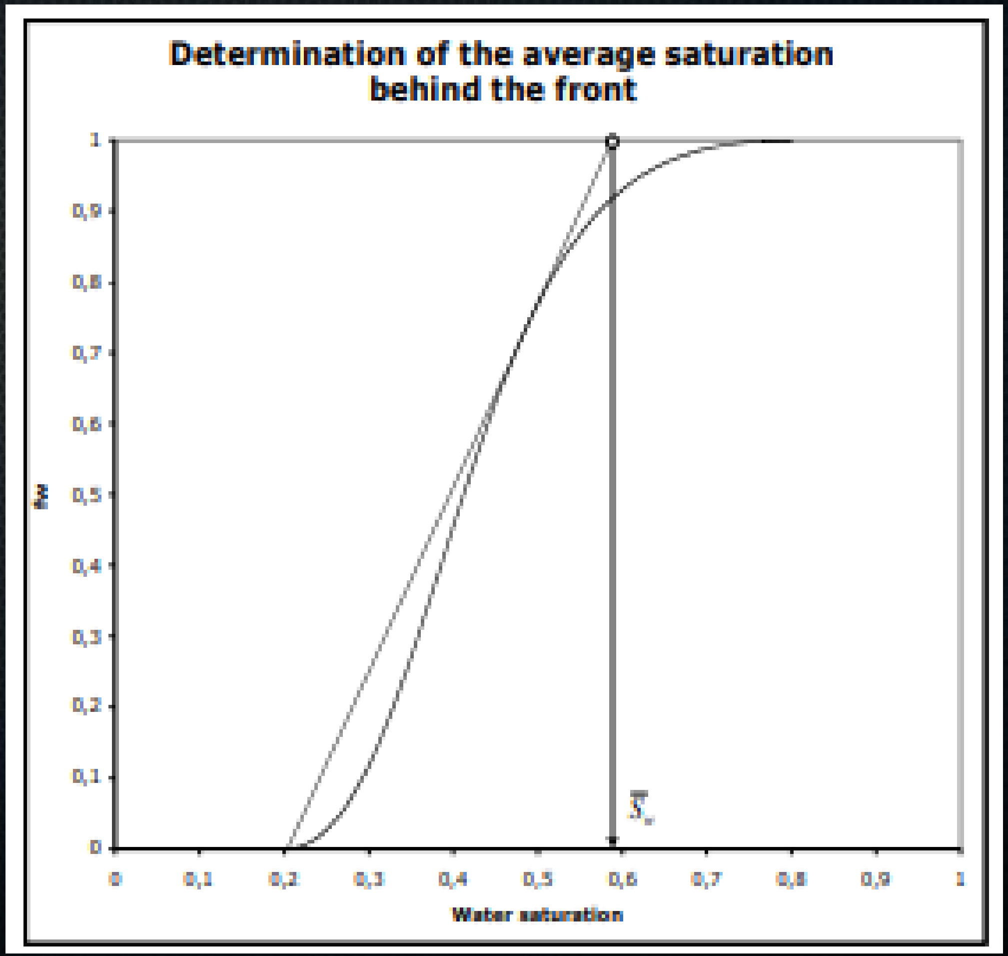
- Volumetric method to estimate the reserves

  $$\text{Reserves} = A\ h\ \varphi\ (1 - S_{wi})/B_o \times RF$$

Each input is given a probability distribution, then samples are randomly drawn from these pre-defined distributions and combined to obtain a possible output for realization.

EOR: Buckley-Leverett Method

- Estimates fluid displacement front in an immiscilble displacement process
- Used for:
- Water displacement
- Surfactant
-



Oil price forecast: SGS – Holmes 2006

- If the price data is not univariate normal, transform the data to obtain normal scored prices.
- Construct a model of temporal continuity or the normal scored data
- Define a random path through all of the months to be simulated, that is each month exactly once
- Use Kriging to determine the mean and variance of the Gaussian conditional probability distribution at a given month. Retain a specified number of neighbouring data to be used as conditioning data.

