

Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа теоретической механики и математической физики

Работа допущена к защите

Директор ВШТМиМФ

д.ф.-м.н., чл.-корр. РАН

_____ А. М. Кривцов

«__» _____ 20__ г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА
ВЫБОР ОПТИМАЛЬНОГО МЕТОДА ИНТЕНСИФИКАЦИИ
ДОБЫЧИ НЕФТИ ПРИ ПОМОЩИ МАШИННОГО ОБУЧЕНИЯ

По направлению подготовки

01.03.03 «Механика и математическое моделирование»

Профиль

01.03.03_03 Математическое моделирование процессов нефтегазодобычи

Выполнил

Студент гр. 5030103/90301

Блащук О.Д.

Руководитель

Доцент ВШТМиМФ, к.т.н.

Курта И.В.

Консультант

Ассистент ВШТМиМФ

Ершов А.Д.

Санкт-Петербург

2023

**САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО**
Физико-механический институт
Высшая школа теоретической механики и математической физики

УТВЕРЖДАЮ

Директор ВШТМиМФ

А.М. Кривцов

«__»_____2023г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студенту Блащук Ольге Дмитриевне, гр. 5030103/90301

1. Тема работы: Выбор оптимального метода интенсификации добычи нефти при помощи машинного обучения.
2. Срок сдачи студентом законченной работы: 05.06.2023
3. Исходные данные по работе: актуальные научные публикации по теме работы, исследования и перечень технических требований к применению методов интенсификации добычи нефти. Набор данных о проводимых геолого-технических мероприятиях и дебите скважин за период их функционирования, геолого-физическая характеристика нефтяных пластов.
4. Содержание работы (перечень подлежащих разработке вопросов): анализ методов интенсификации добычи нефти, составление математической модели, прогнозирующей наиболее подходящий метод интенсификации добычи нефти на участке с определенными геолого-физическими параметрами, оценка эффективности предполагаемого метода.
5. Перечень графического материала (с указанием обязательных чертежей): не предусмотрено.
6. Консультанты по работе: А.Д. Ершов, ассистент ВШТМиМФ
7. Дата выдачи задания 27.02.2023

Руководитель ВКР _____ И.В. Курта, доцент ВШТМиМФ, к.т.н.

Задание принял к исполнению 27.02.2023

Студент _____ О.Д. Блащук

РЕФЕРАТ

На 41 с., 10 рисунков, 3 таблицы

МЕТОДЫ ИНТЕНСИФИКАЦИИ ДОБЫЧИ НЕФТИ, КЛАССИФИКАЦИЯ, МАШИННОЕ ОБУЧЕНИЕ, ГРАДИЕНТНЫЙ БУСТИНГ

В данной работе были изучены методы интенсификации добычи нефти и предложен способ прогнозирования результатов проведения этих методов на скважинах с выбранными геолого-физическими характеристиками. В качестве показателя эффективности метода интенсификации выбрано среднее значение дебита, предварительно нормированного относительно значения дебита до применения метода, фиксируемое в течение года после применения метода. Были определены опорные геолого-физические характеристики пласта, проведена предварительная обработка данных и составлена модель машинного обучения, в основе которой лежит решение задачи классификации алгоритмом градиентного бустинга. После создания модели была предложена методика по ее использованию и оценке качества ее прогноза. Результатом данной работы является модель машинного обучения, позволяющая спрогнозировать наиболее эффективный метод интенсификации добычи нефти для нефтяного пласта по его геолого-физическим характеристикам. Прогноз эффективности метода, полученный с помощью данной модели, может нести рекомендательный характер при составлении планов проведения геолого-технических мероприятий, сократить время исследования и уменьшить риски, возникающие вследствие неправильно подобранного метода.

THE ABSTRACT

41 pages, 10 pictures, 3 tables

OIL STIMULATION METHODS, CLASSIFICATION, MACHINE LEARNING, GRADIENT BOOSTING

In this paper, methods of oil production intensification were studied and a method for predicting the results of these methods on wells with selected geological and physical characteristics was proposed. As an indicator of the effectiveness of the intensification method, the average value of the flow rate, previously normalized relative to the value of the flow rate before the application of the method, fixed for a year after the application of the method, was chosen. The basic geological and physical characteristics of the formation were determined, preliminary data processing was carried out and a machine learning model was compiled, which is based on the solution of the classification problem by the gradient boosting algorithm. After the model was created, a methodology was proposed for its use and evaluation of the quality of its forecast. The result of this work is a machine learning model that allows predicting the most effective method of intensifying oil production for an oil reservoir based on its geological and physical characteristics. The forecast of the effectiveness of the method obtained using this model can be of a recommendatory nature when drawing up plans for geological and technical measures, reduce the time of research and reduce the risks arising from an incorrectly selected method.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1. АНАЛИЗ МЕТОДОВ ИНТЕСИФИКАЦИИ ДОБЫЧИ НЕФТИ.....	8
1.1. Понятие методов интенсификации добычи	8
1.2. Механизмы и особенности методов интенсификации добычи на примере сравнения с методами увеличения нефтеотдачи	9
1.3. Обзор рассматриваемых методов.....	11
ГЛАВА 2. ОПРЕДЕЛЕНИЕ ПОДХОДА ДЛЯ РЕАЛИЗАЦИИ МОДЕЛИ.....	16
2.1. Предпосылки к использованию методов машинного обучения	16
2.2. Определение типа задачи и показателя эффективности метода	18
2.3. Постановка задачи классификации	21
ГЛАВА 3. СОЗДАНИЕ МОДЕЛИ.....	24
3.1. Структура входных данных	24
3.1. Предобработка и анализ данных	28
3.2. Решение задачи классификации с помощью алгоритма градиентного бустинга.....	30
ГЛАВА 4. ПОСТРОЕНИЕ МЕТОДИКИ И ОЦЕНКА РЕЗУЛЬТАТОВ.....	34
4.1. Методика применения моделей машинного обучения	34
4.2. Оценка результатов прогнозирования	35
ЗАКЛЮЧЕНИЕ	37
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	39

ВВЕДЕНИЕ

Во всем мире проявляется тенденция к исчерпанию легко извлекаемых запасов нефти. Согласно Проекту Стратегии развития минерально-сырьевой базы РФ до 2030 г., доля запасов трудноизвлекаемой нефти в России достигает до 65% от общего числа запасов [21]. К таким ресурсам можно отнести нефть, заключенную в сложных низкопроницаемых коллекторах, высоковязкие нефти с низкой подвижностью в пласте, отложения нефти, залегающие на больших глубинах с высокой пластовой температурой и т.д. Если ранее затраты на извлечение остаточных запасов превышали поступления от их реализации и делали дальнейшую эксплуатацию объекта экономически нерентабельной, то в настоящее время разработка нетрадиционных типов месторождений и добыча нефти на зрелых истощенных месторождениях становится все более востребованной. Но без применения нетрадиционных технологий задача по увеличению ежедневной нефтеотдачи за счет вовлечения трудноизвлекаемых запасов в активную разработку затруднена.

В связи с этим в рамках плана по реализации мер по освоению нефтяных месторождений и увеличению объемов добычи нефти в РФ предусматривается стимулирование добычи нефти на месторождениях, содержащих трудноизвлекаемые запасы нефти [22]. Таким образом, важнейшей задачей становится создание условий, подходящих для развертывания неклассических технологических решений, и формирование критериев применимости методов интенсификации добычи нефти на месторождениях, ранее трудных и нерентабельных для разработки.

В силу того, что каждый объект разработки обладает специфическими геолого-физическими и географическими свойствами, подобрать подходящий метод интенсификации добычи нефти становится трудоемкой и затратной задачей. Критерии подбора скважин-кандидатов, а также сами технологические параметры по применению МИД варьируются от месторождения к месторождению и постоянно меняются по мере накопления промыслового опыта. Над согласованием использования метода работает

множество специалистов нефтегазового дела во главе с главным инженером и геологом. Однако ситуации, вследствие которых применение того или иного МИД оказывается неуспешным из-за неправильно подобранной технологии, случаются не редко. Стоит отметить, что в таком случае нежелательным исходом будет считаться не только поломка скважины или неконтролируемый прорыв водоносных или газоносных интервалов, но также падение уровня дебита или вынужденная остановка добывающих скважин [13]. Необходимо учитывать, что такие отклонения от плана по реализации выбранной технологии могут принести компании значительные финансовые потери, вплоть до дальнейших решений о нерентабельности такой скважины. Ввиду всех представленных рисков приобретает значимость задача по оптимизации и грамотной оценке применимости и эффективности предполагаемого МИД на выбранном объекте.

