International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 4
ISSN: 2088-5334

Automate Short Cyclic Well Job Candidacy Using Artificial Neural Networks–Enabled Lean Six Sigma Approach: A Case Study in Oil and Gas Company

Dedi Wilantara [a,*], Nur Budi Mulyono [a], Doni Winarso [b]

[a] *School of Business and Management, Bandung Institute of Technology, Indonesia*
[b] *Information Systems, Muhammadiyah University of Riau, Indonesia*
Corresponding author: [*]dedi_wilantara@sbm-itb.ac.id

Abstract*—Artificial Neural Networks (ANNs) are a part of Artificial Intelligence (AI) that is commonly used for pattern recognition, regression and classification. This technology allows us to learn historical data and generate patterns from the precedent data. In oil and gas companies, large amounts of data are produced every day. Many accurate decisions in this type of company are made from the data. Cilon Indonesia (CI) Co. Ltd. is one of the oil and gas companies currently operating the largest oil field in Indonesia. This type of company's operation and financial profit depends on oil price, which is affected by global oil supply and demand. If oil prices fall suddenly, all oil and gas companies need to run their businesses more efficiently and effectively. There are many ways to make this kind of company run their business effectively and efficiently by implementing several strategies such as capital cost efficiency, operational cost efficiency and even laying off some employees. One of the major costs in operation in oil and gas companies is the cost for well workover. This well workover does not always produce oil gain. In fact, even it is resulting in oil gain, but not all well workover programs are economical whenever the oil price is low. This condition makes Petroleum Engineer (PE) need to select the best well workover for certain wells. Well candidates for workover are usually selected manually using data from many resources, reports and information. Well candidates are reviewed one by one, and with several criteria, the well is proposed to a certain type of well workover. This research explains how this company improves their selection of well candidates for the most economic workover called Short Cyclic Steam Stimulation (SCSS). The process improvement is done using the hybrid method: lean six sigma method and big data analytics method, which utilize ANNs to predict the oil after workover executed. The result demonstrates how this hybrid method can improve the process with a sustainable solution. Its successful improvement in PE time selects SCSS well candidates from 2 hours to 10 minutes to generate 20 wells per day. Its also improve the success rate of SCSS workover from 61% to 73%.

***Keywords*— Artificial intelligence; artificial neural networks; big data analytics; cyclic steam injection; DMAIC; hybrid method; information technology; lean, machine learning; oil and gas company; short cyclic steam stimulation; six sigma.**

Manuscript received 7 Aug. 2020; revised 25 Nov. 2020; accepted 15 Dec. 2020. Date of publication 31 Aug. 2022.

I. INTRODUCTION

Indonesia is a country full of wealth that comes from natural resources. Indonesia's main export capital is natural wealth [1]. One of the Indonesia's wealth is crude oil. Cilon Indonesia (CI) Co., Ltd. is one of the oil and gas companies currently operating the largest oil field in Indonesia. One of the biggest onshore fields of CI Co., Ltd. is the Vuri field. Around 6000 active wells in this field produce crude oil using the Sucker Road Pump (SRP) to lift oil from the bottom to the surface. Onshore oil wells are dependent on maintenance services such as cleaning, reinstatement, and stimulation [2]. Well maintenance (well service and workover) is an operation needed by the oil company to guarantee the optimum production of its oil well [3]. To maintain oil production from the wells optimally and also to reduce the Loss Production Opportunity (LPO), it is necessary to do some workover jobs to the wells as part of maintenance service. There are many types of well workover jobs. Each type of workover has a different cost and oil gain result because of the decline in oil price from US$110 per barrel in 2014 to US$50 per barrel in 2019, CI Co., Ltd. requires to select the most economical workover job as part of operational cost efficiency strategy. The most economical proactive job in the Vuri field is Short Cyclic Steam Stimulation (SCSS), with a total job cost is about US$750/job [4].

SCSS is part of the cyclic steam injection workover type. Cyclic Steam Injection, also called Huff n' Puff, is a thermal recovery method that involves the periodical injection of steam with the purpose of heating the reservoir near the wellbore, in which one well is used as both injector and producer, and a cycle consisting of 3 stages, injection, soaking and production [5].

The SCSS job is the most favorite workover with the current oil price (US$ 50/barrel). The implementation of this workover increased significantly in the last four years from 30 wells/month to 360 wells/month [6]. Those well workover jobs were selected manually by Petroleum Engineers (PE) one by one from more than 6000 active wells in the Vuri field.

The manual process of SCSS job candidacy begins with collecting the most recent data of the well properties, checking the current well work opportunities, and analyzing data such as the production profile and trend, the surveillance data, pump data, and last workover historical performance [7]. The manual process of SCSS job candidacy is very complex and is usually carried out by the PE within 2 hours per day to get 20 wells of SCSS job candidates because data and information come from various sources and reports. The well candidates also come from many PEs who may have different thoughts in selecting the SCSS candidate, which brings the inconsistent criteria for selecting the best candidate for SCSS workover job. In addition, because of this process, humans sometimes follow the human mood and often encourage human error on selecting the wells, which causes a low success ratio of SCSS workover. Machine Learning is a rising technique of Programming that allows computers to learn like a human brain. Nowadays, most companies and applications take advantage of these techniques to improve their products or services [8] efficiently. The Goal of todays of machine learning is not to create an artificial brain but to assist us with making sense of the world's massive data stores [9].

From the explanation above, there is an opportunity for improvement. Through the lean six sigma approach, the SCSS job candidacy process can be improved following the DMAIC Cycles: Define, Measure, Analyze, Improve, and Control. The Machine Learning (ML) algorithm-Artificial Neural Networks (ANNs) were used to improve the process and achieve the desired targets on increasing the success ratio.

II. Materials and Method

The process improvement of the case above employed the lean six sigma approach through Define, Measure, Analyze, Improve, and Control (DMAIC) [10]. Many papers implement the Lean Six Sigma (LSS) approach in various industries such as in automotive, micro, small-medium enterprises, healthcare, education, financial, and many more [11]. In healthcare, LSS was applied to analyze a clinical pathway that successfully improved hospital stay length (LOS) from 10.66 to 7.8 days [12]. The LSS also applied within the university to improve processes in curriculum delivery, business and auxiliary services, admissions and enrollment management, and research [13]. Zhang *et al.* [14] explained that LSS is also implemented in many logistic industries in Singapore, and they reported varying degrees of cost savings and productivity improvements.

In this research project, the process of improvement using LSS method that was combined with Big Data Analytics as a Hybrid Method for the conceptual research framework. This hybrid method is still seldom discussed and addressed in the literature. The illustration of the hybrid method of this conceptual research framework is as follows:

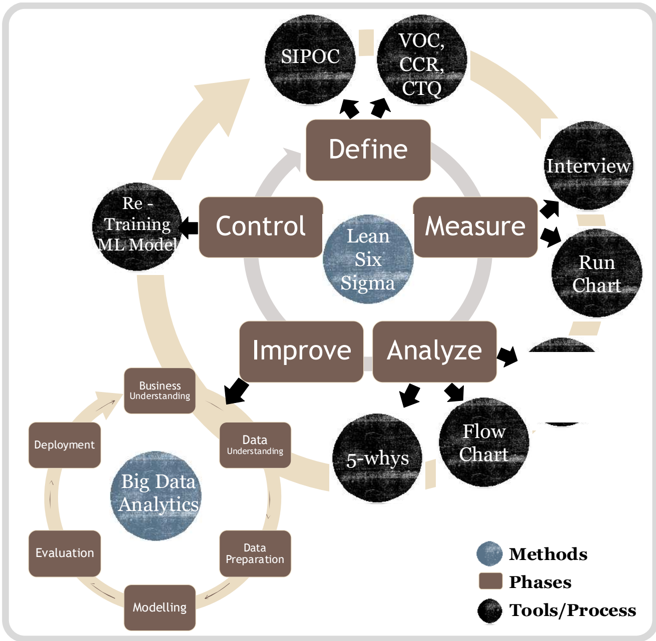

Fig. 1 Hybrid Method for the conceptual research framework

This research starts with LSS with the DMAIC phase in CI Co., Ltd., a simple explanation of general DMAIC activities for each phase in CI Co., Ltd. as follows:

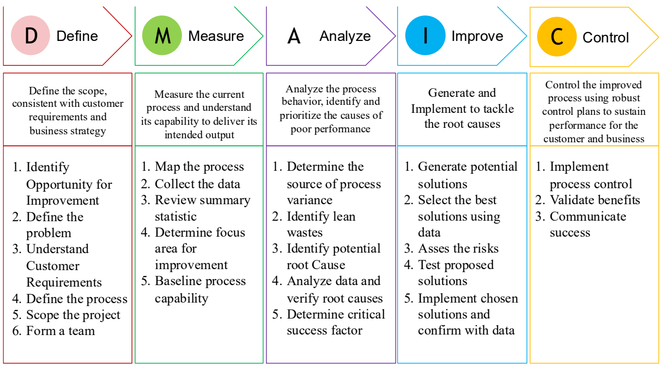

Fig. 2 DMAIC framework at CI Co., Ltd. [15]

A. *Define Phase*

This phase identified several opportunities for improvement by understanding the current process, customer problems, and desire improvement target. SIPOC diagram is a simple diagram to identify the basic elements of a process (boundaries, supplier inputs, steps, customers, and outputs) [16]. It also describes the process snapshot of the current process to help teams and sponsors agree on project boundaries and also the desired targets that need to be achieved. In the SCSS well candidacy process, the SIPOC diagram can be seen in the following figure:

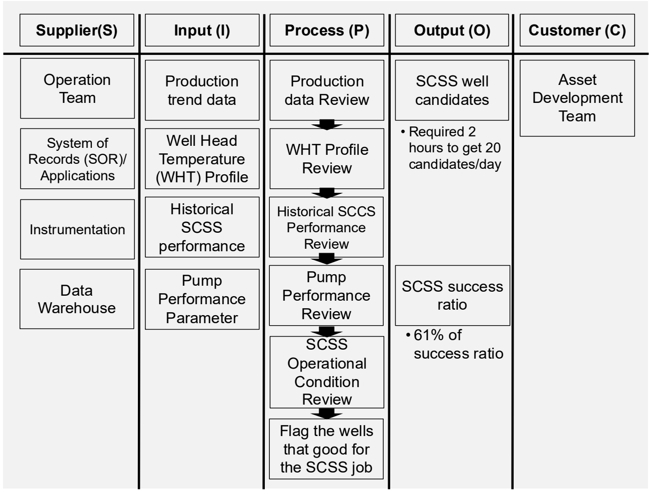

Fig. 3 SIPOC Diagram of SCSS well candidacy process

As mentioned in Figure 3, the process of well candidacy of SCSS is done by reviewing lots of data and information, including production data review, well heat temperature (WHT) review, historical SCSS performance review, pump performance review, and operational condition review. All the above processes provide an output of the list of well candidates for SCSS workover. The list of well candidates was executed and generated how much success ratio. After mapping the current process, this phase tries to collect the Voice of Customer (VOC) and translate them into Critical Customer Requirements (CCR) as well as Critical to Quality (CTQ) for the current SCSS candidacy process.

TABLE I
VOICE OF CUSTOMER (VOC) , CUSTOMER CRITICAL REQUIREMENT (CCR) AND CRITICAL TO QUALITY (CTQ)

Voice of Customer (VOC)	Customer Critical Requirement (CCR)	Critical to Quality CTQ
It takes quite a long to review one by one of well production trending, historical SCSS job performance, and pump performance parameters because of all the information needed in multiple data sources/reports.	Reduce time to collect well data, review, identify and flag it as a good well for SCSS job.	Less than 30 minutes to get up to 20 wells for 20 well candidates of SCSS job.
Each Petroleum Engineer has different criteria to flag the well as good or suitable well for SCSS job	Standardization of SCSS job well candidate criteria	All PEs follow new standards of SCCS job well candidate criteria that programmed to the new tool
It is difficult to prioritize well candidates for SCSS job execution because there are no parameter guidelines to prioritize them.	Provide new parameter to prioritize well candidates for SCSS execution	Require at least 1 parameter to prioritize SCSS well candidates for execution
The success ratio of SCSS job currently still need to be improved.	Increase success ratio of SCSS job	Success Ratio increase at least 5% from the current success ratio

The VOC, CCR and CTQ are collected through face-to-face interviews with the Petroleum Engineers (PE). The following Table I is the VOC, CCR and CTQ of the current SCSS candidacy process. According to the VOC, CCR and CTQ Table I above, it is very clear that 4 opportunities can be improved in the SCSS well candidacy process such as 1) Reduce the time to collect SCCS well candidates with target less than 30 minutes for 20 well candidates 2) Develop standard well candidates criteria that to be followed by all PEs 3) Generate at least 1 new parameter to prioritize SCSS workover candidates 4) Increasing the success ratio at least 5% from the current success ratio of SCSS workover job.

B. Measure Phase

After opportunities for improvement have been identified, this phase aims to fully understand the current state of the process and collect reliable data on process speed, quality, and cost. Also, this phase aims to expose the underlying causes of problems [16]. From the four opportunities that have been identified, some of them can be measured (Quantitative), and some of them cannot be measured (Qualitative) as described in the following table:

TABLE II
QUANTITATIVE AND QUALITATIVE OPPORTUNITIES FOR IMPROVEMENT

No	Opportunity for Improvement	Qualitative/ Quantitative	Unit Measurement
1	Reduce time to review some data for SCSS well candidates	Quantitative	Hours
2	Standardized SCSS well candidates' criteria	Qualitative	Not Applicable
3	Generate new parameter for prioritizing SCSS workover execution	Qualitative	Not Applicable
4	Increase the success ratio of SCSS workover job	Quantitative	Percentage

As mentioned in Table II, there are 2 opportunities of 4 opportunities that improvement: 1) Time to review some data and 2) Success Ratio of SCCS workover Job. This phase collected and measured the reliable data for those two quantified opportunities for improvement:

1) Time to review SCSS candidates: According to interviews with Petroleum Engineers who handle SCSS well candidacy, the time to collect SCSS candidates is as follows.

$\boxed{\textbf{2 hours for 20 candidates per day}}$

2) SCSS workover Success Ratio: The success ratio is calculated based on the oil gain produced by comparing the 90 days of oil cumulative after SCSS job executed with the 90 days of oil cumulative. The assumption is that there is no work over a job well using a 90-day decline rate calculation. The illustration of oil gain produced by SCSS job can be illustrated in the following figure.

