**Подготовил: Хоменко Алексей**

**Дипломная Работа**

Построение модели прогнозирования состояния трехфазного сепаратора (пробкоуловителя) для оптимизации выхода полезного продукта

Оглавление

[1. Актуальность: 2](#_Toc93793544)

[2. Формализованная форма задачи и этапы ее решения: 2](#_Toc93793545)

[Получение исходных данных. Система сбора условного месторождения: 3](#_Toc93793546)

[Получение исходных данных. Описание сценариев. 3](#_Toc93793547)

[3. Математическая модель предсказания. Задача и метрики 4](#_Toc93793548)

[4. Анализ исходных данных: 5](#_Toc93793549)

[5. Обработка исходных данных: 6](#_Toc93793550)

[6. Разделение на обучающую и тестовую выборку: 8](#_Toc93793551)

[7. Итоговая модель 8](#_Toc93793552)

[Описание итоговой модели 8](#_Toc93793553)

[Обучение модели: 9](#_Toc93793554)

[Сравнительные результаты метрик итоговой модели и других алгоритмов: 9](#_Toc93793555)

[8. Возможный вариант внедрения: 10](#_Toc93793556)

[9. Выводы: 10](#_Toc93793557)

[10. Пути развития: 11](#_Toc93793558)

# Актуальность:

В основе построения задачи работы лежит до сих пор не решенная **проблема:**

В течении срока работы месторождения выход скважинного продукта меняется. Так же скважинный продукт может различаться в более короткий период, например, в результате сноса жидкой пробки из скважины или на линейном участке трубопровода.

В результате резкой смены скважинного продукта входные трехфазные сепараторы, установленные на центральном пункте сбора (пробкоуловители) могут не “справиться” (не успеть сдренировать всю жидкость) с резким увеличением компонентов входного продукта. Чем вызовет переполнения входного трехфазного сепаратора, газовые линии которого, не предполагающие (в таком объеме) наличие жидкого компонента, могут быть заблокированы, пропускная способность всего дальнейшего цикла подготовки снижена, а вышестоящее по газовому потоку оборудование может быть деградировано.

Следить за объемом каждого компонента скважинного продукта с каждой скважины или куста финансово не оправдано. Как правило, на выходе с куста устанавливается манометр и термометр для определения давления и температуры скважинного продукта, более технически сложное оборудование, например, проверка состава скважинного продукта (хроматограф), не монтируется на кустовых площадках. Так как для корректной работы последнего требуется очень много данных о входном продукте, получение которых потребует монтажа дополнительного оборудования.

**Была составлена задача:**

Составить прогнозную модель уровня жидкости входного трехфазного сепаратора, в зависимости от определенных паттернов в изменяющихся входных данных. Итоговая прогнозную модель может быть интегрирована в существующую установку сбора. Как возможный вариант реализации: имея предсказание уровня жидкости во входном трехфазном сепараторе контроллер регулирующего клапана, установленного ниже по потоку трехфазного сепаратора, может предикативно подать сигнал на кратковременное уменьшение площади проходного сечения регулирующего клапана. Что вызовет кратковременное снижение объема входного продукта (скорость заполнения трехфазного сепаратора жидкостью будет снижена). Как результат, входной трехфазный сепаратор не переполнится, вышестоящее дорогое оборудование не остановится по превышению порога жидкости, что оптимизирует выход полезного (в данном случае – газа) продукта.