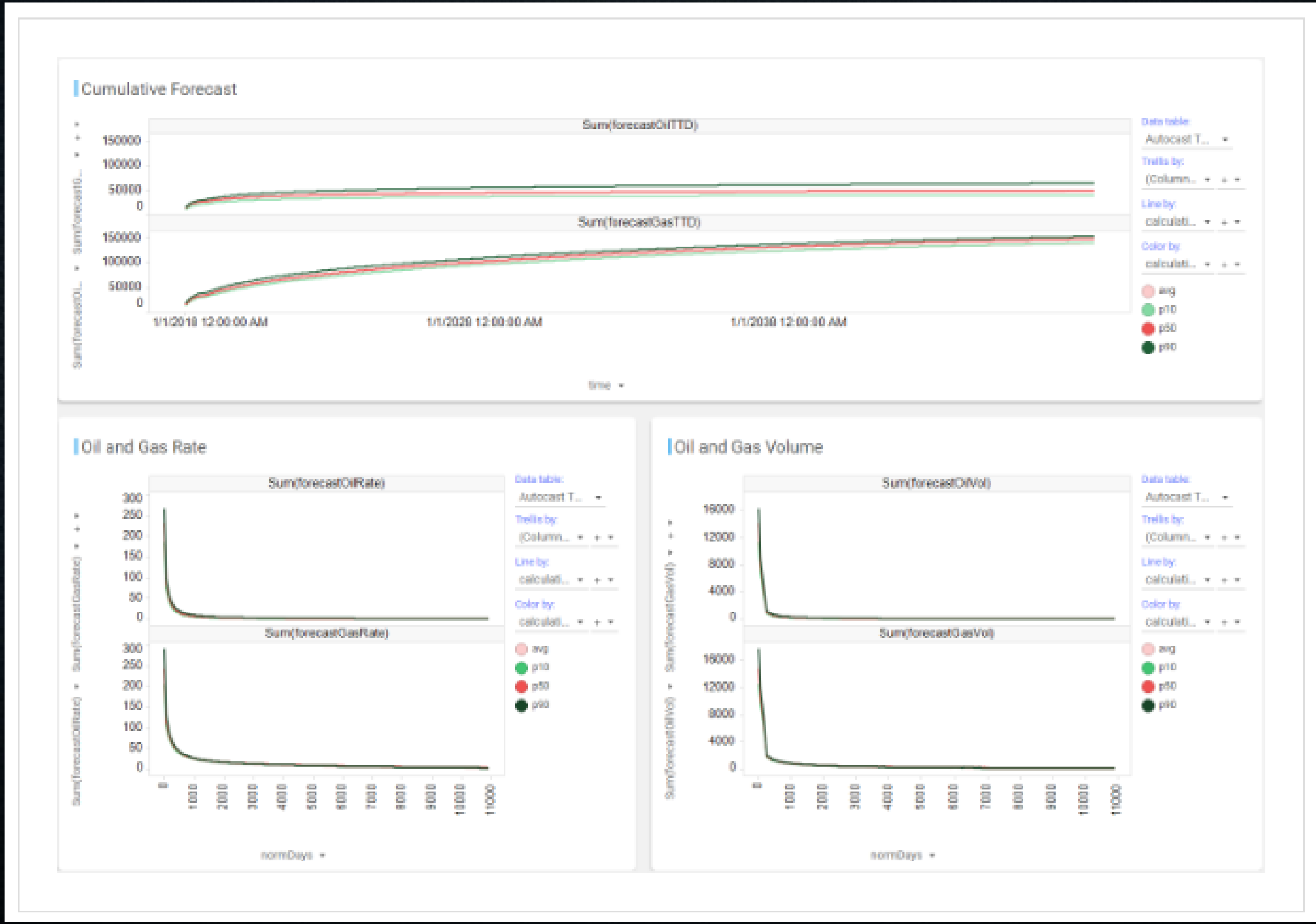
Oil price forecast: SGS – Holmes 2006

- Draw randomly from conditional probability distribution and assign that value to the node being simulated.
- Repeat previous two steps for all simulation minths.
- Back transform the simulate normal values into the original price values using original inflation adjusted price data distribution.
- Repeat steps 3-7 for multiple realizations.