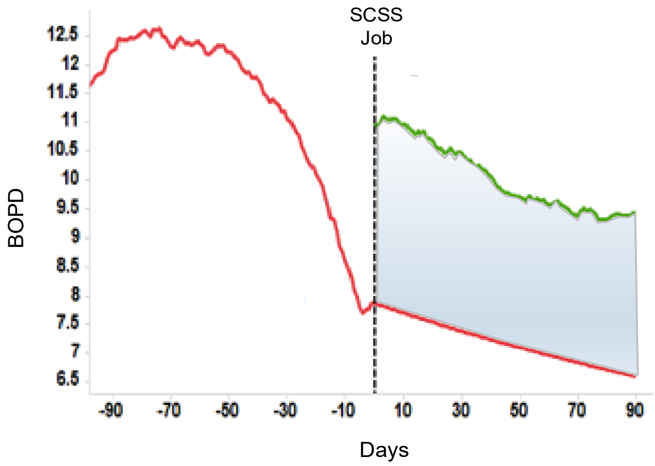

Fig. 4 Illustration of oil gain produce by SCSS job

Figure 4 illustrates the ideal production profile before and after SCSS workover job is executed in a certain well within 180 days. The red line is the actual oil production of the well 90 days before SCSS is executed. If there is no SCSS job is executed, with the calculated decline rate of 90 days, it will draw the line 90 days after. But, as the SCSS job is executed on day 0, the production profile increases, described by the green line. Therefore, the oil gain cumulative 90 days after SCSS job is executed is in the blue area.

The profitability index is the ratio computed by dividing the project's net present value by the required initial investment [17]. The break-even point of the project occurs if the profitability index is 1. Because of SCSS workover total cost is US$750 and the targeted profitability index at CI Co., Ltd. for workover job is 1.2, the target of Present Value (PV) of future cash flow on 90 days from this SCSS workover is US$900. Assuming an oil price is US$50 per barrel of oil, then expected oil gain cumulative for SCSS job in 90 days after the workover executed are **18 barrels or 0.2 barrels oil per day (BOPD).** Therefore, the SCSS workover job can be claimed as a successful workover if it can generate oil gain 18 barrels within 90 days. Based on data from July 2017 – December 2017, the success ratio of SCSS job for the manual process is as follow:

Fig. 5 Run chart success rate of SCSS Job

The average success rate SCSS job is **61%.** Since the research project started in early January 2018, 61% of the success ratio was used as a baseline of the SCSS success ratio that needs to be improved.

C. *Analyze Phase*

This phase aims to pinpoint and verify causes affecting the key input and output variables tied to project goals. Pareto Chart is specialized bar charts that help to focus on the "vital few" sources of trouble, and if we address it, it will have the biggest impact [16]. There are many review processes in the SCSS well candidacy process: Individual Review, Slider Review, Low Inflow Review, Other Review, Peer Heat Review, Drainage Review, Profitability Review, Optima Review. The following Pareto chart shows the contribution of each review process to SCSS well candidates:

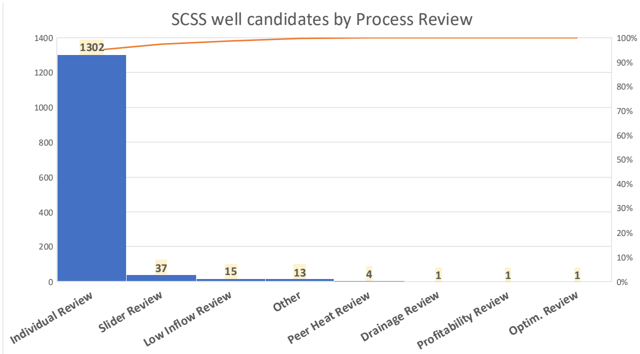

Fig. 6 Pareto chart SCSS well candidates by the review process

According to the Pareto chart figure 6 above, the individual review process contributed significantly to the SCSS well candidacy. However, since there are many Petroleum Engineers (PE) involved in the individual review process, further data analysis is needed. The following Pareto chart describes who has made a major contribution to SCSS candidates during the past 6 months:

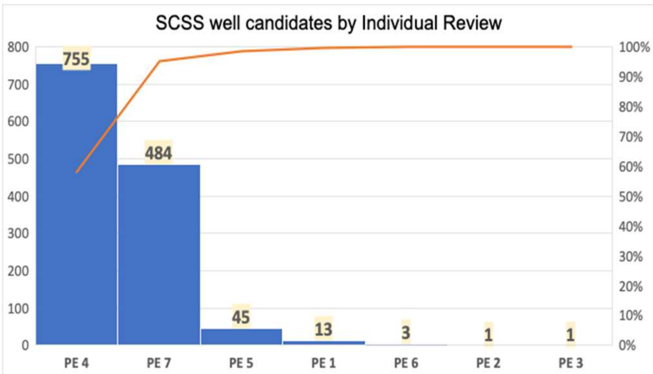

Fig. 7 Pareto chart SCSS well candidates by individual

From the individual Pareto chart figure 7 above, there are currently 2 Petroleum Engineers (PE) who contribute significantly to the SCSS well candidates (PE 4 and PE 7). Therefore, to focus on the root causes of the quantified opportunity of improvement and improve this current process, the discussion involves 2 PEs as Subject Matter Expert (SME). According to the SMEs above, figure 8 is the illustration of the current process on how they review a well and make list of well candidates as part of the individual review.

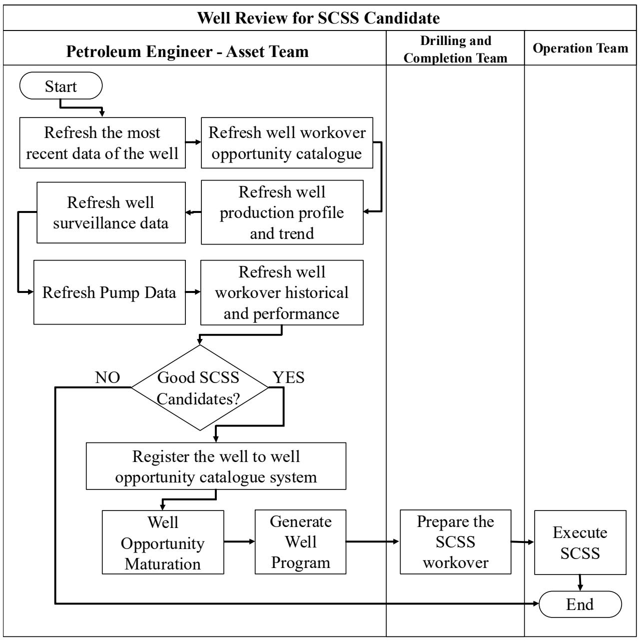

Fig. 8 Well review for SCSS candidate flowchart

The process is started by Petroleum Engineer (PE) in Asset team refreshing. Some of the data such as the most recent data of the well, workover opportunity catalog aims to see any other workover opportunity to the well, production profile and trend. Well surveillance data to see the well heat temperature, pump data and workover historical and performance. From this data, PE has decided whether the well is a good candidate for SCSS or not, then continues to register it in the well opportunity catalog system for the maturation process. After it is maturated, the program was generated and continue to the drilling and completion team for workover preparation. Finally, the SCCS workover job was executed by the operation team.

