Machine Learning in Oil & Gas Industry: A Novel Application of Clustering for Oilfield Advanced Process Control

Conference Paper · January 2019

DOI: 10.2118/194827-MS

2 authors, including:

Rohit S Patwardhan
Saudi Arabian Oil Company
44 PUBLICATIONS **814** CITATIONS

SEE PROFILE

Society of Petroleum Engineers

SPE-194827-MS

Machine Learning in Oil & Gas Industry: A Novel Application of Clustering for Oilfield Advanced Process Control

Kalpesh Patel and Rohit Patwardhan, Saudi Aramco

This paper was prepared for presentation at the SPE Middle East Oil and Gas Show and Conference held in Manama, Bahrain, 18-21 March 2019.

Abstract

Data Analytics is an emerging area that involves using advanced statistical and machine learning algorithms to discover information & relationsips present in different types of data. The work described in this paper illustrates the application of machine learning techniques to an Oilfield Advanced Process Control (APC) project involving deployment of APC at a large onshore conventional oilfield in Saudi Aramco. APC implementation enables better control and optimization of the production from hundreds of oilwells.

APC rollout at the large oilfield involved APC deployment on 300+ oil wells. Using conventional APC implementation methodology, the rollout would be very difficult to manage and would have taken about 3 man years which was not practical. Use of innovative data analytics techniques was essential to ensuring the timely deployment of such a large scale APC project. A machine learning algorithm used to cluster similarly behaving wells, enabled significant (80%) reduction in the engineering effort and operator involvement in developing the models for each well. This allowed the implementation to be completed one year in advance thus realizing the APC benefits earlier than planned.

Introduction

An oilfield consists of hunderds of oil wells. APC implementation in an oil field involves deploying one APC application per oil well. The APC applications automatically manipulate the available handles in an oil well namely the Choke Valve, Electrical Submersible Pump (ESP) speed and volt to speed ratio in order to achieve target production while operating the ESP efficiently within its operating envelope. At the heart of an APC application is an empirical model of the oil well that contains information on how the oil well variables (flow, pressure, temperature, amps, volts etc) change when the above mentioned handles are changed. Developing these empirical models for each of the hundreds of oil well separately takes a significant amount of time resulting in benefit being realized that much later on in future. Reducing the amount of time spent in developing the models is a challenge.

Instead of treating each well separately it would save a lot of time if the wells could be treated in groups. Clustering, which is an unsupervised machine learning algorithm, is well suited for grouping objects with similar characteristics. It lends itself very well to grouping oil wells with similar behavior together and addressing the challenge.

We will first present an overview of the oilfield APC. This is followed by a section on the challenge faced during ACP implementation. We then introduce and propose the use of unsupervised clustering algorithm to address the challenge. This is followed by a discussion on how clustering was used and the results achieved. We then provide a conclusion and future direction.

Oilfield APC

The details on APC deployment on the large oilfield has been published by Patel et.al 2018. This section contains material taken from Patel et.al 2018 publication to introduce Oilfield APC.

An oilfield consists of hunderds of oilwells. Fig. 1 shows a schematic diagram of a typical oil well using ESP as artificial lift, in a conventional oil field. The green circles represent variables that are measured or calculated, while the blue circles represent the variables that can be changed, to achieve flow compliance to target production while operating ESP efficiently within its operating envelope.

