Webinars

A Hybrid Deep Learning Framework for Unsupervised Detection, Assessment and Correction of Real-Time Drilling Sensor Data

Speaker : James Omeke

Team

- Abdulmalik Taj-Liad
- Abdulrasaq Bilau
- Stephen Ibrahim
- Achumba Gabriel

SPE International

Webinars

Content

Introduction

- General overview
- Proposed concepts

Data Preprocessing and Exploratory Data Analysis

- XML to CSV data conversion web app
- Extracted data visualization from Volve field
- Univariate and bi-variate data analysis

Machine Learning (ML) / Artificial Intelligence (AI)

- Screening of ML or AI Algorithm to Apply
- Autoencoders + Isolation Forest
- Principle of Anomaly Detection By Autoencoders

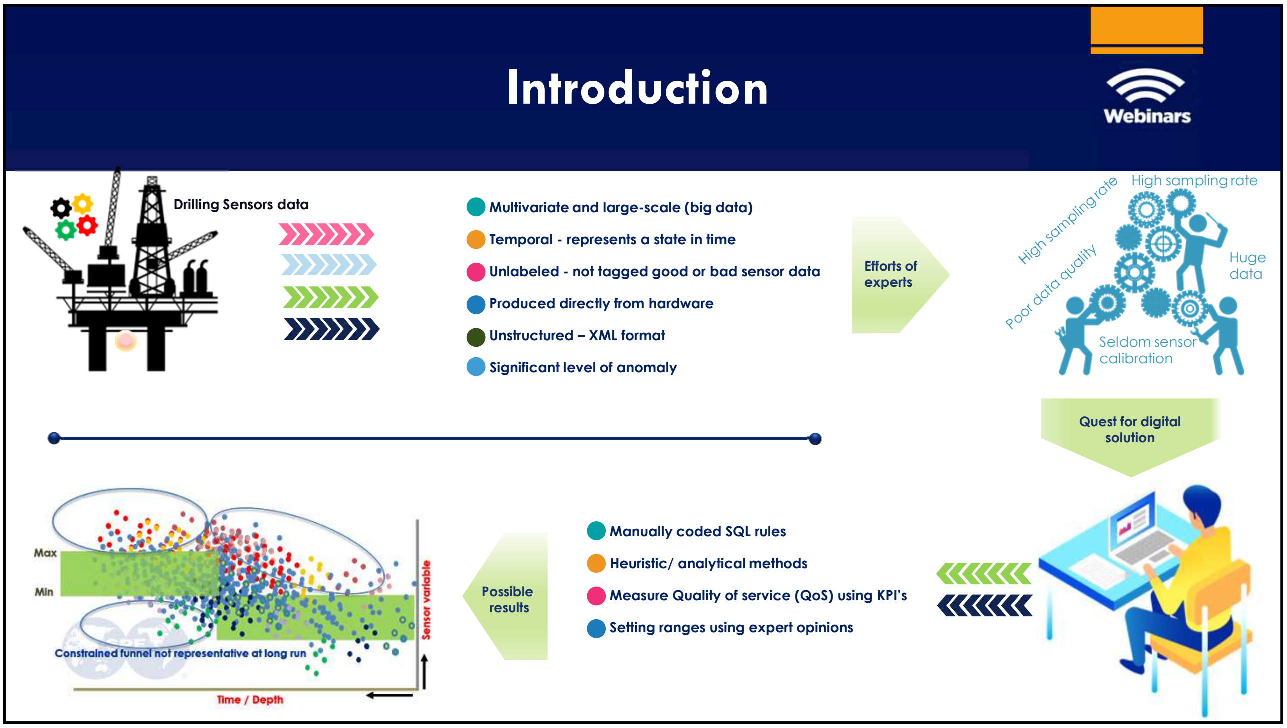
Model Results and Evaluations

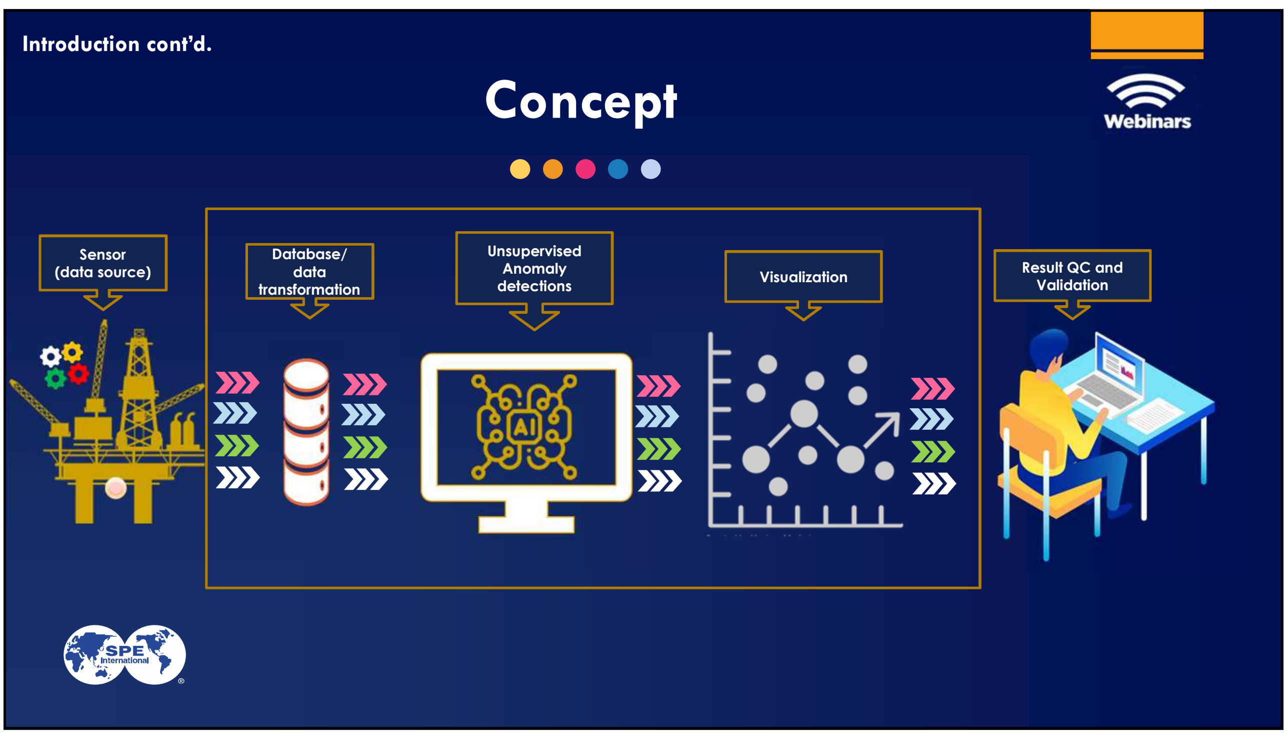
- Training, Testing and Validation of AENN and RNN
- RNN and AENN Model Evaluation
- Training and testing vs prediction
- Uncertainty scoring

