

# Отчет по проекту "Накопленная добыча при закачке смешивающегося углеводородного газа в пласт"

## 1 Постановка задачи

Закачка смешивающегося углеводородного газа в пласт – это один из газовых методов увеличения нефтеотдачи. Обычно процессы 3D фильтрации моделируются в гидродинамических симуляторах (tNavigator, Eclipse, STARS, INTERSECT и прочие). Но поскольку такое полномасштабное моделирование требует большого количества ресурсов как денежных (стоимость лицензии на ПО), так и временных (детальная модель может рассчитываться несколько дней), появляется необходимость построить упрощенной модели, которая будет оценивать суммарную добычу в более короткие сроки.

Будем рассматривать прямоугольный пласт, а именно одну четверть пятиточечной схемы заводнения, при которой нагнетающая скважина находится в центре, а добывающие скважины — в вершинах. Схема пласта представлена на рисунке 1.

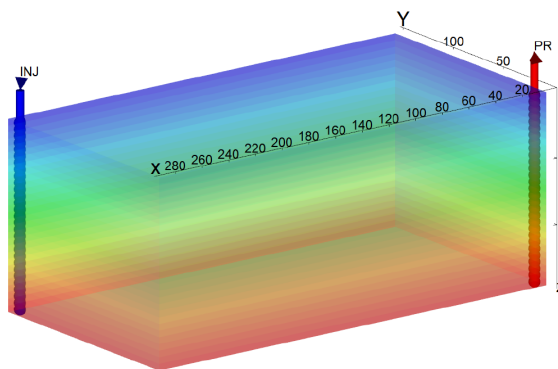


Рис. 1: Схема пласта

Пласт состоит из пяти слоев, и для каждого слоя задаются своя пористость  $poro_i$  и проницаемость  $perm_i$ . Предполагается, что остаточная нефтенасыщенность в пласте равна 0.07.

Таким образом, имеем задачу линейной регрессии с несколькими переменными: необходимо построить алгоритм машинного обучения, который будет предсказывать накопленную добычу описанного пласта после 15 лет разработки по геометрии пласта, пластовому давлению, давлению насыщения, забойному давлению и температуре пласта.

## 2 Описание датасета

Построение датасета осуществлялось в симуляторе tNavigator с использованием композиционной модели. Композиционный состав нефти в пласте и газа в потоке нагнетания фиксировался для всех кейсов. Также фиксировались такие свойства, как коэффициент анизотропии, коэффициенты объемного расширения нефти и воды, плотности нефти и воды, вязкость воды, сжимаемости породы и воды, газовый фактор,

относительные фазовые проницаемости для систем вода-нефть и нефть-газ, проницаемость и пористость для каждого слоя пласта.

В качестве характеристик мы взяли давление насыщения, отношение начального давления к давлению насыщения, температуру пласта, давление на забое и расстояние между скважинами. Целевая переменная – накопленная добычка за 15 лет разработки. Всего было рассчитано 200 кейсов, 120 из которых приходится на обучающий набор, 40 – на тестовый и остальные 40 – на валидационный.

Распределения входных и выходных данных представлены в таблице 1.