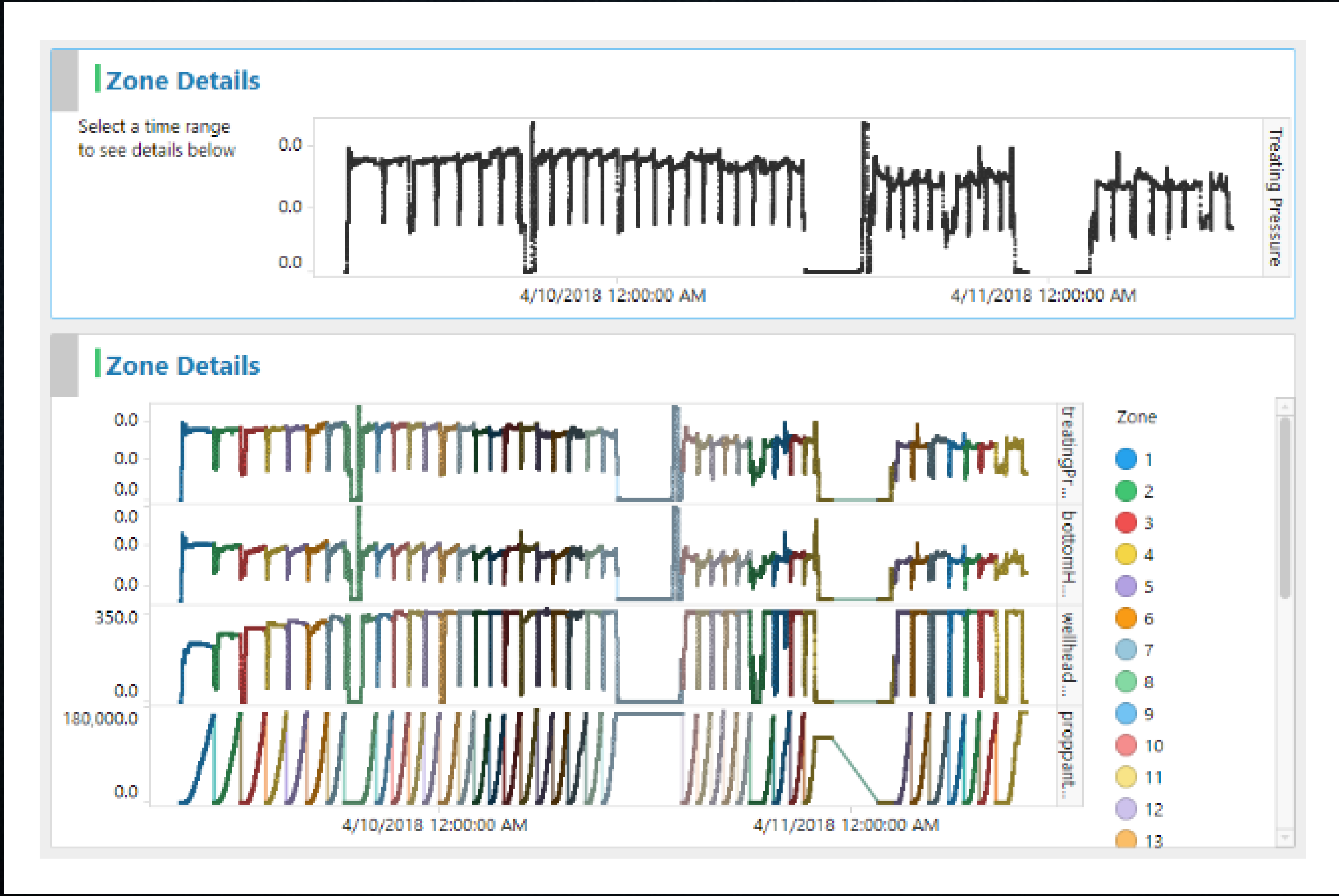
Formation classification: clustering

- Use k-means clustering to break the data collected from well logs into subgroups, which indicate different formations.
- Examples of the properties include: gamma ray, ressistivity, shellity, water saturation, porosity, permeability, bulk density and sonic logs.
- Aftere k-mean clustering algorithm is excecuted, it clearly maps different formations into clusters, which then can be mapped based on TVD.

