## Introduction

The extracting of crude oil to the surface is a very complex, energy-intensive process. More than 90% of oil wells around the world works with the deficiency of oil reservoir pressure, so the effective operation of downhole equipment is a critically important task. This paper will describe the main ideas of the recommendation system that was designed to optimize the operation of electrical submersible pumps (ESPs) in order to increase the oil flow rate from the ESPs to the surface.

The developed AI-based recommendation system (RS) was implemented to one of the onshore Siberian oil deposits in 2018. The deposit is in the plateau period state (average dilution percentage is equal to 42%). The deposit has 700 wells (480 producing wells and 220 injection wells). The 360 wells of producing wells work in a constant mode (with constant ESP engine frequency) and the other 120 wells work in variable mode (with ESP engine frequency always equal to 50Hz, where working periods are alternating with non-working periods). The average flow rate of producing wells is 56 tons per day. There are 6 oil reservoirs developed in this deposit with the depth between 2600-3200 meters.

## Data processing

To solve the task of optimization of the ESPs operation, the available parameters (Table 1) were used. The period of observation and information storage of the current deposit is more than 10 years. However, during the research it was found that the including of measurements that are earlier than the last 1,5 year, doesn’t significantly improve the recommendation system.  For that reason, these measurements weren’t employed in the RS learning process.

*Table 1. Parameters and measurement frequency.*

|  |  |
| --- | --- |
| **parameter** | **measurement frequency** |
| engine frequency (Hz) | once in a minute |
| engine temperature () |
| engine active power (kW) |
| liquid pressure in the pipe (kg/m2) |
| buffer pressure (kg/m2) |
| сurrent (А) |
| voltage (V) |
| collector pressure (kg/m2) |
| engine pressure (kg/m2) |
| rated engine pressure (meters of the water column) (m) |
| ESP on/ESP off | once in 5 minutes |
| dynamic level (m) | once in a hour |
| crude oil density (kg/m3) | once in a day |
| dilution percentage (%) |
| crude oil production (m3/day) |
| rated engine power (kW) | once in a half year or rare |
| reservoir pressure (kg/m2) |
| rated engine productivity  (m3/day) |
| well depth (m) |
| well prolongation (m) |

As shown, TableHoweverТаблица 1 includes measured parameters as well as calculated parameters that were achieved by mathematical modelling of downhole equipment operation. Specifically, «liquid pressure in the pipe» was calculated using following model:

|  |  |
| --- | --- |
|  | (1) |

where, – the dilution percentage, – the crude oil density, – the well depth.

It is important to say that the cooperation of the data scientists and the domain experts played a great role in this project. The domain experts developed simple and high accuracy models of physic processes related to the wells operation. These models allowed to increase significantly the RS accuracy.

Стоит отметить что большую роль при реализации проекта сыграло взаимодействие специалистов по машинному обучению и доменных экспертов, которые смоги разработать точные и, в то же время, простые физические модели, которые смоги существенно поднять точность построения моделей машинного обучения.

During the course of the project, the data mismatch problem was also solved. In some periods of time there was the non-zero oil flow rate for some wells when the pump was switched off, that, of course, is a mistake in recording the pump state. The ML-model of pump state prediction was implemented to such periods of time. This model was trained on the selected data when there was no mismatching between the parameters. Thus, function was developed for these mismatching time periods:

Входе реализации данного проекта так же была решена проблема рассогласование данных. Для некоторых скважин в некоторые периоды времени наблюдались ненулевые дебиты нефти при выключенном насосе, что, конечно, является ошибкой записи состояния насоса. Для таких периодов времени была реализована ML-модель предсказания состояния насоса. Данная модель была натренирована на выделенных определенных времени, когда рассогласований между параметрами не наблюдалось. Т.е. была построена функция :

|  |  |
| --- | --- |
|  | (2) |

where, is the predicted value for the chosen parameter, is the parameters that are used in the prediction model .

где – значение прогнозируемого параметра, – параметры на основании которых строиться прогноз.

Further, handling the historical data, the predicted value can be compared with the historical value , and the predicted value can be employed instead of in periods of large mismatches between and :

Далее, имея исторические данные мы можем сравнить полученное прогнозное значение с фактическим и в периодах большого рассогласования использовать прогнозное значение вместо , т.е.:

|  |  |
| --- | --- |
|  | (3) |

where, is the standard deviation of values in an observation period, s is the acceptable mismatching constant based on the domain expert's physical model.

