

Analysis of Well Head Pressure Sensor Data for Anomaly Detection in Oil Well using IIoT and Unsupervised learning Technique

Shreyas Lele¹

¹B.Tech (Information Technology), Software Engineer

Abstract - Devices with sensors and IIoT analysis became a breakthrough in modern technology. IIoT is used in almost all areas of technology. In the oil and gas industry, IIoT-based analytics is critical. This paper presents our approach where we apply unsupervised learning technique on pressure sensors data to detect the anomaly in Oil well by predicting future pressure values. This document focuses on the number of advances in IIoT technologies thereby helping in analysis of high precision sensor data for future prediction of Oil well parameters. The industrial Internet of things represents the great opportunity in recent decades for the advancement of industrial technology and machine learning. Several engineering techniques that require expertise and data that are difficult to obtain are used currently to predict well performance. In this report, focus is on how advanced machine learning algorithms can be applied on huge amount of sensor data which will greatly ease Oil well maintenance and defect predictions thereby helping the customer to get notified about the issues in the Oil well before hand.

Key Words: IIoT, Prediction, Machine Learning, Unsupervised Learning

1. INTRODUCTION

Oil and Gas industry is one of the prominent and necessary industries. The workflow of these industries is generally categorized in three levels: (a) Upstream, (b) Midstream and (c) Downstream. Upstream part consists of the crude oil production and processing on the site itself and collecting the necessary data for analysis. Midstream part acts as an interface the production site and supplier chain managers. Downstream part is when the processed crude oil and gas reach the end shops and providers for consuming purpose.

There has been gradual rise in demand for IIoT implementation and analytics in Oil and Gas industries. Companies that invest funds to these kinds of industries want a reliable method to ensure progress and upliftment in production stats. Even a small amount of downtime of these oil and gas wells may result in big amount of loss both in terms of oil and gas production and in terms of monetary value industry. Faults may be of various kinds such as leakage of oil or gas, tampering of pipeline, sudden rise in temperature or pressure, pipelines valves malfunctioning causing sudden irregularities in sensor readings. So, companies need to be utmost sure that the wells function properly and if by any chance, faults tend to happen to any machinery equipment like in the ESP (Electronic Submersible Pump) or in the pipelines, we want that to be known beforehand. This is made possible using forecasting mechanisms such that if the actual value does not lie within our defined threshold boundaries (that we predicted) then we can pretend that the actual oil production is not up to the mark and there might be some wear and tear somewhere among the equipment's that may be hampering our oil production. For the gathered time series, we need analysis to be done to forecast the oil gas or liquid generation and study the forecasted visually.

2. LITERATURE SURVEY

Analytics on oil and gas production data is done using variety of algorithm and regression techniques. Exponential smoothing models has been implemented [1] and XGBoost, also known as Extreme Gradient boosting is used [2] to forecast future predictions. Each algorithm is unique in its own domain and one need to analyses the data that is being worked upon before actually finalizing what method to choose. Sometimes Best algorithms tend to perform very poorly on certain datasets due to failure of recognizing existing patterns.

Various visual features that can be seen in time series plotting are:

(1) Trend – Trend can be positive or negative. Positive trend implies the graph has a positive slope over time and negative trend implies the plot goes downwards over time. (2) Seasonality Data is said to be seasonal if it exhibits a periodic trend like daily, monthly, quarterly or yearly. Plot will show rise or fall upon reaching each period. (3) Cycle – Data is said to have cyclic nature if its rises and falls are not of fixed and constant period.

The oil and gas industrial process starts after the crude oil is extracted from the grounds (oilfields) and to use it as a product it must go through the refining process. [4]

The oil and gas industry were at one time, at the forefront of Digital Transformation in industry. Nearly 50 years ago, it was the first industry to use digital distributed control systems (DCS) to control refineries and other downstream plants. Then, with the deployment of the digital oil field concept, it became a leader in adopting digital representations of seismic data representing deposits and reserves shift the way it approaches Asset Performance Management (APM).