5-whys are one of the methods that pushing people to think about root causes. This method is selected for this research project as it is amazingly simple to implement and get the root cause by obvious problems and opportunities. The 5-whys analysis is commonly used in the manufacturing sector. Braglia *et al.* [18] used this 5-whys analysis method to find out the root causes associated with the problem identified form data analysis of production orders of a fashion-luxury company.

To produce the root cause, the 5-whys method begins with the identified problem statement, proceeds to why the problem occurred, and continues back to the second question, why that also happened, and so forth. According to Table II, there are two metrics of quantitative opportunities that need to be verified for what are the root causes using the 5-whys method:

1) The 5 Whys of SCSS review process consume a lot of time: The 5-whys for the root cause of this problem can be described in the following figure:

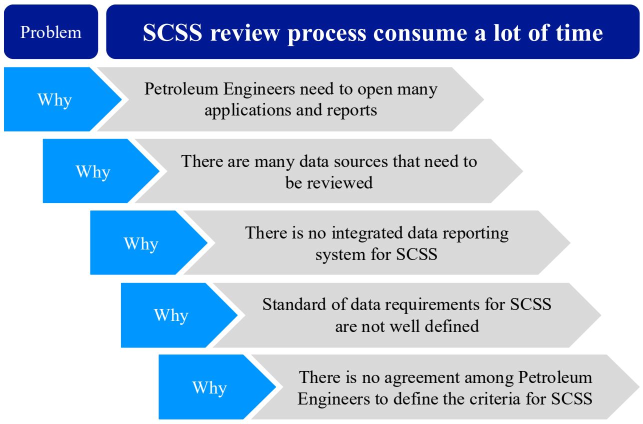

Fig. 9 The 5 Whys of SCSS review process consume a lot of time

2) SCSS workover Success Ratio is still low: The following figure is the root cause of SCSS workover success rate ratio using the 5-whys method.

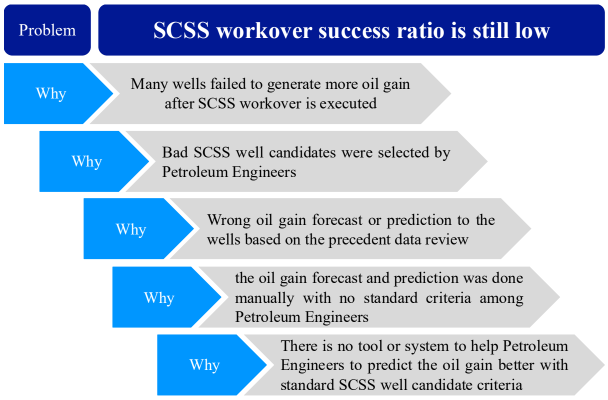

Fig. 10 The 5 Whys of SCSS workover success ratio is still low

According to the 5-whys figure 9 and figure 10 above, to solve the problems of SCSS well candidacy process, 2 root causes need to be addressed:

- There is no agreement among petroleum engineers on the criteria of SCSS well candidates.
- There is no tool or system to help petroleum engineers predict the oil gain better of SCSS workover with standard criteria.

D. Improve Phase

This phase aims to formulate the solution, pilot the solution, and implement the solution in full-scale implementation. According to the root causes identified in the analysis phase, the next process is developing potential solutions and performing solution development. To get some creative ideas, the project team and SMEs conducted a meeting to identify wide range of potential solutions. After generating solution ideas and evaluating the alternatives, then it comes with some tasks to be performed in the following figure 11.

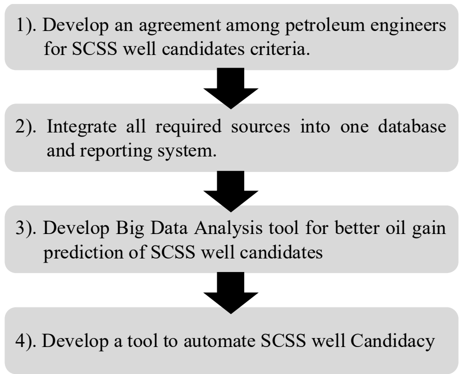

Fig. 11 Prioritized workflow solutions for SCSS well candidacy process improvement

1) Develop an agreement among petroleum engineers for SCSS well candidates' criteria: SCSS well candidates criteria development and agreement meeting were conducted during this phase. All relevant petroleum engineers were invited to this meeting. The meeting resulted in an agreement of SCSS workover well candidate criteria. The criteria are divided into two types of criteria: petroleum engineering criteria and operational criteria. Petroleum engineering criteria are criteria that come from Petroleum Engineers according to their knowledge and senses. Some of the petroleum engineering criteria are as follows:

TABLE III
SCSS WELL CANDIDATE PETROLEUM ENGINEERING CRITERIA AT CI CO., LTD.

No	SCSS Petroleum Engineering Criteria
1.	Well is not horizontal well
2.	No pump issue with the well: last pump fill age data > 95 and pump slip < 10
3.	No proactive workover job completed in last 3 months
4.	Total well test in last 90 days before is more than 2.
5.	Production profile of the wells is declining in last 90 days before
6.	Well Heat Temperature is more than 180 Fahrenheit consecutively in last 6 months
7.	No SCSS workover job was executed in the last 30 days

Operational criteria are criteria that comes from operation constraint or procedures. The following table is the operational criteria:

TABLE IIV
SCSS WELL CANDIDATE OPERATIONAL CRITERIA AT CI CO., LTD.

No	SCSS Operational Criteria
1.	Only producer well with currently active and status is ON.
2.	Only well with T-SPOOL available (T-SPOOL is Facility to inject steam to the production line).
3.	No proactive job backlog which submitted in last 3 months
4.	Exclude area A01 & A02
5.	No Scheduled or in progress SCSS workover job on the well for last 60 days.

Both criteria on Tables III and IV were used by the new tool to generate the well candidates for SCSS workover job.

2) Integrate all required sources into a single database and reporting system: As mentioned previously in analyze phase, petroleum engineers need to open and refresh several reporting systems to get and understand the well performance and flag it as a good candidate for SCSS or not. This solution was integrated to all sources and reports into one integrated database layer and reporting system. Figure 12 shows what is the improve solution in term of database and reporting layer:

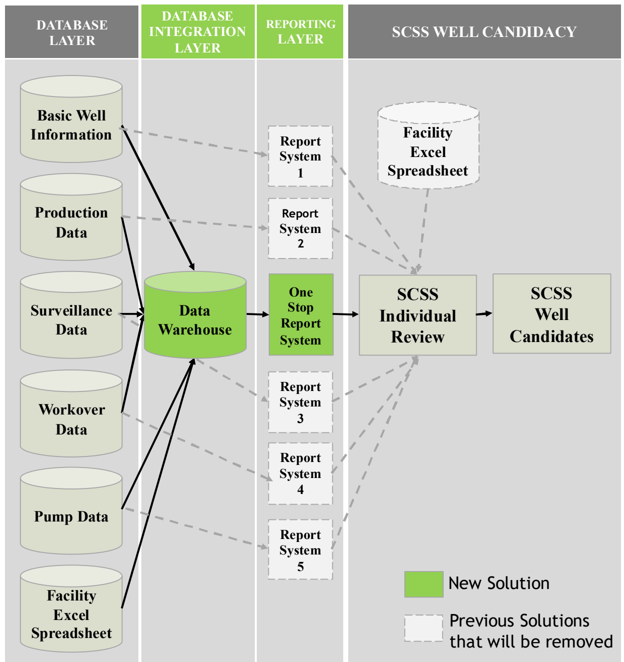

Fig. 12 New database integration layer and one-stop report system for SCSS well candidate review by PE (Individual).