Несмотря на всю сложность текущего согласования применения метода, можно выдвинуть гипотезу, что за счет использования методов машинного обучения, опирающихся на успешный опыт применения технологий, возможно ускорить процедуру исследования и добиться снижения рисков, возникающих вследствие неправильно выбранного метода.

Целью данной работы является создание модели машинного обучения, позволяющей спрогнозировать наиболее эффективный метод интенсификации добычи нефти для пласта по его геолого-физическим характеристикам на основе опыта по применению методов на других скважинах.

Результатом работы модели будет являться оценка прироста дебита после применения различных МИД, которую можно использовать для обоснования выбираемого метода на новом объекте. Предполагается, что данная модель и предоставляемая ею оценка будет носить рекомендательный характер при выборе МИД, и при учетывании ее результатов позволит сократить временные ресурсы специалистов.

Вследствие обозначения цели работы можно выделить основные задачи, которые решаются в данной работе:

1. Изучение представленных методов интенсификации добычи нефти и аспектов их применимости;
2. Выбор определяющих геолого-физических параметров пласта для построения модели машинного обучения и решения поставленной задачи;
3. Построение модели машинного обучения и решение задачи классификации;
4. Проведение анализа результатов модели и оценка эффективности предлагаемого ею метода.

ГЛАВА 1. АНАЛИЗ МЕТОДОВ ИНТЕНСИФИКАЦИИ ДОБЫЧИ НЕФТИ

1.1 Понятие методов интенсификации добычи

Основная задача при разработке нефтяных месторождений заключается в максимально возможном извлечении природных запасов нефти из недр земли. Повышение конечной нефтеотдачи разрабатываемых залежей и увеличение темпов отбора нефти в значительной степени достигаются за счет массового внедрения технологий интенсификации добычи нефти.

Методы интенсификации добычи нефти (МИД) – это методы, направленные на сокращение сроков разработки и эксплуатации нефтяных залежей и извлечение наибольших объемов нефти из пласта. Их предназначением является улучшение фильтрационно-емкостных свойств призабойной зоны скважины и повышение проницаемости коллектора.

Принято выделять несколько типов классификации методов интенсификации добычи нефти: по механизмам воздействия на пласт, по составу используемых жидкостей и использованию оборудованию и т.д. В данной работе выбрана к представлению одна из таких классификаций, основанная на технологии воздействия на пласт, она представлена на рис.1.



Рис 1. Классификация методов интенсификации добычи

1.2 Механизмы и особенности методов интенсификации добычи на примере сравнения с методами увеличения нефтеотдачи

В данной работе предлагается ознакомиться с механизмами и особенностями группы методов интенсификации добычи в ходе их сравнения с другой группой методов воздействия на пласт.

Согласно большинству учебных материалов [3,16], принято разделять методы интенсификации добычи нефти и третичные методы нефтеотдачи (МУН). Несмотря на одинаковые применяемые реагенты и схожие названия некоторых технологий, эти группы методов кардинально отличаются как с точки зрения процесса их реализации, так и механизмов воздействия на пластовую систему, а значит, и достигаемого эффекта. Краткое изложение различий в группах методов приведены в таблице 1.

Таблица 1

Сравнение групп методов МИД и МУН

Критерии сравнения	Методы интенсификации добычи нефти	Методы увеличения нефтеотдачи
Область воздействия	Призабойная зона	Все межскважинное пространство
Объемы затрачиваемых реагентов и стоимость	Низкие	Высокие
Длительность реализации	Низкая	Высокая
Ожидаемый эффект	Увеличение продуктивности скважины и ее дебита	Увеличение коэффициента извлечения нефти

При реализации МИД целевой областью воздействия является околоскважинная зона пласта, размеры которой составляют до нескольких метров в зависимости от геолого-физических характеристик коллектора. В свою очередь МУН оказывают воздействие на всё межскважинное пространство, масштаб расстояния которого составляет сотни метров, то есть воздействие происходит на весь нефтенасыщенный пласт. Отсюда следуют и отличия в технологических параметрах реализации. Так, например, для обработки призабойной зоны пласта объём реагентов обычно изменяется в

пределах от нескольких единиц до десятков кубических метров [3], а основное влияние на этот параметр оказывает толщина обрабатываемого интервала пласта. Для воздействия же на весь пласт необходимо закачать объемы реагентов, сравнимые с поровым объемом пласта, а такая величина может принимать значение от нескольких тысяч до сотен тысяч кубических метров [12]. Такая разница в масштабах воздействия приводит к различной длительности процессов. Так, например, обработка призабойной зоны пласта может длиться всего несколько дней, при этом сам процесс закачки реагентов длится не более нескольких часов, а остальное время затрачивается на спускоподъемные и другие подготовительные операции на скважине [6]. Процесс закачки тысяч кубических метров жидкости с целью увеличения нефтеотдачи пласта же может длиться от нескольких недель до нескольких лет [20]. Всё вышесказанное влечёт за собой и отличие в экономических параметрах технологии. Так, стоимость обработки призабойной зоны пласта может составлять от нескольких сотен тысяч до миллионов рублей, в то время как внедрение методов увеличения нефтеотдачи влечёт за собой значительные капитальные и текущие затраты. По этим причинам экономические затраты для реализации технологии методов увеличения нефтеотдачи могут достигать десятков и сотен миллионов рублей.

И наконец, основное отличие групп методов заключается в ожидаемом эффекте от реализации соответствующих технологий. Так, в методах интенсификации добычи нефти воздействие происходит по большей части на призабойную зону пласта и направлено на удаление загрязнений, восстановление и увеличение проницаемости в ней. Эффективность таких технологий будет заключаться в увеличении коэффициента продуктивности скважины, а значит ее дебита. Таким образом применение МИД в целом приводит в первую очередь к ускорению извлечения запасов, а именно к интенсификации процессов извлечения нефти. В МУН же основной целью является увеличение коэффициента нефтеотдачи.

1.3. Обзор рассматриваемых методов

В данной работе рассматриваются четыре метода интенсификации добычи нефти, три из которых относятся к группе методов гидроразрыва пласта, а четвертый представляет собой кислотную обработку призабойной зоны. Ограниченность числа рассматриваемых методов обусловлена набором используемых в ходе работы входных данных, в которых не зафиксирована информация о применении иных методов на исследуемых скважинах.

Гидравлический разрыв пласта (ГРП) – метод интенсификации добычи, используемый для получения доступа к нефти в плотных формациях. Процесс включает использование воды, песка и химических реагентов под высоким давлением для дробления породы и высвобождения углеводородов. Суть метода заключается в создании в забое скважины давления, превышающего пластовое. Величина этого давления определяется в первую очередь исходя из свойств коллектора, параметров скважины и свойств нагнетаемой жидкости [7,14]. По плоскостям с минимальными напряжениями внутри пласта происходит разрыв, который с продолжением увеличения давления начинает разрастаться, образуя систему трещин. После этого в трещины доставляют раствор с пропантом, который позволяет сохранить трещины раскрытыми даже после снятия избыточного давления. Далее по образовавшейся системе трещин флюид, заключенный в пласте, высвобождается. На рисунке 1 схематично показан процесс ГРП.

Применение метода позволяет не только увеличить проницаемость призабойной зоны, но и увеличить охват пласта и повысить нефтеизвлечение в целом. При успешно проведенном методе возможно добиться прироста дебита на 1–3 тыс. т с одной скважины. Однако при столь внушительных цифрах предполагаемого прироста дебита не стоит не включать во внимания риски, связанные с применением технологии. Вследствие нарушения технологии может произойти снижение продуктивности скважин, а также такие эффекты как неконтролируемое фонтанирование и загрязнение призабойной зоны и пластовых вод. Наиболее частыми причинами повреждения пласта при гидроразрыве являются несовместимость жидкостей,

механические примеси, попадающие в пласт из насосно-компрессорных труб, а также неверное моделирование метода и недостаточный контроль качества работы [9].

Как уже было отмечено, критерии применимости метода варьируются от месторождения к месторождению. На некоторых объектах опорными параметрами принимаются коэффициент проницаемости и толщина нефтенасыщенного пласта, на других – его геолого-физическая структура в целом и особенности сетки расположения скважин [15]. Также определяющими параметрами скважины могут служить тип ее профиля и тот факт, применялась ли на ней ранее процедура ГРП и какой дала результат. Так, к примеру, классическая технология гидроразрыва практически не проводится на горизонтальных скважинах. Стоит учитывать тот факт, что ужесточение или введение дополнительных критериев применимости метода может привести к отсеиванию потенциально подходящих скважин-кандидатов, и в остатке останутся всего единицы.