Conclusion & Recommendation

- Conclusion and recommendation

SPE International


Introduction
Webinars
Drilling Sensors data
Multivariate and large-scale (big data)
Temporal - represents a state in time
Unlabeled - not tagged good or bad sensor data
Produced directly from hardware
Unstructured – XML format
Significant level of anomaly
Efforts of experts
High sampling rate
High sampling rate
Poor data quality
Huge data
Seldom sensor calibration
Quest for digital solution
Max
Min
Constrained funnel not representative at long run
Time / Depth
Sensor variable
Possible results
Manually coded SQL rules
Heuristic/ analytical methods
Measure Quality of service (QoS) using KPI's
Setting ranges using expert opinions


Introduction cont'd.
Concept
Webinars
Sensor (data source)
Database/ data transformation
Unsupervised Anomaly detections
Visualization
Result QC and Validation
AI
SPE International

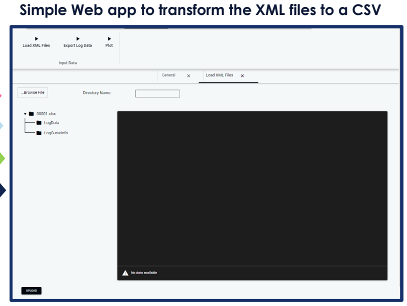
Data Preprocessing and Exploratory Data Analysis

Data Pre-processing

(Real-time drilling data logs obtained from the volve field – Published by Equinor)

Data stored as Extensible Markup Language (XML) in multiple files

Simple Web app to transform the XML files to a CSV

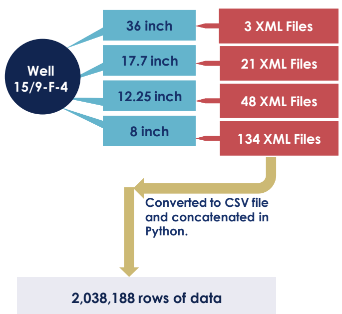


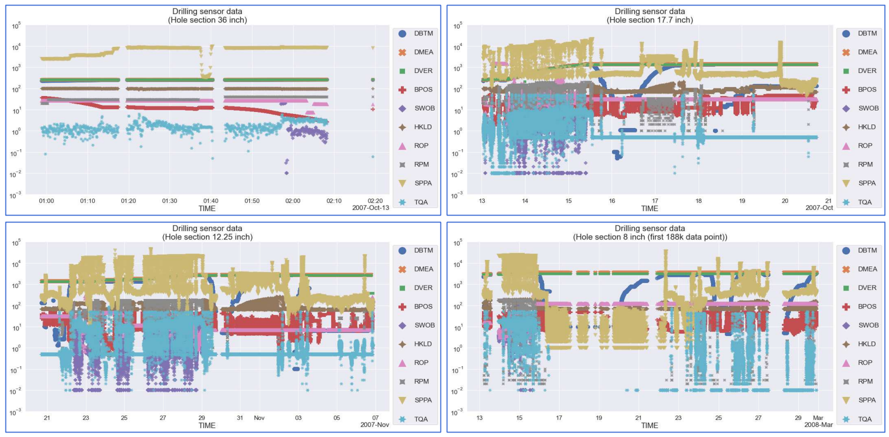
	Variable	Unite	Description
0	DBTM	m	Bit Depth (MD)
1	SWOB	kkgf	Weight on Bit
2	BPOS	m	Block Position
3	RPM	rpm	Average Rotary Speed
4	HKLD	kkgf	Average Hookload
5	DVER	m	Hole Depth (TVD)
6	ROP	m/h	Rate of Penetration
7	DMEA	m	Hole depth (MD)
8	TQA	kN.m	Average Surface Torque
9	SPPA	kPa	Average Standpipe Pressure

Exploratory Data analysis (EDA)

Webinars

Visual Exploration of Real time MWD/LWD time log data from well 15/9-F-4





Visualization of the 10 Sensor Variables vs Time for different hole sections in Well 15/9-F-4

SPE International

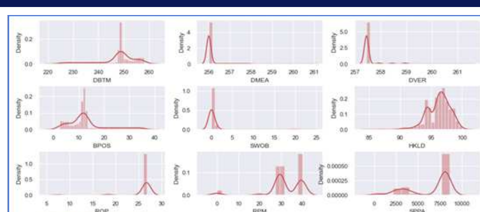
Exploratory Data analysis (EDA)

Webinars

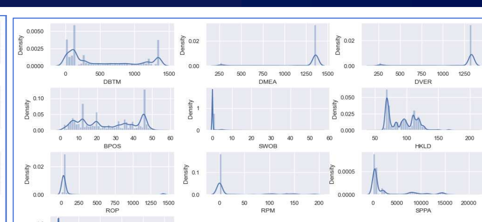
Univariate data analysis (Distribution plots)

The reason for carrying out a univariate data analysis is to establish if there exist a consistent distribution of each of the sensor variable when moving from one hole section to another

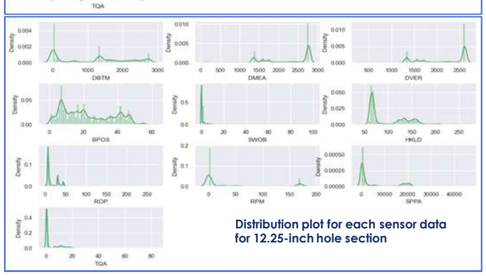
	Variable	Unite	Description
0	DBTM	m	Bit Depth (MD)
1	SWOB	kkgf	Weight on Bit
2	BPOS	m	Block Position
3	RPM	rpm	Average Rotary Speed
4	HKLD	kkgf	Average Hookload
5	DVER	m	Hole Depth (TVD)
6	ROP	m/h	Rate of Penetration
7	DMEA	m	Hole depth (MD)
8	TQA	kN.m	Average Surface Torque
9	SPPA	kPa	Average Standpipe Pressure



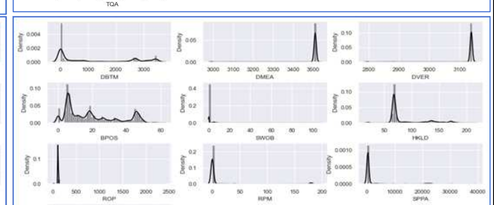
Distribution plot for each sensor data for 36-inch hole section



Distribution plot for each sensor data for 17.7-inch hole section



Distribution plot for each sensor data for 12.25-inch hole section



Distribution plot for each sensor data for 8-inch hole section

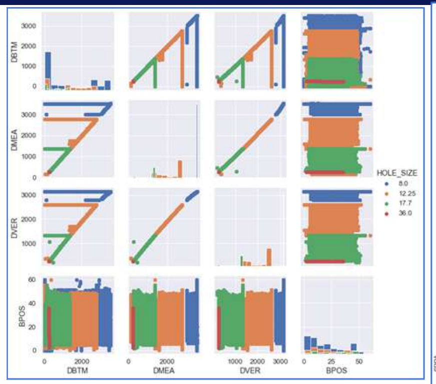
Exploratory Data analysis (EDA)