3) Develop Big Data Analysis tool for better oil gain prediction of SCSS well candidates: The data source and reporting system have been simplified from the previous process. The next step is how to improve the quality of well candidates for SCSS that can increase the success ratio of the SCSS workover job. Many companies in the world today invest more in big data and machine learning. That also happened at CI. Co., Ltd. has just established a new team called Artificial Intelligence Team. Therefore, the solution comes to Big Data analysis by applying Machine Learning (ML).

According to the open standard process model called Cross-Industry Standard Process for Data Mining (CRISP-DM), the Big Data Analytics process is divided into 6 phases: 1) Business understanding, 2) Data understanding, 3) Data preparation, 4) Modelling, 5) Evaluation, 6) Deployment [19].

1) Business Understanding: As explained in the previous section, this big data analysis aims to have better SCSS well candidates by predicting the oil gain utilizing machine learning.

2) Data Understanding: To predict the oil gain of the SCSS workover job, it is necessary to understand the historical data of SCSS workover job and all precedent data. According to the SMEs. Some of the data that need to be collected and understood the success of SCSS workover job are as follows:

- *Basic well information:* It is very basic data of the well to inform where the location of the well, the area, the sand of the perforation, the age of the well etc.
- *Workover data:* This is very important data to collect all historical SCSS workover job including the number

of times the well was executed for SCSS, what time it was executed, who proposed the well etc.

- *Production data:* This data can be used to see the production profile before and after the SCSS workover job is executed. This provides information on whether the SCSS job is successful or not.
- *Surveillance data:* This is supporting data to see the well performance parameter such as well heat pressure, well heat temperature, pump fill age, pump slip etc.
- *Pump data:* To see what the pump type is, what is the specification of the pump, the pump setting, etc.
- *Facility data:* Without facility such as T-SPOOL to inject steam, steam line, the SCSS workover job cannot be performed.

3) *Data Preparation:* After all required data have been identified, then the next process is data preparation. Some of the activities in this process are:

- *Data Integration:* This activity can integrate all required data into one dataset used by the model for supervising learning. It consists of features data and targeted label data. An illustration of the data integration for the dataset can be seen in the following figure 13.

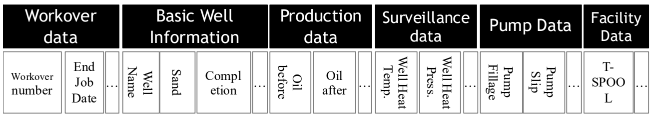

Fig. 13 Illustration of integrated workover history data with other supporting data

- *Feature Selection:* This activity selected which data is as the features and which data as the label or target of prediction. The activity employed the dataset that has been integrated from the previous activity. It starts by looking to the heat map correlation for all numeric variable data that might have correlations each other in the following Figure 14.

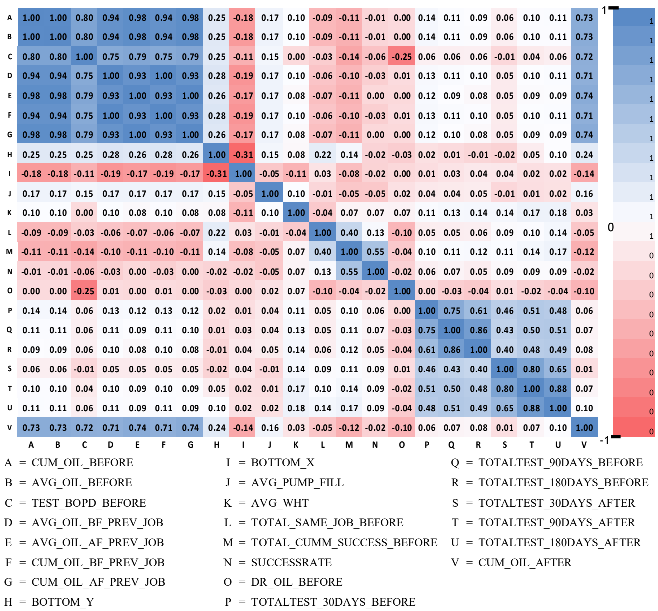

Fig. 14 Feature selection: dataset attributes correlation heat map

The definition of each column attributes as follows:

TABLE V
ATTRIBUTES CODE, NAME AND DEFINITION

Code	Attribute Name	Definition
A	CUM OIL BEFORE	Cumulative 90 days oil before SCSS job is executed
B	AVG OIL BEFORE	Average 90 days oil before SCSS job is executed
C	TEST BOPD BF	Last test oil before SCSS job is executed in Barrel Oil per Day (BODP)
D	AVG THEOR OIL BF PREV JOB	Average 90 days oil before of previous SCSS job is executed.
E	AVG THEOR OIL AF PREV JOB	Average 90 days oil after of previous SCSS job is executed.
F	CUM THEOR OIL BF PREV JOB	Cumulative 90 days oil before of previous SCSS job is executed.
G	CUM THEOR OIL AF PREV JOB	Cumulative 90 days oil after of previous SCSS job is executed.
H	BOTTOM Y	Well subsurface bottom coordinate in axis X
I	BOTTOM X	Well subsurface bottom coordinate in axis Y
J	AVG PUMP FILL	Average 90 days of pump fill age before SCSS job is executed
K	AVG WHT	Average 90 days of well heat temperature before SCSS job is executed
L	TOTAL SAME JOB BF	Total amount of SCSS job that executed on the same well before
M	TOTAL CUMM SUCCESS BF	Total amount of success SCSS job on the same well before
N	SUCCSESSRATE BF	Success rate of SCSS for the well by divided total success SCSS before with total amount of SCSS job before
O	DR THEOR BF	Decline Rate 90 days oil before SCSS job is executed
P	TOTALTEST 30DAYS BEFORE	Amount of well test surveillance in 30 days before
Q	TOTALTEST 90DAYS BEFORE	Amount of well test surveillance in 90 days before
R	TOTALTEST 180DAYS BEFORE	Amount of well test surveillance in 180 days before
S	TOTALTEST 30DAYS AFTER	Amount of well test surveillance in 30 days after
T	TOTALTEST 90DAYS AFTER	Amount of well test surveillance in 90 days after
U	TOTALTEST 180DAYS AFTER	Amount of well test surveillance in 180 days after
V	CUMM THEOR OIL AFTER	Cumulative 90 days oil after SCSS Job is Executed.