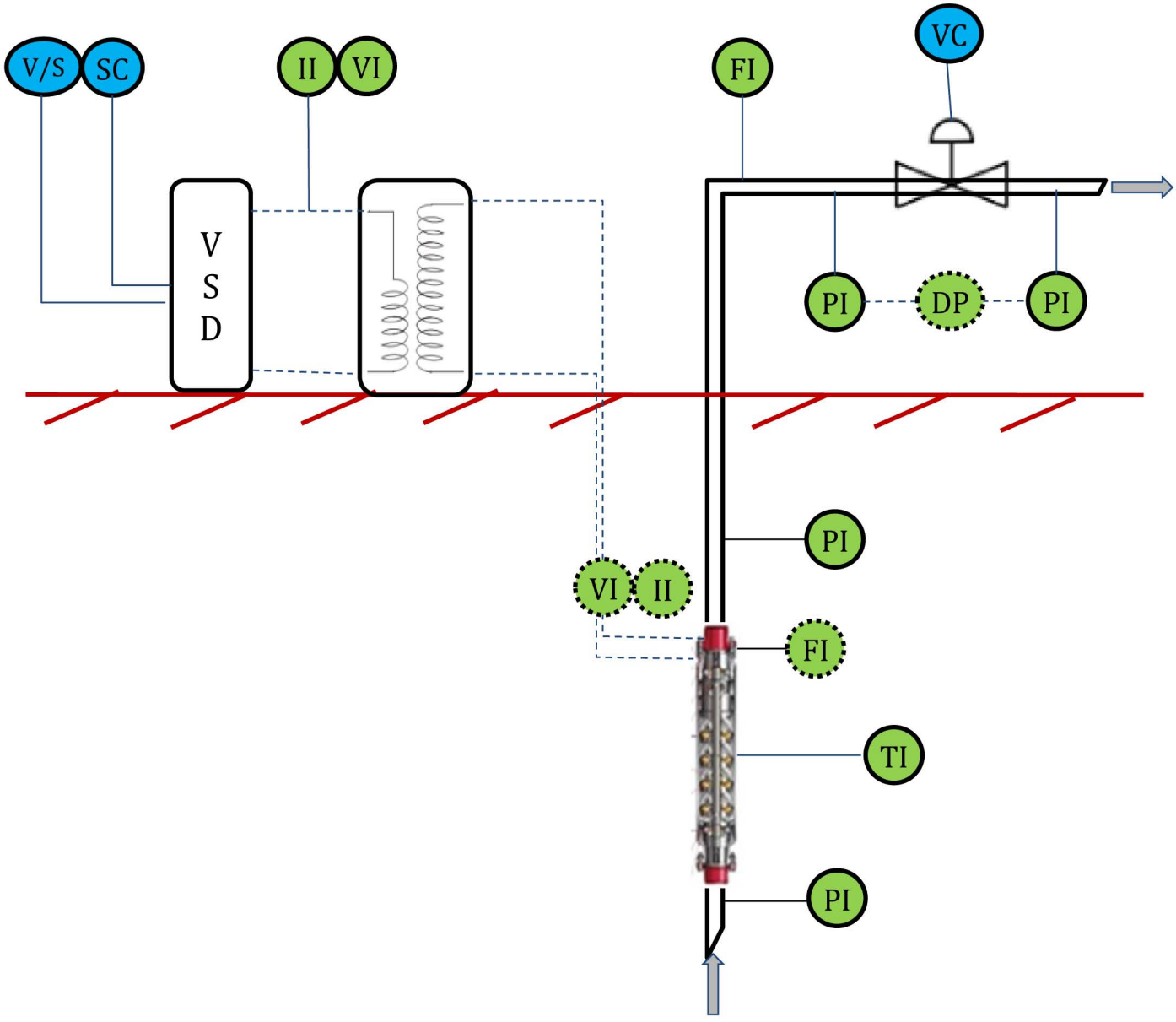


Figure 1—Schematic diagram of a typical oil well with ESP

The oil well consists of an ESP installed thousands of feet underground along with pump intake & discharge pressure and motor winding temperature measurements. The ESP is powered by a Variable Speed Drive (VSD), through a step-up transformer installed at the surface, which allows the ESP speed to be changed from the surface and remotely from the control room. In addition to speed, the VSD may allow the Volts to speed ratio to be changed. The 3-phase voltage and current outputs of the VSD are measured, and used to calculate the ESP motor voltage and current. The ESP pumps the fluid, which rises to the surface and passes through a choke valve whose opening can be set remotely from the control room. The choke upstream and downstream pressure measurements are used to calculate the differential pressure across the choke valve. The production flow is usually measured at the surface by a multiphase flow meter.

The operational objective for individual oil wells is to achieve production flow compliance to the assigned well target, and operate ESP as efficiently as possible and within its operating envelope. This was achieved by implementing one APC application per oil well.