# Формализованная форма задачи и этапы ее решения:

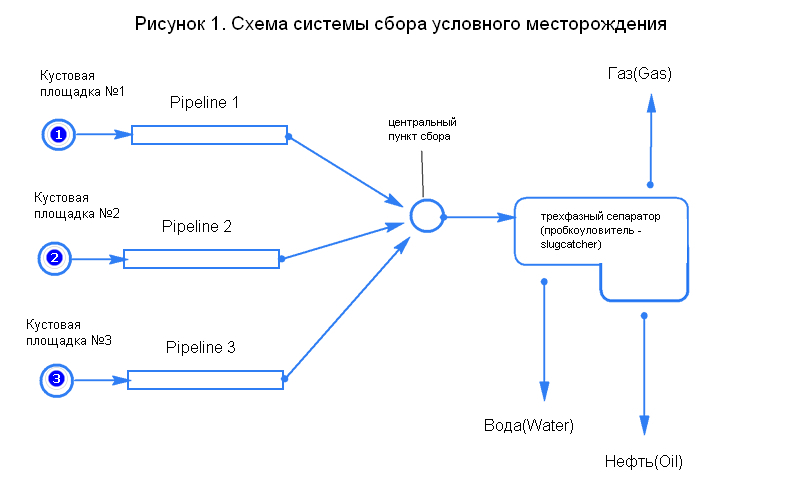
Формализованная **задача**: Составить математическую модель предсказания, в зависимости от изменяющихся входных данных – признаков.

За **исходные данные** принята модель сбора условного месторождения(**рисунок 1**). Условное месторождение состоит из 3 кустовых площадок, центральным пунктом сбора, из центрального пункта сбора скважинный продукт поступает на подготовку: очистку, переработку, транспортировку (на схеме не показано).

Скважинный продукт условного месторождения имеет в своем составе **нефть,** **газ** и **воду** в разных пропорциях. Соотношение каждого из компонентов меняется со временем работы месторождения, а также во время выноса газожидкостной пробки. Во время выноса газожидкостной пробки происходит резкое изменение пропорции каждого из компонентов скважинного продукта, что приводит к росту давления. Следовательно, показания манометров на кустовых площадках будут использованы как признаки, по которым будет строится предсказание.

## Получение исходных данных. Система сбора условного месторождения:

Система сбора состоит из 3-х кустовых площадок, на каждой кустовой площадке всегда устанавливаются и всегда в работе манометры и термометры с каждой скважины и всегда в работе манометры и термометры на выходе из кустовой площадки. Кустовая площадка соединена с центральным пунктом сбора трубопроводом. Профиль трубопровода принят неровным - приближен к реальности, имеющий несколько нижних точек. На центральном пункте сбора все линии соединены к трехфазному сепаратору (смотри **Рисунок\_1** ). Перед трехфазным сепаратором устанавливается регулирующий клапан. Регулирующий клапан для ограничения переполнения сепаратора регулирует объем входного потока в зависимости от показателей уровня и давления сепаратора. Продукт после трехфазного сепаратора поступает на переработку. На **рисунке 1** представлена схема сбора условного месторождения.

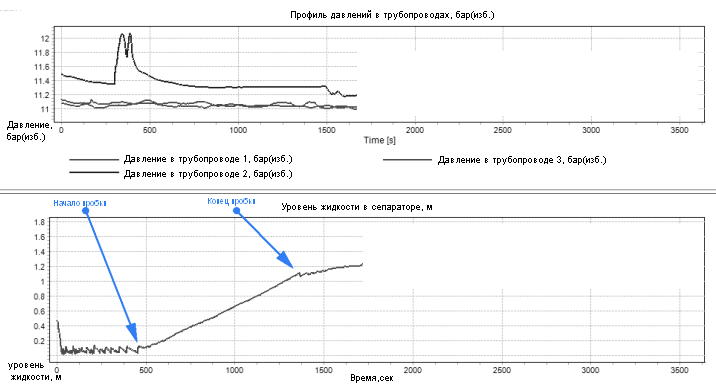


#### Принятые допущения для исходной модели месторождения:

Так как цель настоящего задания состоит в построении модели предсказания (прогнозирования), то характеристики месторождения условные. Принято, что скважинный продукт и характер работы системы сбора квазистационарный.

## Получение исходных данных. Описание сценариев.

В случайный момент времени моделируется выброс газа затем жидкости (снос пробки), газ, в свою очередь, выносит жидкость из скважины и линейного трубопровода.

Пример исходных данных для модели предсказания (смотри **Рисунок\_2** ) 

На рисунке видно, что резкий скачок давления (смотри верхний тренд **Рис.2**) примерно на 300секунде, приводит к росту жидкости в сепараторе (смотри нижний тренд **Рис.2**) примерно на 500секунде.

# Математическая модель предсказания. Задача и метрики

З**адача математической модели:** Составить модель предсказания, в зависимости от изменяющихся входных данных – признаков.

Признаки - исходные данные тренда давлений (см. **Рис.2** верхний тренд) предсказать значения жидкости в сепараторе (см. **Рис.2** нижний тренд). Поскольку задача является задачей регрессии, то метрики качества модели приняты **MSE** (Mean Squared Error) и **Функция потерь Хьюбера (при delta=1 является** SmoothL1 Loss)

Формулы каждой из **метрик**:

**MSE:**

Latex:

l\_n = \left( x\_n - y\_n \right)^2,

**Функция потерь Хьюбера (**приdelta =1 является SmoothL1 Loss**):**

Latex:

l\_n = \begin{cases}

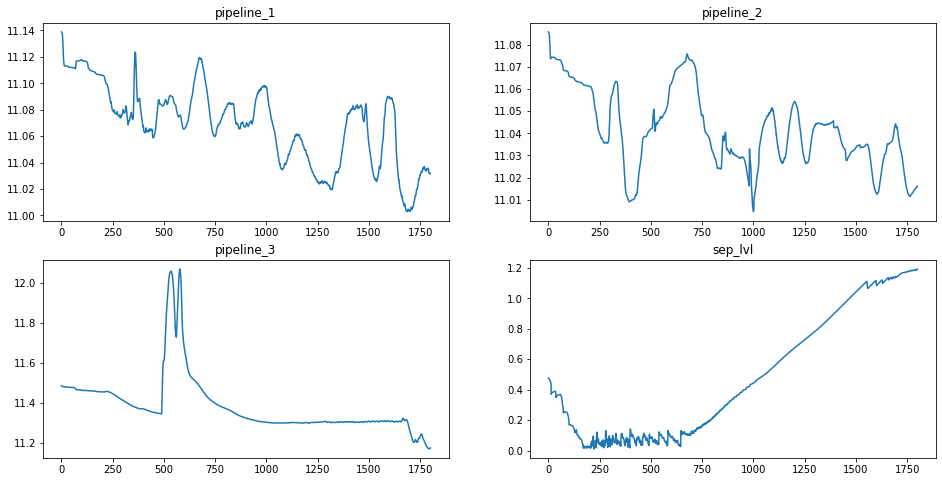
0.5 (x\_n - y\_n)^2, & \text{if } |x\_n - y\_n| < delta \\

delta \* (|x\_n - y\_n| - 0.5 \* delta), & \text{otherwise }

\end{cases}

# Анализ исходных данных:

Как выглядели входные данные. Для примера возьмем **сценарий\_001** (**Рис.3**):



Время, сек.

Время, сек.

Время, сек.

Время, сек.

Давление, бар(изб.)

Давление, бар(изб.)

Давление, бар(изб.)

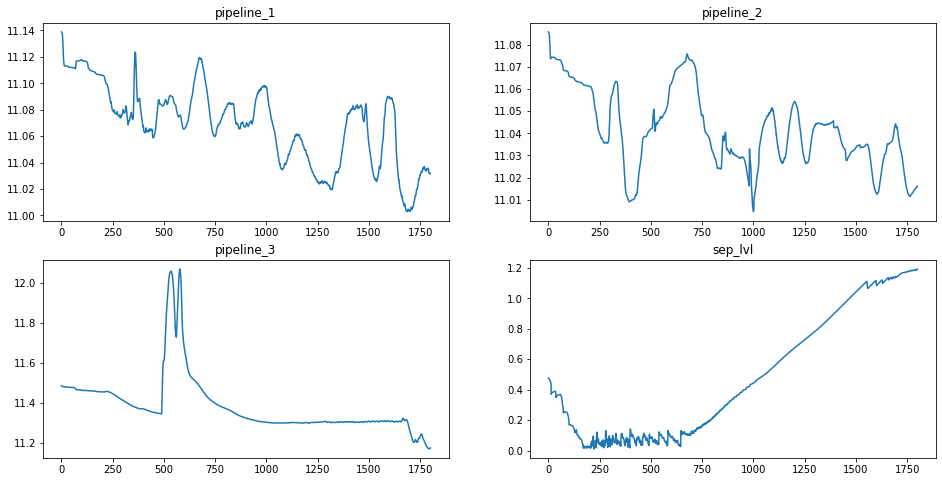
Уровень жидкости, м

По входным данным видно:

* данные приходят очень **хорошего** качества: шаг по времени 0.1сек
* на трубах **1 и 2** границы давлений 0.12 и 0.07, то есть в данном сценарии давление на трубах **1 и 2** квазистационарное.
* По графикам видно, что давления на ‘pipeline 1’ и ‘pipeline 2’ не менялись, и только ‘pipeline 3’ вызвало рост ‘sep\_lvl’ (так и должно было быть, именно в этом кейсе, приход пробки на сепаратор только от "pipeline\_3").
* В столбце ‘sep\_lvl’ очень много колебаний, на которые модель может ошибаться false triggering. Напоминание, цель задачи – **предсказывать** ожидаемый рост, а не отловить все колебания. Стало быть – все колебания являются шумом, нужно будет его обработать.
* Из анализа данных также определили, что Не во всех моделях начало роста давления и начало роста жидкости одинаковое, стало быть задача обработки будет состоять из определения: **когда именно** начался рост давления, **когда именно** начался рост уровня жидкости в сепараторе.
* Момент, когда жидкость приходит в сепаратор для математической модели предсказания будет является **target**-ом или ‘y’.

# Обработка исходных данных:

Пример обработки исходных данных для **сценария\_001** (**Рис.3 )**:



Время, сек.

Время, сек.

Время, сек.

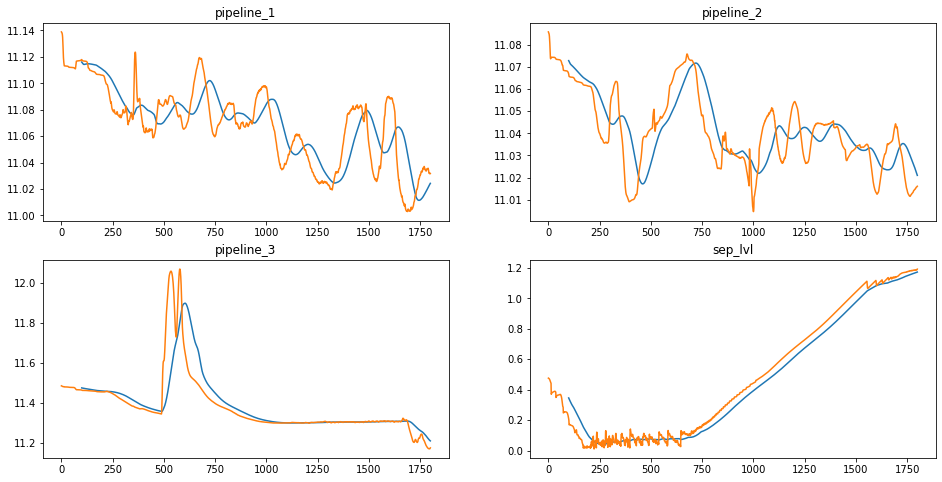
Время, сек.

Давление, бар(изб.)

Давление, бар(изб.)

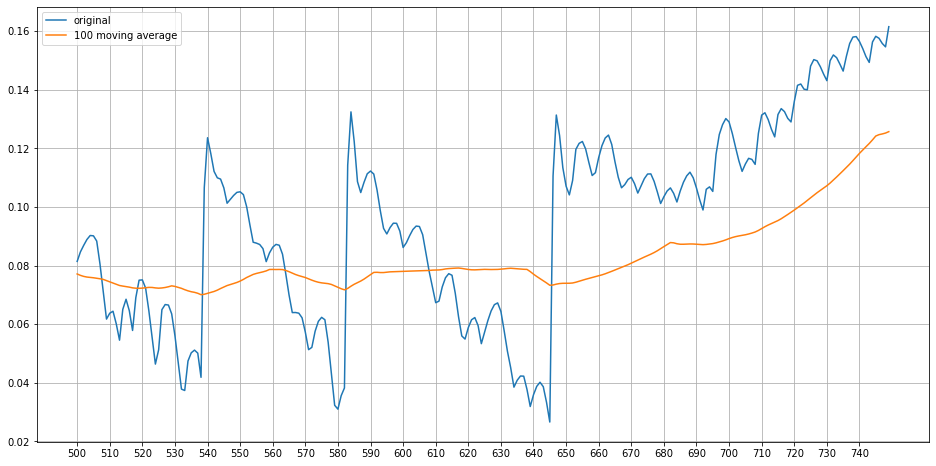
Давление, бар(изб.)

