Predicting Failure in Electric Submersible Pump by Utilizing Machine Learning Based on Real-Time Sensor Data

Heksa Budi Santoso[*1], Marda Vidrianto[2], Silvya Dewi Rahmawati[1] and Utjok Wello Risma Siagian[1]

[1]Petroleum Engineering, Institut Teknologi Bandung
[2]Pertamina Hulu Energi Offshore South East Sumatra
* Email: heksa.bs@gmail.com

Abstract. Electric Submersible Pump has been used widely for oil wells around the world. The investment of ESP installation is costly, therefore the high return is expected from the implementation. This research has the objective to build a machine learning model to predict the possibilities of failure in electric submersible pump installation. By making an early prediction, the user can prepare to do the action needed to overcome the problem. Therefore, it will reduce the loss of fluid production.

The artificial intelligence will use data from a real-time sensor installed on the Electric Submersible Pump. There are six parameters that can be retrieved: ampere, intake pressure, discharge pressure, intake temperature, motor temperature, and vibration. The liquid rate then to be calculated using an analytical equation. The data are prepared by picking a set of data for every three hours from the raw data. This is decided because the interval of the data recorded by the sensor are not always at the same interval. The prepared data then to be transferred into a database. The evaluation will be conducted by developing the slope of the three days data and called as set data. Every data set then labelled according to the type of ESP's failure. The method that is used to develop a proxy model came from comparing several methods, such as logistic regression, decision tree analysis, etc.

There are thousands of data set that already filtered from over 400,000 raw data. The data came from four types of pumps that are commonly used in the observed fields. The proxy model resulting in several kinds of events that can be read, such as low PI, pump wear, tubing leak, increase in frequency, increase in water cut, etc. Those events have individual distinct parameter's characteristics. The model has an accuracy rate of over 70% with the proportion of 75% training set and 25% test set. The implementation of this proxy model will tell the user the most possible event that will happen according to the latest data set.

The utilization of a machine learning to ease the work of a user will be the most efficient way in this era of digitalization. This is the first machine learning approach that is implemented to predict Electric Submersible Pump failure application, and the outcome is to be able to predict the problem in the early stage so the fluid loss due to the problem is decreased and more importantly help prevent the well to be shut-in.

Keyword: Electric Submersible Pump, Machine Learning, Artificial Intelligence, Pump Failure

©2020 IATMI. All rights reserved.

1 Introduction

1.1 Background

As the produced wells age and undergo pressure decline from time to time, there is always a point where the natural pressure of the reservoir is no longer enough to flow and lift up the reservoir fluids to the surface. The solution taken to overcome this problem is to use a tool to add pressure energy at the bottom of the well. One widely used tool is the usage of electrical submersible pump (ESP) as a part of artificial lift that adds up the energy of the fluid column in the wellbore.

The installation of ESP increase production cost. However, high economic return is expected from the ESP implementation, as it will maintain or increase oil production rate. In order to ensure this objective of the ESP implementation is achieved, the ESP operation needs to be maintained and supervised to minimize the problems in the ESP system.

ESP is equipped with a device that is fitted to the most bottom part of ESP structure for monitoring the fluid lifting condition. By equipping the sensor with a real-time surveillance system such as liftwatcher, the user can read the real-time data of parameters that can be retrieved from the sensor reading i.e. average amps, intake pressure, discharge pressure, intake temperature, motor temperature, and vibration. These data can be used as the indicators of the existence of certain problems inside the well which can be seen from the changes in the value of the parameters.

Those data will be saved in an online database where the user can read the data in real-time. The data will be saved according to the reading time, where the time interval of the data reading can be as small as once per second. This system is built for saving data in a long-time window but it is not designed for giving an analysis based on the available data.

Late problem identification can increase the loss of fluid production and damage the ESP system itself. Therefore, to speed up and simplify the sensor data analysis for the user, it needs a system that can be run automatically, read the sensor data in real-time, and give user the most possible problem according to the sensor data.

A similar concept has been established by Bermudez et.al. (2014) and Adesanwo et.al. (2016) in implementing an artificial intelligence named fuzzy logic, which improved by Grassian et.al. (2017). In this paper an improvement has been made, that is the implementation of machine learning. The usage of machine learning has the purpose to make the model to be improvable alongside data increasing.

This paper divided into six main parts, the first one is the introduction which explains about the background and the objectives of this paper, followed by basic theory which explains the knowledge behind the making of this paper, methodology which explains the workflow of this study, case study which explains the implementation process of this study, result and discussion which explains the outcome of this study, and conclusion which explains the whole idea of this study.

1.2 Objectives

The main objectives of this research are to:

1. Identify problems that happened in the past for the database of the machine learning model
2. Build a machine learning model to predict the most possible problem that will happen according to the sensor data

1.3 Basic Theory

1.3.1 Electrical Submersible Pump

ESP is a centrifugal pump that moves with an electric motor that is designed in a way so that it can be submerged into the fluid. ESP can lift a high volume of fluid from the wellbore. It can be used both in onshore and offshore operations and is also considered compatible with deviated wells. Some of the downsides of ESP are that it requires high voltage electricity and is costly for repairments if any damage happened.

The components of the ESP are separated into two, i.e. surface components and subsurface components. The surface components are transformer, VSD, and junction box. Meanwhile, the main subsurface components are pump, gas handler, pump intake, protector, motor, and the sensor.

1.3.1.1 Downhole Sensor

The downhole sensor is located below the motor, making it the most bottom part of ESP. The parameters that can be monitored through the sensor are average amps, intake pressure, discharge pressure, intake temperature, motor temperature, and vibration. The data is sent to the surface through the power cable and can be seen in the motor controller. The user can read the data through supervisory control and data acquisition system by using a real-time surveillance system.

1.3.2 Machine Learning

Machine learning is an artificial intelligence (AI) application that provides the systems with the ability to learn automatically and improve the model over time without being explicitly reprogrammed. The learning begins with data observation to look for patterns in data to be able to do decision-making in the future based on the data provided. The main point of machine learning usage is to allow computer to learn automatically with minimum human intervention or assistance

Generally, machine learning is divided into two, supervised learning and unsupervised learning. For the supervised learning, the user provides the algorithm, which includes the input and the desired output, then the algorithm will find a way to make the desired output from the input given. It called supervised because the user needs to provide supervision to the algorithm by giving the desired output for the example that will be used for making the database.

The second is unsupervised learning, this type of learning is when the input data is the only known factor, while the output is unknown. This kind of learning is harder to evaluate because it depends on the input parameters that are given to the algorithm.

One of the techniques used in machine learning is classification. The purpose of classification is we want to divide the dataset into several groups. However, different from the clustering, in classification, the algorithm for data division is set manually, or in other words humans as the user will teach the machine how to divide the data into groups. There are several classification techniques, such as logistic regression, k-nearest neighbors, support vector machine, naive bayes, decision tree classification, and random forest classification.