In the 1980s, Oil and Gas companies began to adopt digital technologies, with a focus on better understanding a reservoir's resource and production potential, improving health and safety, and boosting marginal operational efficiencies at oil fields around the world. [7]

There are many problems including too many data sources, complex data types, and the large amount of statistical work by hand in production engineering planning, and that influences the working efficiency. Therefore, there is a need to develop oil-production program planning software supporting systems, to achieve the forecast of workload of oil-production program, improve the efficiency of oil-production program planning.[6]

Information management methods of oil production plant of the traditional way are backward. All the departments are in the stand-alone mode, lacking a unified information network platform, generating large amounts of information islands. This backward working model has seriously hindered the development of oil production plant. [5]

In the literature some oil and gas well monitoring systems based on Wireless sensor networks have been proposed. In this section we describe the most recent research efforts in this regard. S. Sava al. [5] have designed a monitoring system for production well which is based on WSNs. This system is comprised of data acquisition terminal and surveillance centre and is devised to transmit and management decision of oil and gas well data. The data acquisition terminal is used to collect, transmit and exchange information. The service centre is used to send and receive the collected information, visualization and statistical analysis. Authorized personnel can remotely send instructions to well using the system and abnormal activities are notified through Short Message Service (SMS).

R. Zhansheng et al. [7] have proposed a sensor network based intelligent control system to avoid manual monitoring. Using this system, health and condition of multiple oil wells and oil storage tanks can be efficiently and remotely monitored and information about the wells and tanks is transmitted wirelessly to an administrator located in a remote location. The sensors involved are level sensors, temperature and gas sensors.

It has proposed an efficient wireless monitoring system which uses ZigBee technology and microchip (PIC microcontrollers) to collect data from a remotely located oil or gas wells through wireless connectivity and it is then sent to the data gathering station (Onshore CCR or process platform) [5]

The author has proposed simulator in which three electronic cards integrated into the wellhead control panel (WHCP) and each contains a XBee radio module (router). Each wellhead microcontroller is coded using micro environment. The feedback from wellhead safety valves position is gathered by the microcontroller and transmitted through XBee module to remote location.

3. SIGNIFICANCE

This paper presents the study of how analytics can monitor the health of the Oil well, and it will notify the customer's that when their Oil well will start detecting anomaly.

This will help users to take corrective actions.

4. Data Pre-Processing and Training Algorithms:

4.1 Principal Component Analysis:

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors (each being a linear combination of the variables and containing n observations) are an uncorrelated orthogonal basis set.

Consider a data matrix, X , with column-wise zero empirical mean (the sample mean of each column has been shifted to zero),

where each of the n rows represents a different repetition of the experiment, and each of the p columns gives a particular kind of feature (say, the results from a particular sensor).

PCA can be thought of as an unsupervised learning problem. The whole process of obtaining principle components from a raw dataset can be simplified in six parts:

1. Take the whole dataset consisting of $d+1$ dimension and ignore the labels such that our new dataset becomes d dimensional.
2. Compute the mean for every dimension of the whole dataset.
3. Compute the covariance matrix of the whole dataset.
4. Compute eigenvectors and the corresponding eigenvalues.
5. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $d \times k$ dimensional matrix W .
6. Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace.

4.2 Mahalanobis Distance:

The Mahalanobis distance measures distance relative to the centroid — a base or central point which can be thought of as an overall mean for multivariate data. The centroid is a point in multivariate space where all means from all variables intersect. The larger the MD, the further away from the centroid the data point is.

The most common use for the Mahalanobis distance is to find multivariate outliers, which indicates unusual combinations of two or more variables.

The Mahalanobis distance between two objects is defined as:

$$d(\text{Mahalanobis}) = [(x_B - x_A)^T * C^{-1} * (x_B - x_A)]^{0.5}$$

Where:

x_A and x_B is a pair of objects, and

C is the sample covariance matrix.