Уровень жидкости, м

* на трубах **1 и 2** границы давлений 0.12 и 0.07, после обработки давления будут стационарные со средним значением (то есть если диапазон < threshold, то давление = **mean**()
* В столбце ‘sep\_lvl’ очень много колебаний – все (не ключевые) колебания являются шумом, будут сглаживаться усреднением по скользящему окну с окном 100( сравнивалось несколько вариантов ширины окна – **100 самое стабильное**) (**Рис.4. Скользящее окно 100**)
* 

Время, сек.

Уровень жидкости, м

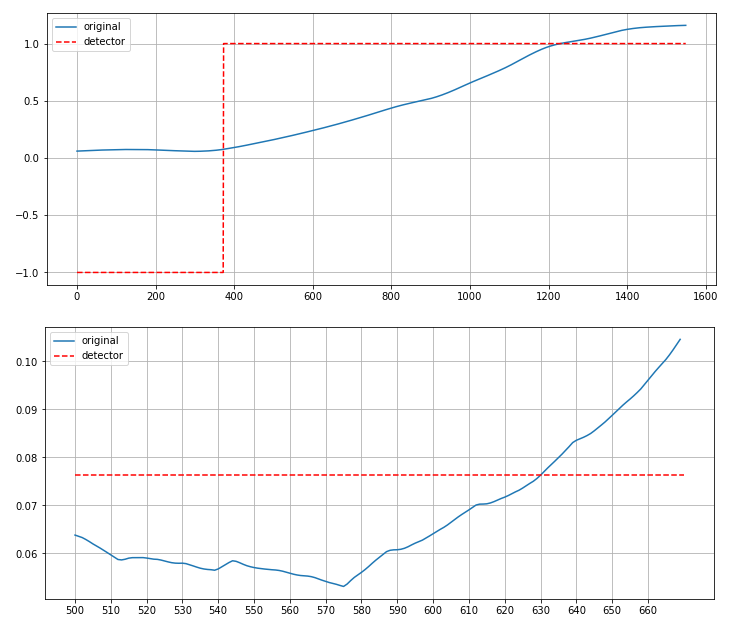
* Далее обработчик определяет, когда именно начался рост **давления**, потому что по всем сценариям рост давления может начаться рандомно (для повышения стабильности модели).
* Далее обработчик определяет, когда начался **рост уровня жидкости** в сепараторе, потому что этот параметр зависит от очень многих факторов, например, давлению, диаметру трубы, неровности профиля трубопровода, откуда пришла пробка. Для примера на **Рис.5** видно, что волна жидкости "приходит" на **645 сек**, но это не всегда.
* 

Время, сек.

Уровень жидкости, м

Обработчик входных данных будет использовать OneClassSVM так как он позволяет выловить самые незначительные аномлии в временном ряде для примера рис.6.

**Рисунок 6.1. SVM One class detector. Тренд роста объема сепаратора.**



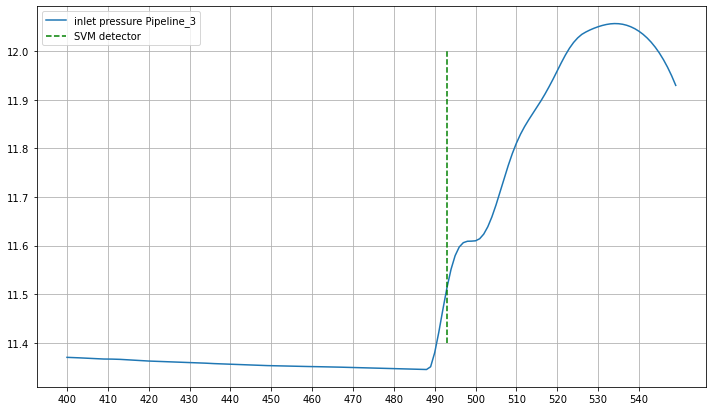
Время, сек.