Webinars

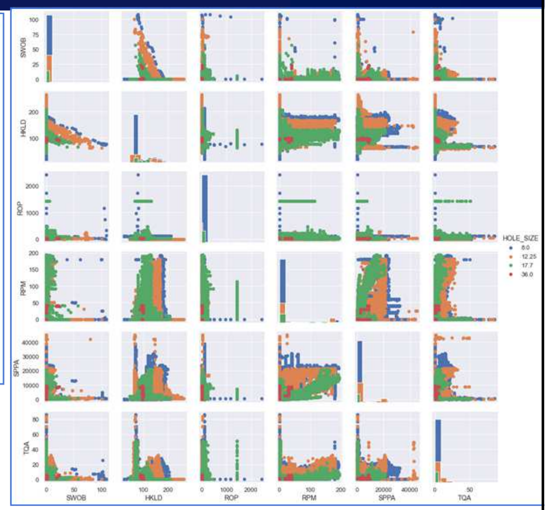
Bivariate data analysis (Grid plots)

The purpose of the bi-variate analysis is to establish if there is a relationship between two sensor variables or if they are truly independent

	Variable	Unite	Description
0	DBTM	m	Bit Depth (MD)
1	SWOB	kkgf	Weight on Bit
2	BPOS	m	Block Position
3	RPM	rpm	Average Rotary Speed
4	HKLD	kkgf	Average Hookload
5	DVER	m	Hole Depth (TVD)
6	ROP	m/h	Rate of Penetration
7	DMEA	m	Hole depth (MD)
8	TQA	kN.m	Average Surface Torque
9	SPPA	kPa	Average Standpipe Pressure



Bivariate analysis of all depth and position related data for different hole sections

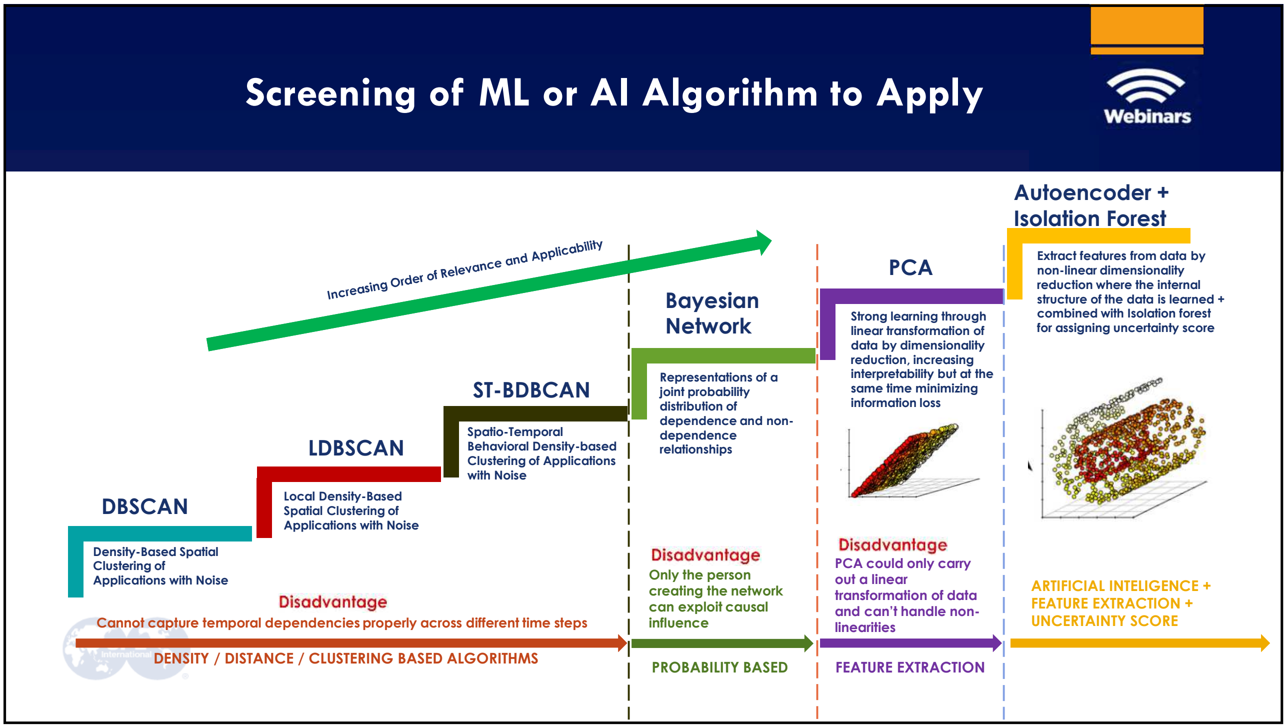


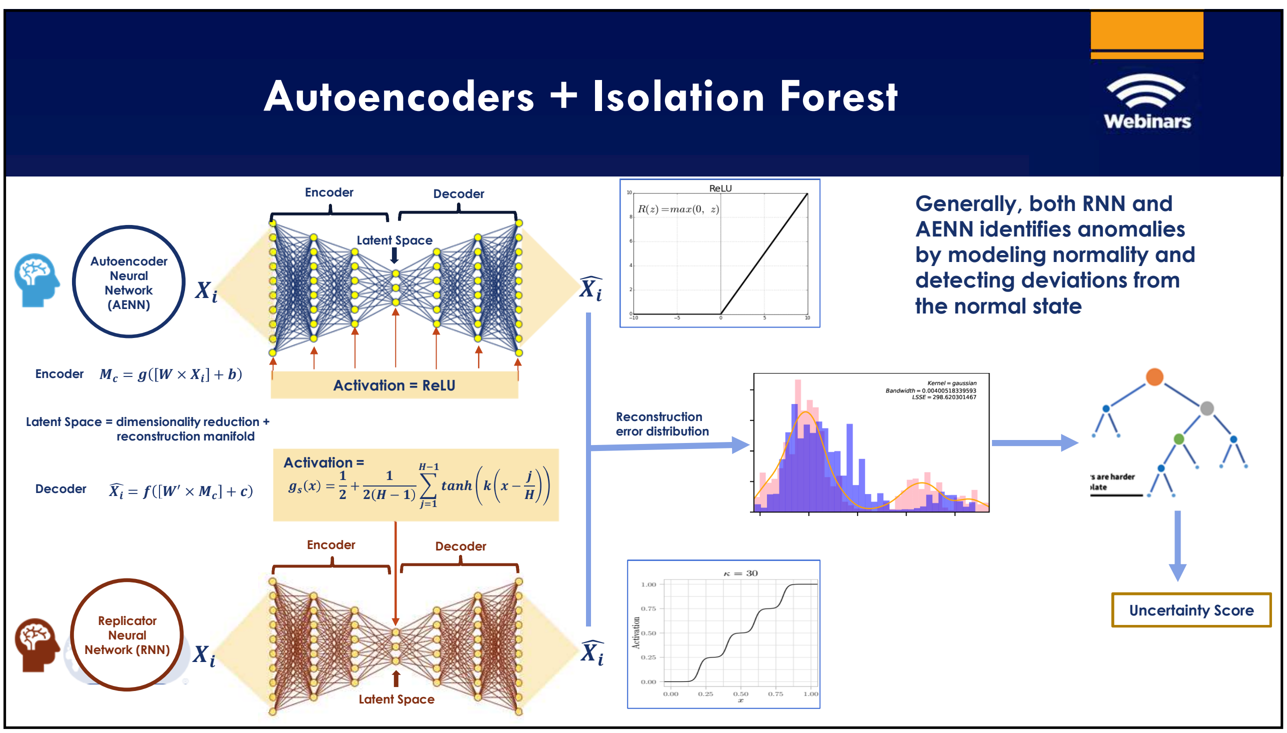
Bivariate analysis of other sensor variables for different hole section