Another version of the formula, which uses distances from each observation to the central mean:

$$d_i = [x_i - \bar{x}]^t C^{-1} (x_i - \bar{x})^{0.5}$$

Where:

x_i = an object vector

\bar{x} = arithmetic mean vector

4.3 Isolation Forest

Isolation forest separates observation by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. It builds an ensemble of isolation trees for a given data and declares anomalies among instances which have shortest average path length on the trees.

The isolation characteristics of isolation trees enables them to build partial models and exploit sub sampling to an extent that is not feasible in existing methods. Since a large part of an isolation tree that isolates normal points is not needed for anomaly detection, it does not need to be constructed.

A small sample size produces better isolation tree because the swamping and masking effects are reduced.

Let T be a node of an isolation tree. T is either an external-node with no child, or an internal-node with one test and exactly two daughter nodes (T_l, T_r).

A test consists of an attribute q and a split value p such that the test $q < p$ divides the data point into T_l and T_r .

Given a sample of data $X=\{x_1, \dots, x_n\}$ of n instances from a d -variate distribution, to build an isolation tree, recursively divide X by randomly selecting an attribute q and a split value p , until either:

(i) The tree attains a height limit, (ii) $|X|=1$ or (iii) all data in X have the same values. Isolation forest calculates anomaly score for each instance using path length $h(x)$.

Since isolation trees have an equivalent structure to binary search trees, the estimation of $h(x)$ for external node terminations are the same as the unsuccessful search in BST.

Given a data set of n instances, the average path length of unsuccessful search in BST can be calculated as

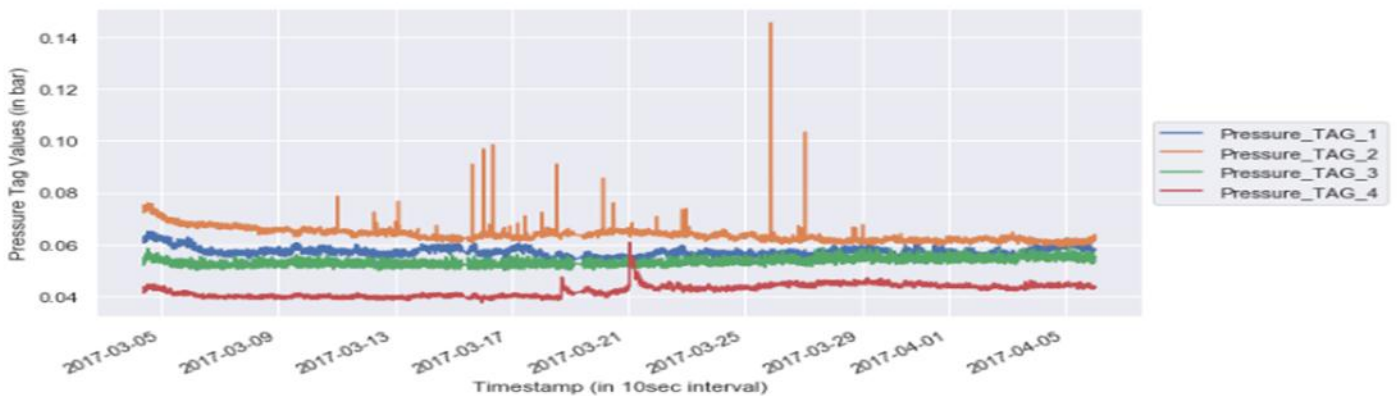
$$c(n)=2H(n-1)-(2(n-1)/n),$$

where $H(i)$ is the harmonic number. Therefore, anomaly score of an instance x is defined as