1.3.2.1 Logistic Regression

Logistic regression is a regression technique that the purpose is to divide the dataset into two groups as the data value in the dataset always binary, whether its value is zero or one. There is no data with the value between zero and one. In other words, logistic regression predicts whether something is true or false. As can be seen from an example picture of the logistic function of logistic regression in Figure 1, if a certain value of x is >50%, then the logistic regression will classify it as y = 1. Otherwise, the logistic regression will classify it as y = 0.

In logistic regression how the line is fit to the data is different from linear regression, with linear regression the line fitted using "least squares", in other words, it finds the line that minimizes the sum of the squares of the residual, the residual also used to calculate the R2. Logistic regression does not have the same concept of a residual, so it can't use least squares and it can't calculate R2. Instead, it uses something called maximum likelihood.

1.3.2.2 Decision Tree

Decision tree analysis is a predictive modeling that uses a tree-like model of decisions. By using this analysis, the purpose is to predict the outcome from several input parameters by keep dividing every parameter into several possible outcomes.

In general, a decision tree asks a question and then classifies the data based on the answer. The classification itself can be in a form of category or numeric.

In the decision tree, the very top of the tree is called the "root", then the root branches into parts called the "nodes", the last parts of a decision tree that illustrates the outcome are called the "leaves"

A simple example of decision tree analysis in oil and gas industry can be seen in Figure 2, where the "root" is "B" where the decision-maker chooses to drill or not to drill. If the decision-maker chooses not to drill, the "leaves" which is the outcome is $0 in value, which is no money gained or lost. The "nodes" is "A", if the decision-maker chooses to drill, then there will be a budget of $100,000 that is given, and the "leaves are the outcome based on the answer from the "root" and the "nodes", the "leaves" categories in the example are Dry Hole (-$50,000) which has 70% probability, 2 Bcf of gas which has 20% of probability and will bring profit of $100,000, and 5 Bcf of gas which has 10% probability and will bring profit of $250,000.

1.3.2.3 K-Nearest Neighbors

K-nearest neighbors is another algorithm used to classify data, if there is already a database that defined all of the data types, it can decide which types of data if there is new data. The k-nearest neighbors classify the new data by looking at the nearest annotated data, in other words, the "nearest neighbors".

The "K" in "K-nearest neighbors" is the number of neighbors that the algorithm used to define the new data's categories

From the example in Figure 3 it can be seen the plot of using k-nearest neighbors with the value of k = 3, the types of the new data will be decided using the nearest three neighbors.

1.3.2.4 Synthetic Minority Oversampling Technique (SMOTE)

SMOTE is an oversampling technique to address an imbalance in the datasets. The way how SMOTE works is by making a line between two data in the same categories from the dataset, then draw a new sample along the line until the desired number of data is fulfilled. The most common way to use this method is by oversampling the minority class, so the amount of data that the minority class has will be the same as the majority class.

1.3.3 Rate Calculation

The rate is calculated using an analytical equation. The method used is based on approach proposed by Camilleri et.al. (2015), the equation is described as follows.

$$\frac{DP \times Q_p}{58847 \times \eta_p} = \frac{V_m \times I \times PF \times \eta_m \times \sqrt{3}}{746} \ldots (1)$$

DP is the differential pressure between the intake pressure and discharge pressure, both of those data can be retrieved from the sensor.
Qp is the flow rate. This is what the outcome that is expected from the calculation.
Vm is the downhole motor voltage. This is measured by subtracting the surface voltage with the voltage loss in the power cable. The voltage loss can be estimated from the cable resistance properties.
I is the motor current. This value can be seen directly from the supervisory control.
PF is the power factor. Power factor is the ratio of the real power to the apparent power.
ηm is the motor efficiency. The value of the motor efficiency is assumed using the maximum efficiency that obtained from the pump's profile
ηp is the pump efficiency. This parameter value comes from the pump profile and uses the assumption that the pump efficiency will always be the same.

1.3.4 Slope Calculation

The slope is required to see the value changes between data. The value of slope is calculated using the equation as follows.

$$m = \frac{y_2 - y_1}{x_2 - x_1} \ldots (2)$$

The slope is calculated by looking for the ratio of the differential value of the y-axis in accordance with the differential value of the x-axis.

2 Methodology

This study is conducted based on the workflow as follows:

1. Data acquisition
   Retrieving data from the cloud system that will be used as the database for the machine learning model.
2. Data preparation
   Filtering data in the wanted time and calculating the value of flow rate for every dataset.

3. Data processing
   Accumulating three days' worth of prepared data and make it into one set of data. Calculate the slope value of every parameter then convert it into a percentage, and also, label the data according to the event.
4. Model building
   Building the machine learning model by comparing the result from several methods and then build the graphic user interface (GUI).
5. Model prediction
   Making the latest sensor data as the new input, the machine learning model with giving an output of the prediction based on several time windows.
6. Model validation
   Validating the model created by making sure the prediction that has been made by the model is correct.

The workflow of this study is described in Figure 4

2.1 Data acquisition

The data come from twelve different well that equipped with liftwatcher. There are four different types of pumps that are used in those well, those four types of pumps are RC1000, RC2500, SN2600, and SN3600. The raw data that comes from the liftwatcher have different time intervals, which is why they need to be prepared before the processing.
The raw data have the smallest interval of every second until the biggest is every one hour. There are six types of data that retrieved from the sensor, those are average amps, intake pressure, discharge pressure, intake temperature, motor temperature, and vibration.
The time window that the data is taken is from 1st January 2019 until 29th February 2020 with the total amount of raw data is 711,654 data. The example of the raw data can be seen from Table 1.

2.2 Data preparation

The raw data needs to be filtered before processed. The way it filtered is by picking the data every three hours, this is done because not all of the data has the same time interval. By picking the data every three hours, all of the data can be ensured to have the same time interval but we still can see the data's trend clearly.
The data picked is the data at 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00. Therefore, in a day there are eight data that can be processed.
The second thing that is done in data preparation is calculating the rate for each time by using the Eq. 1. The value of downhole motor voltage, motor efficiency, and pump efficiency came from the assumptions which the values is the same with the voltage, the motor efficiency, and the pump efficiency from the pump profile. The power factor data of this field's came from the motor profile given from the supplier. The prepared data can be seen such as in Table 2.

2.3 Data Processing

The process of data processing takes 24 set of data that comes from three days' worth of data then make it into one set of data. Then calculate the value of slope using Eq. 2 with the time as the x-axis and the sensor parameter as the y-axis, do this calculation for every parameter.
After that, convert the value of slope to a percentage, this done by dividing the slope value with the first value in that particular set of data.
Lastly, every set of data needs to be labeled based on the company's guideline with the help of some reference from early studies done by Awaid et.al. (2014), Adesanwo et.al. (2016) and Grassian et.al. (2017). The example of the data that has been processed can be seen in Table 3 and Table 4.
There are nine types of problem that found from the process of labeling every set of data, those problems are low PI, pump wear, tubing leak, higher PI, increase in frequency, open choke, increase in water cut, sand ingestion, and closed valve (SSSV/surface valve), the condition of every parameter that lead to those problems are written on Table 5 and the amount of data recorded for every kind of problems is written on Table 6.