где - среднеквадратичное отклонение величины в обучающей выборке, –величина допуска рассогласования прогнозных и фактических значений, подбирается на основе физической модели доменного эксперта.

The proposed approach is not always applicable, and it can be difficult to build a predictive model for some parameters. In such kind of cases with abnormal values, these periods were excluded from consideration. Moreover, the periods of time when the key parameters were not observed at all were also excluded.

It is obviously, that the main part of parameters do not effect on the predicted parameter immediately, so to get better prediction quality lag parameter was generated. Lag parameter is the same parameter as the initial, but the value of lag-parameter equal to value initial parameter in previous time moment (shifted by the lag): . Furthermore, smooth-lag parameters was used. The mean-lag parameters is the lag-parameter smoothed by the selected time base: .

Приведенный выше подход оказывается не всегда применим, для некоторых параметров построение прогнозной модели может быть затруднительно, в этом случае периоды времени на которых наблюдались аномальные значения некоторых параметров были исключены из рассмотрения. Так же были исключены периоды времени, когда ключевые параметры не наблюдались вовсе.

Furthermore, the big problem of the RS implementation was insufficient frequency of the crude oil flow rate measurement. The flow rate meter is not related to specifically one well, but is related to several wells. The oil flow rate measurement periods are not associated with the pump condition. The meter measures oil flow rate for 1-2 hours, then it is switched to another well and so on. Let’s imagine that the well works constantly, but would be stopped for a period of meter measurement, so the zero daily flow rate for this well would be obtained, that, of course, is a mistake. To eliminate this mistake, the calculation of a "true" (calculated) flow rate was implemented. The algorithm for calculating the "true" flow rate is quite complex and deserves a separate report, so here it is a simplified model for the case when only one measurement in a day was made and ESPs engine frequency was constant:

Так же большой проблемой для реализации рекомендательной системы было недостаточная частота измерения дебита нефти. Счетчик дебита нефти не относиться к скважине, а относиться к ГЗУ, при этом периоды замеры дебита нефти никак не ассоциированы с состоянием насоса. Замеры производится в течении 1-2 часов, далее счетчик переключается на другую скважину. Если предположить, что скважина работает постоянно, но на период замера была остановлена то мы получим в результате замера нулевой суточный дебит для данной скважины, что, конечно, не соответствует действительности. Для устранения этого недостатка был реализован расчет «истинного» дебита. Алгоритм расчета «истинного» дебита довольно сложен заслуживает отдельного доклада, здесь приводиться упрощенная модель расчета для случая, когда был произведен только один замер за день:

|  |  |
| --- | --- |
|  | (4) |

where, – the «true» daily oil flow rate, - the measured daily flow rate, - the time period of flow rate measurement, - the percentage of time when a pump is switched on to the total measurement time, - the percentage of time when a pump s switched on to the daily time.

где – истинный дневной дебит, – измеренный на ГЗУ почасовой дебит нефти, – период измерения дебита на ГЗУ. – процентное отношение времени, когда насос был включен к общему времени замера, – процентное отношение времени когда насос был включен к суточному времени.

## Problem

The main purpose of this work is to increase the volume of lifted crude oil, avoiding the risk of well drying and preventing an accident at a well.

To solve this problem it is necessary to determine the acceptable limits of safe and efficient wells functioning. The statistical researches that were conducted and the domain experts knowledge allow us to draw up a table of restrictions (table 2). Following the specified limits guarantees safety and constant oil production process.

Основной целью данной работы было увеличение объема поднимаемой нефти и избежание при этом риска высыхания скважины, остановки работы скважины вследствие возникновения аварии на скважине.

Для решения этой задачи нам необходимо определить допустимые границы безопасного и эффективного функционирования скважины. Проведенные статистические исследования и экспертные знания технологов на производстве позволили нам составить таблицу ограничений (таблица Таблица 2), нахождение в рамках которых позволяет нам обеспечить безопасность и постоянство добычи нефти.