Webinars

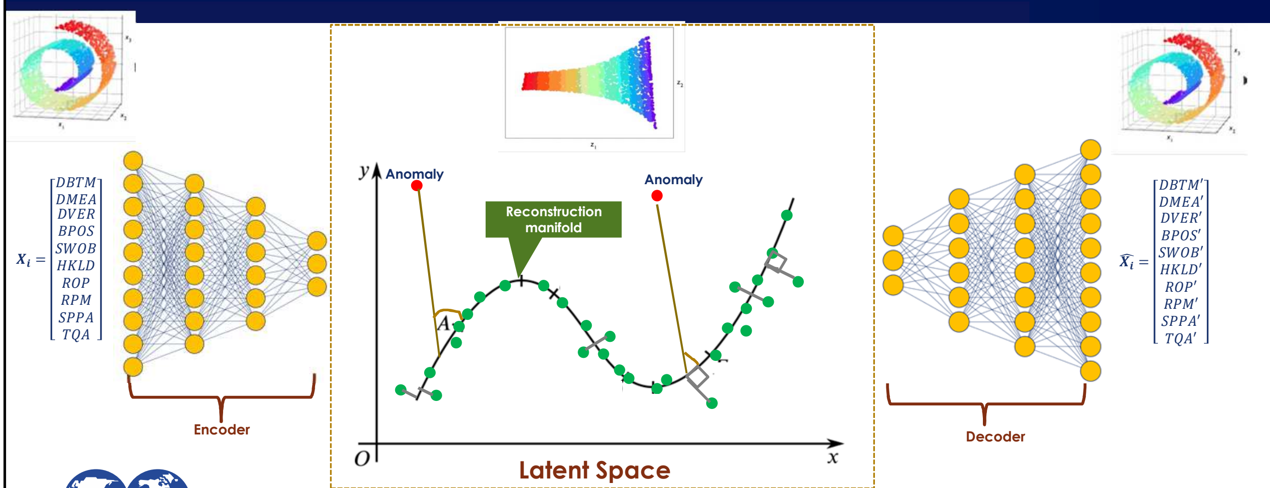
Machine Learning (ML) / Artificial Intelligence (AI)

SPE International


Screening of ML or AI Algorithm to Apply
Webinars
Increasing Order of Relevance and Applicability
DBSCAN
Density-Based Spatial Clustering of Applications with Noise
LDBSCAN
Local Density-Based Spatial Clustering of Applications with Noise
ST-BDBCAN
Spatio-Temporal Behavioral Density-based Clustering of Applications with Noise
Disadvantage
Cannot capture temporal dependencies properly across different time steps
DENSITY / DISTANCE / CLUSTERING BASED ALGORITHMS
Bayesian Network
Representations of a joint probability distribution of dependence and non-dependence relationships
Disadvantage
Only the person creating the network can exploit causal influence
PROBABILITY BASED
PCA
Strong learning through linear transformation of data by dimensionality reduction, increasing interpretability but at the same time minimizing information loss
Disadvantage
PCA could only carry out a linear transformation of data and can't handle non-linearities
FEATURE EXTRACTION
Autoencoder + Isolation Forest
Extract features from data by non-linear dimensionality reduction where the internal structure of the data is learned + combined with Isolation forest for assigning uncertainty score
ARTIFICIAL INTELIGENCE + FEATURE EXTRACTION + UNCERTAINTY SCORE


Autoencoders + Isolation Forest
Webinars
Autoencoder Neural Network (AENN)
X_i
Encoder
Decoder
Latent Space
Activation = ReLU
$\widehat{X}_i$
Encoder $M_c = g([W \times X_i] + b)$
Latent Space = dimensionality reduction + reconstruction manifold
Decoder $\widehat{X}_i = f([W' \times M_c] + c)$
Activation = $g_s(x) = \frac{1}{2} + \frac{1}{2(H-1)} \sum_{j=1}^{H-1} tanh\left(k\left(x - \frac{j}{H}\right)\right)$
Replicator Neural Network (RNN)
X_i
Encoder
Decoder
Latent Space
$\widehat{X}_i$
ReLU
$R(z) = max(0, z)$
$\kappa = 30$
Activation
Reconstruction error distribution
Generally, both RNN and AENN identifies anomalies by modeling normality and detecting deviations from the normal state
Uncertainty Score

Principle of Anomaly Detection By Autoencoders



Two main reasons that drive the ability of autoencoders to detect anomalies:

- Anomalies are not projected orthogonally onto the reconstruction manifold,
- Anomalies often have a larger distance to the reconstruction manifold than normal observation

Model Evaluation and Results

Training, Testing and Validation of AENN and RNN

Webinars

Autoencoder Neural Network (AENN)

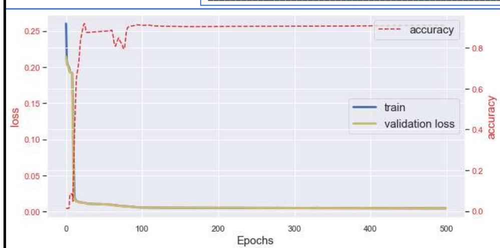
mean square error (mse) = 0.025

```
Layer (type)                 Output Shape              Param #
=================================================================
input (InputLayer)           (None, 10)                0
_________________________________________________________________
1st_encode_layer (Dense)     (None, 8)                 88
_________________________________________________________________
2nd_hiden_layer (Dense)      (None, 6)                 54
_________________________________________________________________
latent_layer (Dense)         (None, 3)                 21
_________________________________________________________________
4th_hiden_layer (Dense)      (None, 6)                 24
_________________________________________________________________
5th_hiden_layer (Dense)      (None, 8)                 56
_________________________________________________________________
output_decode (Dense)        (None, 10)                90
=================================================================
Total params: 333
Trainable params: 333
Non-trainable params: 0
_________________________________________________________________
```