Многостадийный ГРП – технология последовательного проведения нескольких гидроразрывов пласта вдоль ствола одной скважины, позволяющая увеличить темпы выработки нефти и извлечь наибольший экономический эффект. Многостадийный ГРП включает в себя три этапа: инициацию трещин, их распространение и стадию обратного потока [19]. В результате применения метода в пласте формируются высокопроводящие каналы трещин, по которым флюид поступает в скважину.

Это относительно новая технология, на первых годах применения дававшая на скважинах дебит, в полтора раза превышающий плановые показатели. В настоящее время широко распространено использование компьютерного моделирования и симуляторов для уточнения дизайна технологии. При этом проводящаяся с инженерной стороны оптимизация количества стадий ГРП, их интервала и распределение проппанта во время процедуры позволяет максимизировать прибыль компании еще до этапа проведения. Некоторые исследования говорят о возможности увеличения уровня дебита до 15% за счет применения технологии. Однако следует также

учитывать и негативные последствия, возникающие при неверно соблюденной технологии, самым частым явлением из которых является значительное повышение обводненности скважины.

В отличие от традиционной процедуры гидроразрыва пласта, многостадийный ГРП используют преимущественно на горизонтальных скважинах. Сейчас для применения многостадийного ГРП инженерные расчеты в своем большинстве должны опираться на местоположения скважин по отношению к фронту закачиваемой воды, механических свойств породы, толщину и литологию пласта, а также эффективную нефтенасыщенность и запас пластовой энергии. В целом, верным будет суждение, что многостадийный ГРП имеет более строгие критерии применимости, нежели одинарное применение гидроразрыва пласта.

ГРП с изоляцией водопритока – технология гидроразрыва пласта, предполагающая комплексный подход для изоляции активных подошвенных и/или кровельных вод. Этот способ характеризуется особой последовательностью шагов для реализации: в горизонтальный участок скважины спускается колонна насосно-компрессорных труб (НКТ) с установкой пробки мостовой, разбуриваемой ниже интервала ГРП, затем подъемом НКТ и последующим спуском НКТ с установкой пакера выше интервала участка обработки скважины. Для этого метода предполагается использования геля специфического состава для замещения жидкости глушения, при этом основной концепт метода ГРП не изменяется [8]. Технология метода составлена таким образом, что максимально не снижать проницаемость по углеводородам, но в то же время препятствовать увеличению обводненности скважины.

Как и предыдущие реализации ГРП, этот метод предназначен для повышения дебита скважин, а также продления безводного периода эксплуатации после проведения работ. Но в отличие от описанных выше технологий ГРП и многостадийного ГРП, при прорыве трещины в обводненный горизонт происходит изоляция пластовых вод. Этот эффект достигается за счет специального состава геля, образующего на поверхности

каналов тонкую полимерную пленку, выполняющую роль блокирующего экрана. При этом такая поверхность оказывается гидрофобной, она снижает фильтрационное сопротивление и увеличивает фазовую проницаемость для нефти.

Несмотря на то, что применение метода имеет не такие жесткие ограничения по литологии и расположению пласта относительно водонасыщенных горизонтов, как предыдущие, для реализации данного метода стоит учитывать температуру и диаметр поровых объемов пласта, ввиду содержания полимерного состава в геле.

Кислотная обработка скважины – один из видов интенсификации добычи нефти, при котором кислота закачивается в пласт ниже давления разрыва пласта, в целях увеличения проницаемости породы. В основе механизма метода лежит реакция кислотного раствора с породой нефтенасыщенного коллектора и твердыми веществами минерального состава. За счет этого кислотная обработка показывает свою эффективность при наличии твердых частиц, фильтратов бурового раствора или жидкостей глушения, оказывающих негативное влияние на продуктивность добывающих скважин, поскольку несет в себе дополнительную функцию очистки призабойной зоны от загрязнений.

Максимальный эффект метода может также достигаться при комбинировании его с другими физико-химическими методами. С повышением обводненности добываемого флюида рекомендуется применять кислотные составы с различными композициями поверхностно-активных веществ.

Применимость и эффективность кислотного метода интенсификации добычи зависит от множества факторов, включающих в себя степень обводненности продукции, фильтрационно-емкостные свойства коллектора, а также уровень нефтенасыщенности и толщину нефтенасыщенного пласта [17]. Как и в случае методов гидроразрыва пласта, необходимо дополнительно учитывать количество примененных ранее кислотных обработок, а также

временной интервал между ними для контроля за концентрацией кислотного содержания в пласте и флюиде.

Далее в представленной работе будет определяться эффективность рассмотренных выше четырех методов для разных наборов геолого-физических характеристик пласта.

ГЛАВА 2. ОПРЕДЕЛЕНИЕ ПОДХОДА ДЛЯ РЕАЛИЗАЦИИ МОДЕЛИ

2.1. Предпосылки к использованию методов машинного обучения

В современном мире наблюдается тенденция к цифровизации и оптимизации как бизнес-процессов, так и производственных мероприятий, а повсеместная информатизация приводит к накоплению огромных объемов данных в науке. Одним из способов, позволяющих проводить анализ и обработку больших объемов данных, а также автоматизировать большую часть операций, является использование методов машинного обучения в ходе решения производственных задач [11].

Машинное обучение – это раздел теории искусственного интеллекта, предметом которого является поиск методов решения задач путем обучения в процессе решения сходных задач. В основе таких методов лежат элементы математической статистики, численных методов, методов оптимизации, теории вероятности и других технических наук. Алгоритмы машинного обучения тренируются на поданных им на вход данных, вычисляя их главные особенности и находя между ними скрытую от ручного анализа корреляцию. Таким образом, машинное обучение позволяет анализировать данные, запоминать информацию, строить прогнозы, воспроизводить готовые модели, а также выбирать наиболее подходящий вариант из предложенных.

Машинное обучение принято разделять на две основные категории:

- Обучение с учителем — характеризуется тем, что модель обучается на основе помеченных данных, где каждый пример данных имеет соответствующую метку. Метка представляет собой правильный ответ для данного примера.
- Обучение без учителя — это задача машинного обучения, при которой модель обучается на основе данных, где каждый пример не имеет метки. Задачи обучения без учителя включают в себя кластеризацию, снижение размерности и обнаружение аномалий. Задачи обучения с учителем включают в себя классификацию и регрессию.

Суть данной работы заключается в прогнозировании наиболее эффективного МИД для выбранных геолого-физических характеристик пласта на основе опыта его применения на других скважинах. Это означает, что создаваемая модель будет делать прогноз на основании имеющихся данных об эффективности или не эффективности того или иного метода, примененного на скважинах с определенным набором параметров. В этом случае примером для модели выступает информация о скважине, а в качестве метки – успешность/неуспешность примененного МИД. Таким образом, исходя из постановки задачи, в данной работе для реализации модели выбран тип обучения с учителем.

При обучении с учителем задается X – множество описаний объектов, Y – множество допустимых объектов и существует неизвестная целевая зависимость f , являющаяся отображением множества X на Y , иначе говоря:

$$X, Y \exists f, f: X \rightarrow Y.$$

Значения целевой функции $y_i = f(x_i)$, известны только на конечном числе объектов из обучающей выборки $x_i \subset X, i = 1 \dots m$. Основной задачей является построение такого алгоритма α :

$$\alpha: X \rightarrow Y,$$

который приближал бы неизвестную целевую зависимость как на элементах выборки, так и на всем множестве X . Для измерения точности ответов вводят определенно заданный функционал качества.

В зависимости от направленности задачи машинное обучение с учителем, в свою очередь, разделяются на два типа:

- Регрессия – тип обучения, при котором обучающая выборка x_i характеризуется вектором вещественнозначных признаков. Исход y_i также принадлежит множеству вещественных чисел. Иными словами, в регрессии целевые данные являются непрерывными числами, то есть на выходе получается численное значение. При этом множество значений y_i не обязано быть конечным.

- Классификация – тип обучения, при котором множество объектов разделяется на классы на основе некоторого качественного свойства. Классификация объекта означает указание номера или наименования класса, к которому он относится. При этом множество возможных значений y_i конечно и определяется количеством заданных классов.

2.2. Определение типа задачи и показателя эффективности метода

Для определения типа задачи, которую необходимо решать, требуется провести анализ данных, на основании которого делать выбор в пользу одного из типов. В ходе работы требуется определить наиболее эффективный метод интенсификации добычи, при этом число методов интенсификации добычи, используемых при решении задачи, как было сказано в Главе 1, ограничено.

В связи с тем, что количество МИД ограничено, а постановка задачи требует выбора одного наиболее эффективного метода, было принято решение взять за основу задачу классификации: для каждого набора геолого-физических параметров пласта требуется определить, к какому классу относится каждый МИД на этой скважине: к типу эффективных или не эффективных и с какой вероятностью (рис.2).