$$s(x,n)=2^{-(E(h(x)))/(c(n))}$$

5. RESULTS AND ANALYSIS.

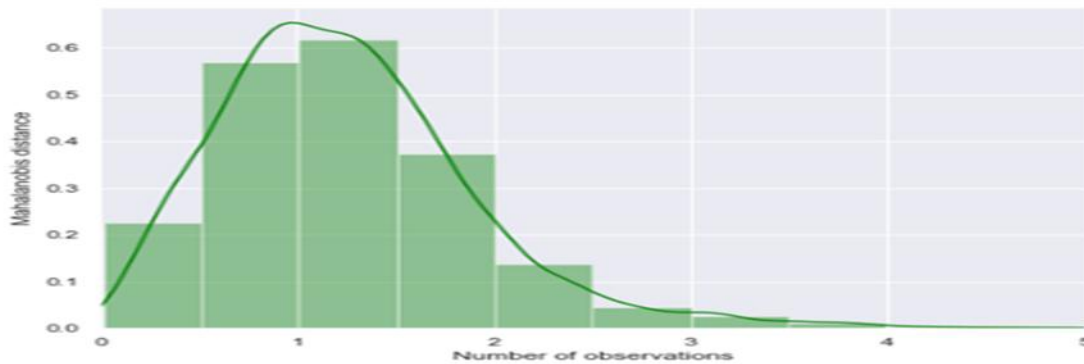
(i) Time series Plot for Well Head Pressure



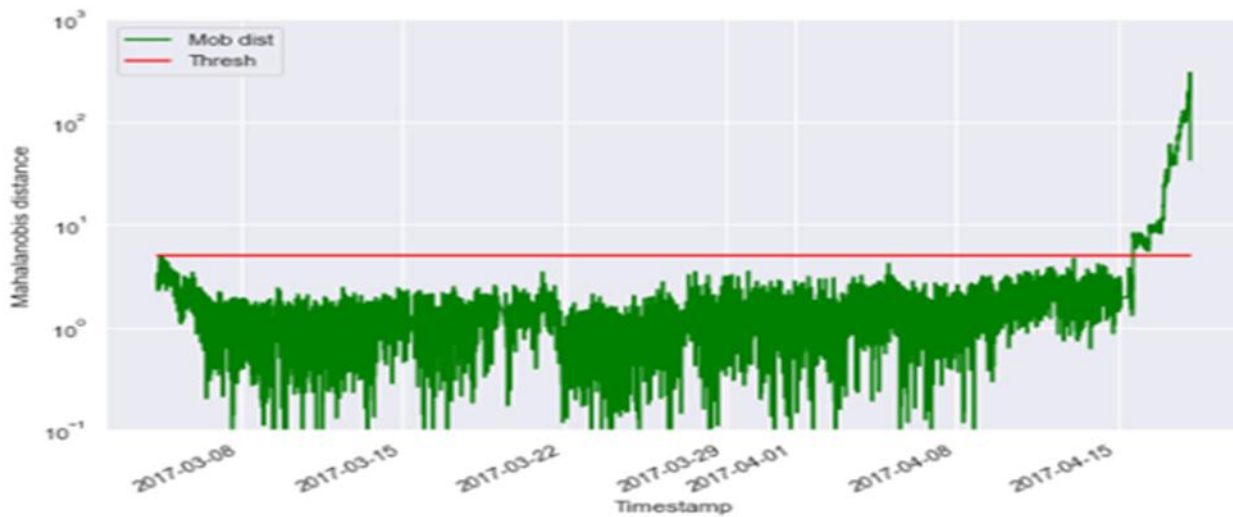
(ii) After applying PCA



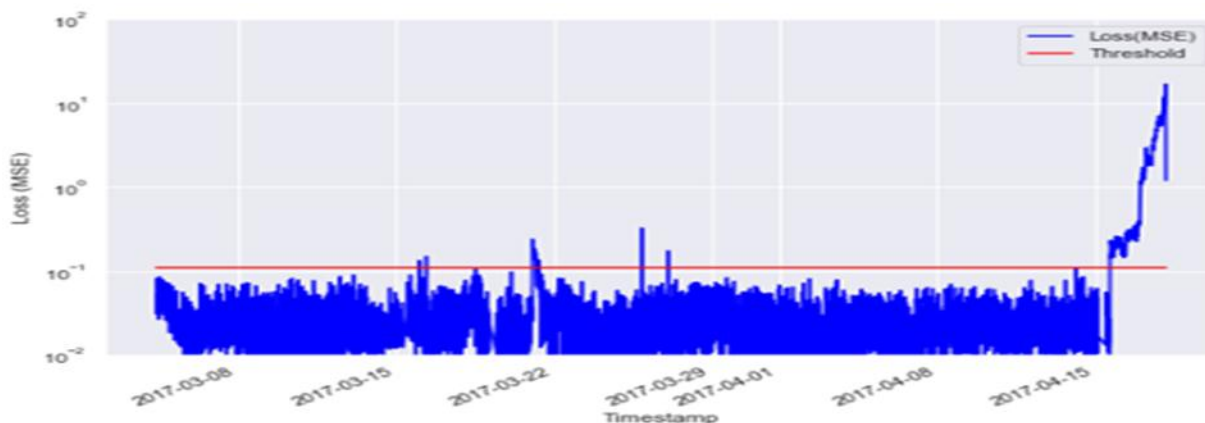
(iii) Mahalanobis Distance using Histogram plot



(iv) Anomaly Threshold detected using Mahalanobis



(v) Anomaly detection using Isolation Forest



5. CONCLUSIONS

With the use of IIOT and machine learning technique de-ployed on cloud, the visual system developed for analysis based on historical data will provide appropriate visualization that displays the current production and what-if analysis to simulate the best production optimization possible.

It will also allow users to assess production performance against forecasted production and analyze how distant actual production is from the forecast and examine the cause for the deviation along with the estimate of the loss of oil per barrel also helps in study and comparison of live data.

The future of the IIoT must integrate these proven applications into larger systems-of-systems that bring the power of cloud analytics and business intelligence to industrial systems. This is the core vision of the Industrial Internet. DDS is the right standard protocol to fulfill that vision because it provides both the extreme capabilities required by intelligent machines and the needed integration to extend to cloud-based analytics and optimization. Can predict and forecast the values for oil, gas, liquid, water, and many other tags.

The risk assessment of IIOT equipment's are of utmost im-portance to industries because malfunction for even a fraction of time in the Oil and gas drilling equipment's such as ESP (electronic submersible pumps) and pipelines may lead to heavy loss both from production as well as economic point of view.

This paper showed a decent approach to prevent such scenario by analysis the data sent by sensor like temperature, well head pressure etc. and watch out for anomalies in real time by fore-casting using Isolation Forest. Now if at any time in the fore-casted future, we see an anomaly among predicted data values, we can alert the site engineer prior to any malfunction that could occur in the equipment's so that corrective measures may be taken to prevent possible risk associated with the onsite machinery

REFERENCES

- [1] Hanyeu Zhang, Cheng Fei. "Design of Oil Well Monitor-ing Information Management System Based on IOT Technol-ogy" Springer Publication, December2015
- [2] Mohammed Y Aalsalem, Wazir Zada Khan, Wajeb Gharibi. "An intelligent oil and gas well monitoring system based on Internet of Things", 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)
- [3] "Outlier prediction for Temporal Data in Oil and Gas": A Survey Manish Gupta, Jing Gao, Charu C. Aggarwal, Jiawei Ha, 2014
- [4] P. Shoubridge, M. Kraetzl, and D. Ray, "Detection of Ab-normal Change in Dynamic Networks," IEEE Information, Decision and Control, 1999,
- [5] S. Savazzi, U. Spagnolini, L. Goratti, D. Molteni, M. Latva-aho and M. Nicoli, "Ultra-wide band sensor networks in oil and gas explorations," in IEEE Communications Magazine, April 2013.
- [6] "A Review on Anomaly Detection Methods for Optimiz-ing Oil Well Surveillance" Mohd Hilmi Hasan, Azlinda Abdul Malik and Mazuin Jasamai, IJCSNS International Journal of Computer Science and Network Security, Nov 2017
- [7] Wen ZHANG, Zhansheng SONG, Qingping Wang, Wei-dong HAO The Forecasting of Workload of Oil Production Program,2011 International conference of business manage-ment, pp.452-456.

BIOGRAPHIES



Shreyas Lele, Graduated in Information Technology from VJTI, Working as Software Engineer.