Replicator Neural Network (RNN)

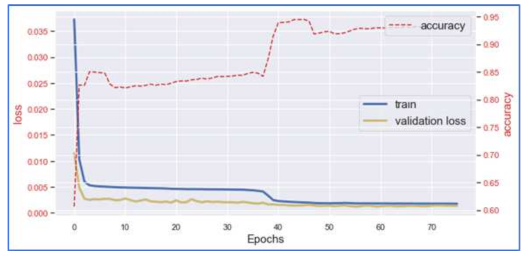
mean square error (mse) = 0.0015

```
Layer (type)                     Output Shape              Param #
=================================================================
rnn_input (InputLayer)           [(None, 10)]              0
_________________________________________________________________
1st_encode_rnn_layer (Dense)     (None, 8)                 88
_________________________________________________________________
2nd_hiden_layer (Dense)          (None, 6)                 54
_________________________________________________________________
latent_rnn_layer (Dense)         (None, 3)                 21
_________________________________________________________________
6th_decode_rnn_layer (Dense)     (None, 6)                 24
_________________________________________________________________
5th_hiden_layer (Dense)          (None, 8)                 56
_________________________________________________________________
output_rnn_decode (Dense)        (None, 10)                90
=================================================================
Total params: 333
Trainable params: 333
Non-trainable params: 0
_________________________________________________________________
```



Training Performance AENN

mse will not be used as measure of performance



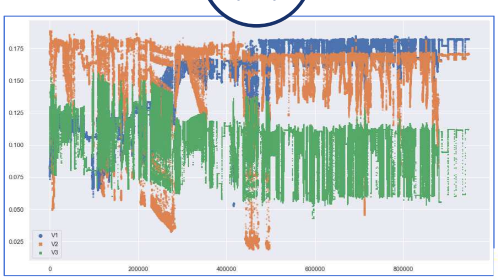
Training Performance RNN

RNN and AENN Model Evaluation
(Latent Space Manifold plot)

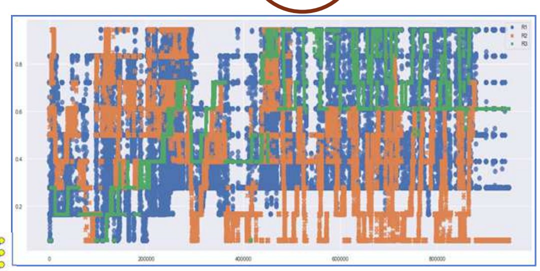
Webinars

Autoencoder Neural Network (AENN)

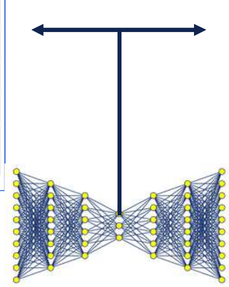
Replicator Neural Network (RNN)



The three-dimensional latent space manifold created by AENN & plotted against data points to see its evolution



The three-dimensional latent space manifold created by RNN & plotted against data points to see its evolution



	V1	V2	V3
V1	1.000000	0.420164	-0.224869
V2	0.420164	1.000000	-0.224476
V3	-0.224869	-0.224476	1.000000

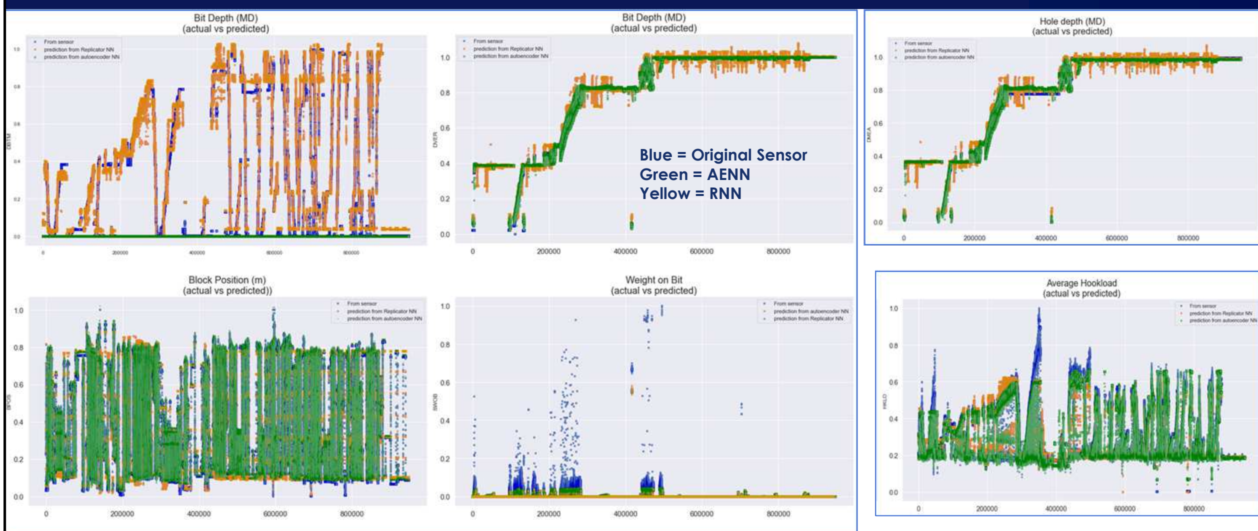
$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

$\sigma_X \sigma_Y = standard\ deviation\ of\ X\ and\ Y.$

	R1	R2	R3
R1	1.000000	0.030927	0.150011
R2	0.030927	1.000000	0.051378
R3	0.150011	0.051378	1.000000

Sensor Training Data Vs Prediction Of RNN And AENN

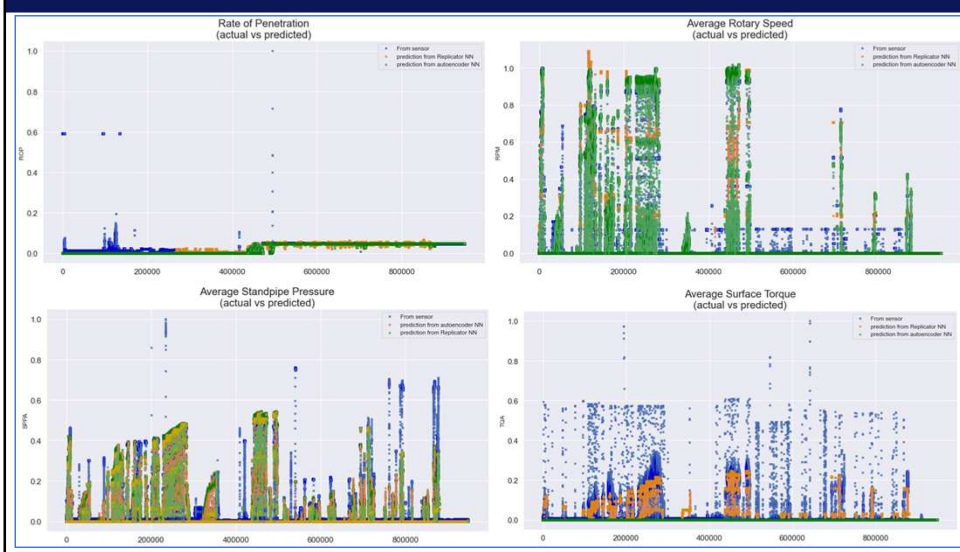
(Assessing the similarity in data structure learned by both algorithm)



