IATMI22-034

# Early Electric Submersible Pump Failure Detection Using Artificial Intelligence

Faradillah Hamid Bayagub[1]
[1]Pertamina Hulu Energi OSES
* Email: faradillahbayagub@gmail.com

**Abstract.** Electric Submersible Pump (ESP) is a highly effective artificial lift method for boosting oil production in both onshore and offshore fields. The ESP maintenance will be conducted regularly as other artificial lift to prevent costly production disruptions due to unexpected pump failures. Many diagnostic methods have determined the ESP system's status by using the automation system; however, these methods usually only provide backward-analysis after failure events have occurred. This paper involves acquired real-time data to establish an analytical methodology to detect impending ESP failures**.**
The classification will be done on ten minutes-interval data forecasting performance, which shaped up into slope. This is primarily achieved using Supervised Learning Technique; Logistic Regression, Random Forest, Decision Trees, K-Nearest Neighbors, and Recurrent Neural Network Technique: Long-Short Term Memory.
The models will be built based on individual distinct parameter's characteristics of nine status consists of; low PI, pump wear, tubing leak, higher PI, increase in frequency, open choke, increase in water cut, sand ingestion, and closed valve with an accuracy rate over 90%. These automation and control systems require constant surveillance by a human operator to verify that all processes are running normally. Furthermore, the abnormal behavior is identified in advance, and the operators can early determine the best corrective action to avoid an ESP's failure built upon the recommendations attached. It is the human operator's responsibility to react to any alarm conditions that occur during operation.
This introduced technology is an effective way to monitor the ESP system that leverages Artificial Intelligence. The operator can rely on the built surveillance system's ability to detect abnormal behavior, allowing the operator to focus on higher priority tasks. Moreover, an engineer's significant advantages are taking pre-emptive action to avoid failure and generating billions of revenues.
**Keyword(s):** ESP; Artificial Intelligence; Early Failure Prediction; Machine Learning



## 1. Introduction

Over 90% of wells go on the artificial lift at some point during their lifecycle. Artificial lift techniques are employed when reservoirs do not have sufficient energy to naturally produce oil or gas to the surface or at desired economic rates. Among artificial lift options, ESPs are often considered efficient and reliable for pumping high volumes from greater depths and higher temperatures. ESPs represent a significant CAPEX and OPEX cost item for operators [1].
Conventional ESP installations typically do not have downhole flowmeters installed as part of the equipment configuration. Some ESPs have downhole sensor subassembly used to monitor the health and performance of the ESP. An increased level of monitoring comes by connecting the downhole sensor to a SCADA (Supervisory Control and Data Acquisition) system. The system continuously monitors well behavior to provide a basic level of automated control in the oil and gas industry for the last decade [2].

Despite their design, engineering, and manufacturing to be rugged and highly reliable, ESP pumps can fail and do so unexpectedly. When that happens, given the complexities and high costs associated with exchanging a failed sub-surface installed pump with a new one, the result is typically prolonged downtime and associated production losses.

The solution to overcome this problem is to use big data, which has been a trend in the previous decade. Accumulation of data from years of spreadsheet database archiving has made industries gather valuable patterns and work processes that led to success stories [3]. By utilizing this data, industries have advanced in handling their problems; avoid ESP shutdowns by moving from a supervised approach towards failure mitigation to a more practical approach based on early predictive analysis and prevention. In this paper, an automated early predictive analysis built upon a real-time dashboard is proposed to monitor and safeguard ESP operations by focusing on patterns that deviate from expected normal behavior. Such a system can maximize equipment availability, save millions of dollars in maintenance and lost production, and eliminate the need to deploy many field personnel and instruments to monitor and investigate ESP operations [4].

## 2. Methodology

Several parameters significant to ESP operation were used as input variables for the analytical model. These included well inflow parameters, such as fluid pressure and temperature at the pump intake, tubing pressure, flow rate, and tubing diameter; pump performance parameters, such as discharge pressure, pump setting depth, stages, pump type; and motor diagnostic parameters, such as vibration and motor temperature.

Two different data records were tagged for the study. These are:

1. Data records containing time-series information of various parameters on downhole gauges. The data was recorded at a one-minute interval for one well.
2. Data records containing information on the time when a trip or failure occurred in that well.

Those data come from seventeen different wells equipped with lift watchers provided by four different types of ESP. This record was used to study the behavior of the patterns based on the selected parameters obtained from the historian long before, immediately before, and precisely during the trip or failure [4].

### 2.1 Downhole Sensor Parameters Forecasting

This step aims to predict the downhole sensor parameters; motor temperature, current consumption, motor vibrations, temperature & pressure for both the intake & discharge sections of the pump for the 1-day ahead. The plenteous amount of raw data from the lift watcher with different time intervals leads to the appearance of noises needed to be down sampled for smoothing the generated line.