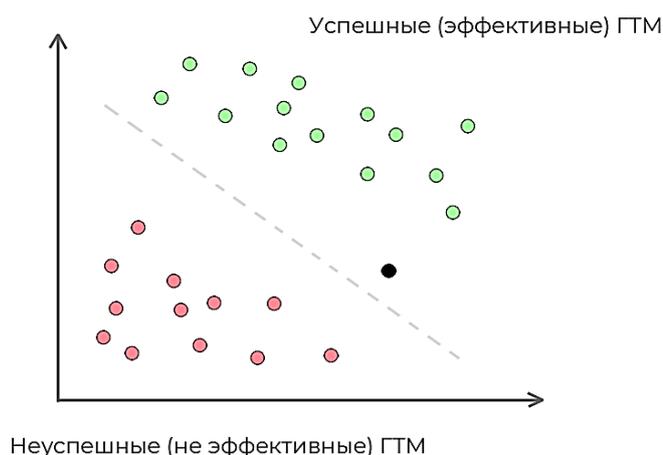


Рис.2. Визуализация задачи классификации в рамках постановки

Внутри каждой нефтедобывающей компании определяются свои критерии эффективности метода: зачастую, они основаны на проектных планах и отталкиваются от рентабельности проводимых геолого-технических

мероприятий. В связи с этим встает задача по определению унифицированного понятия эффективности метода.

Ясно, что динамика дебита скважины меняется после применения метода интенсификации добычи. Положительной динамикой в таком случае справедливо назвать в общем случае возрастающую кривую дебита, а отрицательной – убывающую кривую, которую можно интерпретировать как явное снижения уровня добычи на скважине с течением времени, даже несмотря на проведенные геолого-физические мероприятия. Понятно, что строгого возрастания кривой дебита ожидать нельзя, как минимум ввиду конечности доступных для извлечения запасов, которые со временем исчерпываются. Исследовать аналитически кривые на предмет возрастания или убывания тоже не представляется возможным, в виду того, что исследуемое значение дебита дискретно и фиксируется в лишь в определенные моменты времени. Однако визуально можно определить общую тенденцию кривых. Особенно выражено выделяются две категории кривых дебита, представленные на рисунка 3 и 4. Предварительно данные по оси ординат были отнормированы относительно первого значения.

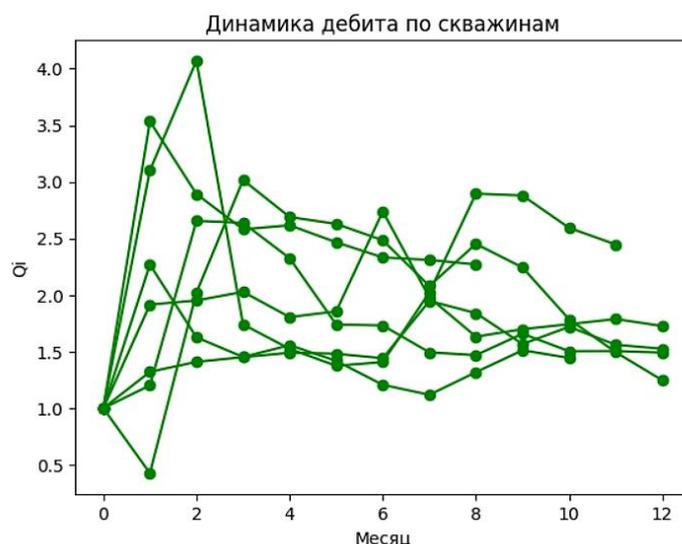


Рис.3. Положительная тенденция кривых дебита после применения МИД

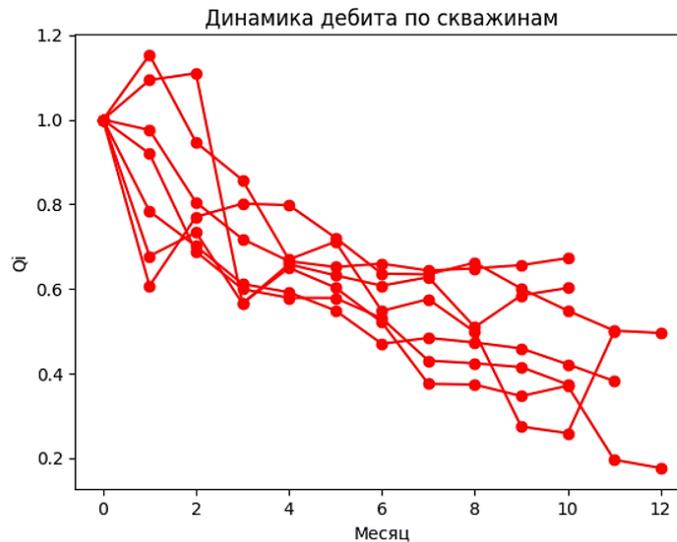


Рис.4. Отрицательная тенденция кривых дебита после применения МИД

В данной работе эффективным выбрано считать метод интенсификации добычи, после применения которого среднее значение дебита, рассчитанное в течение года спустя применения МИД, не упадет ниже значения дебита, зафиксированного на данной скважине за месяц до применения метода. Также с целью унификации данных все значения дебита, рассчитанные за период наблюдения, нормированы относительно значения дебита за месяц до применения метода. За счет этой операции была проведено своеобразное масштабирование признаков [5]. Тогда значение дебита до применения метода Q_0 принимается равным единице, а сам критерий эффективности будет выглядеть следующим образом:

$$\text{МИД} \rightarrow \begin{cases} \text{Эффективен,} & \text{if } Q_{\text{средн}} > (Q_0 = 1) \\ \text{Не эффективен,} & \text{if } Q_{\text{средн}} \leq (Q_0 = 1) \end{cases}$$

где $Q_{\text{средн}}$ — среднее значение дебита, рассчитанное за год спустя применения метода, определяемое по формуле:

$$Q_{\text{средн}} = \frac{\sum_{i=1}^{12} Q_i}{12},$$

где Q_i — значения дебита со скважины, зафиксированные в каждый месяц года спустя применения МИД, отнормированные относительно Q_0 . Иначе

говоря, количественное значение добытой нефти за месяцы, нормированные относительно значения Q_0 .

Таким образом, вычисляя среднее значение нормированного дебита и сравнивая его с единицей, можно отнести МИД, примененные на скважине, к классу эффективных на этом наборе геолого-физических параметров или классу неэффективных. Так, поставленная задача определения оптимального МИД сводится к решению задачи бинарной классификации. Для наглядности принцип определения эффективного метода интенсификации добычи представлен на рисунке 5.

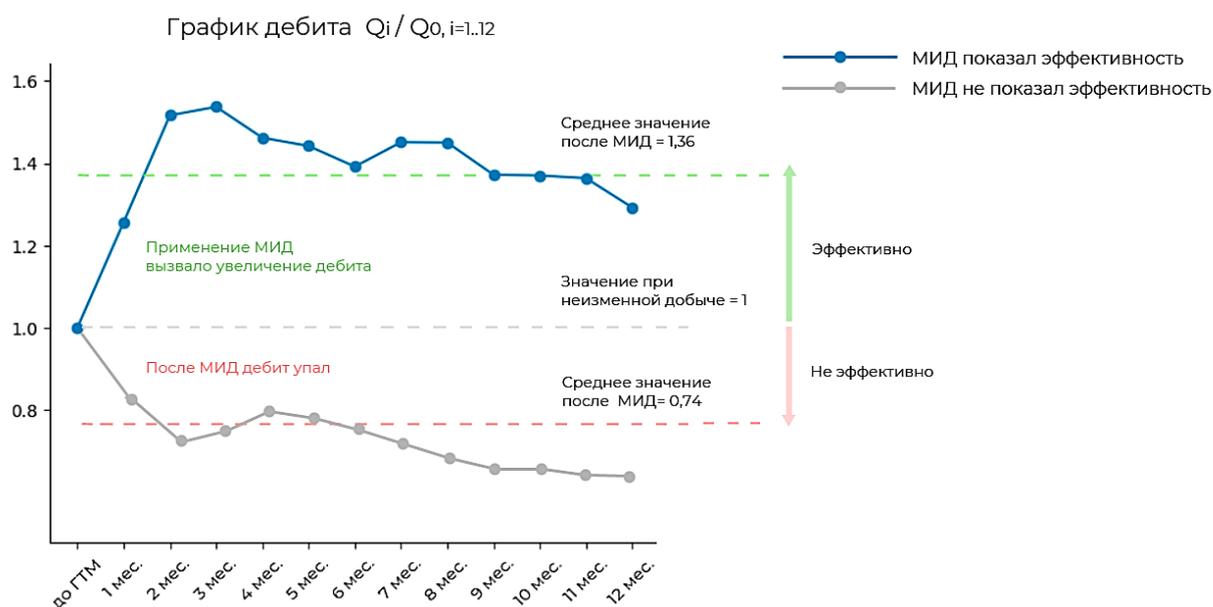


Рис.5. Визуализация принятого критерия эффективности МИД

2.3. Постановка задачи классификации

Как уже было отмечено ранее, в ходе данной работы будет решаться задача бинарной классификации. Данный алгоритм в начале вычисляет вероятность того, что наблюдение будет отнесено к первому классу, а затем определяет, к какому из двух классов отнести это наблюдение на основе того, пересекла ли вероятность пороговое значение. По умолчанию, пороговое значение равно 0,5.