Blue = Original Sensor, Green = AENN, Orange = RNN

Sensor Training Data Vs Prediction Of RNN And AENN – continued

(Assessing the similarity in data structure learned by both algorithm)

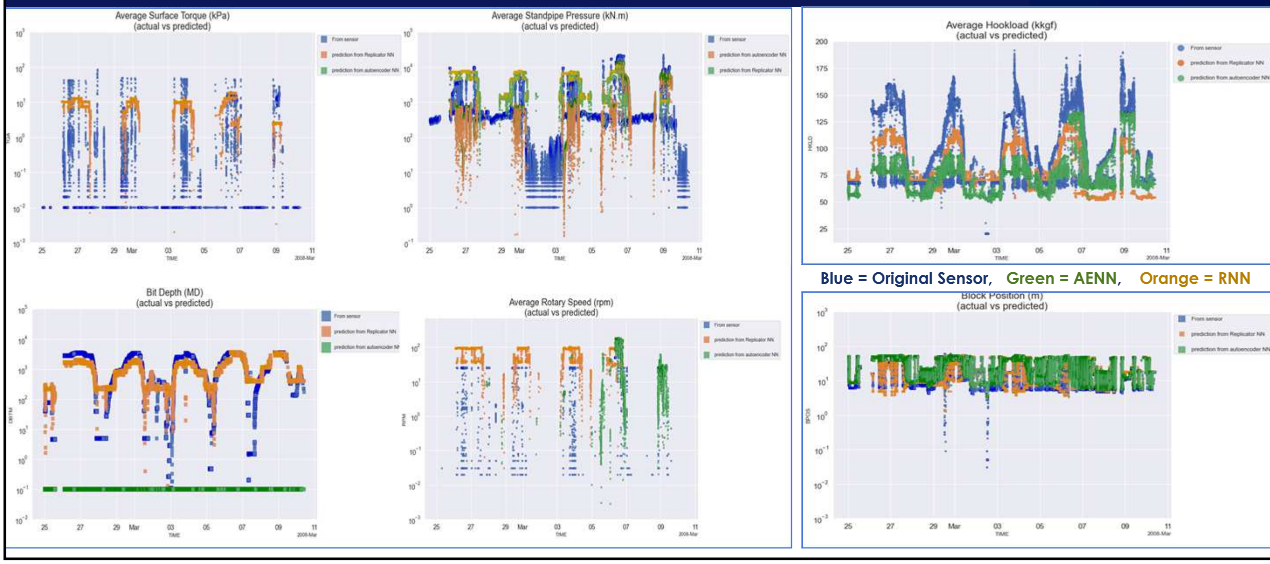


Blue = Original Sensor
Green = AENN
Orange = RNN

Sensor Test Data Vs Prediction Of RNN And AENN

(Assessing the similarity in data structure learned by both algorithm on Test data)

Webinars



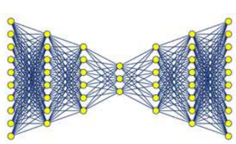
Reconstruction Error Produced

Webinars

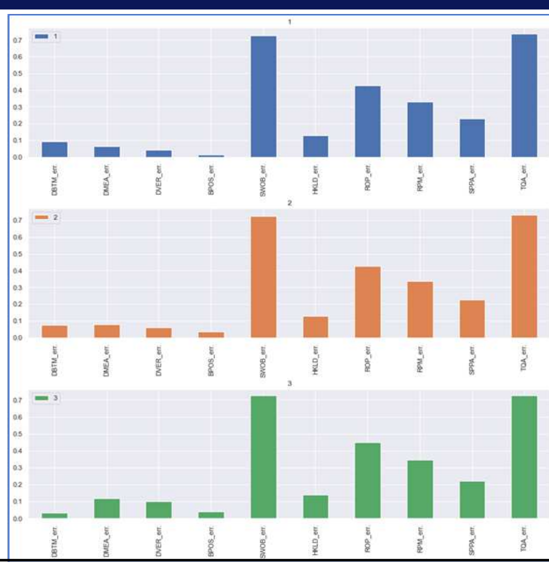
$$mse = mean(actual\ data\ - reconstrcted\ data)^2$$

$$mse = mean(\boldsymbol{X_i} - \widehat{\boldsymbol{X_i}})^2$$

$$\boldsymbol{X_i} = \begin{bmatrix} DBTM \\ DMEA \\ DVER \\ BPOS \\ SWOB \\ HKLD \\ ROP \\ RPM \\ SPPA \\ TQA \end{bmatrix}$$



$$\widehat{\boldsymbol{X_i}} = \begin{bmatrix} DBTM' \\ DMEA' \\ DVER' \\ BPOS' \\ SWOB' \\ HKLD' \\ ROP' \\ RPM' \\ SPPA' \\ TQA' \end{bmatrix}$$



1st rows of data with error value for each variable (vertical axis)