Records of discharge pressure with the time interval of 11$^{th}$ April 2020 02.10 PM, precisely 3 months before trip happened, are selected to be analyzed for further forecasting to enable the model gains an understanding of the well characteristics. Then, the data will be split into a 70:30 ratio of training and test datasets to learn the relationship between independent variables and the target variable in terms of mathematical function or is captured as a set of rules. In this case, the program will use Long-Short Term Memory (LSTM), supports multivariate time series; and Seasonal Autoregressive Integrated Moving Average (SARIMA), supports univariate time series data with a seasonal component.

### 2.2 Flow Rate Reconstruction

Flow rate is a component that needs to be considered for each value on downhole sensor parameters reading on failure prediction of ESP. Infrequent well tests carried in the field causing the flow rate will be estimated using the provided model from the related company for each point of sensor data.
The data needs for the flow rate reconstruction includes static data as well profile and dynamic data which may be continuous and time-series data. ‘Static’ implies that there is no change in the data unless under particular circumstances for an extensive period in the life of ESP well. It consists of pump installation data and well-completion information. While the information of dynamic data is natural as it represents the well’s behavior and pump operation, which consists of:

1. Downhole Gauge Data: Pump discharge pressure, pump intake pressure, ampere reading, and intake temperature.
2. Surface Pressure Gauges Data: Casing pressure and tubing pressure.

The pre-processing data involves 121,462 raw data, split into an 80:20 ratio of training and test datasets, and started with scaling all the numerical features into range 0 and 1, then performing hyperparameter tuning to determine the optimal values for a given model.
The k-nearest neighbor, a chosen model to regress the virtual flow rates, is one of the algorithms that the learning is based on “how similar” a data from others and goes together with a large amount of data.

### 2.3 Early ESP Failure Prediction

Machine learning algorithms tends to perform well in the majority class but poorly in the minority class on imbalanced data. In this model making, the method of oversampling used is Synthetic Minority Oversampling Technique (SMOTE); the minority class with fewer data recorded can be matched up with the majority class with the most data which illustrated on Figure 1.
The well-data-integration engine will apply multivariate analytical techniques on identifying patterns and correlations between the variables. These steps will be done using supervised learning; logistics regression, random forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and decision tree. The model accuracy using logistics regression, random forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and decision tree are 86.16%, 91.96%, 90.88%, 93.8%, and 90.85% respectively. The obtained accuracies are quite similarly excellent; therefore, the confusion matrix for those machine learnings is generated to observe what kind of errors they are making. The result seems similar for random forest, k-nearest neighbor, and decision tree, they are reaching a 100% rate in the majority of the categories shown in Figure 2.
A common problem at model application stage is that the prepared data consists of multiple units corresponding to each real-time parameter of the well. Accordingly, the data transformed into slope using the percentage change between the current and a prior element. If the current well data show a trend that matches the pattern for a historic failure, the algorithm generates an alarm for the specific problem and delivers it to the field technicians any time of day. There are nine types of problems that can be predicted; low Productivity Index (PI), higher Productivity Index (PI), pump wear, tubing leak, increase in frequency, open choke, increase in water cut, sand ingestion, and closed valve (SSSV/Sub-Surface Safety Valve). The algorithm allows for developing an agreed-upon, mathematically tested pattern for all of the common ESP failures.

## 3. Results and Discussion

### 3.1 Downhole Sensor Parameters Forecasting

The process will be started using Auto SARIMA provided by Python module. The result of model selection is detailed with (1) as p, (1) as the difference, and (0) as the q, and (0,1,0,4) as the seasonal element. Followed by the next model, namely Long Short-Term Memory, with two hyperparameters of (1) as the batch size and (50) as the maximum epoch. Epoch is entire dataset is passed forward and backward through the neural network, while batch size is total number of training examples present in a single batch.
The prediction will be evaluated using Root Mean Squared Error (RMSE) to check how close the observed data points are to the model's predicted values [5]. The obtained value of Root Mean Square Error (RMSE) from SARIMA and LSTM are 21.4 and 2.5888 respectively, less than the standard deviation value with number 121.6, indicating the model acceptance for further forecasting. Later on, the forecasting results will be cohered with the original dataset on the corresponding time to measure the forecasting accuracy using percentage error to validate the program's credibility. The obtained value of Mean Absolute Percentage Error from SARIMA and LSTM are 6.1% and 0.889% respectively. According to Criteria for Model Evaluation [6], this is categorized as ‘highly accurate forecasting’, which means the reliability of this program is pretty much competitive when compared to commercial software; the result is also eligible to be used for further analysis.
The program will be used Long Short-Term Memory (LSTM) as the selected model as it provides better forecasting accuracy and enables an automate multivariate analysis to the sensor data. The forecasting data points, which illustrates in Table 1 are convenient to the model’s predicted values with percentage error also lies at 0.6% evenly, while the vibration has the highest error amongst all with 7%. It happens because the vibration has low values, so a slight difference would lead to a considerable percentage error. Thereafter, the forecasting results will be used as input at the flow rate reconstruction represented in Table 2.