Таблица 2. Key parameters and their limits

|  |  |
| --- | --- |
| **параметр** | **ограничение** |
| «input pressure» | > 28 (кг/см2) |
| «engine temperature» | < 110 () |
| «isolation resistance» | > 8 000 Ом. |
| «buffer pressure» | < 40 (kg/m2) |
| «сurrent» | < 200А |

The ESP engine frequency is the only one control parameter in this maximization task, that can variate between 40-60 Hz.

Hence, the task can be reformulated as follows: for each ESP, it is necessary to determine the maximum of engine frequency at which all the parameters from table 2 will stay in within their limits.

The given formulation assumes the construction of predictive model for each parameter from table 2. During this project, several attempts were made to build such predictive models. Unfortunately, all of them were unsuccessful. General problem of predictive models constructing is unknown reservoir state at a perforation area. The reservoir pressure is usually measured twice in a year, furthermore, the reservoir pressure cannot fully describe crude oil flowing from reservoir to the well hole.

Единственным управляющим параметром в данной задаче максимизации является частота работы двигателя ЭЦН, которая может варьироваться в пределах 40-60 Гц.

Таким образом задача может быть переформулирована следующим образом: Для каждого погружного насоса необходимо определить такую максимальную частоту вращения при которой все параметры из таблицы Таблица 2 не выходят за своих ограничения.

Данная формулировка подразумевает построение прогнозной модели для каждого параметра из таблицы Таблица 2. В рамках данной работы были предприняты несколько попыток построения таких прогнозных моделей. К сожалению, все они были неудачны. В основном, проблемы построения прогнозных моделей в данной задаче заключаются в проблеме неизвестного состояния пласта в точке перфорации. Пластовое давление меряется, обычно, раз в год, кроме того пластовое давление не может полностью описать процесс притока нефтяной смеси из пласта.

## Reservoir pressure problem

As mentioned above, there were several unsuccessful attempts to build a predictive model for reservoir pressure and other key parameters. In all these attempts we tried to build model (type(2)) based on raw and aggregated parameter. Let’s look at the simple measurement model of the “reservoir pressure” parameter:

Попытки построить прогнозную модель для предсказания дебита нефти и остальных ключевых параметров, используя «исходные» параметры (значение параметров в каждый момент времени) была не удачной. Рассмотрим простую модель измерения параметра «Давление пласта» (, для этого параметра корректно следующее представление:

|  |  |
| --- | --- |
| , | (5) |

where, is the measured value of reservoir pressure at the moment , is the true reservoir pressure at the moment ; - the high-frequency part of a measurement error; - the constant part of a reservoir pressure measurement error. The part is the main error part, it changes slowly, but because the reservoir pressure measurements are very rare, it turns out that we do not know exactly the reservoir pressure at any given time. In order to get rid of this component of the error, the approach "conversion from the absolute measurement to the incremental measurement" was implemented. This method consists of choosing a certain time base , where can be considered as constant:

где –измеренное значение давления пласта в момент времени *t*; – истинное давление пласта в текущий момент времени; – высокочастотная часть ошибки измерения; – постоянная часть ошибки измерения пластового давления. Основная проблема заключается в составляющей . медленно меняется, но т.к. измерения пластового давления проводятся очень редко, получается, что в каждый момент времени мы в точности не знаем пластового давления. Для избавления от этой составляющей ошибки был реализован прием, который в математике называется «переход от поля измерений к полю приращений измерений». Данный прием заключается в выборе некой временной базы, в рамках которой можно считать постоянным и использованию вместо исходного параметра разницы параметра по данной временной базе:

|  |  |
| --- | --- |
| = | (6) |

Since, is constant within the selected time base, it is possible to reduce this component, so we get:

Т.к. в рамках выбранной временной базы постоянно – можно сократить эту составляющую, получаем:

|  |  |
| --- | --- |
|  | (7) |

As we can see, we managed to get rid of the constant component of the error. Although, the high-frequency component has doubled, it is not significant because this part is very small. It is very important to choose the correct time base . On one hand, too small leads to an increase in the error, on the other hand, a big value leads to an additional error related to the difference between and

The transition from the initial parameters values to their increments for all other parameters was implemented in the same way.