Рассматривается обучающая выборка $X_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$, где каждый набор данных содержит вещественнозначные значения:

$$(x_i, y_i) \in X^m \times Y, \quad i = 1..m,$$

и где каждому набору данных вектора X , отвечающему за информацию о геолого-физических параметрах пласта, соответствует значение Y , заданное одним из двух возможных целых чисел:

$$Y = \{0, 1\}.$$

В этой постановке предлагается обозначать эффективно примененный метод за единицу, в противном случае за ноль.

Полагается, что все объекты независимы и взяты из некоторого неизвестного распределения $(x_i, y_i) \in P(x, y)$, $i = 1..m$. Иными словами, что информация с каждой скважины не зависит от другой.

Поэтому требуется определить, во-первых, вероятность отнесения метода к одному из двух классов, а во-вторых, сделать это наиболее достоверно. Для оценки качества модели вводится понятия функционала потерь.

Функционал потерь указывает, насколько близка вероятность предсказания к соответствующему истинному значению (0 или 1 в случае двоичной классификации). Чем больше прогнозируемая вероятность отклоняется от фактического значения, тем выше значение функции потерь. Таким образом, требуется с помощью алгоритма найти такое распределение исходов, чтобы вероятности соответствующих исходов были максимальны, а выбранный функционал потерь – минимален.

В качестве функционала потерь в работе выбран функционал бинарной поперечной энтропии LogLoss . Он измеряет расхождение между фактическими метками классов и предсказанными вероятностями принадлежности к классу 1. Чем меньше значение LogLoss , тем лучше качество модели.

Отсюда можно основной принцип работы алгоритма:

$p_i = \alpha(x_i, a)$ – результат работы алгоритма α на i -ом наборе данных с набором признаков a ,

$$P(y|X, a) = \prod_i p(y_i | x_i, a) \rightarrow \max$$

$$\text{LogLoss} = -\frac{1}{N} \sum_{i=1}^n (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \rightarrow \min$$

Предлагается, что геолого-физические свойства пласта могут описываться как количественными, так и качественными признаками. Обучающая выборка составляется из прецедентов двух классов: параметров пласта и скважины, на которых МИД был успешен и тех, на которых не был успешен. Данный подход к решению задачи заключается в поиске закономерностей в массиве данных, который содержит меньше объектов, чем признаков. Классические статистические методы могут плохо отработать в этой ситуации. В процессе решения задачи выделяются короткие наборы признаков, которые наилучшим образом разделяют классы и обладают наибольшей информативностью. Это является важным побочным результатом исследования, который представляет значительный интерес для геофизиков и геологов.

ГЛАВА 3. СОЗДАНИЕ МОДЕЛИ

3.1. Структура входных данных

Структура имеющихся входных данных для исследования состоит из трех таблиц, содержащих следующую информацию:

- Данные о дебите нефти Q со скважин по месяцам;
- Параметры пласта и тип профиля для каждой из скважины;
- Даты применения геолого-технических мероприятий, проводимых на скважине и их длительность.

Из всех проводимых на скважинах мероприятий требуется отобрать методы интенсификации добычи. Как уже упоминалось ранее, в данных зафиксирована информация только о четырех таких методах: ГРП, многостадийный ГРП, ГРП с изоляцией водопритока и кислотная обработка.

По постановке задачи модель не учитывает эффект от последовательного применения двух МИД, а значит для корректности решаемой задачи выбирается информация лишь о тех скважинах, дебит которых не находится под эффектом примененного ранее метода интенсификации добычи.

В качестве дополнительного параметра скважины, оказывающего непосредственное влияние на выбираемый метод, используется тип ее профиля. В исходных данных он представлен следующими категориями:

- Вертикальные скважины (Vertical),
- Наклонно-направленные скважины (S-shaped, J-shaped, S/J-shaped),
- Горизонтальные скважины (Horizontal, 80 deg Horizontal).

Вертикальные скважины – скважины, имеющие вертикальный ствол, который просверливается от поверхности земли до глубины, залегания продуктивного пласта.

Наклонно-направленные скважины – скважины, для которых проектом предусматривается определенное отклонение забоя от вертикали, а ствол проводится по заранее заданной траектории. Скважины с наклоном могут

иметь различные типы профилей, в зависимости от глубины скважины и особенностей геологического разреза [10]. Два из таких профилей представлены в имеющихся данных и напоминают по форме латинские буквы S и J.

Горизонтальные скважины в свою очередь состоят из направляющей и горизонтальной частей. Под направляющей частью понимается часть ствола от устья до точки с заданными координатами на кровле и в самом продуктивном пласте. Направляющая часть ствола является началом горизонтального участка под заданным углом. Горизонтальные скважины бывают собственно горизонтальными (когда наклонный и горизонтальный участки являются продолжением обычных вертикальных скважин) и боковыми горизонтальными стволами, которые бурятся из стволов ранее пробуренных скважин.

Представленные в работе типы профилей для наглядности представлены на рисунке 6.

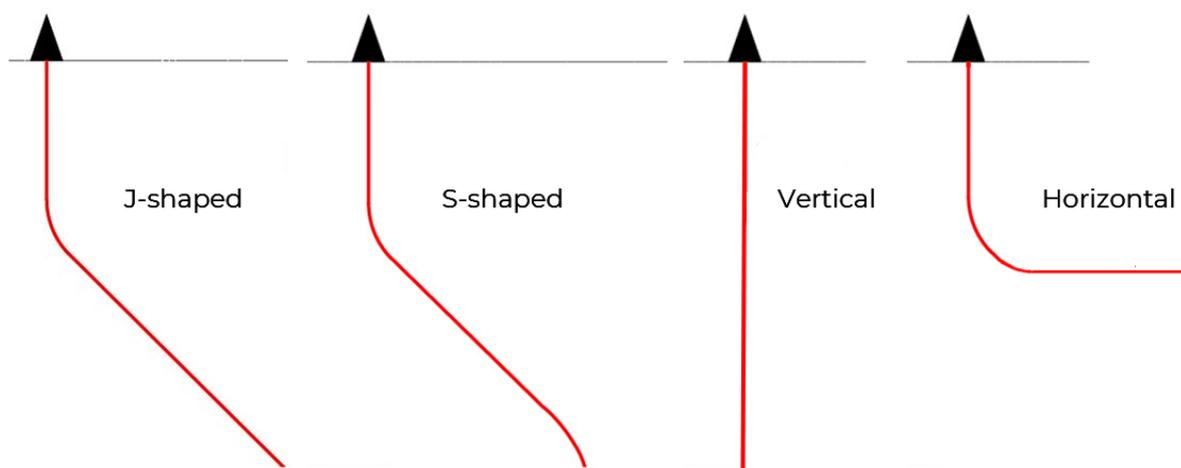


Рис. 6. Типы профилей скважин

Как указывалось в Главе 1, на практике от типа профиля скважины зависит многообразие допустимых для применения на ней методов. Отсюда возникает потребность в обязательном порядке учитывать это при построении модели.

В качестве параметров пласта во входных данных содержатся следующие физические величины:

- Мощность горных пород, м
- Эффективная мощность коллектора, м
- Протяженность пласта, м
- Глубина залегания пласта, м
- Коэффициент пористости пласта, %
- Нефтенасыщенность пласта, %.
- Коэффициент проницаемости пласта, Д

Под мощностью горных пород понимается толщина пласта или комплекса геологических отложений, замеренная по нормали между его кровельной и подошвенной частью [18]. В свою очередь, абсолютная мощность коллектора означает суммарную мощность прослоев коллекторов в пласте. Обе эти величины выражаются в метрах, так же как глубина пласта и его протяженность.