2.4 Model Building

Due to an imbalance in the data recorded, an oversampling is needed. In this study the method of oversampling used is SMOTE. By using SMOTE, the minority class that has fewer data recorded can be matched up with the majority class that has the most data. Here the majority class is the running condition where there is no problem within the well, so all of the problem categories are oversampled to match the amount of the running data.
After the data are oversampled, the machine learning method is chosen from the result of comparing several methods, those methods are logistic regression, decision tree analysis, and k-nearest neighbors.
The data divided into two sets of data, 75% train data and 25% test data. Even though the test data is only 25%, the model created already have sample across all of the categories, it means that the model can recognize all of the categories that exist. The model accuracy using logistic regression, decision tree analysis, and k-nearest neighbors are 85.72%, 99.72%, and 99.8% respectively. From the matrix plot of the logistic regression method in Figure 5, the machine learning can predict the majority of the problems with the accuracy above 70% except for the problems with the index of two and eight, which are pump wear and sand ingestion. For the decision tree analysis and k-nearest neighbors, the matrix plots are similar, as can be seen in Figure 6 and Figure 7, both of the methods can make a prediction really accurate, even reaching a 100% rate in the majority of the categories.
Despite being the method with the lowest accuracy, the chosen method is the logistic regression. This is because for the datasets that need to be oversampled, accuracy close to 100% is said to be overconfident.
This is proven by using a set of data as the input for three models with three different method, then the prediction result compared to the result from company's guideline. The input for this method comparison can be seen in Table 7. From the company's guideline, the problem happening from those three days data is Low PI, then the prediction result using logistic regression, decision tree analysis, and k-nearest neighbors can be seen in Figure 8, Figure 9, and Figure 10 respectively. The prediction result for three days data using logistic regression happens to be Low PI, which matches with the company's guideline.

However, the result using decision tree analysis is said to be running, where it means that there is no problem in the well. Meanwhile, the result from k-nearest neighbors is known to be open choke. Both the result from decision tree analysis and k-nearest neighbors are considered to be incorrect, and this proves that the models with accuracy rate close to 100% where the database was built using oversampling method, are said to be overconfident.

After the model's method selected, next is to tune the parameter to increase the accuracy of the model because previously the model has been made using default parameter. After hyperparameter tuning has been done by changing the penalty, C, and solver parameter of the logistic regression method, from the hyperparameter tuning result the best combination is by using the value of 10 as the C parameter, l1 as the penalty parameter, and saga as the solver parameter, by doing so the accuracy of the model that previously is 85.72%, now it is becomes 92.91%.

After the model is done, the next step is to design the GUI, the GUI will show a graph of three days' worth of sensor data and also giving a prediction result based on the last three hours data, last one day data, and last three days data.

2.5 Model Prediction

The inputs are the latest data from the sensor, which already in the same format as the prepared data. Then the machine learning will retrieve those data and run it into the predictive model and giving the output of some prediction in different time windows. The time windows are the last three hours, last one day, and last three days. Each time there is a new data in the input document, the GUI will show the updated graph and prediction.

2.6 Model Validation

The validation of the model created is done by picking some data that are used to build the database, and then copy that data into another document that will be read by the model as the input of the predictive model. Then the result from the predictive model is compared with the result from the company's guideline that have already been done prior the database making.

The validation done several times using data from different wells, and all of the results from the predictive models are the same compared to the result from the company's guideline.

3 Result and Discussion

From the available data, there are nine types of problems that can be detected. Those problem are low PI, pump wear, tubing leak, higher PI, increase in frequency, open choke, increase in water cut, sand ingestion, and closed valve (SSSV/surface valve). Each of those problems have different patterns that can be seen in Table 5.

Because there is an imbalance in the data, it needs to be oversampled first, after the process of oversampled, the chosen method is the logistic regression method with 85.72% accuracy, the reason why the decision tree analysis method and the k-nearest neighbors method are not the ones selected because of the accuracy for both method are overconfident, with 99.72% and 99.8% accuracy respectively, this is overconfident given that the database is made through an oversampling method, the data that made from SMOTE method will be too ideal for the model.

Because of that, part of the test data and training that used by the machine learning to build the model also come from SMOTE, and it can cause overfitting.

Overfitting means that the model that created is too dependent with the training data, this can cause an overconfidence in accuracy. When there are new data that are not in the training data, the model with overfitting will have bigger error and will be more difficult to adjust when there is additional data.

To increase the accuracy of the selected method then hyperparameter tuning needs to be done where the model that previously has 85.72% accuracy, it becomes 92.91%.

For the purpose to make a data driven model, the chosen method is the lowest accuracy where the model will be easier to adjust to additional data, and alongside the accuracy of the model also increase.

The GUI that made can show the user two things, the first is the graph that show the graph of sensor data for the last three days, the second one is the prediction based on the last three hours data, last one day data, and last three days data. A smaller time window for the prediction can be made as long as the input data from the user is in a smaller time interval and the time interval is always the same throughout the whole input data. The example of the data as the input can be seen in Table 8 and the GUI graph can be seen in Figure 11 and the legend can be seen in Figure 12, then the prediction result can be seen in Figure 13. Figure 14 is the full GUI that the user will see from this program usage.

4 Conclusion

The conclusions that can be learn from this study are listed as follows:

1. There are nine recorded problems based on past sensor data that made into the database, those problems are:

a. Low PI
b. Pump wear
c. Tubing leak
d. Higher PI
e. Increase in frequency
f. Open choke
g. Increase in watercut
h. Sand ingestion
i. Closed valve (SSSV/surface valve)

2. The machine learning model is made using logistic regression method with 92.91% accuracy rate. The GUI shows the graph of sensor data in the last three days and the prediction based on the last three hours data, last one day data, and last three days data.

5 Recommendation

Many considerations have to be done prior to preparing the database, since using the approach by Camilleri et.al. (2015) for calculating rate has many assumptions (constant downhole motor voltage, constant power factor, constant motor efficiency, and constant pump efficiency). These assumptions make the values of the rate are not fully correct compared to the well test. Therefore, in the future, a method to calculate the flow rate based on the sensor data needs to be developed further.

Moreover, in the process of building the database, the data is limited because not all wells are equipped with real-time surveillance system in the ESP sensor. Hence, the number of wells that can be studied are also limited. If possible, an addition in real-time surveillance system for other wells will help in increasing the number of collected data.

The last thing that needs to be mentioned is that the sensor readings are not in constant time intervals. If the sensor readings are set to be run in the exact same time intervals, the process of time filtering is no longer needed and it can shorten the process of database making.

References

[1] Adesanwo, M., Denney, T., Lazarus, S., & Bello, O. (2016). Prescriptive-Based Decision Support System for Online Real-Time Electrical Submersible Pump Operations Management. SPE Intelligent Energy International Conference and Exhibition. Aberdeen: Society of Petroleum Engineers.