Как мы видим нам удалось избавиться от постоянной составляющей ошибки. Хотя высокочастотная составляющая увеличилась в два раза, это не существенно т.к. шумы измерений очень малы. Очень важно правильно выбрать временную базу приращения . С одной стороны слишком малое приводит к возрастанию ошибки , с другой стороны большое значение приведет к возникновению дополнительной ошибки связанной с разницей в разные моменты времени.

Аналогичным образом происходит переход от исходных значений к приращениям всех остальных параметров, используемых в прогнозной модели.

## Prediction model

Assuming previous paragraph consideration, we used model (8) for building predictions instead of classical model (2).

Резюмирую рассуждения предыдущего параграфа, скажем что вместо классической предсказательной модели типа (2) была применена модель вида:

|  |  |
| --- | --- |
| *,* | (8) |

where – the predicted value, – the value of the same parameter in time , – the prediction of changing parameter value, – the predictive function that implements the forecasting depending on current state , last changing - and changing engine frequency - .

где – значение прогнозируемого параметра в предыдущий момент времени, – изменение прогнозируемого параметра,  *–* прогнозная функция реализующая прогноз изменения параметра при изменении параметров за выбранную базуи изменение частоты ЭЦН - .

From all the history of measurements, the moments when the ESP engine frequency changed were selected to build a predictive model . Machine learning methods - training with a teacher - were used to implement the predictive model. The measurement value before switching were used as parameter values and the value parameters after switching were used as .

Для построения прогнозной модели в истории измерений были выбраны моменты времени переключения частоты ЭЦН. Для реализации прогнозной модели использовались методы машинного обучения – обучение с учителем.

Unfortunately, there were not enough switches for every ESP in history to ascertain how engine frequency depends on key parameters. The common predictive model was built for these cases, so the full predictive model consists of two parts:

1. Common model – that was learned using all switches for all ESPs. This model is based on method, called «Random forest», with a lot of branches and a small depth for each tree.
2. Individual model – that was learned using only current ESP data. The linear regression model was used in this case.

Для каждой скважины были использованы два основных типа моделей:

1. Общая модель – обученная на истории изменений частоты ЭЦН на всех скважинах. В основе этой модели лежит метод «случайный лес».

2. Персональная модель - построенная на основе истории изменений частоты ЭЦН на текущей скважине, реализована посредством линейной регрессии.

The final forecast was built as follows:

Итоговый прогноз строился следующим образом:

|  |  |
| --- | --- |
|  | (9) |

where, is the individual forecast, is the common forecast, – the confidence of individual forecasting (). The algorithm of calculation is not given here, because of its complexity. Let's just say that is close to 1 if we are confident in individual model forecasting and it is 0 otherwise.

где – прогноз изменения параметра по построенной персональной модели, - прогноз общей модели изменения параметра.

Forecast -perdiction

## Recommendation system

The recommendation system is as follows:

Firstly, the prediction model builds a forecast for each parameter from Table 2 for each well. The forecast contains an error. Let's try to estimate this error by selecting all the moments of time when the ESP frequency was changing. Let’s consider equation (7) one more time but with forecast error:

Рекомендательная система строиться следующим образом:

Для каждой скважины, строиться прогноз каждого параметра из таблицы Таблица 2. Мы знаем, что прогноз может содержит ошибку. Попробуем оценить эту ошибку выбрав и истории все моменты времени, когда происходило изменение частоты ЭЦН.

|  |  |
| --- | --- |
|  | (10) |

where is the parameter value after engine frequency changing, is the parameter value before engine frequency changing, is the prediction of changing value after engine frequency changing, is the prediction error.

где – значение параметра *d* после изменения частоты ЭЦН, – значение параметра *d* до изменения частоты ЭЦН, – прогноз изменения параметра *d* при изменении частоты, – ошибка прогнозирования.

It is very important for to stay within range (table Таблица 2) in order to keep safe functioning of mining equipment. Consequently, it is necessary to evaluate . Let’s take all engine frequency switches and build a forecast of «input pressure» changing, then take a difference between and заменить на d. Following this way, we will get the histogram on Figure 1 1.