Пористость пласта — это доля объема пор в общем объеме пористого тела. Она характеризует наличие в породе пустот и может быть разделена на общую, открытую и закрытую пористости. Общая пористость — это весь объем пустот в породе, открытая - объем связанных поровых каналов, по которым может фильтроваться жидкость или газ, а закрытая пористость — это объем изолированных пустот. Для количественной характеристики пористости используется коэффициент пористости, который равен отношению объема пустот образца породы к объему всего образца, выраженному в процентах.

Нефтенасыщенность пласта характеризует содержание нефти в коллекторе. Коэффициент нефтенасыщенности может быть определен как отношение объема нефти к суммарному объему порового пространства горной породы в процентах. В процессе разработки месторождения принято разделять среднюю нефтенасыщенность для коллектора, а также нефтенасыщенность в зонах активного дренирования. Изменение этой характеристики пласта во

время проведения геолого-технического мероприятия играет большую роль для контроля за разработкой месторождения. В исходных данных же содержится средний коэффициент нефтенасыщенности к началу применения МИД, исходя лишь из которого прогнозировать с помощью модели изменение этого параметра не представляется возможным.

Проницаемости пласта – способность горных пород пропускать через себя флюиды при наличии перепада давления. Для оценки проницаемости горных пород применяется линейный закон фильтрации Дарси, который был установлен в 1856 году. Он устанавливает зависимость скорости фильтрации жидкости от градиента давления:

$$k = \mu \frac{q_{\phi} L}{\Delta P F}$$

Где k – проницаемость пласта, м²;

μ – динамическая вязкость флюида, Па · с;

q_{ϕ} – объемный расход флюида (иначе говоря, дебит), $\frac{\text{м}^3}{\text{с}}$;

L – длина образца породы, м;

ΔP – перепад давления, Па;

F – площадь фильтрации, м².

Другой принятой в нефтегазовой отрасли единицей измерения проницаемости является Дарси, в честь открывателя вышеуказанного закона фильтрации: $1 \text{ Д} = 1,02 \cdot 10^{-3} \text{ мкм}^2 = 1,02 \cdot 10^{-12} \text{ м}^2$. Отсюда коллекторы, плохо фильтрующие сквозь себя жидкости или газы при незначительном перепаде давления, называют низкопроницаемыми, их количественное значение проницаемости лежит в интервале 0–100 мД. К таким породам можно отнести глинистые сланцы, песчаники с глинистыми пропластками, мергели и т.д. Такие запасы можно отнести к трудноизвлекаемым, и именно для улучшения фильтрационно-емкостных свойств пласта, в том числе проницаемости, применяются такие методы как ГРП и кислотные обработки.

Таким образом, в качестве параметров пласта выступают 7 вещественных характеристик. Тип профиля скважины является категориальным признаком, так же, как и применённый на ней МИД.

3.1. Предобработка и анализ данных

Перед построением модели машинного обучения требуется проводить обязательную предобработку данных. Это необходимо для улучшения качества модели и предотвращения ошибок. Предобработка данных используется для решения перечисленных ниже задач.

- Уменьшение объема данных: агрегирование данных позволяет уменьшить размер таблиц и ускорить процесс обучения модели.
- Обработка пропущенных значений и шума: пропущенные значения и выбросы могут привести к ошибкам в обучении модели.
- Избежание несогласованности и дубликатов: удаление конфликтующих между собой данных или расхождений улучшает качество модели.

Для улучшения выборки вводится следующий порядок действий, позволяющий повысить качество входных данных. Первым этапом является получение скважин, содержащих достаточную информацию о дебите для проверки критерия эффективности:

1. Из рассмотрения убираются скважины, имеющие большие простои после применения МИД, то есть фиксированное значение дебита на них больше месяца равнялось нулю.
2. Из рассмотрения убираются скважины, значение дебита по которым фиксировалось меньше, чем спустя 8 месяцев после применения МИД. Полагается, что данные с таких скважин не могут дать оценку эффективности метода по выбранному ранее критерию;
3. Данные о дебите со скважин, в которых фиксировался дебит в течение лишь 8–11 месяцев после применения МИД (то есть,

содержащие от 8 до 11 значений), аппроксимируются до 12 значений дебита.

На всех оставшихся наборах данных возможно достоверно проверить критерий успешности.

Далее для всех подходящих скважин собирается информация о геолого-физических параметрах пласта. Строки, содержащие пропущенные значения какого-либо параметра, удаляются из выборки.

Таблица с данными, содержащая 176 примеров проведения ГРП, 28 примеров проведения многостадийного ГРП, 71 пример проведения ГРП с изоляцией водопритока и 41 пример использования кислотной обработки, представлена на рисунке 7.

Профиль	Глубина	Нефтенас.	Протяж.	Абс.мощн.пласта	Абс.мощн.колл	Пор-ть	Проницаемость	Эффект.	ГТМ
Horizontal	3367.446667	0.549000	2949.590000	19.190000	16.190000	0.152000	3.795667	1	Многостадийный ГРП
J-Shaped	2342.713333	0.111333	2337.025333	13.228000	6.349333	0.153333	4.924000	1	Кислотная обработка
S-Shaped	2787.620833	0.252083	2775.627917	20.345417	6.771250	0.114583	18.232917	1	ГРП
J-Shaped	2844.211765	0.238824	2832.638235	14.812941	3.815294	0.122941	19.713529	0	ГРП с изоляцией
S-Shaped	2719.909524	0.306190	2710.951905	14.859048	5.104286	0.123810	41.005714	0	ГРП
...
S-Shaped	2379.241667	0.175000	2372.752500	14.376667	5.636667	0.132500	15.282500	0	ГРП с изоляцией
S-Shaped	2854.925000	0.171667	2845.235417	15.458333	3.561250	0.097500	16.137917	1	ГРП
Horizontal	3217.630000	0.563000	2907.090000	9.090000	6.200000	0.164000	6.445000	1	Многостадийный ГРП
S-Shaped	2844.523529	0.234706	2836.411765	14.107647	5.812353	0.124706	26.157059	1	ГРП
S-Shaped	2430.370000	0.410500	2421.310000	16.480500	8.538000	0.120000	27.002500	1	ГРП с изоляцией

Рис.7. Таблица с обработанными данными

Разброс параметров пласта, на которых будет обучаться модель, влияет на состоятельность прогноза для вводимых тренировочных параметров. Стоит отметить, что алгоритм для предобработки данных, а также сам используемый метод машинного обучения не будет давать негативный вклад в точность прогнозирования. Однако при большом различии в параметрах тренировочной и тестовой выборок модель может не распознать, к какому классу относится нововведенный разброс чисел. Отсюда возникает необходимость провести анализ диапазона значений, на которых модель будет давать наиболее точный прогноз без необходимости пополнения обучающей выборки. Результаты анализа отражены в таблице 2.

Таблица 2

Статистическая информация об используемых моделью данных

	Глубина, м	Протяженность, м	Нефтенасыщенность, %	Абс. мощность пласта, м	Абс. мощность коллектора	Пористость, %	Проницаемость, Д
Среднее значение	2811,7825	2770,4747	0,2595	52,9662	7,0259	0,1280	14,7430
Стандартное отклонение значений	369,2969	318,7801	0,1265	260,9940	3,3182	0.0161	9,3402
25-й процентиль	2517,8259	2506,1669	0,1812	12,0892	4,6150	0,1168	7,7259
Медиана	2801,6091	2785,8347	0,2347	16,3827	6,1545	0,1256	12,388
75-й процентиль	3021,9379	2989,7712	0,2747	20,2228	8,5190	0,1371	21,0786
Минимальное значение	2201,1636	2190,3682	0,1113	1,0900	2,3900	0.0975	1,3467
Максимальное значение	4251,2078	3787,4000	0,8683	2212,4000	19,8632	0,1750	41,0057

После изучения структуры данных и их предобработки можно приступить к описанию алгоритма. Как было указано ранее, постановка работы свелась к решению задачи бинарной классификации.

3.2. Решение задачи классификации с помощью алгоритма градиентного бустинга

Существует множество алгоритмов машинного обучения, направленных на решение задачи классификации. В данной работе выбор был сделан в пользу алгоритма градиентного бустинга.

Градиентный бустинг (Gradient boosting) — это метод машинного обучения, который используется для задач классификации и регрессии. Он

работает по принципу построения ансамбля слабых моделей, как правило, деревьев решений. Эти модели последовательно обучаются на ошибках предыдущих моделей, что позволяет улучшать качество предсказаний. Схема алгоритма изображена на рисунке 8.

