**“StudentDigitalFest”**

# Optimization of the well production collection system using machine learning methods


**Astana IT University**

Sultan Khajidursunogly

Yerkezhan Edgenova

Ulpan Kudaibergen

Batyrkhan Shutenov

**Instructor:**

Zhakiyev Nurkhat


2022, Nur-Sultan

**Agenda**

## Abstract

Oil has been considered one of the world's main sources of energy for a long time. The most important uses of oil include energy for transportation, industrial power, heating and lighting, and fertilizer production. Due to the fact that production efficiency is a priority for oil production enterprises, much attention is paid to short-term operational planning. While operational planning aims to maximize productivity on a daily basis, rapidly changing conditions during the production process mean that even a daily plan may not always be up-to-date. In this regard, the relevance and effectiveness of the application of machine learning methods in optimizing the system for collecting well production is undeniable. Our strategy is to maximize oil using historical data and machine learning methods. But in order to reach the use of these algorithms, the data must be analyzed and go through several preparation stages such as determining the outliers, identifying correlation parameters, and examining distribution. During the working process, the inverse proportionality between gas and oil was noticed, implying that we can find when oil would be maximized, and gas minimized. Based on the further data, we learned that the "drop pressure" parameter is not so strongly correlated with the maximization of oil and the minimization of gas. As a result, using important inputs like WellGOR and Well Pressure, and based on such common methods as backward_propagation and chain_rule, the invented algorithm is able to offer the most optimal result with the minimum error. Therefore, the application of the proposed machine learning model to optimize the routing and maximize the production of oil in the industry guarantees the regular implementation of oil production plans and reduces the amount of work required to extract the needed volume of raw materials.

## Introduction

### Project relevance

Oil is one of the world's main sources of energy, as well as the main ingredient in a wide range of products used in everyday life. Among the most important uses of oil are energy for transportation, industrial power, heating and lighting, and fertilizer production. In geosciences, complex forward problems met in geophysics, petroleum system analysis, and reservoir engineering problems often require replacing these forward problems with proxies, and these proxies are used for optimization problems.

Due to the fact that production efficiency is a priority for oil production enterprises, much attention is paid to short-term operational planning. While operational planning aims to maximize productivity on a daily basis, rapidly changing conditions during the production process mean that even a daily plan may not always be up-to-date and may not always correspond to provided conditions. Thus, operators may need to update and re-solve the planning problem several times a day. Moreover, the problem of optimizing well routing is complicated by the interconnection between the well collection system and process units, multilevel wells, and system limitations. The use of models based on conversion physics to solve the problem is limited and time-consuming.

In this regard, it is impossible to deny the relevance and effectiveness of the application of machine learning methods in optimizing the system for collecting well production. The use of machine learning for data cleaning, integration, data transformation, application development and analysis of oil well data is a new scientific approach to solving the optimization problem.

**Ecological justification**

While much of the world depends on the production or trading of oil to fuel their economies, these activities can cause serious environmental damage, knowingly or unintentionally. The oil industry carries a great potential for danger to the environment and can affect it at different levels: air, water, soil and, consequently, all living beings on our planet. In this context, the most common and dangerous consequence of the oil and gas industry is environmental pollution. Pollution is associated with almost all activities at all stages of oil and gas production, from exploration to refining. Wastewater, gas emissions, solid wastes and aerosols generated during drilling, extraction, processing, and transportation make up more than 800 different chemicals, among which oil and oil products certainly predominate.

Other detrimental impacts on the environment include increased greenhouse effect, acid rain, deteriorating water quality, and groundwater pollution. The oil and gas industry can also contribute to the loss of biodiversity as well as the destruction of ecosystems, which in some cases may be unique.

Wastewater from refineries and petrochemical plants dumps tons of toxic waste into nearby waters. Gas and oil pipelines block many streams and rivers, flooding pastures and arable land. In addition, entire bays and coastlines have been polluted by oil spills and runoff of toxic chemicals.