APC is a multi-input multi-output (MIMO) technology that optimizes operations by monitoring multiple input variables, predicting the future behavior of the process variables, and manipulating multiple output variables simultaneously to achieve operating objectives consistently.

The monitored input variables are called Controlled Variables (CV) shown in green circles in Figure 1, while the manipulated output variables are called Manipulated Variables (MV) shown in blue circles in Figure 1. Both MVs and CVs have limits within which APC operates the process. The MV limits are hard limits, which are always honored, whereas CV limits are soft limits, which APC attempts to operate within but may not be able to all the time. An internal empirical model consisting of relationship between MVs and CVs is used to predict future behavior, as well as to optimize and calculate the changes in MVs.

When applied to a process, APC results in stabilization of the process, thereby reducing the variability in the CVs. Once the process is stable, the operating targets can be moved closer to the equipment limits or product specifications (as shown in Fig 2), thereby increasing profitability by either maximizing revenue, minimizing energy consumption, or both.

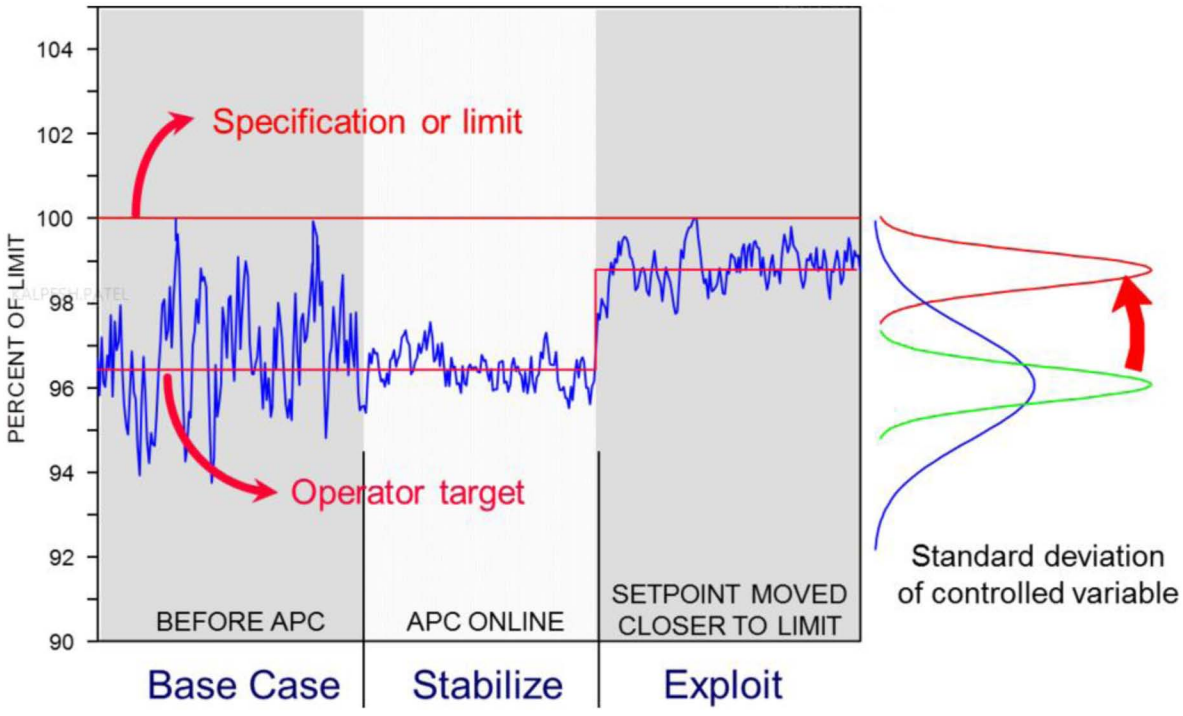


Figure 2—Benefit of APC im-plementation

The design of the APC application when implementing APC on an oilwell is summarized in Table 1.