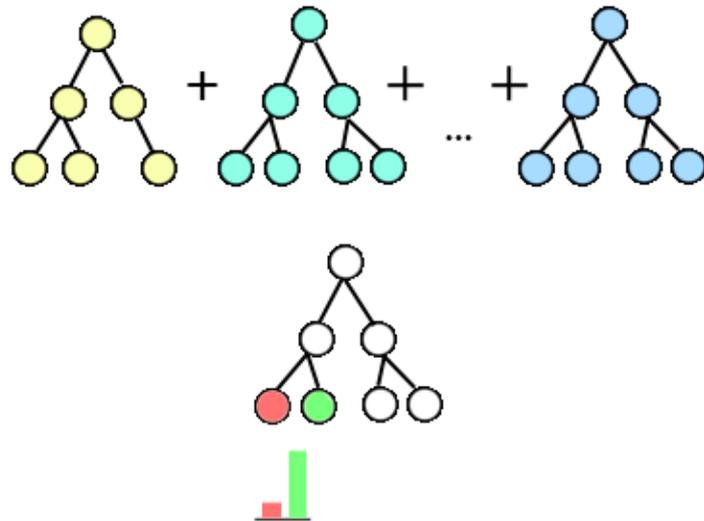


Рис.8. Схема градиентного бустинга

Градиентный бустинг является итеративным алгоритмом, который позволяет приближать оптимальные параметры обучения с каждой итерацией. Это позволяет получать более точные результаты, чем алгоритмы машинного обучения, которые используют одну модель. Он не требует жесткой типизации данных, в отличие от некоторых других алгоритмов, а также хорошо работает на разнородных данных.

Алгоритм градиентного бустинга реализуется следующим образом. Для этого требуется вернуться к определенным ранее выражениям: $\{x_i\}_{i=1}^m \in X$ – обучающая выборка, $y_i \in Y$ – истинные значения меток. Необходимо построить распознающий оператор, который наиболее приближен к предполагаемой зависимости f и точнее всего предсказывает метки для каждого нового объекта $x \in X$.

Пусть задано некоторое семейство базовых алгоритмов H_α , каждый элемент $\alpha(x, a) \in H: X \rightarrow R$ которого задан некоторым вектором параметров $a \in A$.

Требуется найти алгоритм классификации в виде композиции

$$F_N(x) = \sum_{n=1}^N b_n \alpha(x; a_n), \quad b_n \in R, \quad a_n \in A.$$

Выбор оптимального набора параметров $\{a_n, b_n\}_{n=1}^N$ является трудоемкой задачей, отсюда встает необходимость строить такую композицию путем жадного наращивания, иначе говоря, каждый раз добавляя в сумму самый оптимальный алгоритм в качестве слагаемого.

Полагается, что уже построен классификатор F_{n-1} длины $n - 1$. Тогда задача будет сведена к нахождению пары $\{a_n, b_n\}$ для классификатора n -ой длины.

$$F_n(x) = F_{n-1}(x) + b_n \alpha(x; a_n), \quad b_n \in R, \quad a_n \in A.$$

Вводится функция потерь $L(y_i, F_n(x_i)), i = 1..M$ и минимизируется ранее введенный функционал потерь *LogLoss*, обозначенный далее как Q :

$$Q = \sum_{i=1}^M L(y_i, F_n(x_i)) \rightarrow \min$$

Отмечается, что функционал Q – вещественная функция, зависящая от точек $\{F_n(x_i)\}_{i=1}^M$ в пространстве M -ой размерности. Тогда предлагается решать задачу по минимизации Q методом градиентного спуска [2]. Рассмотрим F_{n-1} в качестве точки, для которой будет искаться оптимальное приращение. Берется градиент функционала ошибки:

$$\nabla Q = \left[\frac{\partial Q}{\partial F_{n-1}}(x_i) \right]_{i=1}^M = \left[\frac{\partial (\sum_{i=1}^M L(y_i, F_{n-1}))}{\partial F_{n-1}} \right]_{i=1}^M = \left[\frac{\partial L(y_i, F_{n-1})}{\partial F_{n-1}}(x_i) \right]_{i=1}^M$$

Тогда в силу метода градиентного спуска наиболее выгодно добавить новое слагаемое в классификатор таким образом [4]:

$$F_n = F_{n-1} - b_n \nabla Q, \quad b_n \in R$$

где b_n подбирается линейным поиском по множеству вещественных чисел \mathbb{R} :

$$b_n = \underset{b \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^M L(F_{n-1}(x_i) - b \nabla Q_i)$$

Поскольку ∇Q является не базовым алгоритмом из семейства H_α , определенным при $\forall x \in X$, а лишь вектором оптимальных значений для каждого объекта x_i , требуется найти такой $\alpha(x; a_n) \in H_\alpha$, наиболее схожий с ∇Q [1]. Для этого проводится повторная минимизация функционала ошибки, основанного на принципе явной максимизации отступов:

$$a_n = \underset{a \in A}{\operatorname{argmin}} \sum_{i=1}^M L(\nabla Q, \alpha(x_i, a)) \equiv \text{обучить}(\{x_i\}_{i=1}^M, \{\nabla Q_i\}_{i=1}^M),$$

что и соответствует базовому алгоритму обучения. Затем с помощью линейного поиска находится коэффициент b_n :

$$b_n = \underset{b \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^M L(F_{n-1}(x_i) - b \alpha(x_i, a_n)).$$

Таким образом, значения алгоритма на обучающей выборке подбираются как можно близкими к вычисляемому вектору ∇Q . Так итерационный процесс проходит по всем n начиная с 1 вплоть до N . В конечном счете определяется конечный вид композиции, характеризующий оптимальный алгоритм.

ГЛАВА 4. ПОСТРОЕНИЕ МЕТОДИКИ И ОЦЕНКА РЕЗУЛЬТАТОВ

4.1. Методика применения моделей машинного обучения

Построение алгоритма градиентного бустинга, описанное в предыдущей Главе, позволяет решить задачу классификации и сделать прогноз, будет ли примененный на скважине с заданными параметрами пласта метод интенсификации добычи давать положительный результат. Теперь необходимо переопределить задачу: выбрать, какой из методов дает положительный результат с наибольшей вероятностью. Для этого предлагается следующая методика:

1. Требуется создать и обучить 4 модели (по количеству представленных МИД), решающие задачу классификации алгоритмом градиентного бустинга. Каждая из моделей будет обучаться на наборах данных, соответствующих скважинам, на которых применялся определенный МИД.
2. Объединить все тестовые данные.
3. На каждом наборе данных из объединенной тестовой выборки использовать последовательно все 4 модели. Так, для каждого набора параметров пласта будет предсказана эффективность каждого из 4 МИД с некоторой вероятностью.
4. Из тех МИД, которые по прогнозу окажут положительный эффект на данных геолого-физических параметрах, для соответствующего набора выбирается оптимальным тот метод, который обеспечен наибольшей вероятностью успеха.

Таким образом, результатом данной методики является оценка успешности методов интенсификации добычи, которую можно использовать для обоснования применяемых методов. Пример интерпретации результатов модели показан на рисунках 9 и 10.

	ГРП	Многостадийный ГРП	ГРП с изоляцией водопритока	Кислотная обработка
Успешен с вероятностью	0,94	0,12	0,67	0,75
Не успешен с вероятностью	0,06	0,88	0,33	0,25

Рис.9. Интерпретация результатов методики

На рисунке 9 показано, что при выбранных геолого-физических параметрах пласта лишь по результатам прогнозирования лишь три метода окажутся эффективными. Из них будет выбран тот, чья вероятность оказать положительный эффект будет наибольшая. Сравнение вероятностей успеха трех полученных МИД может быть проиллюстрировано лепестковой диаграммой и представлено на рисунке 10.

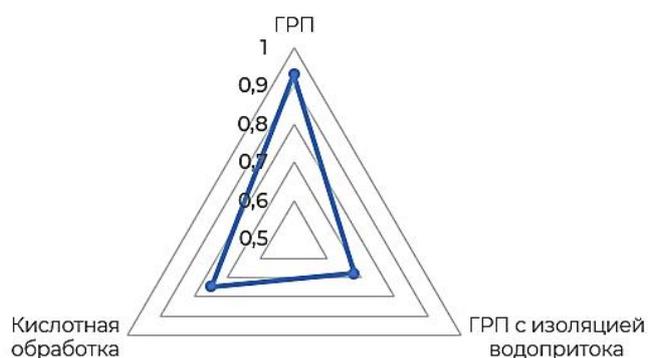


Рис.10. Интерпретация результатов методики

4.2. Оценка результатов прогнозирования

Для валидации результатов прогнозирования необходимо подобрать отдельный алгоритм, поскольку оценивать точность прогноза эффективности МИД, которые в реальности не были применены на соответствующих параметрах пласта (т.е. данных об этом не содержится в исходной информации) представляется сложной задачей.