Время, сек.

Уровень жидкости, м

Уровень жидкости, м

**Рисунок 6.2. SVM One class detector. Тренд роста давления.**

****

Время, сек.

Давление, бар(изб.)

Напишем обработчик входных данных и передадим ему все исходные данные:

**Обработчик** данных находится в папке ‘./scripts\_d/**data\_preparator**.py’

Данные **до** обработки : ‘./data/**raw\_data**/’

Данные **после** обработки : ‘./data/**prepared\_data**/’

# Разделение на обучающую и тестовую выборку:

Так как задача прогнозированиянапрямую завязана на поиск и отлавливание только одного “скачка” давления, то нет никакого смысла разбивать один сценарий на train и test. Для test выборки использовались отдельные сценарии на них отмечались метрики качества одели и устойчивость(robust).

# Итоговая модель

## Описание итоговой модели

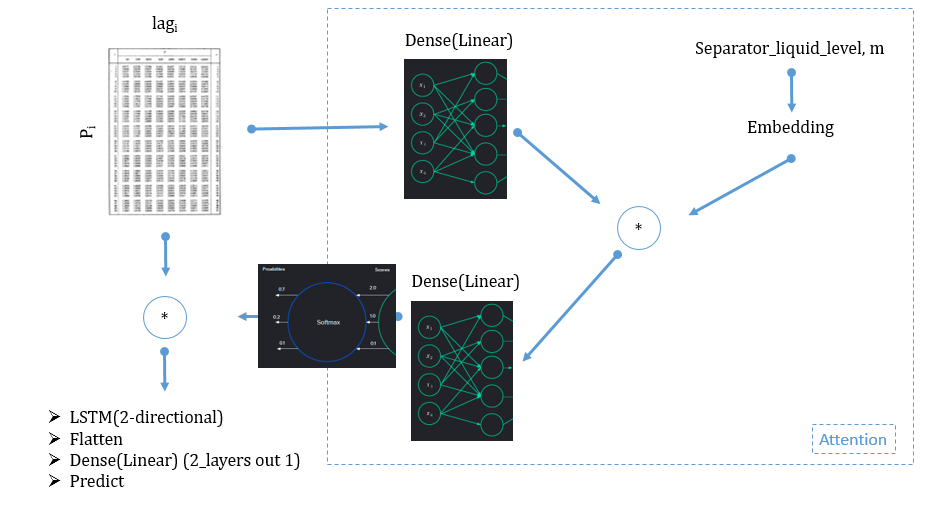
Исходными данными для итоговой модели вытянутые в строку значения lag\_j:

, где - это давление в момент времени i

- это отступ от давления в момент времени i на j шагов

1. Все lagij создаются для каждого трубопровода и конкатенируются горизонтально. Сами же абсолютные значения давлений в таблице удалены
2. Далее идет блок attention, суть которого определить на какую строку обращать внимание
3. Полученный softmax (attention) умножается на исходные lagij
4. Далее LSTM(2-directional)
5. Вытягивание в один вектор (Flatten)
6. Два полносвязанных слоя на выходе последнего 1 нейрон c активацией – linear

Граф вычислений итоговой модели представлен на Рис.7.



**Итоговая модель** находится в папке ‘./scripts\_d/**Attention\_Model**.py’

## Обучение модели:

1. Модель просчитывает весь тренд давлений делая предсказания.
2. Полученные предсказания сравниваются реальными данными, считается **функция ошибки HuberLoss (delta=1 - SmoothL1 Loss)**
3. Считается градиент. Данный подход качественнее и в 10 раз быстрее чем считать ошибку для каждого значения предсказания.