$$error\ in\ DBTM = (DBTM - DBTM')^2$$

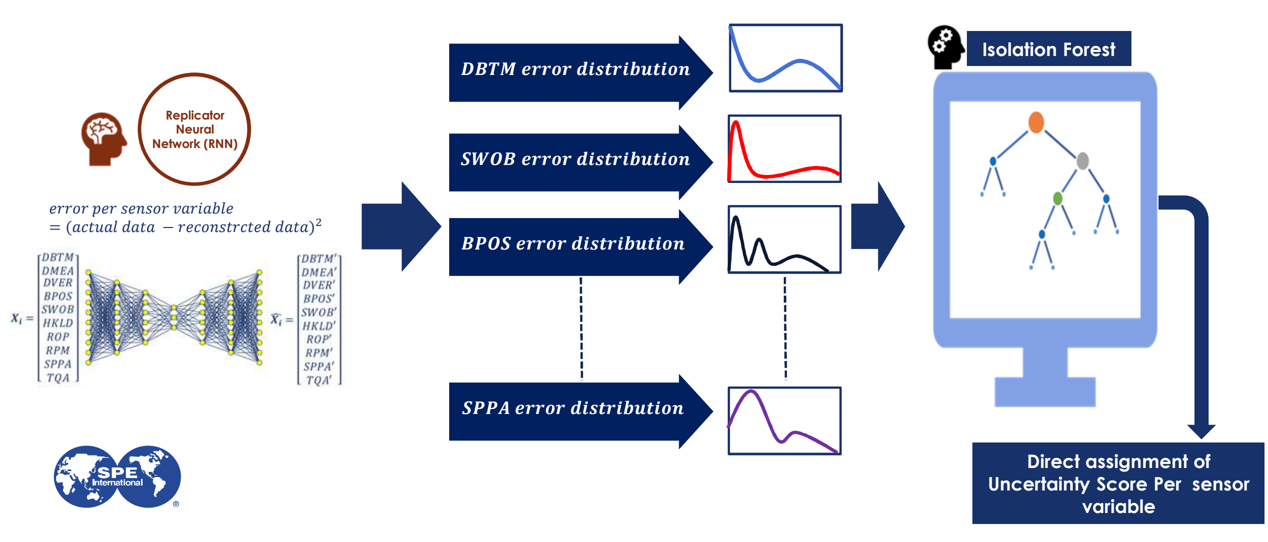
2nd row

3rd row

SPE International

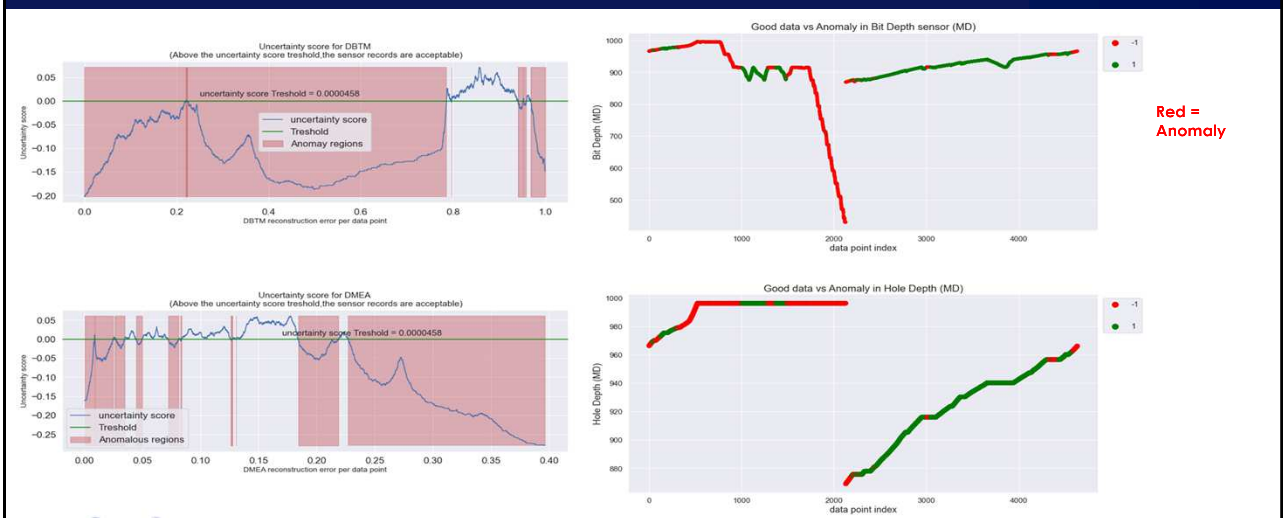
Setting Threshold/ Uncertainty Score Using Isolation Forest Algorithm

Webinars



Anomaly Visualization Based on Uncertainty Score
(Bit Depth & Hole Depth Sensor)

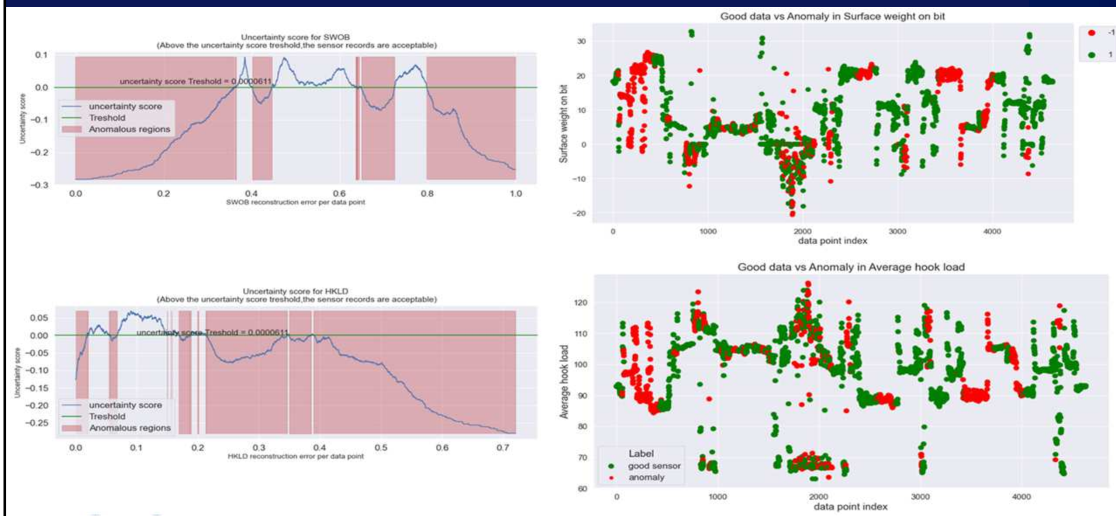
Webinars



Left figures represent the Uncertainty score generated by isolation forest showing red regions as regions where anomaly falls in. Above the threshold line (The unshaded region) lies the good data. The plot on the right are actual sensor data where the color red represents anomalies and greed represents the good data. Example of hole depth

Anomaly Visualization Based on Uncertainty Score
(Surface Weight on Bit & Average Hook Load Sensor)

Webinars

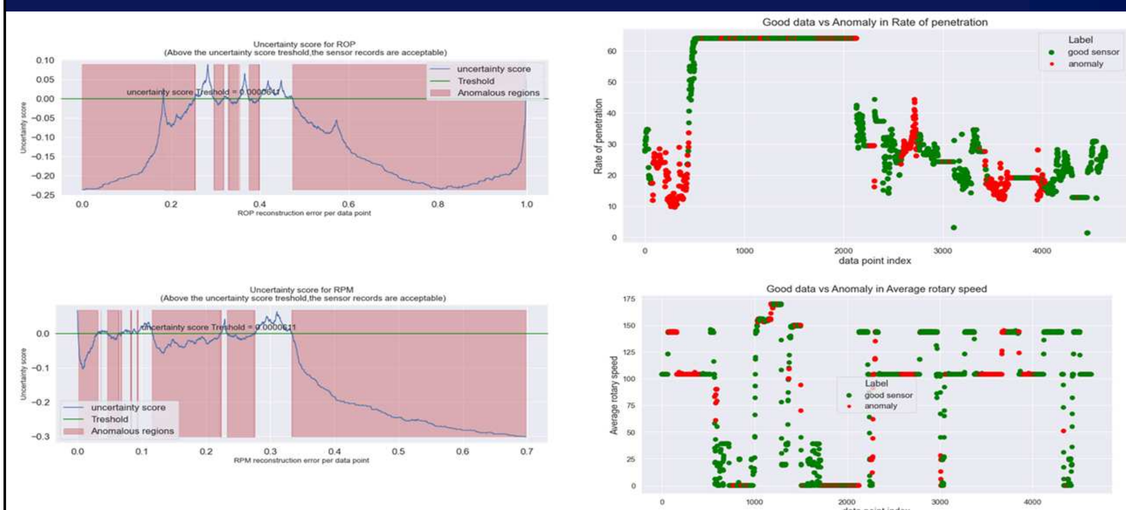


Uncertainty score and anomalies in Hook load and weight on bit sensor

Red = Anomaly

Anomaly Visualization Based on Uncertainty Score
(Rate of Penetration & Average Rotary Speed Sensor)