[2] Awaid, A., Al-Muqbali, H., Al-Bimani, A., Yazeedi, Z., Al-Sukaity, H., Al-Harthy, K., & Baillie, A. (2014). ESP Well Surveillance using Pattern Recognition Analysis, Oil Wells, Petroleum Development Oman. International Petroleum Technology Conference. Doha: International Petroleum Technology Conference

[3] Bermudez, F., Carvajal, G.A., Moricca, G., Dhar, J., Adam, F.Md., Al-Jasmi, A., Goel, H.K., & Nasr, H. (2014). Fuzzy Logic Application to Monitor and Predict Unexpected Behavior in Electric Submersible Pumps. SPE Intelligent Energy Conference and Exhibition. Utrecht: Society of Petroleum Engineers.

[4] Camilleri, L., El Gindy, M., Rusakov, A., & Adoghe, S. (2015). Converting ESP Real-Time Data to Flow Rate and Reservoir Information for a Remote Oil Well. *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. Abu Dhabi: Society of Petroleum Engineers.

[5] Coopersmith, Ellen, Dean, Graham, McVean, Jason, & Storaune, Erling. (2001). Making Decision in the Oil and Gas Industry. USA: Schulmberger

[6] Grassian, D., Bahatem, M., Scott, T., Olsen, D. (2017). Application of a Fuzzy Expert System to Analyze and Anticipate ESP Failure Modes. *Abu Dhabi International Petroleum Exhibition & Conference*. Abu Dhabi: Society of Petroleum Engineers.

[7] Muller, Andreas C., & Sarah, Guido (2017). Introduction to Machine Learning with Python. Sebastopol: O'Reilly Media Inc.

[8] Scikit-Learn Developers (2019). *Logistic Function*. Retrieved from Scikit Learn: https://scikit-learn.org/stable/auto_examples/linear_model/plot_logistic.html

[9] Vidrianto, M., & Utami, T. S. (2018). *Analisis Efisiensi Energi Electrical Submersible Pump pada Lapangan Minyak Tua di Laut Jawa*. Paper presented at Prosiding Full-Paper Seminar Nasional Teknologi dan Manajemen Teknologi (SNTMT) UI 2018, Depok, Indonesia.

List of Figure

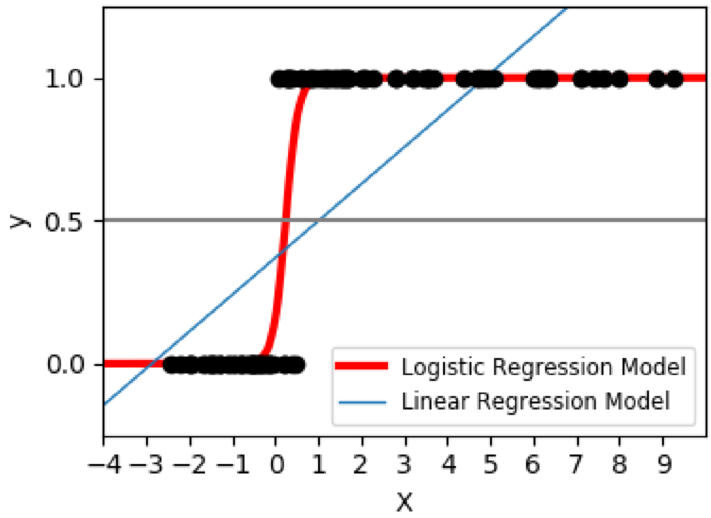

Figure 1. Logistic Regression Model (Scikit Learn Developers, 2019)

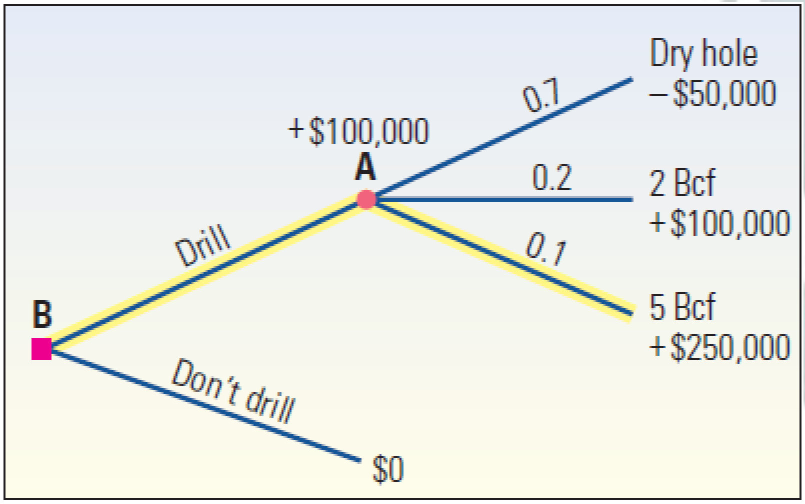

Figure 2. Decision Tree Analysis (Coopersmith, Ellen et.al., 2001)

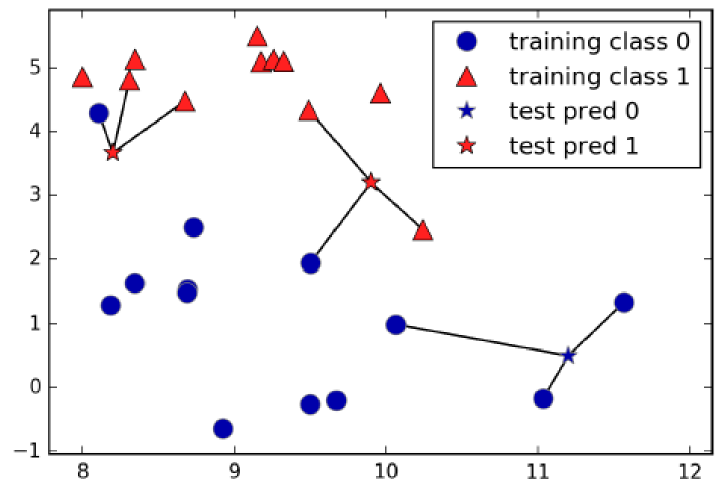

Figure 3. K-Nearest Neighbors Model (Muller, Andreas C., & Guido, Sarah, 2017)