Table 1—List of MVs and CVs for an oil well

Process variable	Variable Type	APC philosophy
Choke Valve opening	MV	Maximize
ESP Speed	MV	Minimize
ESP Volt@60Hz (representing the volts to speed ratio)	MV	Minimize
Oil production flow	CV	Hold at target
Choke Valve DP	CV	Minimize to a low limit
Wellhead pressure	CV	Keep below high limit
Downhole flow	CV	Keep within upthrust and downthrust limits
VSD output volts	CV	Keep below Max VSD output
Motor volts x 3	CV	Keep within motor volt rating
Motor amps x 3	CV	Keep within motor amp rating
Motor winding temperature	CV	Keep below motor rating
Pump intake pressure	CV	Keep above bubble point pressure
Pump discharge pressure	CV	Keep below high limit

A tabular matrix representation of the empirical model used in APC application for a fictious oil well is shown in Table 2. It is also called the steady state gain matrix. The elements of the steady state gain matrix represent the final effect of a unit change in MV on all CVs. The first column of values in the table, starting with 100 and ending with -4, indicate the magnitude of change in the respective CV happening simultaneously due to a 1% increase in Choke Valve MV. Similarly the second column of values in the tbale, starting with 300 and ending with 6, indicate the magnitude of change in the respective CV happening simultaneously due to a 1Hz increase in ESP speed MV. The actual model used in APC includes information on how the CVs change dynamically which is not shown in Table 2.

Table 2—APC model steady state gains

CVs ↓ / MVs →	Choke Valve (%)	ESP Speed (Hz)	Volt@60Hz (volts)
Oil production (bpd)	100	300	
Choke Valve DP (psig)	-5	5	
Wellhead pressure (psig)	-5	5	
Downhole flow (bpd)	125	400	
VSD output volts (volts)		6	1
Motor volts x 3 (volts)		7	6
Motor amps x 3 (amps)	1	8	0.1
Motor winding temperature (DegC)	0.5	3	
Pump intake pressure (psig)	-1	-5	
Pump discharge pressure (psig)	-4	6	

The challenge

The conventional method of implementing an APC application on a well is as follows

1. Perform a response test on the well. This includes setting up data collection and step testing each MV to characterize the dynamic as well as steady state gains between the MVs and the CVs.
2. Identifying a model for the well. This includes analysing the data collected during response test and developing a time series model for the well that captures the dynamic and steady state gains of the process.
3. Configuring the APC application by incorporating the identified model.
4. Commissioning the APC application online.

It takes about 2 days to perform all the steps mentioned above for a well. 1 man day for engineering work related to testing and modeling the wells and another man day for configuraiton and commissioning. Figure 3 shows the time taken by each step.

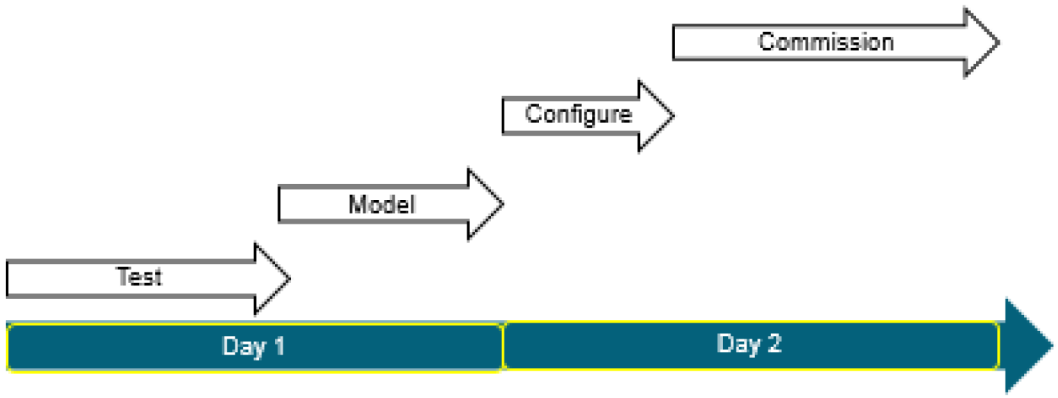


Figure 3—Typical APC implementation time for a well

As an oilfield has hundreds of production wells, the amount of effort in implementing APC controllers for all production wells is significant. Shown below is the calculation of time it would take to implement APC in an oilfield.