Webinars

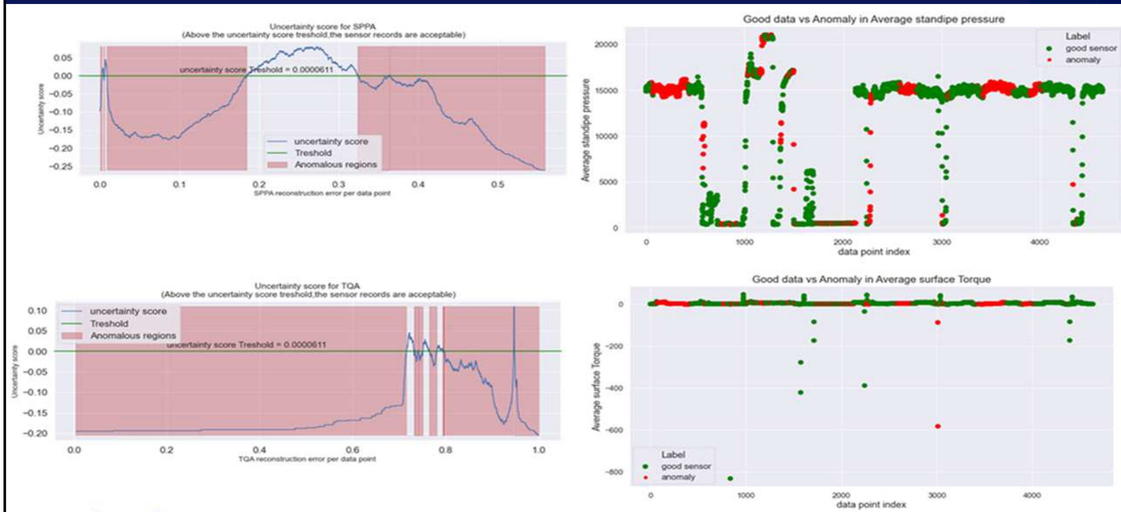


Uncertainty score and anomalies in ROP and Average rotary speed sensor

Red = Anomaly

Anomaly Visualization Based on Uncertainty Score
(Average Standpipe Pressure & Surface Torque Sensor)

Webinars

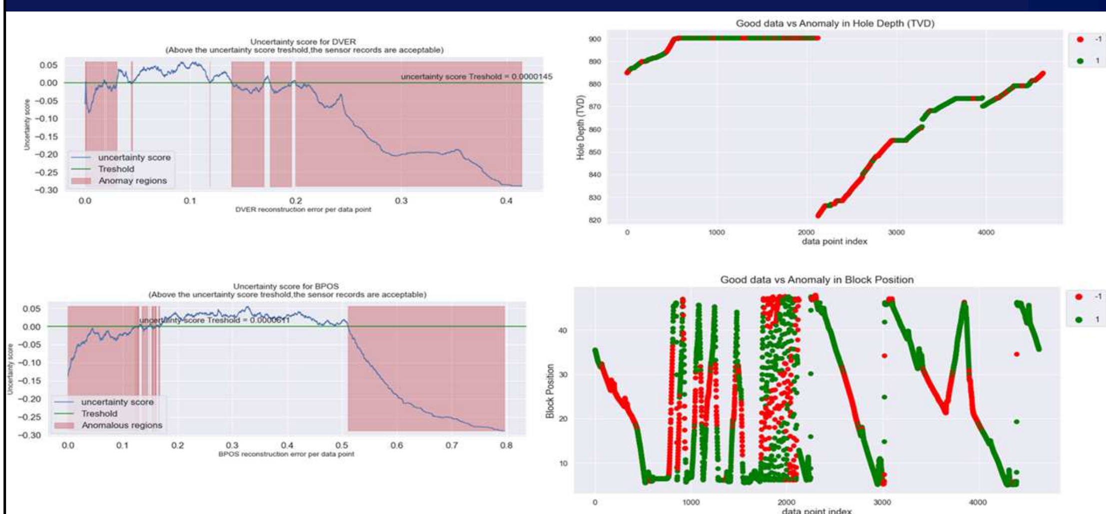


Red = Anomaly

Uncertainty score and anomalies in TQA and SPPA sensor

Anomaly Visualization Based on Uncertainty Score
(Hole Depth & Block Position Sensor)

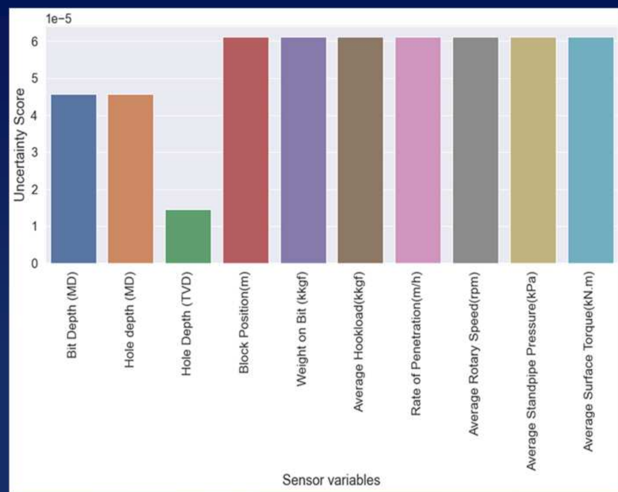
Webinars



Red = Anomaly

Uncertainty score and anomalies in Block position sensor

Summary



Below this score for a given Variable, the sensor data is in an anomaly region

Conclusion / Recommendation

- Auto-associative neural networks combined with clustering algorithm have proven to be useful in this Project
- It is recommended that to validate the developed frame work, a ground truth, with good sensor label should be used in a supervised learning approach.
- Expert / Domain Knowledge is required for result quality checking

Webinars

Thank You

Questions ??

SPE International

Webinars

Back Up

SPE International

Exploratory Data analysis (EDA)

Webinars

Pairwise Pearson Correlation

- The Pearson correlation coefficient is a measure of linear correlation between two sets of data
- Quantifying the Dependencies or Independencies

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

SPE International