Because the target is to predict the oil gain, it will use the cumulative oil after (v) attribute as the label where the oil gain can be calculated later by subtracting the cumulative oil after by cumulative oil before. The correlation heat map in figure 14; 1 means that it has a very strong correlation, and -1 means it has an opposite correlation. Therefore, the heat map tells which features that have a strong correlation to the cumulative oil after (v) as the target or label. The high correlation attributes to the label are cumulative oil before (A), average oil before (B), last well test BOPD before (C), average oil before previous SCSS job (D), average oil after previous SCSS job (E), cumulative oil before previous SCSS job (F), cumulative oil after previous SCSS job (G). Other features that correlate more than 0.1 that still need to be considered are coordinate bottom Y (H), average pump fill (J), total test in 180 days (U). Some of the features have an opposite correlation with a value less than -0.1 as coordinate bottom x (I), total cumulative success before (M), and decline rate oil before (O).

- *Data Filtering and Cleaning:* This activity is an important process. Not all data can be used for modeling. And also need to filter several types of data to make the model narrow to the results. It also usually reduces the number of model errors. Some filters are needed as follow:

TABLE VI
DATA FILTER OF THE DATASET

No	Data Feature	Filter Description
1.	SCSS workover date	Filter only 3 years data to make sure the first model gets the most recent data
2.	Total well test 90 days before	Filter data has more than 2 tests in 90 days before the SCSS job was executed to make sure the decline rate calculation can be performed.
3.	Total well test 90 days after	It also required 2 tests in 90 days after SCSS was executed to make sure the new production profile is more accurate
4.	All column attributes	Drop any null or empty data for all feature attributes to ensure no error when running the model.

The last step in this activity is cleaning up the oil gain resulted in 90 days of SCSS job by removing the outlier data. The tool that used for the outlier detection is box plot as follows:

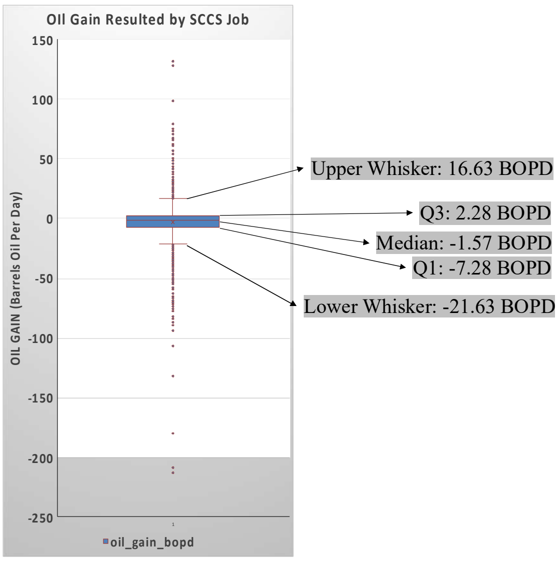

Fig. 15 Average 90 days oil gain box plot from the filtered dataset

Data values that fall far above the upper whisker or below the lower whisker are labeled outliers. Therefore, the dataset used in the model is only SCSS workover history with oil gain between -21.63 BOPD and 16.63 BOPD.

- *Data Transformation:* This activity intends to transform the data by changing the structure, the format, or the values of the data itself. Some data need to be transformed in the dataset of SCSS workover history, especially some categorical data such as area, sand, and well completion type. These data will be used by the model based on SME recommendations. The categorical encoding technique that will be used is One-Hot-Encoder. Following an example of how One-Hot-Encoder transforms the area attribute of the well.

AREA	IS_AREA 1	IS_AREA 2	IS_AREA 3	IS_AREA 4	IS_AREA 5	IS_AREA 6	IS_AREA 7
AREA 1	1	0	0	0	0	0	0
AREA 2	0	1	0	0	0	0	0
AREA 3	0	0	1	0	0	0	0
AREA 4	0	0	0	1	0	0	0
AREA 5	0	0	0	0	1	0	0
AREA 6	0	0	0	0	0	1	0
AREA 7	0	0	0	0	0	0	1

Fig. 16 Transforming the categorical data: area data using One-Hot-Encoder

One-Hot-Encoder will transform the categorical data by creating new column attributes as many as the unique value of the data encoded. In the case above, there are seven areas: One-Hot-Encoder will create new seven columns of area. All the data preparation is almost done. But, before the model uses the dataset for the training process as part of supervised learning, all data need to be normalized to become 0-1 using a min-max scaler to have the same weight.

4) Modeling: All the data preparation is done in the previous activities. The number of total rows of data after all data preparation activities are completed is 4558 records. The next step is developing a model for the SCSS oil gain prediction. The ML algorithm used for the SCSS oil gain prediction model is Artificial Neural Networks (ANNs). ANNs mimic how the human brain works. A neural network in ANNs represents a highly parallelized dynamic system with a directed graph topology that can receive the output information employing a reaction of its state on the input actions. Processor elements and directed channels are called nodes of the neural network [20].

Before the dataset is used for the training process, the dataset is split to train data and test data where the test data is selected 25% from all datasets randomly. The model employed the train data for the training process and the test data will be used to evaluate the model performance. Cross-validation will be conducted by split the dataset randomly several times for the training process. In ANNs training process, hyperparameters are the variables that will determine how the ANNs work through the training process. After doing the due diligence of try and error in the training process, some of best ANNs model's hyperparameters are as follows:

TABLE VII
ANNS HYPERPARAMETER

Hyperparameter	Values
Hidden layer size	78
Activation	Sigmoid
Batch size	200
Learning Rate	0.001
Maximum Epoch	100000
Auto-stop	Yes
n iteration for auto stop	1000

Figure 17 also illustrates the ANNs architecture according to the configured hyperparameter that will be used for the training process.

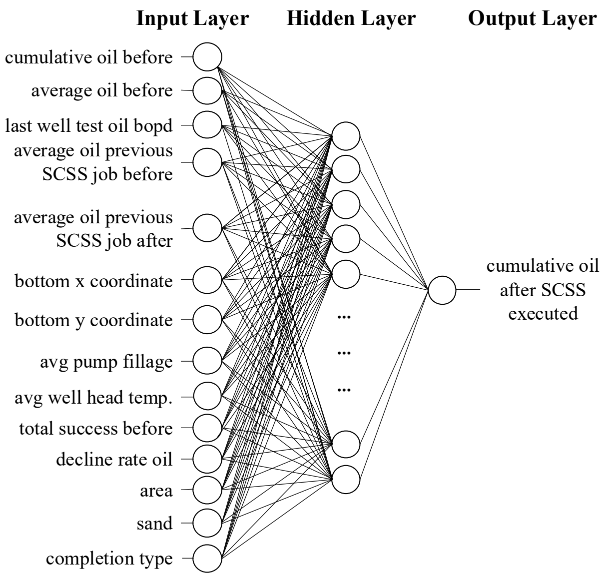

Fig. 17 ANNs Architecture

Because Area, Sand, and Completion type were encoded using One-Hot-Encoder, then the number of input nodes became 50. Since this ANNs architecture also uses only 1 hidden layer with 78 nodes, the final ANNs architecture is 50 – 78 – 1.

5) Evaluation: This activity will evaluate the model performance. Many metrics can be used for model evaluation. Some common metrics such as R Squared and Root Mean Square Error (RMSE). Hence, the following table is the score of that two metrics for the SCSS oil cumulative 90 days after prediction using ANNs algorithm:

TABLE VIII
ANNs Training Metrics

Train Data		Test Data (Blind Data)	
R Squared	**RMSE**	**R Squared**	**RMSE**
82,96 %	553.48	82.92%	544.95

The model can predict the blind data (data that the model has never seen before) with good R squared Score 82.83% and RMSE 546.42. It is also confirmed as a good model by visualizing the actual cumulative oil 90 days after and ANNs prediction of cumulative oil 90 days after in the following figure:

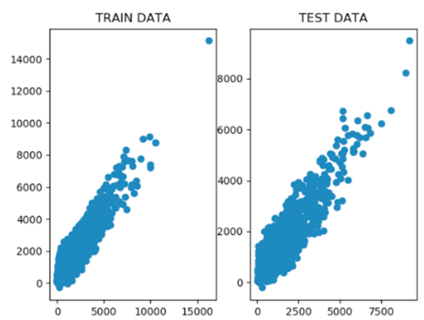

Fig. 18 Scatter Plot of cumulative oil actual and cumulative oil prediction

The scatter plot data almost follow the ideal trend line data for actual cumulative oil (from the lower left corner to the upper right corner). Performance comparison between oil gain prediction from ANNs and manual process by PE cannot be done because there is no data for oil gain prediction in manual process. But performance comparison to another Machine Learning (ML) can be conducted. Some other ML algorithms that have been tried to run and train the model using similar data set and information are: Random Forest Regressor, Extra Tree Regressor, Gradient Boosting Regressor and Adaboost Regressor. The metrics score comparison between the methodologies is as follow:

TABLE IX
Metrics score comparison among ML Algorithm for SCSS oil gain prediction

Machine Learning Algorithm	Train Data		Test Data (Blind Data)	
	R Squared	**RMSE**	**R Squared**	**RMSE**
Artificial Neural Networks	82.96 %	553.48	82.92%	544.95
Random Forest	84.32 %	531.05	82.70%	548.51
ExtraTreesRegressor	99.99 %	0.21	83.06%	542.73
Gradient Boosting Regressor	82.31%	563.96	80.68%	579.53

According to table IX above, if it follows the metric scores, the best ML algorithm will be Extra Trees Regressor. Although the R Squared is 83.06% (highest R Squared) for the blind data, the sense tells it is too overfitting with the train data, which is 99.99%. Therefore, At the end of the evaluation, the project team agreed to go with the second-highest blind data R squared, ANNs Algorithm for the modeling algorithm. Raza *et al.* [21] also compared the performance of ANNs algorithm with logistic regression and support vector machine to assess the health of a strainer located at the suction side of the pump and then finding said as the same as this research project that ANNs proved the better algorithm for a certain dataset.

6) Develop a tool to automate SCSS well Candidacy: After the Big Data Analytics model for SCSS oil gain prediction is completed develop, then the deployment process of auto-generate SCSS well candidates' tool is as follows:

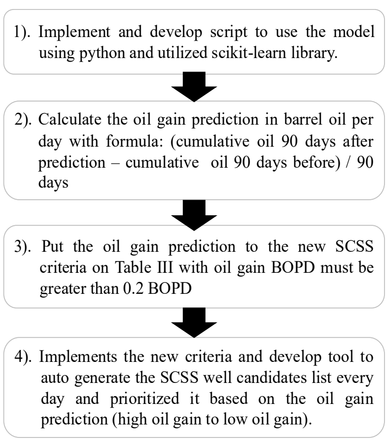

Fig. 19 Auto Generate SCSS Well Candidates Tool Deployment Process

The final product of the tool is a list of SCSS well candidates that are automatically generated by the system and follows all the standard SME's criteria as well as prioritizes them by oil gain prediction from high to low produced by ANNs. This final product is expected to become the solution of Critical Customer Requirement (CCR) and Critical to Quality (CTQ) mentioned in Table I.

III. Results and Discussion

Many process improvements were made successfully using lean six sigma methods in many fields or functions within organizations such as procurement [22], production [23], human resource, information technology, etc. The business issues of SCSS well candidacy process in this research project have also been solved by generating a product through hybrid methods lean six sigma and big data analytics. For the lean process, some of the wastes of well review for SCSS candidates has been eliminated from figure 8. The following figure is the new process:

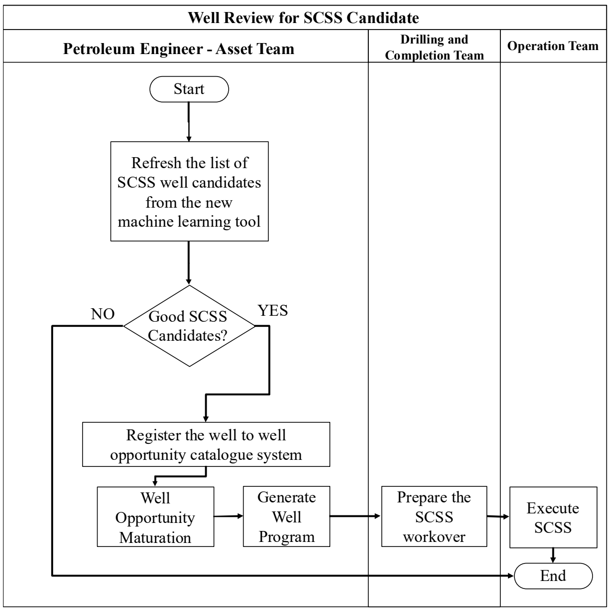

Fig. 20 The new process of well review for SCSS Candidate

There are five waste processes from Figure 8 that have been eliminated by the new machine learning tool in Figure 20. This new machine learning avoids human error and helps petroleum engineers make decisions faster and consistently follow all agreed criteria of SCSS workover job. According to the study done by Alkunsol *et al.* [24] in the manufacturing industry, there is a strong relationship between LSS variables and business performance. It is also proven in this research project that LSS affects the business performance as follow:

1) Reduce time to review SCSS candidate: because of the SCSS candidacy process is done by the new machine learning tool, then the time to get and validate the SCSS well candidates is reduced:

From 2 Hours to 10 minutes for 20 candidates per day

2) Increase SCSS workover Success Ratio: the baseline of the SCSS workover job success ratio, as mentioned earlier, is 61% (Period: July 2017 – December 2017). The tool was launched in January 2018 and conducted a trial for three months. Therefore, as explained in figure 21 below, the SCSS workover success ratio has increased to 73% (Period: March 2018 – April 2020).

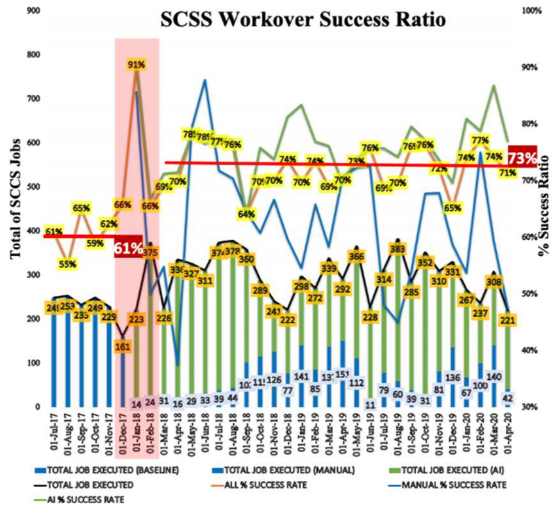

Fig. 21 SCSS Workover Success Ratio Improvement

Although there are still any manual reviews described in the blue bar chart above and blue line, the success ratio of manual process is mostly lower than the aggregate success ratio (the orange line). The number of SCSS workovers selected using the new automated ML tool is still significant as described by the green bar chart above with a higher success ratio (the green line).