Number of production wells	350
Number of days to implement APC	350 x 2 = 700 man-days
Number of days available in 1 year	365 - 52*2 - 10 = 251 man-days (Working days excluding weekends and holidays)
Number of man years to implement APC	700/251 = 2.8 man-years

Implementing a 2.8 man-year effort could easily take up to 5-6 calander years when personnel and operational availability is factored in. Meanwhile in the 5-6 years a lot can change like more wells being added, ESPs being replaced etc which would require additional effort. Practically, the effort is unmanageable using conventional APC implementation methodology.

The challenge was to find an innovative approach to reduce the implementation effort and ensure that the project implementation can be completed in a timely fashion.

Introduction to Clustering

Data Analytics offer a range of machine learning algorithms. These machine learning algorithms can be classified into supervised and unsupervised learning. Supervised learning involes a target or labeled variable which is used to train one or more models. The model developed using the training data set can be used

to infer the target variable on a new data set to generate predictions. Supervised learning can be applied to the following problem types:

- Classification: When the data is being used to predict a categorical variable. This is the case when assigning a label or indicator, for example the state of a compressor – running or shutdown. When there are only two labels, it is called a binary classification problem. When there are more than two categories, it is called mult-class or multinomial classification.
- Regression: When the target variable is a continuous variable, the problem becomes a regression problem. An example is the generation of shift vectors for running planning LPs using simulation data.
- Forecasting: This involves making predictions about the future based on current and past data. An example is future prediction of crude end points on a crude column using historical data and process models.

Unsupervised learning involves completely unlabeled data. In this case the machine learning algorithm is asked to discover patterns present in the underlying data, such as clusters of similar data points, a lower dimensional underlying structure.

- Clustering: Grouping sets of similarly behaving variables according to some criteria. This is often used to segment the whole dataset into several groups. Further analysis can be performed within individual groups to find intrinsic patterns.
- Dimensionality Reducions: Reducing the number of variable under consideration. For example, in many cases due to abundant sensor measurements, the raw data may have high dimensionality features. Reducing the dimensionality helps find the true underlying relationships which are governed by physical laws – mass and energy balances.

Application of Clustering

One way to reduce the APC implementation time is to consider production wells in a group rather than separately such that an identified model for one well can be used for other wells in a group without testing them individually. This would lead to significant reduction in the engineering effort and operator invovement in testing and modeling of wells. The unsupervised machine learning technique of clustering, defined in the previous section, is very well suited to these situations.

The following steps were followed, with some iterations, to cluster production wells with the aim of being able to reuse an identified model for one well to other wells in a cluster. An advanced analytics software (IBM, 2014) was used for clustering the wells.

- Database creation and preparation
- Applying clustering
- Sub-grouping the clusters

An excel database was prepared with static and dynamic data for 237 production wells. Static data consists of information regarding the well that normally doesn't change during the life of the well. It includes the following 15 charateristics for each well, as shown in Table 3, making the database size 237 by 15.

Table 3—Static well data

Static well Charateristic	Description
INST_CO	Installation company
INSTALL_DATE	ESP install date
START_DATE	ESP start date
DEPTH	Depth at which ESP is installed
MODEL	ESP pump model
STAGES	Number of pump stages in ESP
M_TYPE	ESP motor type
M_HP	ESP horse power rating
M_AMP	ESP motor current rating
M_VOLT	ESP motor coltage rating
GROUP	Group number in oil field
S_LONG	Well longitude at surface
S_LAT	Well latitude at surface
D_LONG	Well longitude downhole
D_LAT	Well latitude downhole

Dynamic data consists of process variable measurements that change on a day to day basis. It includes the daily average values of the following 11 measurements for each well, as shown in Table 4, for one year making the database size 237 × 11 × 365 or 237 × 4015.