Для безопасного функционирования погружного оборудования необходимо что бы не вышло за ограничения из таблицы Таблица 2. Для этого, кроме прогноза необходимо оценить ошибку прогноза . Имея исторические данные и построив прогноз для любого ключевого параметра для всех переключений частоты из истории, мы можем построить гистограмму распределения ошибки . Пример такой гистограммы для параметра «Давление на приеме» приведен на рисунке 1.

*figure* *1*

Analyzing «input pressure», the cases when are the most interesting. In such cases there is a risk to cross the bottom limit (28 kg/m^2) and it can lead to well drying. Let’s look at the left tail of histogram (Figure [1]). As it is seen (honestly, it is impossible to see, but could be calculated) that built prediction model works quite good for «input pressure» parameter. The prediction model with a probability of 99% does not produce error more than 2.4 kg/m^2. So, now we know if «input pressure» predicted value will be more than 30.4 kg/m^2, the limit for «input pressure» will be accomplished with probability more than 99%.

При прогнозировании давления на приеме нам интересны те случаи, когда прогноз ошибается в большую сторону, т.е. , именно в этом случаи мы рискуем пересечь нижнюю границу (28 кг/см^2) из таблицы Таблица 2 что, может привести к высушению скважины. Анализируя левый хвост гистограммы ошибок на рисунке 1, можно сделать вывод о том, что прогнозная модель с вероятностью 99% не ошибается в большую сторону более чем на 2.4 кг/см^2. Таким образом, для параметра «давление на приеме» значение не должно быть меньше чем 30.4 кг/см^2.

All other key parameters were processed using the same logic. Note that key parameters values do not change immediately. The «engine temperature», for example, changes in 2-3 hours, after that time it becomes asymptomatic and remains constant until the next engine frequency changing. Moreover, «input pressure» changes in 5-10 days and becomes constant only after that. So it’s very important to choose true to get correct evaluation of .

Аналогичная логика была реализована для всех остальных ключевых параметров. Заметим, что при переключении частоты не все параметры меняются одинаково быстро. Так температура двигателя меняется в течении 2-3 часов, после этого выходит на асимптоту и остается постоянной вплоть до следующего переключения частоты. Давление на приеме при изменении частоты меняется в течении 10 дней, после этого становиться постоянным. Это надо учитывать при оценки из (10) и выбирать правильное приращение для получения корректной оценки ошибки .

Assuming all previous reasoning, RS recommends the biggest engine frequency when none of the key parameters exceeds the sated limit.

В качестве рекомендации, рекомендательная система выдает наибольшую частоту ЭЦН при которой ни один ключевой параметр не превышает заданное для него ограничение.

Unfortunately, ESP can’t be effective enough in wells with low reservoir pressure. In some cases, the electricity cost spent on crude oil upstream can be more than the crude oil cost. To protect against inefficient energy consumption, the following criterion is applied:

К сожалению, ЭЦН бывает недостаточно эффективен и при низком дебите скважин расход электроэнергии может превышать стоимость поднятой нефти. Для защиты от неэффективного расхода электроэнергии применяется следующий критерий

|  |  |
| --- | --- |
| , | (11) |

where is the prediction of daily production of crude oil, is the current daily production of crude oil, is the oil cost, is the dilution percentage, is the prediction of daily energy consumption that will be used in prediction mode, is the current energy consumption, is the energy unit cost.

где – прогноз дневного дебита нефти при выполнении рекомендации, – текущий дневной дебит нефти, – количество электроэнергии (кВт) в день необходимое для работы насоса в рекомендуемом режиме, – текущие дневные траты электроэнергии (кВт), – стоимость электроэнергии (руб/кВт), – стоимость тонны нефти, – процент обводненности.

Using this criterion RS gives the recommendation only if the increase in electricity consumption does not exceed the increase in profit from crude oil production.

Рекомендательная система выдает рекомендации только в том случае, если увеличение трат на электроэнергию не превосходит увеличение прибыли от добытой нефти.

The real example of recommendations proposed by RS can be seen in Table Таблица 3.

Пример рекомендации выданные рекомендательной системой можно видеть в таблице Таблица 3.