Для этого предлагается оценивать точность всех четырех прогнозов по прогнозу лишь того МИД, который в реальности был применен этих взятых в тестовую выборку параметрах. Отслеживаемыми метриками качества модели в этом случае могут быть accuracy, precision и recall.

Метрикой *Accuracy* принято обозначать долю правильно классифицированных объектов относительно всех объектов в тестовой выборке. *Accuracy* показывает, насколько хорошо модель работает в целом.

Precision — доля объектов, которые были верно классифицированы как положительные, относительно всех объектов, которые были классифицированы как положительные. *Precision* показывает, насколько точно модель определяет класс объектов.

Recall обозначает долю объектов положительного класса, которые были верно классифицированы, относительно всех объектов положительного класса в тестовой выборке. *Recall* показывает, насколько хорошо модель находит положительные объекты.

Применения неоптимального МИД, как уже было отмечено ранее, может повлечь за собой множество негативных последствий. Поэтому в данной задаче первостепенно требуется предсказать класс положительных объектов (эффективных) с высокой точностью, а значит для в данной постановке приоритетным помимо *Accuracy* будет отслеживание метрики *Precision*. В таблице 3 представлены эти две метрики по результатам тестирования модели.

Таблица 3

Метрики точности модели

	Модель для ГРП	Модель для МС ГРП	Модель для ГРП с изоляцией водопритока	Модель для кислотной обработки	Общая модель
<i>Accuracy</i>	0,71	0,83	0,78	0,89	0,76
<i>Precision</i>	0,73	0,95	0,69	0,80	0,75

По итогам проведения прогноза на имеющихся тестовых данных методика показала высокий результат: точность предсказания достигает 75-76% по обоим выбранным метрикам.

ЗАКЛЮЧЕНИЕ

В ходе данной работы были изучены представленные методы интенсификации добычи нефти и критерии их применимости. Из имеющегося набора входных данных был выбран набор определяющих геолого-физических параметров пласта для построения модели машинного обучения и в последствие решения поставленной задачи.

Далее было введено понятие эффективности метода интенсификации добычи, позволяющее оценивать его успешность на основании изменения дебита вследствие применения метода. Исследована структура входных данных для модели, проведена их предобработка для улучшения качества последующего прогноза.

Проведено построение модели машинного обучения, решающей задачу отнесения метода интенсификации добычи к классу эффективных или неэффективных на выбранном наборе параметров при помощи алгоритма градиентного бустинга. Составлена методика последовательного применения моделей, позволяющая для каждого набора геолого-физических характеристик пласта спрогнозировать эффективность каждого из представленных методов с определенной вероятностью, на основании которой предлагается выбирать оптимальный метод интенсификации добычи.

Проведен анализ результатов прогнозирования сборной модели и оценена точность ее прогноза. Представленная в данной работе методика предсказывает оптимальный метод интенсификации добычи с точностью до 76%, что является хорошим результатом.

Таким образом, можно сделать вывод, что цель выпускной квалификационной работы была выполнена и поставленные задачи решены полностью. А значит полученная модель машинного обучения может нести рекомендательных характер при составлении планов проведения геолого-технических мероприятий и послужить инструментом для выбора оптимального метода интенсификации добычи, сократив время исследования

и уменьшив риски, возникающие вследствие неправильно подобранного метода.

В дальнейшем возможны следующие направления развития данной работы:

1. Улучшение точности моделей за счет включения в данные дополнительных характеристик пласта и условий проведения геолого-технических мероприятий;
2. Пополнение базы используемых данных по новым скважинам с целью расширения диапазона значений обучающей выборки;
3. Развертывание методики на большем числе методов интенсификации добычи и проверка ее результатов на промышленных кейсах.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы // М.: Наука, 1987, с.38
2. Вержбицкий В. М., Основы численных методов // Учебник. М.: Изд-во "Высшая школа", 2008, с.114–115
3. Коршак А.А., Шаммазов А.М. Основы нефтегазового дела: Учебник для вузов. // 3-е изд., испр. и доп., Уфа.: ООО «ДизайнПолиграфСервис», 2005. 231–232 с.
4. Alfio Quarteroni, Riccardo Sacco, Fausto Saleri. Numerical Mathematics// Springer, 2000 – P.155-156.
5. Géron Aurélien, Hands on Machine Learning with Scikit Learn, Keras and Tensorflow Concepts Tools and Techniques to Build Intelligent Systems // O'Reilly Media, 2019 – P. 102–104
6. Johannes Karl Fink. Enhanced Oil Recovery // Petroleum Engineer's Guide to Oil Field Chemicals and Fluids, 2012, – P. 498–499.
7. Баловацкая А. А. Гидроразрыв пласта в горизонтальных скважинах // Вестник науки и образования. 2019. №8–1 (62).
8. Долгушин В. А., Зозуля Г.П., Голофаст С.Л. и др. Способ гидравлического разрыва пласта с изоляцией водопритока в добывающих скважинах // Патент РФ № 2566345С1 от 27.10.2015, ТюмГНГУ, 2015.
9. Кукарский, С. Н. Критерии выбора объекта и основные аспекты применения технологии гидравлического разрыва пласта (ГРП) // Молодой ученый, 2018. № 50 (236), с. 102–104
10. Мищенко И.Т., Бравичева Т.Б., Ермалаев А.И. Выбор способа эксплуатации скважин нефтяных месторождений с трудноизвлекаемыми запасами // Нефть и газ, 2005, с. 440
11. Панов А.И. Введение в методы машинного обучения с подкреплением: учебное пособие // М.: МФТИ, 2019, с. 52

12. Салимов О.В., Насыбуллин А.В., Сахабутдинов Р.З., Салимов В.Г. О критериях подбора скважин для гидроразрыва пласта// Георесурсы. 2017. Т. 19. № 4. ч. 2. с. 368–373.
13. Хайдар А.М., Борисов Г.А., Горин А.Н., Латыпов И.Д. Анализ и классификация причин преждевременных остановок закачки при проведении гидравлического разрыва пластов // Нефтяное хозяйство. 2008. № 11, с. 38–41.
14. Хайруллин М.Х., Гадильшина В.Р., Шамсиев М.Н., и др. Термогидродинамические исследования вертикальных скважин с трещиной гидравлического разрыва пласта // Теплофизика высоких температур. 2017, т.55, в.1, с.129–133
15. Халилов, И. Р. Анализ эффективности гидравлического разрыва пласта на примере месторождения Западной Сибири // Молодой ученый, 2019. № 46 (284), с. 47–49.
16. Грушевенко Е. Перспективы развития третичных МУН в мире и в России// Центр энергетики Московской шк. управления Сколково, 2021., 16 с. – URL: https://energy.skolkovo.ru/downloads/documents/SEneC/Research/SKOLKOVO_EneC_RU_MYN.pdf (дата обращения: 06.05.2023).
17. Кадралиев Т.А. Анализ эффективности мероприятий по интенсификации добычи нефти на примере воздействия многокомпонентным кислотным составом на западно-асмомкинском месторождении // Материалы X Международной студенческой научной конференции «Студенческий научный форум» URL: <https://scienceforum.ru/2018/article/2018007651> (дата обращения: 29.05.2023).
18. neft.academic.ru, Нефтегазовая микроэнциклопедия, Мощность пласта, URL: https://neft.academic.ru/213/Мощность_пласта. (дата обращения: 20.05.2023)
19. Chen, B., Barboza, B.R., Sun, Y. et al. A Review of Hydraulic Fracturing Simulation. Arch Computat Methods Eng 29, 1–58, 2022.

URL: <https://doi.org/10.1007/s11831-021-09653-z> (дата обращения: 14.04.2023)

20. Mohammed Al-Shargabi, Shadfar Davoodi, David A. Wood, Mohsen Ali, Valeriy S. Rukavishnikov, Konstantin M. Minaev. A critical review of self-diverting acid treatments applied to carbonate oil and gas reservoirs // *Petroleum Science*, 2023. URL: <http://earchive.tpu.ru/handle/11683/74895> (дата обращения: 18.05.2023)
21. Стратегии развития минерально-сырьевой базы Российской Федерации до 2030 г. Проект (Версия от 12.09.2016) // Федеральное агентство по недропользованию МПР РФ; ООО «Минерал-Инфо». М., 2016, с. 5 – URL: <https://www.rosnedra.gov.ru/article/8743.html>
22. Стимулирование освоения нефтяных месторождений и увеличения объемов добычи нефти в Российской Федерации. План мероприятий № 598п-П9 (Версия от 25.01.2019) // Министерство энергетики РФ, М., 2019, с. 141 – URL: https://gkz-rf.ru/sites/default/files/docs/o_vnesenii_izmeneniy_v_metodiku_ekspress-ocenki_zapasov_uglevodorodnogo_.pdf