### 3.2 Flow Rate Reconstruction

The K-Nearest Neighbor (KNN) regressor as the selected model will be fitted on the training dataset, 80% of the data, as a procedure to see whether the model suits the test dataset. It is resulting up to 99.59% accuracy, which possibly overfits. Overfitting means that the created model is too dependent on the training data, this can cause an overconfidence inaccuracy. The blind data test was conducted to validate the model accuracy by predicting another well test data with 43 various wells resulting in 5% error on the average. The percentage error can still be accepted, and the model will be applied to the obtained data in the previous section to predict the virtual rate.
Pressure drops, tubing diameter, and pump installation are the essential components for rate calculation. However, due to the limitations of the conducted well test, the tubing pressure will be estimated from the conversion of discharge pressure from each sensor reading as in the equation (1) and equation (2).

$$T_p = D_p - \Delta P \qquad (1)$$

$$\Delta P = 0.433 \times PSD \times SG \qquad (2)$$

The predicted values, shown in Table 3, were then compared to the nearby flow rate obtained from well test data with a value of 282.5106, which is quite identical.

### 3.3 Failure Prediction

Based on the company's historical data, the possible failure that might be procured on the pump is a sand ingestion because of loss flow happened on the well 3 days after the prediction. The chosen model is k-nearest neighbor because of the compatibility of the predictive result with the company's history data.
Eventually, the surveillance system could predict the possible failure of the Electric Submersible Pump (ESP) by 1st August 2020 at 12 AM, while the company suspected the failure on 3rd August 2020 before the existence of this system. The validation was done several times using data from different wells. All of the results from the predictive models are the same as the result from the company's guidelines.
The logic driving the notification system is based on pattern recognition for event detections and pre-diagnostic applications. These notifications allow for fast and automated operational corrections to maintain optimal pump operation. An illustration of the output of the surveillance system as a whole, including the prescription of preventive action for the specific failure prediction, is shown in Figure 3.

## 4. Conclusions

This study successfully proves the huge impact of leveraging Artificial Intelligence on ESP Monitoring process. Based on the result analysis, it can be concluded:

1. Forecasting the downhole parameters installed on the Electric Submersible Pump (ESP) well is done using LSTM for 1-day ahead with acceptable scale-dependent error and percentage error, which is also categorized as highly accurate forecasting and eligible for further analysis.
2. The virtual flow rates are reconstructed using a k-nearest neighbor regressor based on "how similar" data from the provided well test data with an accuracy rate of 95%.
3. Machine learning with a k-nearest neighbor is chosen to be the surveillance system of detecting abnormal data with 93.8% and generate the predictive results compatible with the company's historical data.

The early failure prediction system then could be built upon SCADA system to provide real-time monitoring of patterns that deviate from expected normal behavior.

## References

[1] Al Maghlouth, A., M. Cumings, M. Al Awajy and A. Amer, "ESP Surveillance and Optimization Solutions: Ensuring Best Performance and Optimum Value," 2013.

[2] Dustin E. Ratcliff, SPE, Marathon Oil Company, and Cesar Gomez, SPE, Ivan Cetkovic, SPE, and Odafe Madogwe, SPE, Weatherford, "Maximizing Oil Production and Increasing ESP Run Life in a Brownfield Using Real-Time ESP Monitoring and Optimization Software: Rockies Field Case Study," in *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 2013.

[3] Lucky Bagus Waskito, Marda Vidrianto, and Bambang Cahyoniarso, "Initiate Digital Oil Field Application at Mature Offshore Oil Field in South East Sumatra, Indonesia," in *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*, Bali, Indonesia, 2019.

[4] Supriya Gupta, Frontender Corporation; Michael Nikolaou, University of Houston; Luigi Saputelli, Frontender, "ESP Health Monitoring KPI: A Real-Time Predictive Analytics Application," in *SPE Intelligent Energy International Conference and Exhibition*, Aberdeen, Scotland, UK, 2016.

[5] Kenney, J. F. and Keeping, E. S., "Root Mean Square Error," in *Mathematics of Statistics Pt. 1, 3rd ed*, NJ, Van Nostrand, 1962, pp. 59-60.

[6] Lewis, C.D., Industrial and Business Forecasting Methods, London: Butterworth Scientific, 1982.

[7] C.E. Ejim, and J. Xiao, "Sensitivity and Uncertainty Analysis of a Downhole Flow Metering System," in *Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, 2017.

[8] Carillo, W., "Prognostics for Oil & Gas Artificial Lift Applications," in *GE Oil and Gas, PHM Conference*, New Orleans, 2013.

## List of Figures and Tables

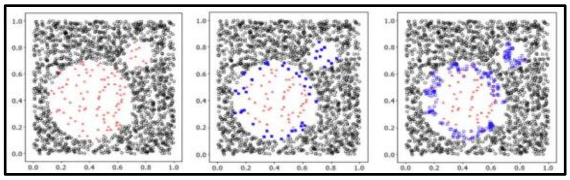
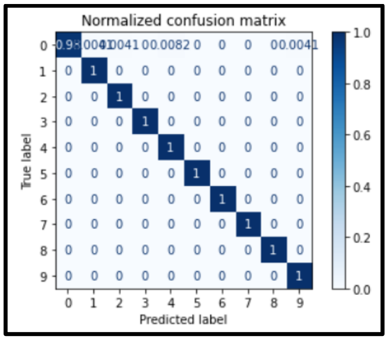


Figure 1 Diagram of Borderline Synthetic Minority Oversampling Technique (SMOTE)

*(a) The original distribution of the circle data set (b) Selection of borderline minority samples (solid blue squares) (c) Generation of borderline synthetic minority examples (hollow blue squares)*



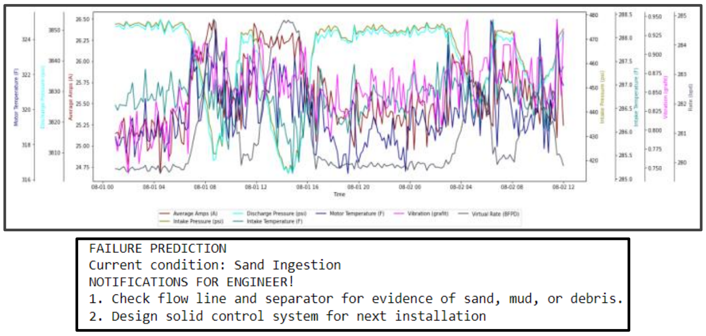




Figure 2 Normalized Confusion Matrix

Figure 3 Output of the Surveillance System

Table 1 Sample of Prediction Results and Error Calculation

| Well | Reading Time | Real Discharge Pressure | Predicted Discharge Pressure | Error (%) |
|---|---|---|---|---|
| FHB-03 | 2020-08-01 01:10:00 | 3839.683 | 3823.097900390625 | 0.4319% |
| FHB-03 | 2020-08-01 01:20:00 | 3848.477 | 3823.31591796875 | 0.6538% |
| FHB-03 | 2020-08-01 01:30:00 | 3854.760 | 3823.573974609375 | 0.809% |

Table 2 The Sample of Forecasting Result as Input for Rate Reconstruction

| Well | Reading Time | Average Ampere | Intake Pressure | Discharge Pressure | Intake Temperature | Motor Temperature | Vibration |
|---|---|---|---|---|---|---|---|
| FHB-03 | 2020-08-01 01:10:00 | 25.14 | 476.02 | 3823.09 | 286.56 | 318.44 | 0.76 |
| FHB-03 | 2020-08-01 01:20:00 | 25.16 | 476.32 | 3823.31 | 286.45 | 317.91 | 0.80 |
| FHB-03 | 2020-08-01 01:30:00 | 25.08 | 476.23 | 3823.57 | 286.57 | 317.94 | 0.75 |

Table 3 Rate Re-Construction Results

| Reading Time | Casing Pressure | Tubing Pressure | PBHP | PBHT | Ampere Reading | Pump Type | Stages | HP | Volt | Amp | PSD | TUBINGID | Virtual Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2020-08-01 01:00:00 | 50.02146 | 615.7779 | 476.0293 | 286.5625 | 25.14661 | low | 359 | 150 | 1672 | 58.7 | 10109 | 2.992 | 279.7837 |
| 2020-08-01 01:10:00 | 50.02146 | 615.9668 | 476.325 | 286.4519 | 25.16164 | low | 359 | 150 | 1672 | 58.7 | 10109 | 2.992 | 279.7756 |
| 2020-08-01 01:20:00 | 50.02146 | 616.1224 | 476.2307 | 286.5748 | 25.0839 | low | 359 | 150 | 1672 | 58.7 | 10109 | 2.992 | 279.7478 |