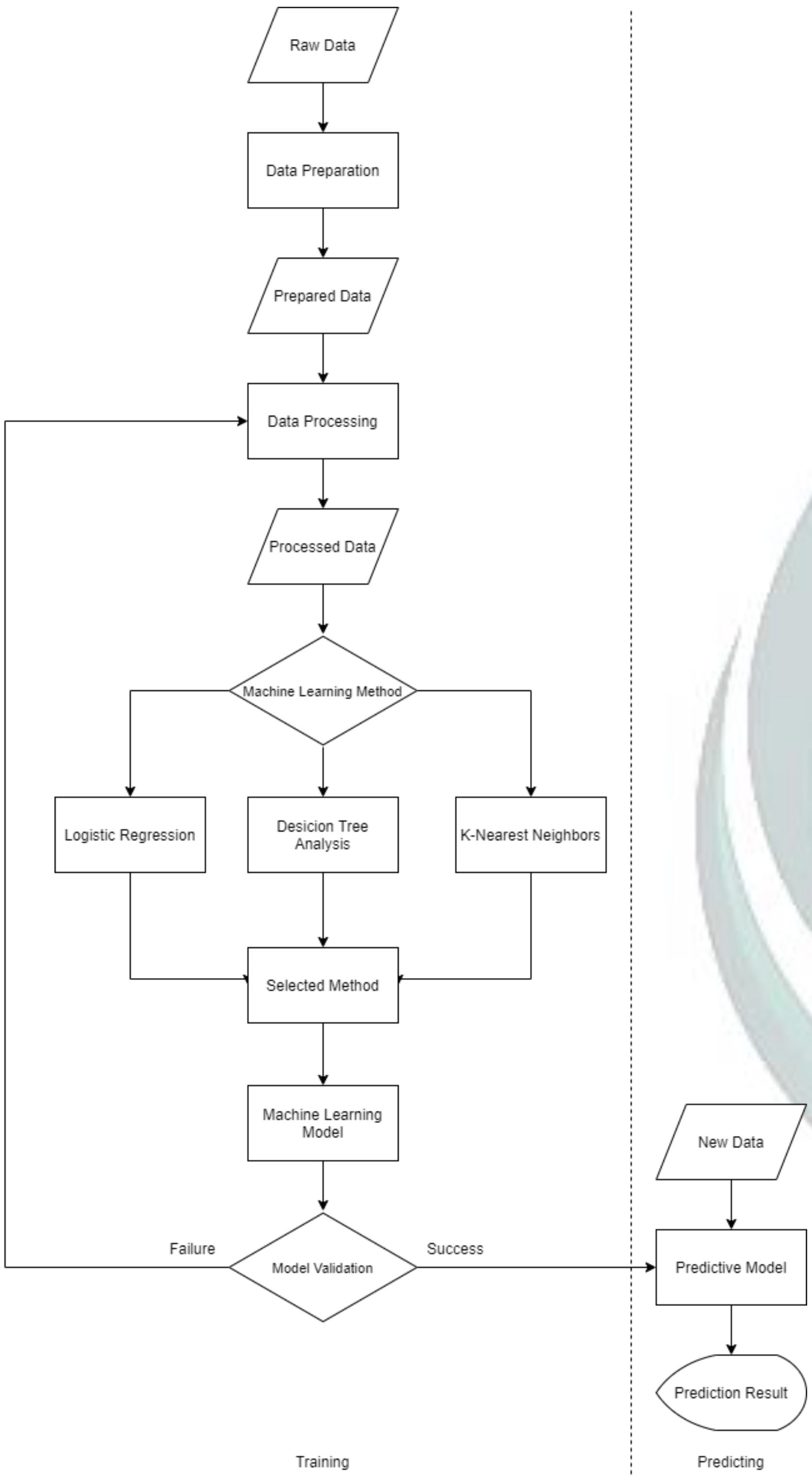

Figure 4. Study Workflow

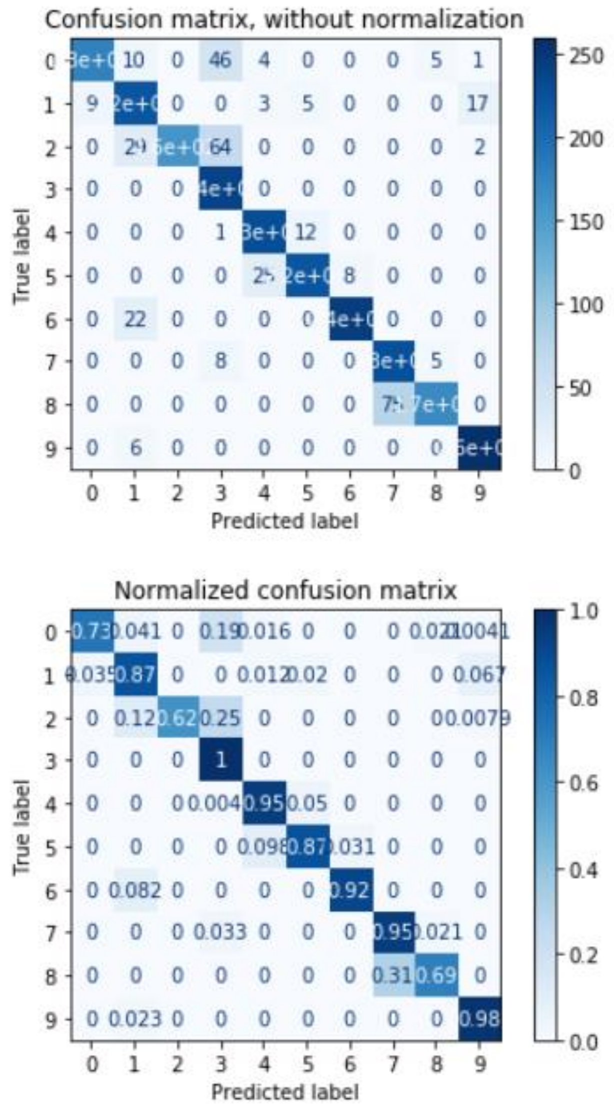

Figure 5. Logistic Regression's Matrix Plot

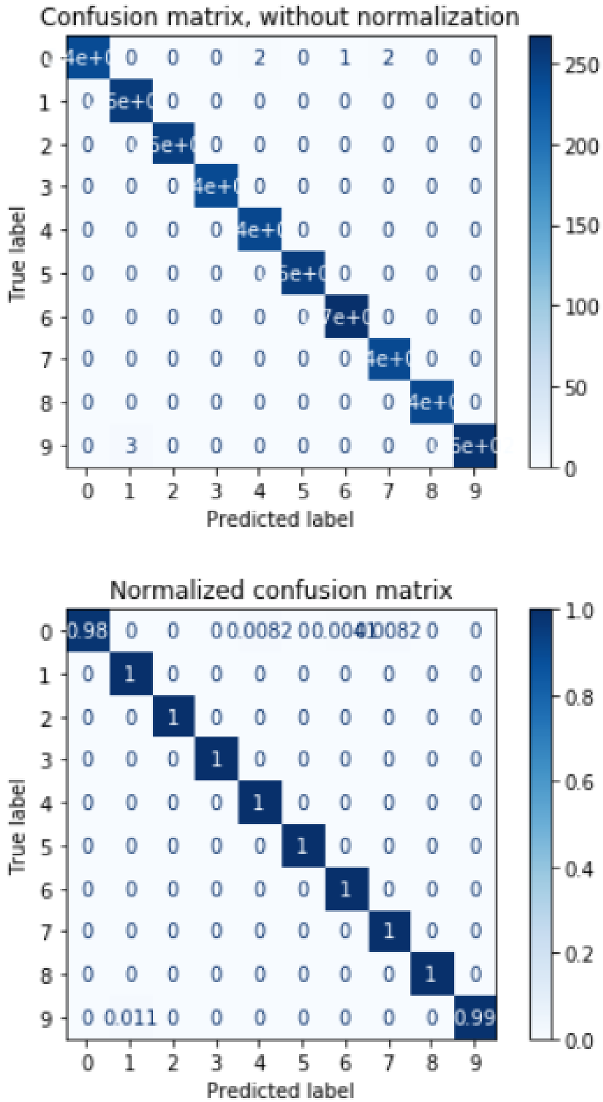

Figure 6. Decision Tree Analysis's Matrix Plot

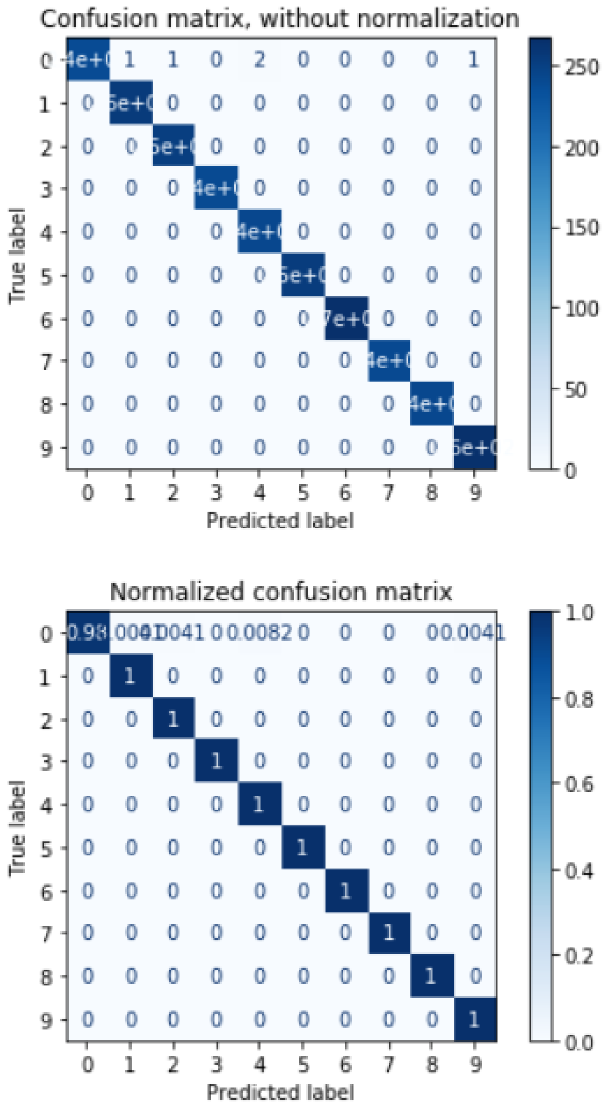

Figure 7. K-Nearest Neighbors's Matrix Plot

Time Window	Condition
Last 3 hours	Open Choke
Last 1 day	Open Choke
Last 3 days	Low PI

Figure 8. Model Comparison Result using Logistic Regression Method

Time Window	Condition