Table 4—Dynamic well data

Dynamic well measurements	Description
SI-xyz	ESP speed
MI-xyz	Choke valve position
FI-xyzA	Oil production
FI-xyzB	Water production
PI-xyzA	Choke valve upstream pressure
PI-xyzB	Choke valve downstream pressure
EI-xyz	VFD volts
II-xyz	ESP Motor current
PI-xyzC	Pump intake pressure
PI-xyzD	Pump discharge pressure
TI-xyz	Motor winding temperature

Combining the static and dynamic databases together for 237 wells, the database size becomes 237 × 365 × (15+11) or 86505 by 26 as seen in the Figure 4.

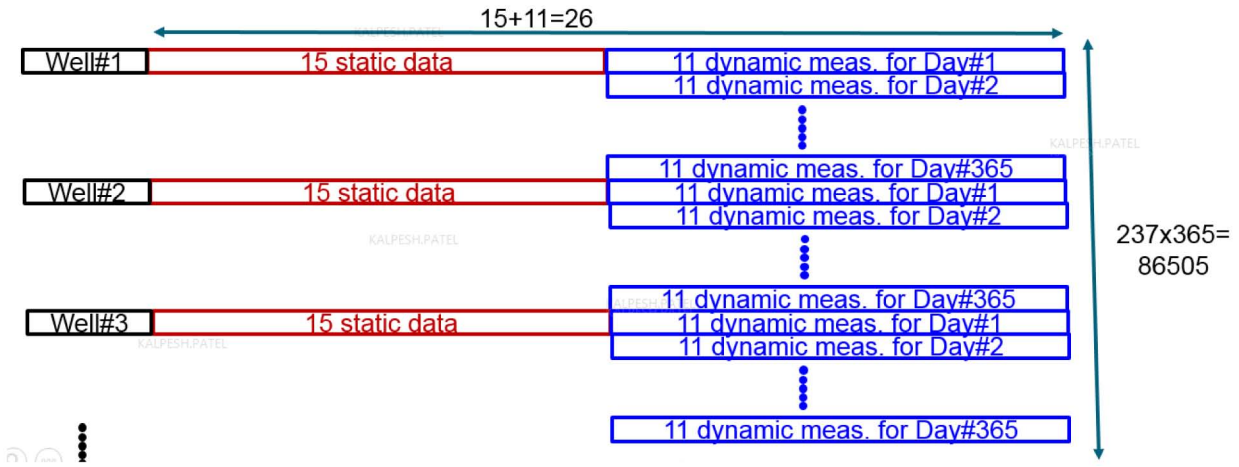


Figure 4—Overall database size

Applying unsupervised clustering directly to the whole database and analyzing the clusters wasn't relevant with respect to the purpose of clustering which was to group wells based on similarity in their models. It was apparent that the dataset size had to be reduced while making it more meaningful for the purpose.

The options considered for reducing the database size are listed in the Table 5.

Table 5—Options for database size reduction

Options	Remarks
Average	This was rejected as averaging would result in loosing the model related information content in the data.
Corelation coeficients between MVs and CVS	This seemed to be a good option as it will provide information on relationship between MVs and CVs which is similar to the model gains. But as the correlation coefficients are scaled between -1 to 1, with 1 or -1 representing good correlation and 0 representing no correlation, the difference between a high flow gain well and a low flow gain well was lost as the correlation coefficients for flow for both the wells got scaled to 1. So this too was rejected.
Model gain calculation	This was used. More details provided below.

The 4 model gains, shown in Table 6, were calculated using the daily averages for the measurements when all measurements were available. These are the 4 gains that the APC application is expected to use the most during normal operation.

Table 6—Options for database size reduction

	Downhole Flow CV	Choke DP CV
Speed MV	✓	✓
Choke MV	✓	✓

After a few iterations of clustering using these calculated gains along with other static and dynamic variables, we found that clustering using only the 4 calculated gains resulted in identification of 5 cluster with good cluster quality as seen in Figure 5. The cluster quality shown is a measure of intra-cluster cohesion and inter-cluster separation.