The environmental damage caused by the extraction and production of oil can also directly affect the lives of people in the region. It may include water pollution and soil pollution. People suffer from environmental devastation because it damages vegetation, livestock, and the physical health of the people themselves.

Thus, the use of oil resources for energy production leads to emissions that damage the environment. As a result, the use of alternative technologies for generating energy from renewable sources is becoming more widespread. However, the International Energy Agency estimates for 2020 that oil and gas still account for about 56% of the world's total energy consumption. Therefore, the environmental issue of reducing the harm caused by the oil industry is still relevant. Unfortunately, this damage cannot be completely eliminated, but it can be reduced through optimization processes. Suppose machine learning techniques are applied to optimise the routing and maximize oil in the industry. In that case, much less work will be required to extract the required number of raw materials, while also reducing the amount of waste and avoiding possible leaks.

**Economic justification**

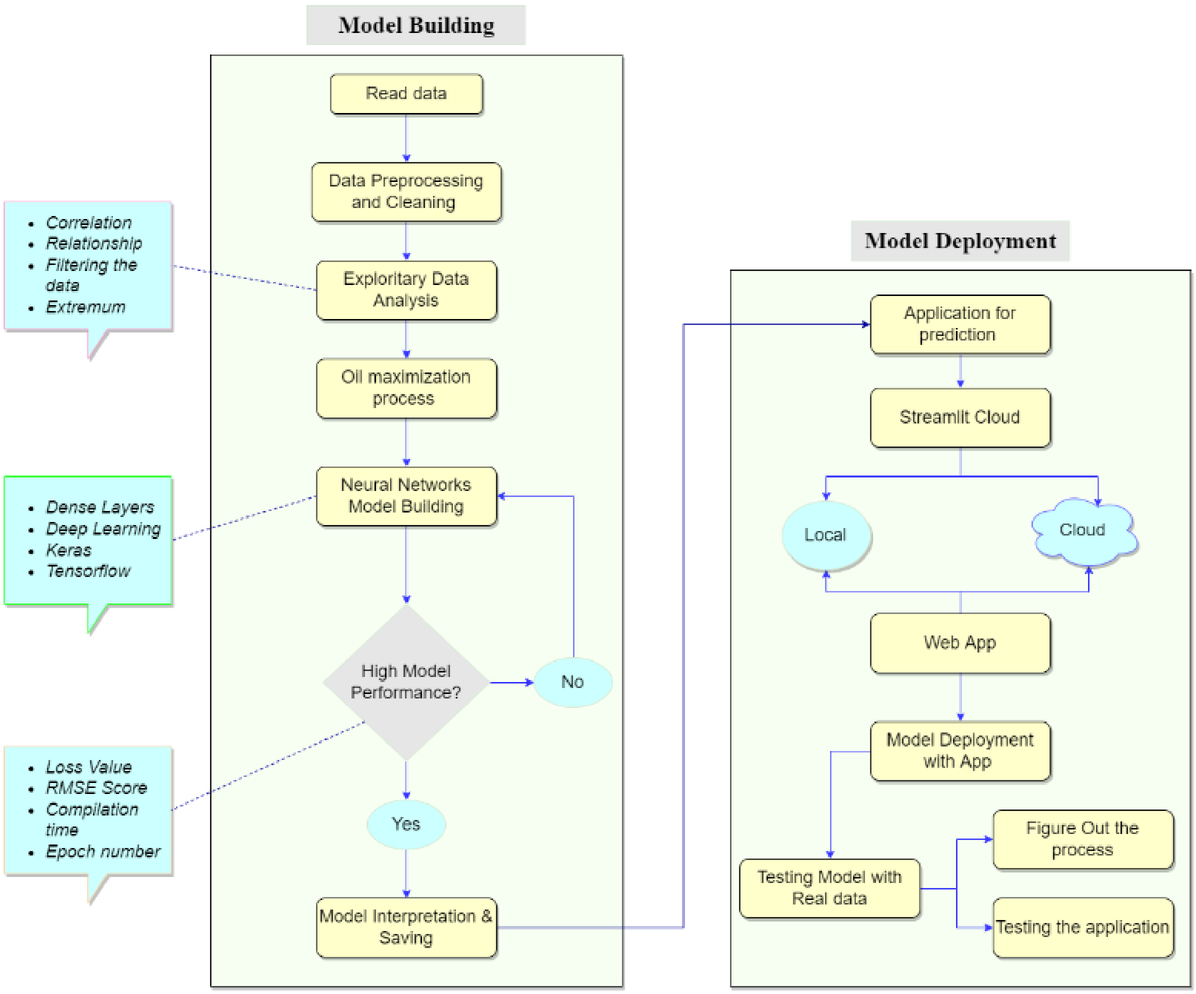
Oil and gas have been the lifeblood of the world economy for over a hundred years, accounting for more than half of humanity's primary energy resources. Due to their high energy density and easy availability of fossil fuels, they have played an important role in some of the largest industries such as chemicals, transportation, energy, petrochemicals, etc. This availability of cheap and abundant energy enables countries to move out of poverty while making energy security become a national priority for most of them. The supply of crude oil and natural gas has become very important, especially with the growing demand for energy for comfort in life and technological development.