Last 3 hours	Running
Last 1 day	Open Choke
Last 3 days	Running

Figure 9. Model Comparison Result using Decision Tree Analysis Method

Time Window	Condition
Last 3 hours	Open Choke
Last 1 day	Open Choke
Last 3 days	Open Choke

Figure 10. Model Comparison Result using K-Nearest Neighbors Method

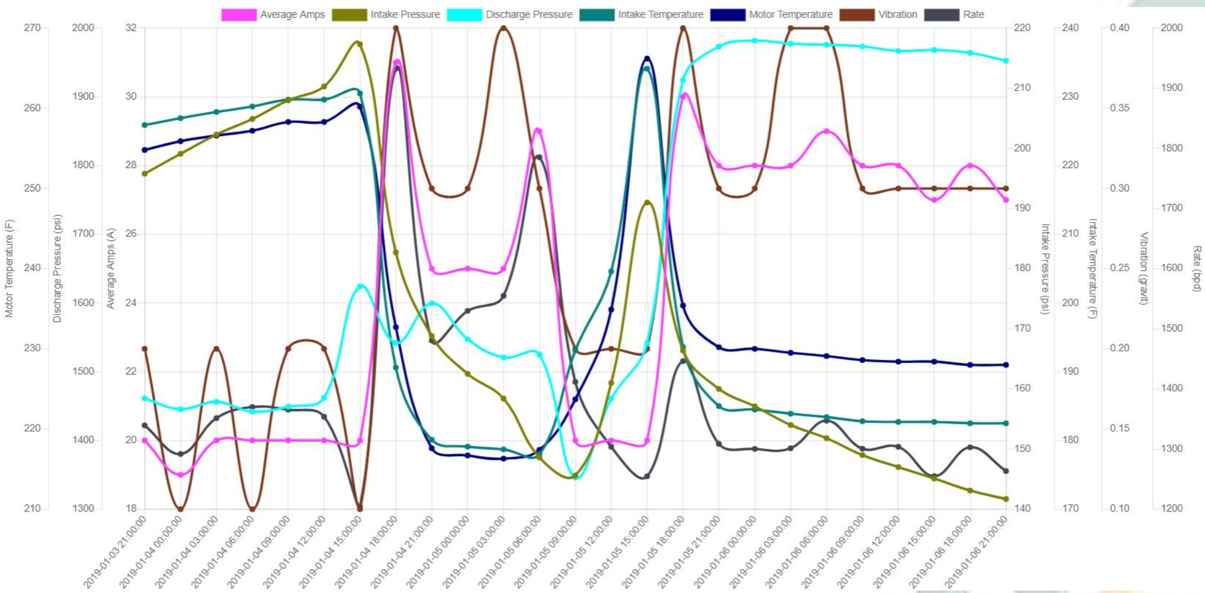

Figure 11. Graphic User Interface of Sensor Data Graph

Figure 12. Sensor Data Graph's Legend

Time Window	Condition
Last 3 hours	Closed Valve
Last 1 day	Low PI
Last 3 days	Sand Ingestion