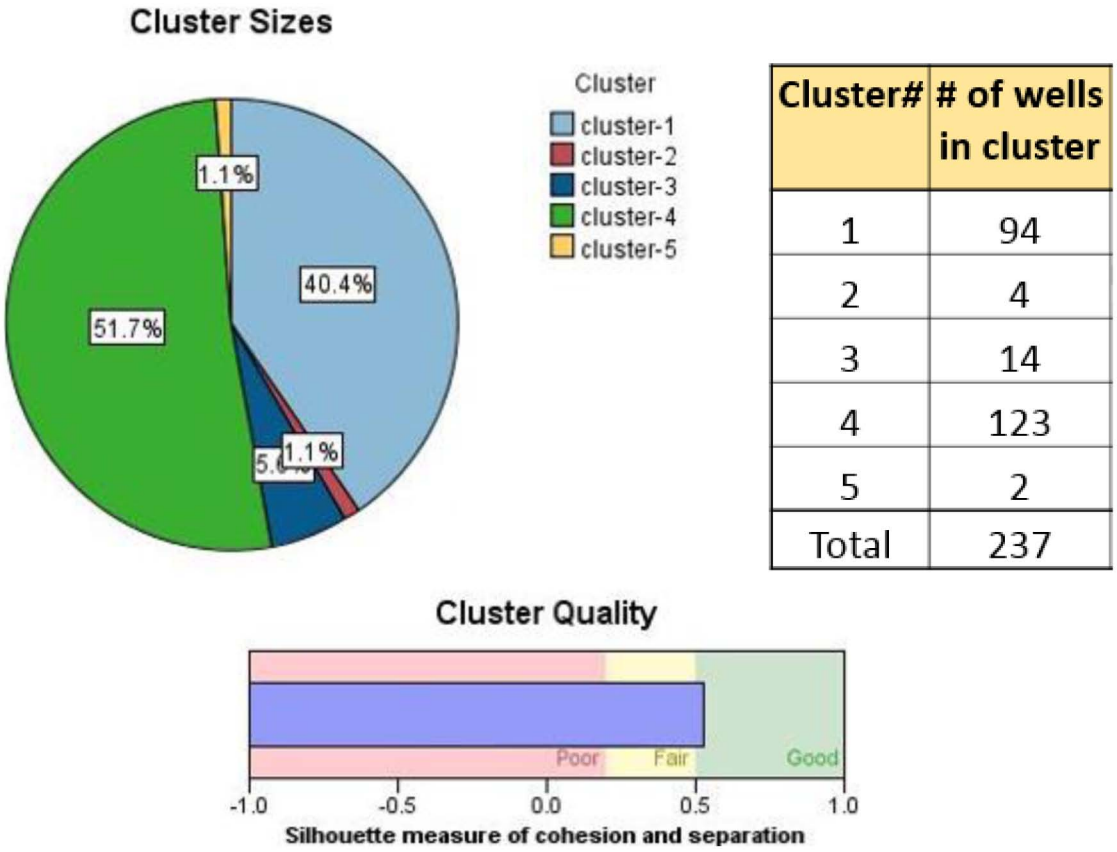


Figure 5—Result of clustering

As many of the wells were clustered in two big clusters, it was decided to have sub-groups in the bigger clusters, such that the wells in each sub-group within a cluster are more similar to each other than to other sub-groups in the same cluster.

This was done by running a supervised multi-class classification problem with C5.1 algorithm which is a decision tree algorithm. It constructs a decision tree by recursively splitting data into 2 or more subgroups defined by predictor fields as they relate to target or categorical variable. The 5 clusters identified earlier were set as the targets or categorical variables and the static data of the wells were set as the input variables. The clusters were categorized with 88% accuracy as seen in Figure 6.

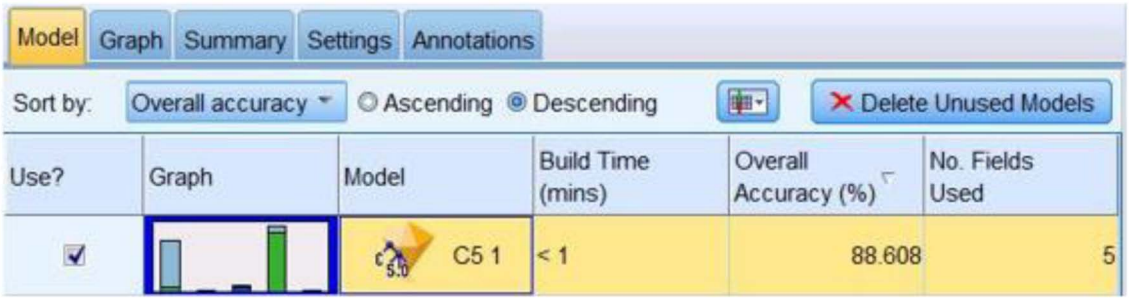
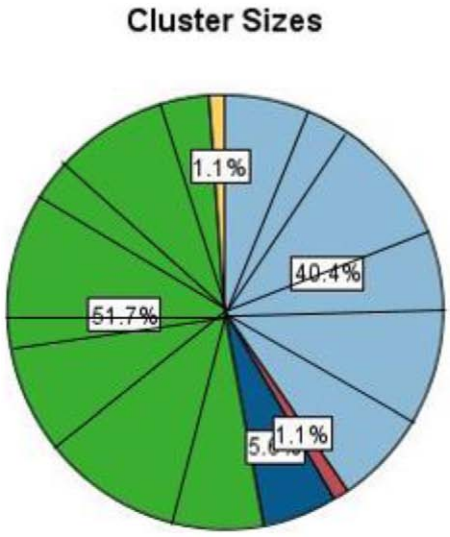


Figure 6—Accuracy of cluster categorization

The major contributors to the categorization, or predictor fields in the resultant decision tree, were found to be pump type and choke valve size which was consistent with the domain knowledge. Based on the pump type and choke valve size, the 5 clusters were manually divided into 42 sub-groups. The result of sub-grouping is shown in Figure 7.



Cluster#	# of wells in cluster	Well to be Tested
1	94	16
2	4	4
3	14	7
4	123	13
5	2	2
Total	237	42

Figure 7—Result of sub-grouping

One well from each sub-group was choosen, based on its availablity, for testing and model identified for use in APC application. The identified model was then reused to configure APC application for other wells within the same sub-group, thus saving significant engineering effort in addition to speeding up the deployment time. Overall, only 42 out of 237 wells had to be tested and identified which represents an 80% reduction in engineering effort involved.

The benefit of reduced engineering effort is calculated as follows

Original implementation time	2.8 man years
Original engineering effort	350 mandays or 1.4 man years
80% reduction in engineering effort	0.8x1.4 man years = 1.12 man years (~1 man year)

Thus the APC implementation and its benefit, which amounts to several million $/year, was achieved 1 year earlier.

Conclusion

Conventional APC imlementation methodology is impractical for an oilfield with hundreds of wells. The engineering effort and operator involvement in building models for each well is significant. The use of machine learning techniques to cluster wells with similar behavior allows using a model identified for one well for all the wells in the same group. This allows the APC implementation to be done significantly faster and keeps the effort manageable.

When dealing with multiple wells/equipment that are structurally similar, the use of machine learning techniques to cluster similar entities and reduce the engineering effort is recommended. This innovative application of advanced analytics illustrates the usefulness of these machine learning techniques. It also serves to reinforce the role of data preparation and domain knowledge when using data analytics. Machine learning should not be considered as a substitute for domain knowledge. The best machine learning and AI applications in the process industry will be those where domain (human) expertise is effectively combined with machine learning/AI tools. Moreover as illustrated by the application here, the business case should drive the use of ML and AI technology and not the other way round.

Acknowledgement

The authors would like to thank Saudi Aramco management for supporting the work conducted in this paper and its publication.

References

1. Patel, K., Bakurji, A., Salloum, H., Kim, H., Winarno, M., and Mubarak, S. 2018. Use of Advanced Process Control for Automating Conventional Oilfield Operations. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, April 2018, Dammam, Saudi Arabia. https://doi.org/10.2118/192393-MS
2. IBM SPSS Modeller user's guide, 2014, IBM Company.