Код приближен к sklearn:

```Python

net = AttentionModel()

net.fit(X,y)

net.predict(X)

```

## Сравнительные результаты метрик итоговой модели и других алгоритмов:

Результаты итоговой модели (после 25 эпох обучения) средняя:

**MSE** = 0.015

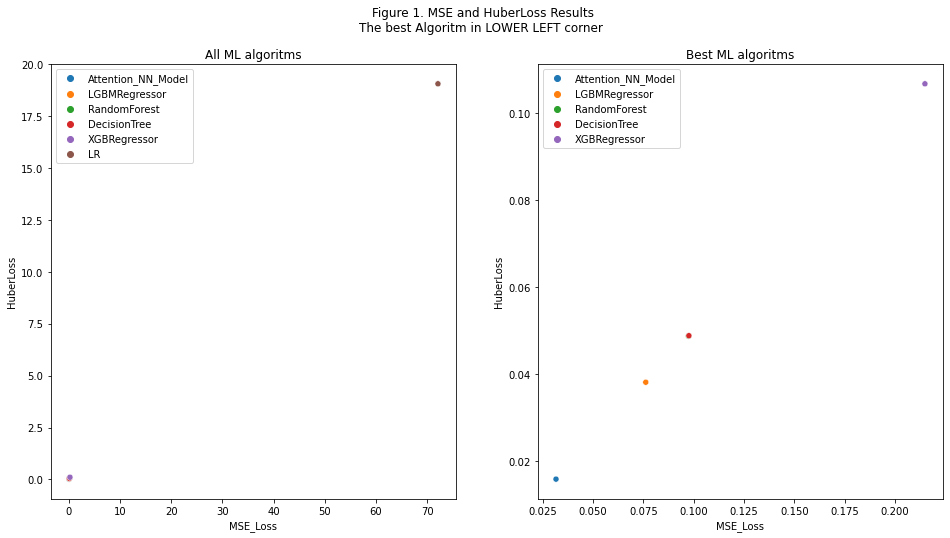
**SmoothL1** = 0.031

Также были обучены:

* Decision Tree Regressor
* RandomForestRegressor
* LinearRegression
* XGBRegressor
* LGBMRegressor

Для всех моделей гиперпараметры считались на Optuna

Сравнительная диаграмма разных алгоритмов **Рис.8**



# Возможный вариант внедрения:

Контроллер регулирующего клапана, установленного ниже по потоку трехфазного сепаратора, в режиме реального времени опрашивает модель, если получает сигнал, о приближении жидкости, то корректирует % открытия регулирующего клапана. Получается что установка не пытается как можно быстрее дренировать пробку ,а уже готова к большой жидкостной пробке и заблаговременно приняла действия. Такой подход снизит время простоя на дренаж, обезопасит от останова вышестоящего по потоку газа дорогого оборудования, что оптимизирует выход полезного (в данном случае – газа) продукта.

# Выводы:

* Была составлена модель подготовки исходных данных, применены методики поиска выбросов для определения начала заполнения сепаратора и выброса газожидкостной пробки
* Robust от шумов модель хорошо показала на различных сценариях
* Проводился сравнительный анализ с другими моделями и ансамблями моделей Machine Learning ( Decision Tree Regressor, RandomForestRegressor, XGBRegressor, LGBMRegressor)
* NN модель не использует абсолютные данные, значит устойчива для разных начальных значений и сценариев
* Модель может быть масштабируема на количество трубопроводов

Сравнительная диаграмма разных алгоритмов (Рис.8)

# Пути развития:

* Остался открытым вопрос изменения в соотношении состава скважинного продукта (содержания воды) и начальных давлений близких к критическим.
* применить подход для других установок, например, для улавливания пробок между скважинами и кустовой площадкой
* применить похожую методику для прогнозирования соотношения дистиллята (продукта верха ректификационной колонны) в зависимости от входных данных (подход будет уникальным так как зависит от конкретной технологической схемы, исходного продукта, количества тарелок, холодильника, теплообменника, количеством фракций дистиллята)