It is clear that the activities of extracting crude oil and natural gas make a huge economic contribution that benefits both entire countries and their citizens. Some of the ways in which well production contributes to the economy include taxes, employment and job creation, gross regional product, local spending on goods and services, contribution to electricity and utilities, the provision of foreign exchange reserves, and the possibility of realizing investments.

Despite such a positive impact on the economy, the energy industry is under constant pressure in all production environments: traditional or non-traditional; on land or at sea; in the extraction of oil or gas. When oil prices fall, the margins of oil companies decrease. The development of new oil fields requires large capital expenditures, which companies face periodically in times of uncertainty. What's more, in the face of rising costs and recent widespread setbacks, exploration and production companies are looking to increase reserves and maximize production while operating safely and avoiding environmental impacts. But inaccurate or redundant solutions often lead to unnecessary downtime, suboptimal performance, and increased maintenance and safety issues. In this regard, the application of machine learning methods to optimize the well gathering system becomes an important step for oil companies on the way to high productivity while simultaneously increasing their contribution to the economic sphere.

**Main Part**

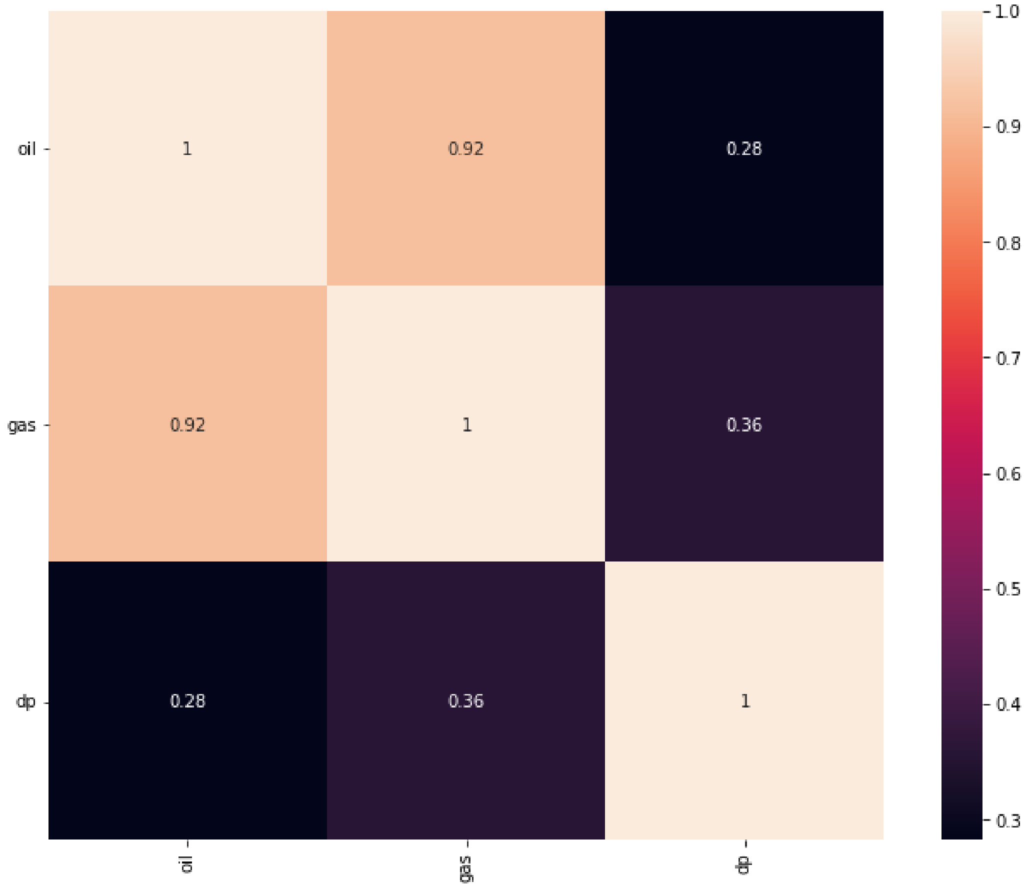
The strategy of the project



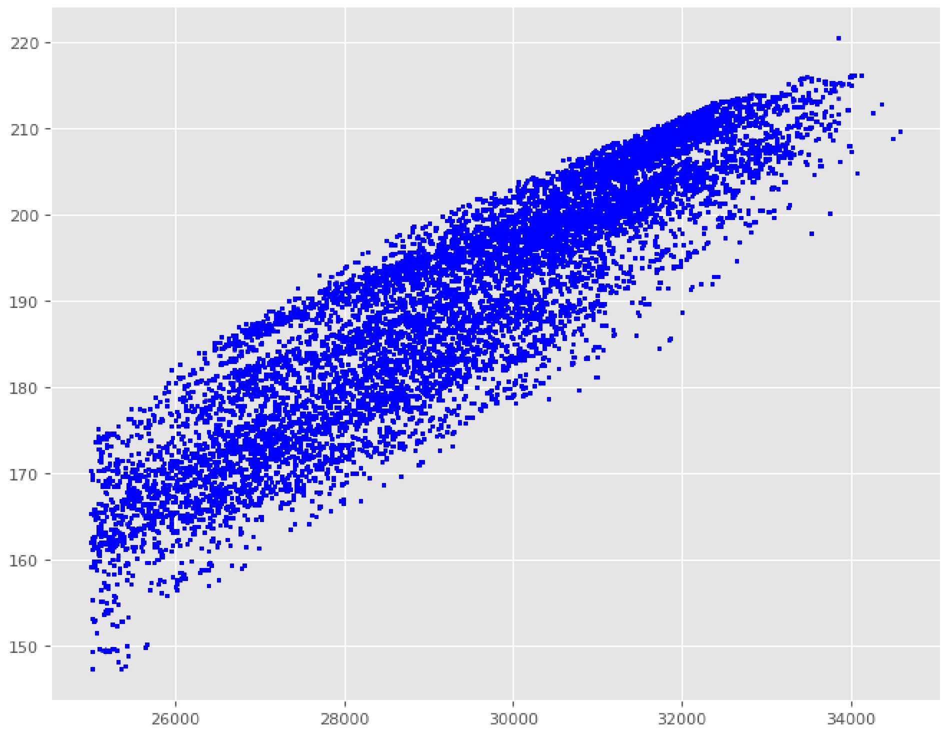

Above, given a short, briefly described flowchart by each stage with sub-stages of our work. All stages of working with data were completed and software was created to show the results in test mode. Our strategy is to maximize oil using historical data using machine learning methods. But in order to reach the use of these algorithms, the data must be analyzed and go through several stages such as determining the outlier, identifying correlation parameters, distribution, and so on. All these steps are necessary in order to understand how the data behaves, and analysis is also needed for the correct selection of parameters such as gas and oil consumption, pressure, and temperature. On the screen you can see our strategy for working with the project:

**Maximization of the oil production (Determine the best combination)**

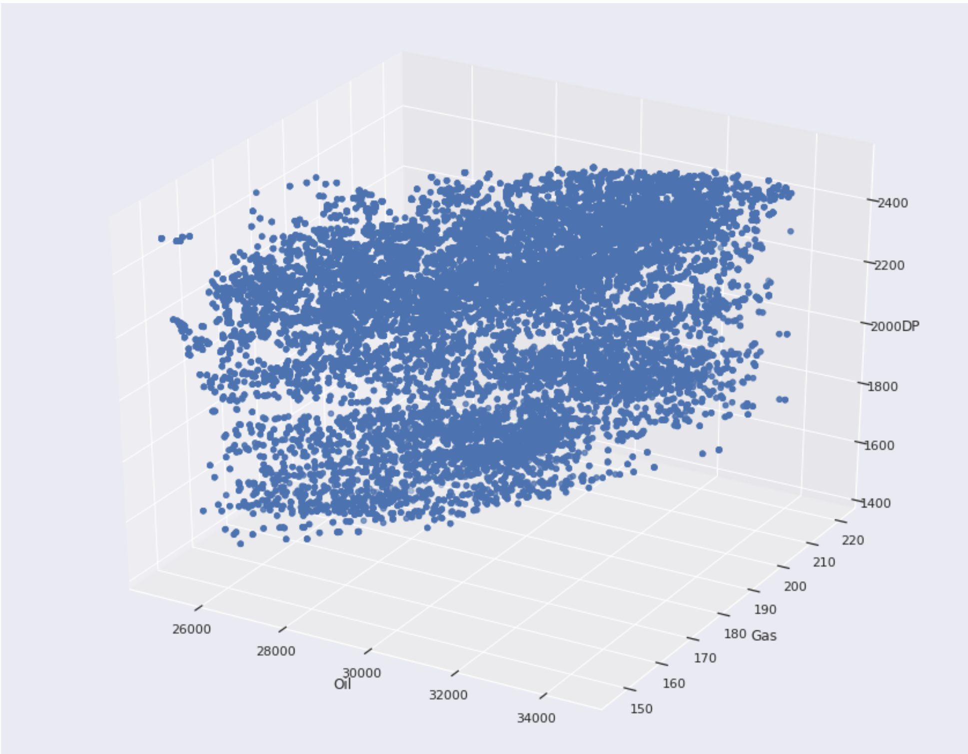
To maximize oil, we used the data ranking method using google collab and the python programming language, after cleaning and aggregating the data, we select the best parameters using correlation. Since all our data is discrete and it has its own local extremum point, we can determine the best several combinations where oil is maximized. In addition to this, the inverse proportionality between gas and oil was noticed, where we can find when oil would be maximized, and gas minimized. Based on the data, we learned that the "drop pressure" parameter is not so strongly correlated with the maximization of oil and the minimization of gas. The fact was proved using a graph of a function with 3 axes where we can see parameters such as oil, gas rate and drop pressure.




**Correlation plot between 3 main parameters**




**Dependency between oil and gas rates**




**Scatter plot with 3 axes (Oil, Gas, Drop Pressure)**

**Understanding Machine Learning models (Model selection process)**

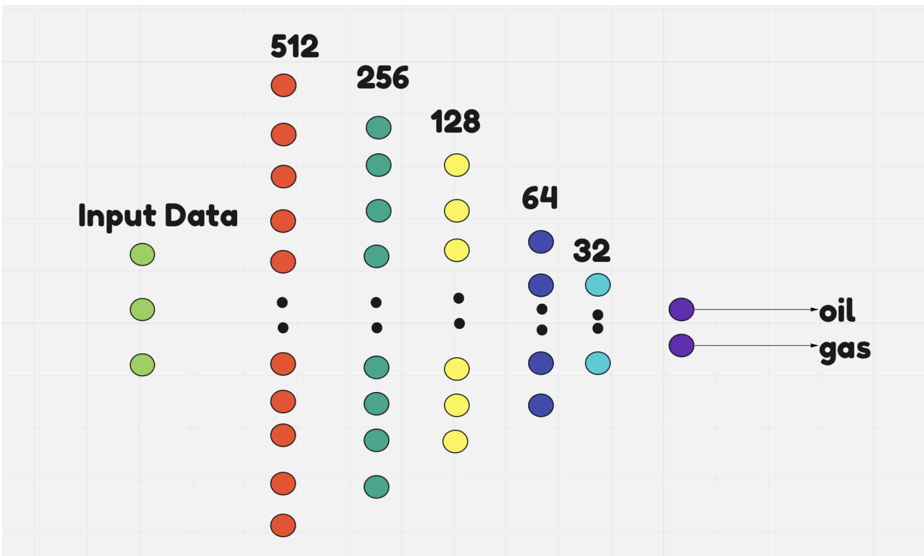
Let's go directly to the algorithm that we used in our project. The name of the project: Optimization of oil wells using machine learning methods. That is, it is the methods and algorithms of machine learning that are needed. Machine learning learns from past data to make the right decisions. In our case, we give our robot historical data so that it understands where oil is maximized and gas is minimized, after more attempts, our robot uses its algorithm to select such parameters to output oil maximized. That is, we enter parameters, then make predictions, and then compare with the real answer.

**Machine Learning has several steps of implementation of the algorithms:**

1. **Preprocess and load data -** As we have already discussed data is the key to the working of neural network and we need to process it before feeding it to the neural network. In this step, we will also visualize data which will help us to gain insight into the data.
2. **Define model -** Now we need a neural network model. This means we need to specify the number of hidden layers in the neural network and their size, the input and output size.
3. **Loss and optimizer -** Now we need to define the loss function according to our task. We also need to specify the optimizer to use with the learning rate and other hyperparameters of the optimizer.
4. **Fit model -** This is the training step of the neural network. Here we need to define the number of epochs for which we need to train the neural network.

After fitting the model, we can test it on test data to check whether the case of overfitting. We can save the weights of the model and use it later whenever required.

**Implementation of the Neural Network model**



In any neural network, a dense layer is a layer that is deeply connected to its previous layer, which means that the layer's neurons are connected to each neuron of its previous layer. This layer is the most commonly used layer in artificial neural networks.

The model accepts GOR, Reservoir Pressure data as input, which initially goes through a linear function, then passes through a Relu activation function, which accepts only positive maximum values.

As we saw in the parameters, we have three main attributes: activation function, weight matrix and displacement vector. Using these attributes, the dense layer operation can be represented as:

- Output = activation(date(input, kernel) + bias)

The model accepts reservoir pressure data as input, which initially passes through a linear function, then passes through the Relu activation function, which takes only positive maximum values.

- $z = wx + b$
- $a = max(0, z)$

The training took 1000 epochs, with each iteration he changed his parameters using back propagation and gradient descent.

- $w = w - \alpha \frac{dL}{dw}$
- $b = b - \alpha \frac{dL}{db}$

The learning rate in the model is 0.001, and the optimizer is Adam.

- $\alpha = 0.001$

The output of 2 outputs is oil and gas, with a linear mathematical equation.

**Performance evaluation**

1. We will need to **load the weights** we saved during the model training and **compile the model** with those weights.
2. **Predict the values** in the test dataset using the model.
3. **Compare** the **predicted** values to the **actual** values in the test dataset.
4. Calculate RMSE Value

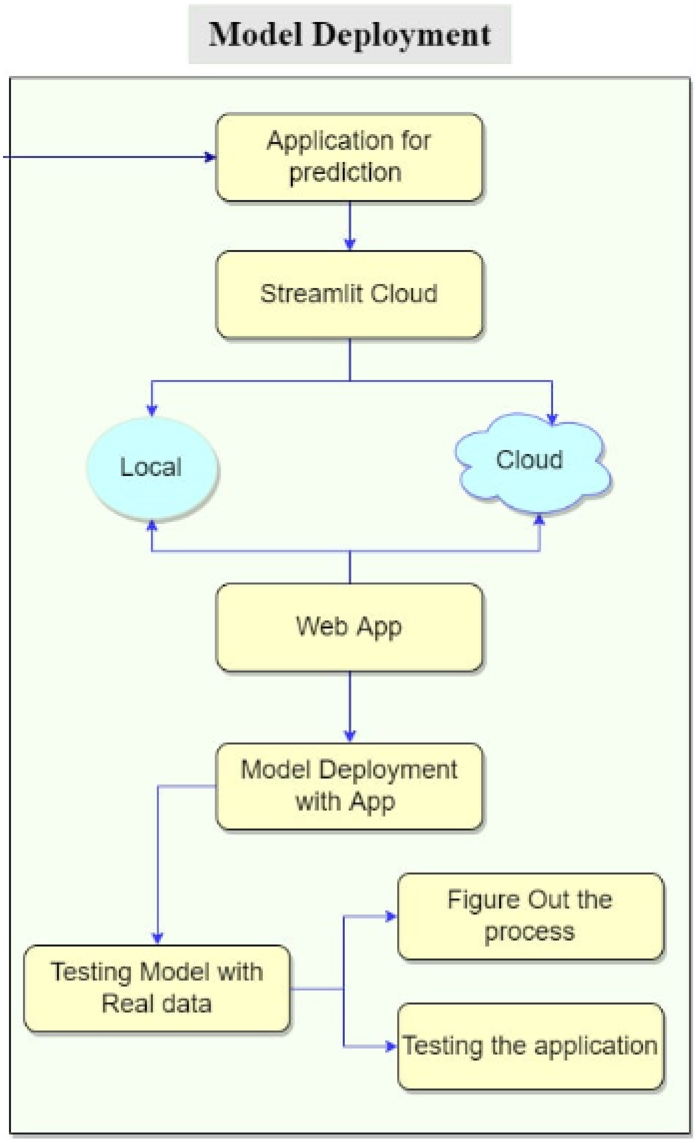
**Software Architecture**

Our software programme demonstrates how we use machine-learning methodology with software and make predictions. For the software technology, we used python programming language with some libraries and trained neural network model.

We have created two applications for two specific algorithms. The first is to predict the total volume of oil and gas, the second is to find the optimal combination of pipelines. Right now on the screen you can watch a demonstration of the application we created to predict the total volume of oil and gas. To use the service, you need to enter some data about the wells provided, namely 18 Gas Oil Rate indicators and 18 Pressure indicators. Then, to get the prediction result, you need to click on the appropriate Predict button. As a result, you get two values: the total volume of oil and the total volume of gas.

The second application, designed to find the optimal combination of pipelines, works on the same principle: you enter the necessary data and press the button to get the desired result. In this case, the result will be a combination of 81 numbers implying pipelines.
It is worth noting that the created applications demonstrate a specific case of using the model, that is, they are designed specifically for the data provided to us about 18 wells. However, it can be easily adapted to a different context.

Now let's move on to the technical part and the application development algorithm itself. The demonstrated applications were created in the Python programming language using the streamlit library. First, the machine learning models themselves were developed. Since the process of training the model took too much time due to the generation of 1000 iterations at each launch, we decided to build the application in such a way that the model is trained only once during creation and at subsequent calls a ready-made saved version in sav format is launched. Later, the web application interface was built using the streamlit library. Streamlit is an open source Python framework used to deploy machine learning models in various web applications. Next, using the Sublime text editor in several files of the py format, we prescribed a link between the machine learning model and the web interface. In conclusion, we uploaded all the source documents and code to the GitHub repository and deployed the created application in open access via a link on the official streamlit workspace.