Таблица 3. Пример рекомендаций рекомендательной системы

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| well id | current engine frequency | recommended engine frequency | current crude oil production | predicted crude oil production | recommendation | comment |
| … | 40.02 | 40.02 | 27.78 | 27.78 | optimal mode installed |  |
| … | 49.17 | 48.17 | 205.76 | 196.19 | decrease engine frequency | Alert!!! «input pressure to low». risk of well drying out |
| … | 52 | 54 | 316.65 | 348.89 | increase engine frequency |  |
| … | 50.04 | 51 | 38.66 | 42.98 | increase engine frequency |  |
| … | 50.07 | 50.07 | 48.18 | 47.57 | optimal mode installed |  |
| … | 49.08 | 50 | 50.97 | 46.29 | increase engine frequency |  |

As you can see in Table Таблица 3, the RS gives recommendation not only to increase engine frequency, but as well to decrease engine frequency to avoid risky situations (engine overheating and etc.) for different wells. In particular, the second row of Table 3 contains the recommendation to decrease the frequency of ESP in order to prevent the risk of well drying.

Как видно из таблицы Таблица 3 рекомендательная система выдает рекомендации не только по увеличению частоты работы ЭЦН, но, так же, выдает рекомендации для предотвращения нежелательных ситуаций (перегрев двигателя и тд.). В частности во второй строке таблицы Таблица 3 приведена рекомендация по снижению частоты ЭЦН для предотвращения риска высыхания скважины.

It is worth mentioning that proposed RS could be applied (and is real applied) not only for ESPs that work in constant mode, but as well for ESPs that work in alternating mode. In alternating mode the ESP engine always works with the same constant frequency (50Hz), it works for some time , then it stops for time 0 ,then it starts working again for time and so on. In this case, RS does not operate the engine frequency, but it operates the and length. The extension of the period instead the period also makes it possible to increase the oil flow rate.

Стоить отметить что разработанная рекомендательная система может быть применена (и применяется) так же для скважин, работающих в переменном режиме. В этом режиме ЭЦН всегда работает с номинальной частотой (50Гц) при этом ЭЦН работает в течении времени , затем ЭЦН останавливается на время , затем снова включается на время , и тд. В данном режиме работы рекомендательная система управляет не частотой ЭЦН, а периодом работы. Увеличение периода за счет периода , так же позволяет увеличить дебит нефти.

## Summary

The proposed in this article prediction and recommendation systems have shown high accuracy and reliability throughout the entire period of using. The authors relate this result not only to the high quality of the ML models, but also to the close interaction between the data scientists and the domain experts. The proposed RS based on the above models was deployed into production in mid-2018 and so far has shown a stable increase in crude oil production of 1.5%. Also, during the implementation of the RS, the increase in the reliability of ESP functioning was noted. The potential cost of ESP failure extremely higher than one warning to stop or correct the ESPs engine frequency. This result can be connected to the recommendations to reduce the engine frequency in critical cases. The implementation of RS based on ML approaches does not require additional equipment (high performance servers, sensors etc.) and can be easily implemented on any available technical base. It is worth noting that the proposed RS can also be adapted to the rod pumps.

Приведенные в данной работе прогнозные и рекомендательные модели показали достаточную точность и надежность на протяжении всего периода применения. Авторы связывают этот результат с тесным взаимодействием между специалистами по машинному обучению и доменными экспертами. Рекомендательная система на основе приведенных моделей была введена в эксплуатацию в середине 2018 года и до настоящего момента показывает стабильный прирост добычи нефти на уровне 1.5%. Так же при внедрении рекомендательной системы был отмечен рост надежности функционирования погружного оборудования, связанный с рекомендациями о снижении частоты в критических случаях. Реализация рекомендательной системы на основе ML-подходов не требует дополнительного оборудования, может быть легко реализована на любой имеющейся технической базе. Стоит отметить, что реализованная рекомендательная система так же может быть легко адаптирована для применения на rod pump.

References

• The potential cost of ESP failure extremely higher. ESP’s are placed at the bottom of oil and gas wells, operate in the extremely harsh environment, are hard to monitor, expensive to replace, and can lead to significant lost of production if “problems” go undiagnosed or unnoticed. Failure rate ~ 0.209[1].

1. <https://cdn.osisoft.com/corp/en/media/presentations/2012/UsersConference2012/PDF/UC2012_IndustrialEvolution_Tom_Hosea_ESPOptimization.pdf>
2. ---