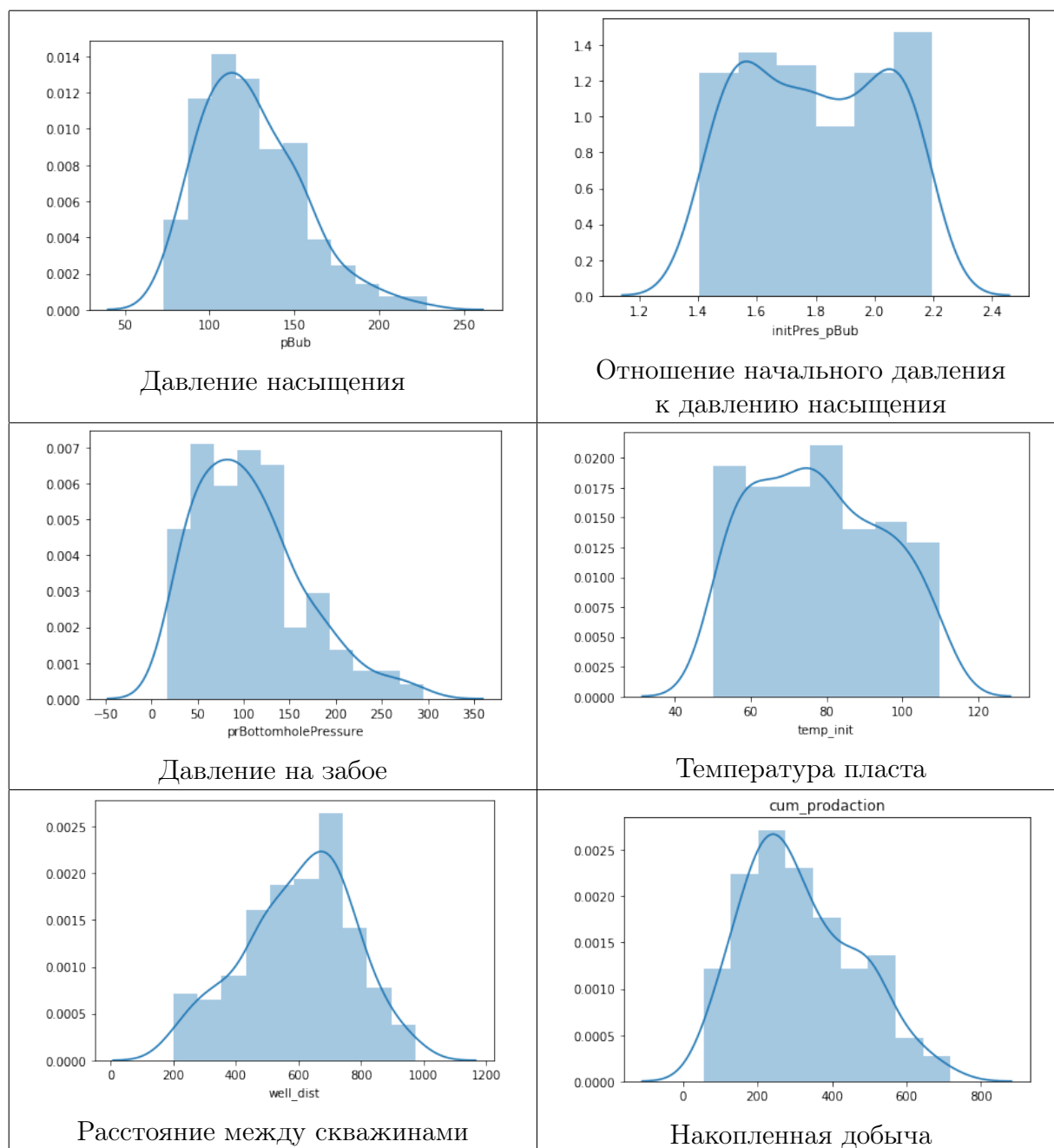


Таблица 1: Распределение входных и выходных данных

Матрица корреляции представлена на рисунке 2. Из нее видно, что суммарная

добыча зависит в основном от расстояния между скважинами, которое соответствует количеству начальных запасов.

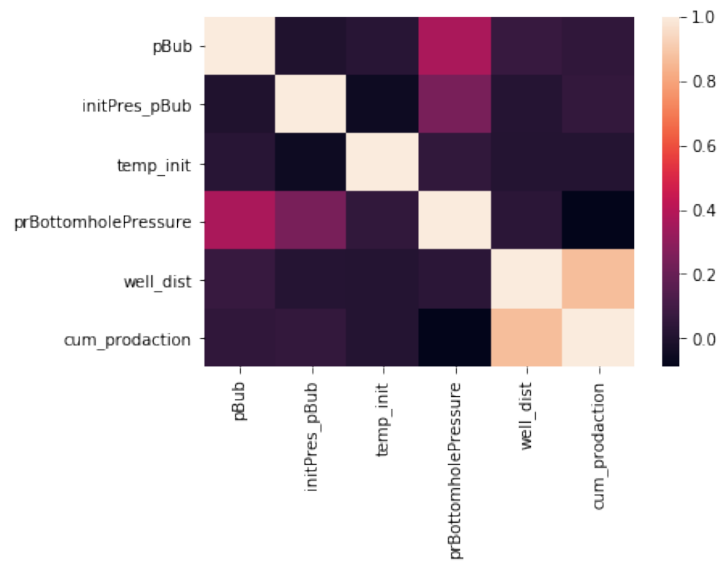


Рис. 2: Распределение накопленной добычи

### 3 Описание подхода

Для решения поставленной задачи мы применили 4 разных алгоритма: линейную регрессию LinearRegression, линейную регрессию с регуляризацией LassoRegression, градиентный бустинг на деревьях XGBoost и метод опорных векторов SVR.