Difficulties in application development:
1. New development environment (streamlet), learning new syntax
2. The process of training the model takes too much time since 1000 iterations are generated
3. Interface development, design thinking
Solving the existing difficulties:
1. Viewing online tutorials, reading articles
2. In order not to train the model at each launch, we decided to build the application in such a way that the model is trained once during creation and a ready-made saved version is launched at subsequent calls
3. Exploring existing templates

Expected result and benefits of the project

As mentioned earlier, due to the recent widespread failures of oil production plans, companies involved in the exploration and production of good products are striving to increase reserves and maximize the extraction of the required resource, while ensuring maximum profit and avoiding environmental impact. This gives rise to two important factors that must be taken into account when planning activities: environmental and economic. Firstly, the oil industry carries a great potential for danger to the environment and can affect it at different levels: air, water, soil and, consequently, all living beings on our planet. Second, inaccurate or excessive economic decisions often lead to unnecessary downtime and costs, suboptimal performance, and increased maintenance and safety issues, especially in the face of falling oil prices and labour shortages.

It is expected that the model proposed during the “Deep Learning Methods for Pipeline Combinations” project will be able to address these issues. The work of the model is thought out based on information about existing wells. Using important inputs like WellGOR and Well Reserv. Pressure, and based on such common methods as backward_propagation and chain_rule, the invented algorithm is able to offer the most optimal result with the minimum error. Therefore, the application of the proposed machine learning model to optimize the routing and maximize the production of oil in the industry guarantees the regular implementation of oil production plans and reduces the amount of work required to extract the required volume of raw materials. At the same time, from the environmental side, it will be possible to reduce the amount of waste and avoid possible leaks. This will also have a positive impact on the economy in connection with the opening of additional jobs related to the field of automation and the possibility of new investments.

In general, in the field of optimization, machine learning can be used to find patterns in data, which can then be used to make decisions and make predictions. The algorithm used for this becomes more accurate as it processes more data. Machine learning is especially useful for datasets that are too large for manual analysis. More accurate results mean more efficient decisions and processes that benefit safety, people, the environment, production and profits.

## Conclusion & Recommendation

Real-time optimization of well production is a difficult and complex task. Rapidly changing conditions require an immediate and appropriate solution, while complex reservoir dynamics lead to multi-level and highly intricate models. The uncertainty of the problem requires the decision maker to consider the trade-off between maximizing oil production and maintaining feasibility against constraints in a given environment.

**References:**

Bikmukhametov, T., & Jäschke, J. (2019). *Oil Production Monitoring using Gradient Boosting Machine Learning Algorithm*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2405896319302009

Zhang, W., Liu, T., & Leng, S. (2021, December). Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0960852421013535?casa_token=jSXR3ACaY8sAAAAA:STqlpB4qWLQ44j9JCq7q9hqyQk2Eul1X2HyVtDaykSRlQNLs-u5IB85Je9ZSSHwOjwv2PXo_r4CN

Teixeira, A., & Secchi, A. (2019). Machine learning models to support reservoir production optimization. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2405896319301971

Wang, S., Chen, Z., & Chen, S. (2019, August). Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. Applicability of Deep Neural Networks on Production Forecasting in Bakken Shale Reservoirs. https://www.sciencedirect.com/science/article/pii/S0920410519303523?casa_token=xtcTC4nwvfMAAAAA:SMs8IR-B-3EbfEBfqTaOM_bpFiohDrhbcn4YUJBCN7CuLAUZ9n8TCM2l9wOWHcfKYQg7deHrL_9O

Wei, W., Rezazadeh, A., Wang, J., & Gates, I. (2021). *An analysis of toe-to-heel air*

*injection for heavy oil production using machine learning*.
https://www.sciencedirect.com/science/article/pii/S0920410520311633

Motlaghi, S., Jalali, F., & Ahmadabadi, N. (2008, November 4). *An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework*. ScienceDirect.
https://www.sciencedirect.com/science/article/abs/pii/S0957417407003697