Figure 13. Graphic User Interface of Sensor Data Prediction Result

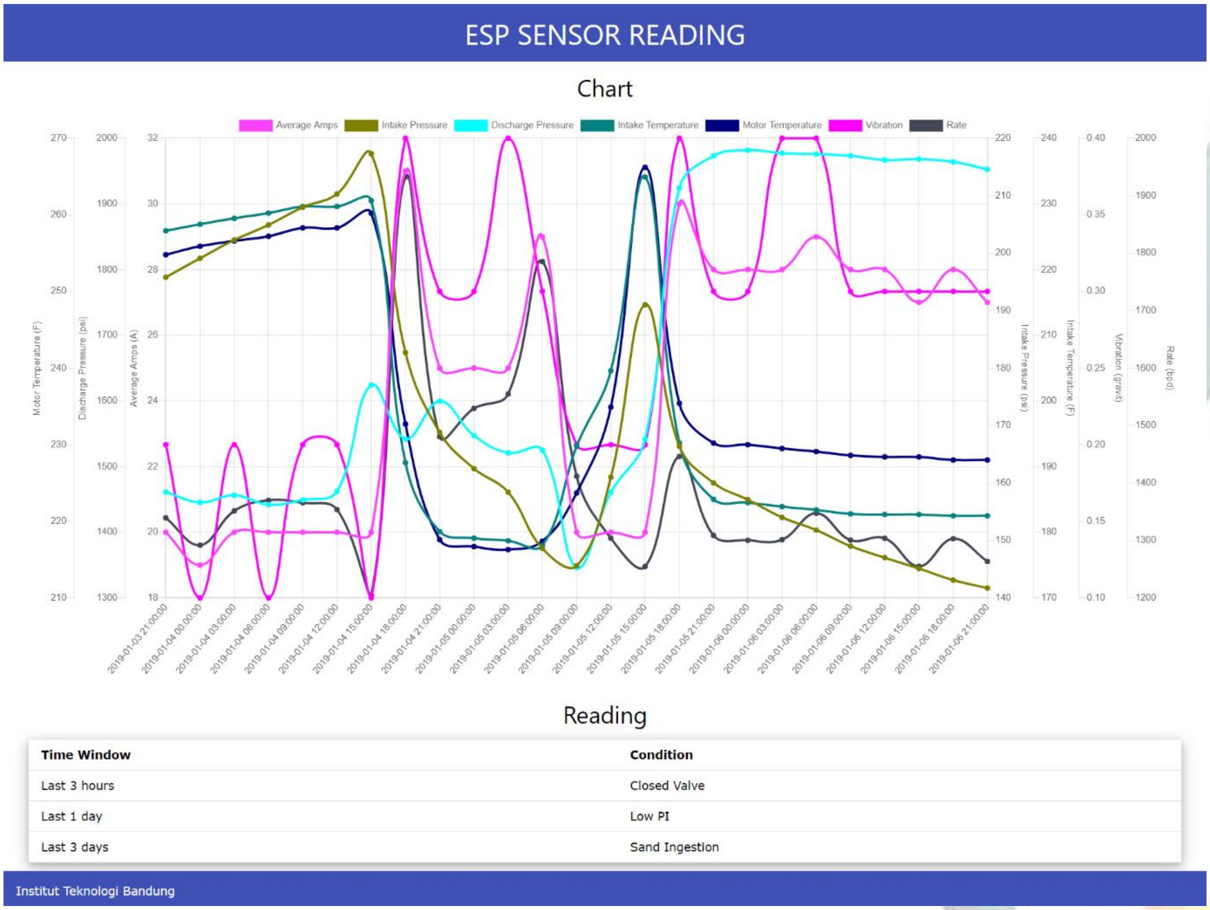

Figure 14. Graphic User Interface

List of Tables

Table 1. Raw Data

UWI	Reading Time	Average Amps (A)	Intake Pressure (psi)	Discharge Pressure (psi)	Intake Temperature (F)	Motor Temperature (F)	Vibration (gravit)
CBU-02	01/01/2019 00:00	57	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 00:08	56	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 00:49	56	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 01:00	57	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 02:00	56	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 03:00	57	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 04:00	57	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 05:00	57	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 06:00	56	4331.6	1599.7	284.2	6023.7	0
CBU-02	01/01/2019 07:00	56	4331.6	1599.7	284.2	6023.7	0

Table 2. Prepared Data

TIME	AVERAGE AMPS	INTAKE PRESSURE	DISCHARGE PRESSURE	INTAKE TEMPERATURE	MOTOR TEMPERATURE	VIBRATION	RATE
18/07/2019 09:00	39	761.9	3889.4	256.8	285.1	0.1	1163.945899
18/07/2019 12:00	39	745.6	3880.5	257	285.1	0.2	1161.198379
18/07/2019 15:00	38	737.5	3882.2	257	285.1	0.2	1127.898143
18/07/2019 18:00	39	731	3885.6	257.2	285.3	0.2	1153.946871
18/07/2019 21:00	38	745.3	3965.1	256.8	284.5	0.2	1101.590562
19/07/2019 00:00	38	751.5	3958.7	257.2	285.1	0.1	1105.918337
19/07/2019 03:00	38	747.7	3927.9	257.4	285.3	0.1	1115.30762
19/07/2019 06:00	38	742.9	3942.1	257.5	285.8	0.1	1108.683825
19/07/2019 09:00	38	736.2	3919.4	257.2	285.4	0.2	1114.2565
19/07/2019 12:00	39	713.1	3813.6	258.1	286.2	0.2	1174.081858

Table 3. Processed Data

WELL	START DATE	END DATE	SLOPE OF AVERAGE AMPS	A	SLOPE OF INTAKE PRESSURE	IP	SLOPE OF DISCHARGE PRESSURE	DP
CBU-02	15/03/2019 00:00	17/03/2019 21:00	-0.660869565	-1.3%	4.736	1.1%	-2.970434783	-0.1%
CBU-02	16/03/2019 00:00	18/03/2019 21:00	-2.069565217	-4.9%	32.96278261	14.5%	-35.3613913	-1.2%
CBU-02	17/03/2019 00:00	19/03/2019 21:00	-1.596521739	-3.3%	49.49217391	21.4%	-33.20173913	-1.0%
CBU-02	18/03/2019 00:00	20/03/2019 21:00	-0.699130435	-1.4%	52.22295652	18.3%	-10.52730435	-0.3%
CBU-02	19/03/2019 00:00	21/03/2019 21:00	1.213913043	2.5%	25.11791304	7.2%	56.47373913	1.9%
CBU-02	20/03/2019 00:00	22/03/2019 21:00	2.020869565	5.5%	23.67686957	6.5%	43.92626087	1.4%
CBU-02	21/03/2019 00:00	23/03/2019 21:00	-1.391304348	-2.8%	22.50678261	5.9%	-47.42956522	-1.4%
CBU-02	22/03/2019 00:00	24/03/2019 21:00	-1.325217391	-2.7%	5.654956522	1.4%	-13.58434783	-0.4%
CBU-02	23/03/2019 00:00	25/03/2019 21:00	1.631304348	4.2%	-3.586782609	-0.8%	65.97982609	2.2%
CBU-02	24/03/2019 00:00	26/03/2019 21:00	0.831304348	1.7%	-8.903304348	-2.2%	25.4306087	0.7%

Table 4. Processed Data (continue)

SLOPE OF INTAKE TEMPERATURE	IT	SLOPE OF MOTOR TEMPERATURE	MT	SLOPE OF VIBRATION	V	SLOPE OF RATE	R	TRIP
-2.399304348	-0.9%	-3.515826087	-1.2%	-0.025043478	-12.5%	-17.46566446	-1.1%	0
0.432	0.2%	-0.005565217	0.0%	0.020521739	20.5%	-34.60744067	-2.5%	2
0.574608696	0.2%	0.079304348	0.0%	0.007652174	3.8%	-12.88881974	-0.9%	2
0.558608696	0.2%	0.702608696	0.2%	-0.013913043	-7.0%	8.283170455	0.5%	3
0.353043478	0.1%	1.285565217	0.5%	0.011826087	5.9%	23.32830964	1.4%	4
0.223304348	0.1%	-0.43826087	-0.2%	0.020173913	20.2%	55.90660734	4.3%	4
0.289391304	0.1%	-0.511652174	-0.2%	0.017043478	17.0%	-10.31272812	-0.7%	2
0.240347826	0.1%	1.305391304	0.5%	-0.003826087	-3.8%	-34.90742	-2.3%	2
0.018086957	0.0%	0.045565217	0.0%	-0.000695652	-0.3%	17.72158176	1.3%	5
-0.116869565	0.0%	-0.491130435	-0.2%	0.010434783	5.2%	9.979327614	0.7%	0

Table 5. Pattern Recognition Rule Sets

Well Condition	Average Amps	Intake Pressure	Discharge Pressure	Intake Temperature	Discharge Temperature	Vibration	Rate	Index
Low PI	I/D	D	D	-	-	-	D	1
Pump Wear	D	I	D	-	-	I	D	2
Tubing Leak	C	I	D	I	I	-	D	3
Higher PI	I/D	I	I	-	-	-	I	4
Increase in Frequency	I	D	I	-	I	-	I	5
Open Choke	I	D	D	-	C	-	D	6
Increase in Watercut	I	I	I	-	C	-	D	7
Sand Ingestion	I	D	I/D	-	I/D	I/D	D	8
Closed Valve (SSSV/Surface Valve)	D	I	I	I	I	-	D	9

I = increase, C = constant, D = decrease, - = has no effect

Table 6. Data Recorded

Well Condition	Data Recorded
Low PI	12
Pump Wear	19
Tubing Leak	3
Higher PI	16
Increase in Frequency	9
Open Choke	24
Increase in Watercut	6
Sand Ingestion	7
Closed Valve (SSSV/Surface Valve)	5

Table 7. Input Data for Model Comparison

TIME	A	IP	DP	IT	MT	V	R
29/03/2019 21:00	48	391	3396.8	261.1	290.8	0.2	1490.550392
30/03/2019 00:00	37	414	2975.3	261.1	287.8	0.2	1348.362856
30/03/2019 03:00	38	411.9	2976.7	261.1	288.1	0.2	1382.915351
30/03/2019 06:00	50	438.5	3418.7	261.1	284.2	0.2	1565.994022
30/03/2019 09:00	49	408.4	3434.8	261.1	287.1	0.2	1511.246324
30/03/2019 12:00	48	374.4	3390.2	261.1	290.1	0.1	1485.607921
30/03/2019 15:00	49	389.9	3423.1	261	288.3	0.2	1507.858327
30/03/2019 18:00	49	402.8	3428.1	261.1	286.9	0.2	1511.795814
30/03/2019 21:00	49	400.2	3424.7	261.1	287.8	0.2	1512.195694
31/03/2019 00:00	38	408.5	2978.5	261	286.2	0.2	1380.117234
31/03/2019 03:00	37	411.6	2981.5	261	285.4	0.2	1343.85065
31/03/2019 06:00	50	402.5	3422.9	260.8	285.1	0.2	1545.151432
31/03/2019 09:00	50	412.8	3430.9	261	284.5	0.2	1546.328943
31/03/2019 12:00	45	354.4	3380.5	260.6	290.1	0.2	1388.016868
31/03/2019 15:00	49	402.4	3426.7	260.6	285.8	0.2	1512.295697
31/03/2019 18:00	49	376.8	3414.6	260.4	288.1	0.1	1505.575047
31/03/2019 21:00	38	401.9	2977.4	260.4	283.3	0.2	1377.169983
01/04/2019 00:00	47	339.6	3378.5	260.4	288.9	0.2	1443.60027
01/04/2019 03:00	49	354.7	3383.9	260.4	284.9	0.2	1509.849424
01/04/2019 06:00	49	356	3380.8	260.1	283.8	0.2	1512.045714
01/04/2019 09:00	36	336.7	2983.8	260.4	288.5	0.2	1269.397558
01/04/2019 12:00	49	344.7	3401.4	260.1	286	0.2	1496.265867
01/04/2019 15:00	37	351.1	2966.8	260.2	285.4	0.2	1320.320291
01/04/2019 18:00	44	320	3334.1	260.2	288.1	0.2	1362.575342
01/04/2019 21:00	48	355.8	3025.8	260.2	284	0.2	1678.013621

Table 8. Input Data for Predictive Model

TIME	A	IP	DP	IT	MT	V	R
03/01/2019 21:00	20	195.8	1461.1	225.9	254.8	0.2	1339.383
04/01/2019 00:00	19	199.1	1445.4	226.9	255.9	0.1	1291.812
04/01/2019 03:00	20	202.3	1456.3	227.8	256.6	0.2	1351.452
04/01/2019 06:00	20	204.9	1442	228.6	257.2	0.1	1369.914
04/01/2019 09:00	20	208	1449.1	229.6	258.3	0.2	1365.499
04/01/2019 12:00	20	210.3	1462.2	229.6	258.3	0.2	1353.719
04/01/2019 15:00	20	217.3	1624.2	230.5	260.2	0.1	1204.578
04/01/2019 18:00	31	182.7	1541.9	190.6	232.7	0.4	1932.62
04/01/2019 21:00	25	168.8	1599.5	180.1	217.6	0.3	1480.674
05/01/2019 00:00	25	162.5	1547.2	179.1	216.7	0.3	1529.863
05/01/2019 03:00	25	158.4	1520.8	178.7	216.3	0.4	1554.904
05/01/2019 06:00	29	148.6	1525	178	217.4	0.3	1785.342
05/01/2019 09:00	20	145.6	1346.1	193.1	223.7	0.2	1411.679
05/01/2019 12:00	20	161	1460.8	204.6	234.9	0.2	1303.832
05/01/2019 15:00	20	191	1541.7	234.1	266.2	0.2	1254.698
05/01/2019 18:00	30	166.4	1924.2	193.6	235.4	0.4	1446.172
05/01/2019 21:00	28	160	1973.1	185	230.2	0.3	1308.593
06/01/2019 00:00	28	157.1	1981.8	184.5	230	0.3	1300.274
06/01/2019 03:00	28	154	1977.3	183.9	229.5	0.4	1301.272
06/01/2019 06:00	29	151.8	1975.8	183.4	229.1	0.4	1347.229
06/01/2019 09:00	28	149	1973.3	182.8	228.6	0.3	1300.559
06/01/2019 12:00	28	147	1966.8	182.7	228.4	0.3	1303.775
06/01/2019 15:00	27	145.1	1968.2	182.7	228.4	0.3	1254.936
06/01/2019 18:00	28	143.1	1964	182.5	228	0.3	1302.987
06/01/2019 21:00	27	141.7	1952.6	182.5	228	0.3	1263.39