Spotfire examples: Decline Curve Analysis

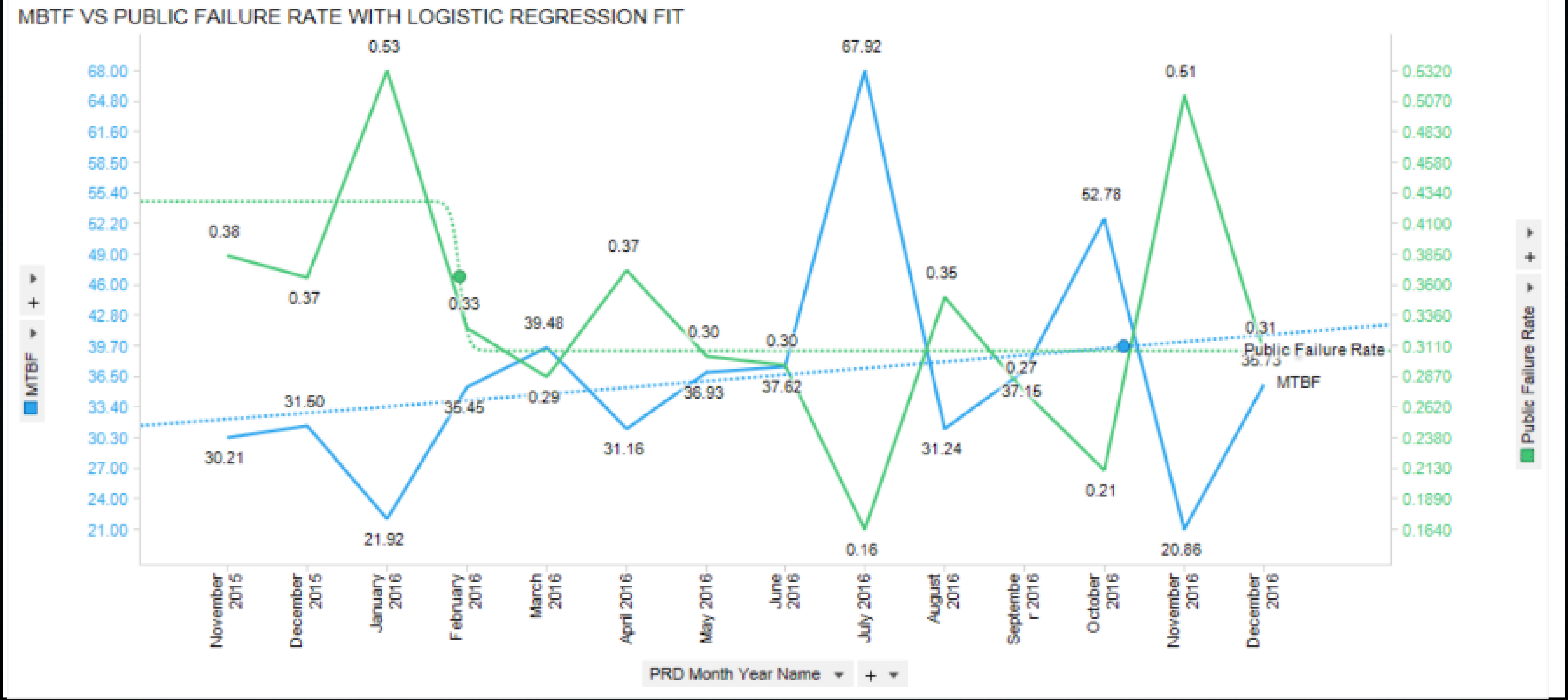


Spotfire Exampes: Completions Book



2020

LOE: Lease Operating Expenses – Well Failure Rates



Thank you