In addition, other qualitative opportunities for improvement mentioned in table II were also resolved. The individual review of SCSS candidates has standard criteria on table III and Table IV with additional criteria such as the oil gain prediction greater than 0.2 BOPD. The new parameter to determine the priority of the SCSS workover candidate execution has been agreed as well. The new parameter is the oil gain prediction.

E. Control Phase

The purpose of this phase is to complete project work and create some process control plans. One of the control plans is how to sustain the ANNs model can predict the oil gain after as good as the first performance. Therefore, it is required to do retrain the model by certain time. Project team agreed that the re-train process will be conducted every 6 months. Following table is the result of re-training process within 2 years (January 2018 – January 2020) of implementation:

TABLE X
Retraining Model Evaluation

Dataset Period	Total Records	Train Data		Test Data (Blind Data)	
		R Squared	RMSE	R Squared	RMSE
Jan 2015 -Jan 2018	4558	82.96 %	553.48	82.92%	544.95
Jan 2015 -Jun 2018	6179	84.05%	508.46	82.13%	519.61
Jan 2015 -Jan 2019	7895	82.98%	489.03	83.65%	503.48
Jan 2015 -Jun 2019	9304	82.86%	476.82	82.20%	490.15
Jan 2015 -Jan 2020	11283	80.89%	479.23	82.64%	482.30

From the 5 times of retraining processes, the model still gives R Squared above 70% and RMSE score also continues to decrease. Therefore, it means the model is still good on predicting the SCSS oil gain.

IV. Conclusion

The improvement of SCSS well candidacy process in CI Co., Ltd. has been presented. In this study the process improvement was done using lean six sigma approach through DMAIC Cycles and big data analytics. Process improvement begins with identifying the problems and desire targets in the define phase and then measuring the process in the measure phase. The root cause of the problem was identified in detail in the analyze phase, and the solution was conducted in the improve phase. The machine learning developed in this study could serve as a robot to analyze oil well data and predict oil gain after SCSS job with better results compared to the manual process. This prediction will help petroleum engineers to make the decision faster to do SCSS well workover or not and save their time for other productive activities. Further studies could also be conducted to apply a similar approach to another type of oil well workover. There are also opportunities to use other machine learning algorithms to improve prediction accuracy.

Acknowledgment

The authors are grateful to the Department of Analytic Support Team of CI Ltd., headed by Teguh Handjoyo and also the Department of Upstream Application of CI Lt.d, headed by Triatmojo Rosewanto, for approval study permission. We are obliged to the school of business and management, Bandung Institute of Technology, which gives this research paper the opportunity.

References

[1] A. P. Windarto, L. S. Dewi, and D. Hartama, "Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports with BP Algorithm," International Journal of Recent Trends in Engineering & Research (IJRTER), vol. 3, no. 10, pp. 1–12, 2017.

[2] Ribeiro G M, Mauri G R and Lorena L A N (2011). A simple and robust simulated annealing algorithm for scheduling workover rigs on onshore oil fileds. Comput Ind Eng 60: 519-526, 2011.

[3] Kromodihardjo, S., Kromodihardjo, E.S. Modeling of Well Service and Workover to Optimize Scheduling of Oil Well Maintenance. AMM 836, 311–316, 2016.

[4] PT. CI, *Workover Cost and Oil Gain Performance – Internal data.* Riau, Indonesia, 2020.

[5] Alvarez, J. and Han, S. *Current Overview of Cyclic Steam Injection Process*. J. Pet. Sci. Res. 2 (3): 116–127, 2013.

[6] PT. CI, *Short Cyclic Jobs – internal data.* Riau, Indonesia, 2020.

[7] PT. CI, *Short Cyclic Steam Stimulation Well Selection – internal document.* Riau, Indonesia, 2020.

[8] ΣΕΡΣΕΜΗΣ, Αθανάσιος Κωνσταντίνου. Machine Learning Strategies to Handle Medical Data. A comparison between popular ML models using Python. Aristotle University of Thessaloniki, 2020.

[9] Lantz B. Machine learning with R. 2nd ed. Birmingham: Packt Publishing; 2015.

[10] Jacobs F.R., Chase R.B., *Operation and Supply Chain Management.* New York, USA: McGraw-Hill, 2018.

[11] Singh, M. and Rathi, R., *A structured review of Lean Six Sigma in various industrial sectors*, International Journal of Lean Six Sigma, Vol. 10 No. 2, pp. 622-664, 2019.

[12] Improta, G., Balato, G., Ricciardi, C., Russo, M.A., Santalucia, I., Triassi, M. and Cesarelli, M., *Lean Six Sigma in healthcare: Fast track surgery for patients undergoing prosthetic hip replacement surgery*, The TQM Journal, Vol. 31 No. 4, pp. 526-540, 2019.

[13] Hess, J.D. and Benjamin, B.A., *Applying Lean Six Sigma within the university: opportunities for process improvement and cultural change*, International Journal of Lean Six Sigma, Vol. 6 No. 3, pp. 249-262, 2015.

[14] Zhang, A., Luo, W., Shi, Y., Chia, S.T. and Sim, Z.H.X., *Lean and Six Sigma in logistics: a pilot survey study in Singapore,* International Journal of Operations & Production Management, Vol. 36 No. 11, pp. 1625-1643, 2016.

[15] PT. CI, *Lean Six Sigma Training – internal document.* Riau, Indonesia, 2020.

[16] Michael L.G., Rowlands D., Price M., Maxey J., *The Lean Six Sigma Pocket Toolbook.* New York, USA: McGraw-Hill, 2005.

[17] Garrison R.H., Nooren E.W., Brewer P.C.. *Managerial Accounting*. New York, USA: McGraw-Hill, 2010.

[18] Braglia, M., Marrazzini, L., Padellini, L. and Rinaldi, R., *Managerial and Industry 4.0 solutions for fashion supply chains*, Journal of Fashion Marketing and Management, Vol. ahead-of-print No. ahead-of-print, 2020.

[19] Shearer C., *The CRISP-DM model: the new blueprint for data mining*, J Data Warehousing (2000); 5:13—22.

[20] Galushkin A.I., *Neural Networks Theory.*Berlin, Germany: Spinger-Verlag, 2007.

[21] Raza, J., Liyanage, J.P., Al Atat, H. and Lee, J., *A comparative study of maintenance data classification based on neural networks, logistic regression and support vector machines*, Journal of Quality in Maintenance Engineering, Vol. 16 No. 3, pp. 303-318, 2010.

[22] Saputro F.P., *The Improvement of Procurement Process in Oil Refinery Maintenance by Using Lean Six Sigma*. Jakarta, Indonesia. 2017.

[23] A.R. Iskandar, *Adoption of Six Sigma – DMAIC Methodology to Optimize Service Level at PT. Ultra Prima Corrugator.* Bandung, Indonesia. 2019.

[24] Alkunsol, W.H., Sharabati, A.-A.A., AlSalhi, N.A. and El-Tamimi, H.S. *Lean Six Sigma effect on Jordanian pharmaceutical industry's performance*, International Journal of Lean Six Sigma, Vol. 10 No. 1, pp. 23-43, 2019.