### 4 Эксперименты и результаты

В таблицах 2 и 3 представлены результаты численных экспериментов. Можно увидеть, что алгоритм XGBRegressor показывает лучший результат.

	Linear Regression	Lasso regression	XGBRegressor	SVR
valid_set	0.076	0.154	0.075	0.080
test_set	0.099	0.207	0.088	0.080

Таблица 2: Средняя абсолютная ошибка

	Linear Regression	Lasso regression	XGBRegressor	SVR
valid_set	0.367	1.019	0.257	0.268
test_set	0.295	0.668	0.268	0.308

Таблица 3: Средняя относительная ошибка

В таблице 4 представлена зависимость средней абсолютной ошибки от коэффициента регуляризации для алгоритма Lasso Regression. Видно, что с уменьшением коэффициента регуляризации ошибка уменьшается.

коэффициент регуляризации	0.01	0.1	1
valid_set	0.072	0.154	0.154
test_set	0.111	0.207	0.154

Таблица 4: Средняя абсолютная ошибка для алгоритма Lasso Regression с различными коэффициентами регуляризации.

На рисунке 3 представлена диаграмма с R2 метриками, отображающими точность разных моделей. Можем увидеть, что алгоритм XGBRegressor демонстрирует лучшую точность для данной задачи.

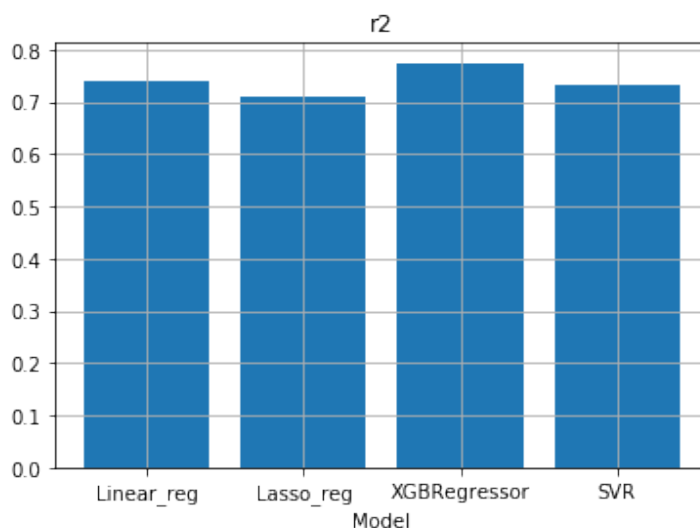


Рис. 3: Точность различных алгоритмов.

На рисунке 4 предстален график зависимости предсказанной добычи от реальной, который показывает, что точки, соответствующие XGBRegressor, лежат ближе к прямой  $x = y$ .

## 5 Заключение

Проанализировав полученные результаты и принимая во внимание небольшой размер датасета, мы можем видеть, что все алгоритмы демонстрируют нормальный результат. Также в ходе численных экспериментов была выявлена сильная зависимость точности алгоритма Lasso Regression от значения коэффициента регуляризации.

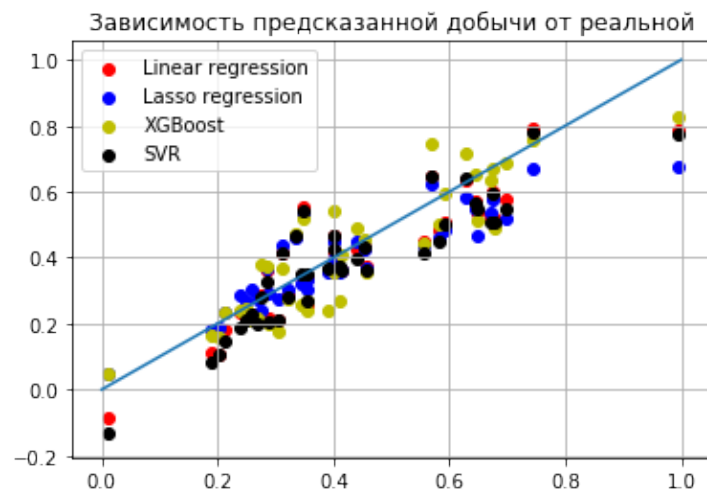


Рис. 4: