

ТОМСКИЙ
ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ



На правах рукописи



Давуди Шадфар

**ГИБРИДНАЯ ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ДЛЯ ОПЕРАТИВНОГО
ОПРЕДЕЛЕНИЯ СВОЙСТВ БУРОВОГО РАСТВОРА НА ОСНОВЕ
МАШИННОГО ОБУЧЕНИЯ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

2.3.1 – Системный анализ, управление и обработка информации,
статистика

Томск – 2023

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский Томский политехнический университет».

Научный руководитель:

Рукавишников Валерий Сергеевич

PhD, доцент отделения нефтегазового дела
Инженерной школы природных ресурсов Томского
политехнического университета

Официальные оппоненты:

Бурнаев Евгений Владимирович

Доктор физико-математических наук, профессор,
автономная некоммерческая образовательная
организация высшего образования «Сколковский
институт науки и технологий», г. Москва, директор
центра прикладного искусственного интеллекта

Хасанов Марс Магнаевич

Доктор технических наук, профессор, публичное
акционерное общество «Газпром нефть», г. Санкт-
Петербург, директор по науке

Защита состоится 19 марта 2024 года в 16:00 часов на заседание диссертационного совета ДС.ТПУ.15 при ФГАОУ ВО «Национальный исследовательский Томский политехнический университет» по адресу: 634034, г. Томск, ул. Советская, 84/3, ауд. 214.



С диссертацией можно ознакомиться в научно-технической библиотеке Томского политехнического университета и на сайте dis.tpu.ru при помощи QR-кода.

Автореферат разослан «____» _____ 2024 г.

Ученый секретарь
диссертационного совета ДС.ТПУ.15,
доктор технических наук
Ученый секретарь

A handwritten signature in blue ink, consisting of stylized, overlapping strokes, located to the left of the name.

Пак Александр
Яковлевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. При бурении нефтяных и газовых скважин одним из наиболее технически сложных аспектов является работа с буровым раствором. Для повышения эффективности бурения скважин и минимизации потенциальных проблем, связанных с буровым раствором, инженеры по бурению должны определить оптимальную программу промывки скважины и определить компонентный состав раствора для каждой секции скважины. Им также необходимо периодически контролировать и измерять технологические свойства раствора в процессе бурения скважины. Для этого обычно используются экспериментальные методы контроля свойств бурового раствора путем прямого измерения в полевых лабораториях по месту бурения скважин. Такой подход может быть трудоемким и дорогостоящим, но он является необходимым в настоящее время для контроля технологического процесса бурения скважин. Среди всех свойств буровых растворов наиболее важными являются фильтрационные и реологические свойства. В связи с этим тщательное проектирование и приготовление буровых растворов с соответствующими реологическими и фильтрационными свойствами в сочетании с оперативным мониторингом имеют большое значение для успешного выполнения буровых работ. В процессе выполнения всего цикла буровых работ необходимо регулярно получать информацию относительно всех ключевых параметров буровых растворов. Как правило, реологические и фильтрационные свойства бурового раствора проверяются не чаще, чем один или два раза в день. Тем не менее, другие свойства бурового раствора, такие как плотность бурового раствора ($\rho_{бр}$), содержание твердой фазы (ТФ) и условная вязкость (УВ), как правило, контролируются несколько раз в час в силу простоты измерений.

Методы машинного обучения могут служить перспективным инструментом для оценки реологических и фильтрационных свойств бурового раствора на основе других легко и часто измеряемых параметров бурового раствора, таких как $\rho_{бр}$, ТФ и УВ. Метод наименьших квадратов опорных векторов (LSSVM) и метод многослойного экстремального машинного обучения (MELM) - два метода машинного обучения, которые нашли свое успешное применение для решения различных задач регрессии и классификации. Эти методы могут быть использованы для разработки прогнозных моделей для точного определения вышеупомянутых свойств буровых растворов. Однако эти методы имеют свои недостатки, которые негативно влияют на точность прогнозирования и обобщаемость. В этом отношении автономная модель LSSVM страдает от локальной оптимизации, используемой для определения оптимальных гиперпараметров модели, которая склонна к падению до локальных минимумов, а автономная модель MELM требует трудоемкого анализа методом проб и ошибок для определения правильной архитектуры модели и страдает от случайного назначения гиперпараметров. Эти проблемы могут быть решены путем объединения автономных моделей с методами глобальной оптимизации для

определения оптимальной архитектуры модели MELM и оптимальных гиперпараметров для обеих моделей MELM и LSSVM.

Наконец, стоит отметить, что, получение подобной информации о фильтрационных и реологических свойствах растворов в режиме реального времени позволит избежать осложнений, связанных с изменением свойств буровых растворов во время бурения, путем раннего выявления и своевременного принятия необходимых корректирующих действий. Решение данной задачи также приближает нас к достижению полной автоматизации прогнозирования реологических и фильтрационных свойств буровых растворов непосредственно в полевых условиях.

Степень разработанности темы. Данную тематику изучали такие исследователи, как А. Алсабаа (Саудовская Аравия, 2021), С. Элькататный (Саудовская Аравия, 2017), А. Говида (Саудовская Аравия, 2020), О. Томива (Нигерия, 2019), С. Гюль (США, 2020), А. Лекомцев, К. Абдельгавад (Россия, 2022), И. Гомаа (Саудовская Аравия, 2021), М. Рази (Иран, 2013), А. Третьяк (Россия, 2022), Р. Голами (Норвегия, 2022). Постоянный мониторинг и регулярное измерение реологических и фильтрационных свойств буровых растворов крайне важны для оптимизации буровых работ. Своевременное получение информации об этих свойствах позволяет буровой бригаде быстро принимать необходимые меры по регенерации бурового раствора. Существует ряд исследований, направленных на разработку интеллектуальных моделей для прогнозирования вышеупомянутых свойств, с ограниченной практической применимостью по причине ограничений по конкретным типам буровых растворов, а также по причине недостижения оптимального решения задачи. Кроме того, в настоящее время не разработано модели машинного обучения, способной прогнозировать объем фильтрации буровых растворов на основе наиболее часто отслеживаемых параметров ($\rho_{б.р}$, ТФ и УВ). Таким образом, в настоящее время существует актуальная проблема разработки прогностических моделей с высокой обобщающей способностью для прогноза трех целевых свойств, которые могут быть использованы в качестве системы помощи в принятии решений.

Цель исследования. Целью данного исследования является разработка и оптимизация моделей машинного обучения для точного определения трех критических параметров буровых растворов, а именно водоотдачи, пластической вязкости и динамического напряжения сдвига, на основе других легко и регулярно измеряемых свойств бурового раствора.

Для достижения поставленной цели в диссертации необходимо решить следующие задачи:

- Провести обзор существующих методов для прогнозирования трех целевых параметров бурового раствора, включая методы в основе которых лежит искусственный интеллект.
- Разработать две автономные модели машинного обучения (MELM и LSSVM) для прогнозирования целевых параметров буровых растворов. Комбинировать разработанные модели машинного обучения с методами оптимизации для

определения оптимальной структуры и гиперпараметров моделей с целью повышения точности прогнозирования и обобщаемости моделей.

- Провести серию экспериментов и анализов на этапе обучения моделированию гибридных моделей для повышения их точности прогнозирования и обобщаемости.
- Применить предложенные автономные и гибридные модели для прогнозирования целевых параметров бурового раствора; провести оценку эффективности разработанных моделей.

Объектом исследования в данной работе являются важнейшие фильтрационные и реологические свойства буровых растворов на водной основе: водоотдача, пластическая вязкость и динамическое напряжение сдвига.

Предметом исследования является прогнозная интеллектуальная модель для определения свойств бурового раствора на водной основе.

Методы исследования. Проведенное в данной диссертации исследование основано на методах машинного обучения, методах глобальной оптимизации, математическом и статистическом анализе, а также аналитических и численных экспериментах. Все вычисления проводились с использованием вычислительной платформы MATLAB.

Научная новизна работы.

1. Предложена, обоснована и экспериментально проверена прогнозирующая модель, основанная на многослойном экстремальном обучении (MELM), обеспечивающая повышение точности прогнозирования свойств бурового раствора с помощью определения оптимального количества скрытых слоев и входящих в них нейронов, а также нахождения оптимальных значений весов и смещений, приписываемых каждому нейрону и скрытому слою соответственно.
2. Предложена, обоснована и экспериментально проверена прогнозирующая модель, основанная на методе опорных векторов (LSSVM), обеспечивающая повышение точности прогнозирования свойств бурового раствора за счет определения подходящей функции ядра и организации поиска гиперпараметров, при которых достигается глобальный минимум среднеквадратического отклонения (RMSE).
3. Предложено применение разработанных прогнозирующих моделей для формирования набора гибридных моделей, обеспечивающих получение точных оценок водоотдачи, пластической вязкости и динамического напряжения сдвига бурового раствора на основе измеряемых параметров: плотности, условной вязкости и содержания твердой фазы бурового раствора.

Теоретическая значимость работы. Теоретические положения, сформулированные автором, способствуют развитию гибридных моделей машинного обучения, позволяющих делать точные прогнозы с высокой обобщаемостью. Задача разработки таких гибридных прогностических моделей выполнена с использованием методов глобальной оптимизации и получила

научное обоснование в результате проведения аналитического описания и многочисленных вычислительных экспериментов над предложенными моделями.

Практическая значимость работы. Практическая значимость работы заключается в том, модели, предложенные в данной диссертации, могут быть применены при бурении скважин для частого определения трех критических параметров бурового раствора, а именно пластической вязкости, динамическое напряжение сдвига и водоотдачи. Применение таких моделей, генерирующих надежные прогнозы целевых параметров бурового раствора, снижает требования к проведению большого количества трудоемких экспериментальных измерений.

Положения, выносимые на защиту.

1. Предложенная прогнозирующая модель, основанная на многослойном экстремальном обучении, обеспечивает повышение точности прогнозирования свойств бурового раствора с применением оптимальной архитектура MELM на 35-62 % в зависимости от целевого параметра по сравнению с традиционным подходом.
2. Предложенная прогнозирующая модель, основанная на методе опорных векторов LSSVM, обеспечивает повышение точности прогнозирования свойств бурового раствора на 19-45 % по сравнению с традиционным подходом.
3. Разработанные гибридные модели машинного обучения позволяют определять водоотдачу, пластическую вязкость и динамическое напряжение сдвига буровых растворов с погрешностями (2,95 – 4,27) %, (1,27 – 1,73) %, и (1,83 – 2,42) % соответственно.

Достоверность результатов. Достоверность результатов, представленных в диссертационном исследовании, обеспечивается путем оценки эффективности прогнозирования разработанных моделей на тестовом наборе данных. Предложенные прогностические модели, примененные к тестовому набору данных, обеспечивают достоверные прогнозы трех целевых параметров.

Внедрение результатов работы. Результаты настоящего диссертационного исследования внедрены в учебный процесс Инженерной школы природных ресурсов ТПУ, а также используются в компании ООО "Инновационные Технологии" для нужд мониторинга процесса бурения. Акты внедрения приложены к диссертационной работе.

Апробация работы. Основные положения диссертационного исследования докладывались и обсуждались на следующих конференциях:

- 3-й Восточно-Средиземноморский семинар EAGE (1-3 декабря 2021 года, Ларнака, Кипр);
- XXVI Международный научный симпозиум имени академика М.А. Усова «Проблемы геологии и освоения недр» (4-8 апреля 2022 года, Томск),
- XXVII Международный научный симпозиум имени академика М.А. Усова «Проблемы геологии и освоения недр» (3-7 апреля 2023 года, Томск), и
- Тюменская нефтегазовая конференция (4-8 сентября 2023, Минеральные Воды).

Публикации по теме диссертации. По результатам диссертационного исследования опубликовано 12 работ, в том числе 9 статей в журналах первого квартиля (Q1), проиндексированных в базах данных Web of Science и Scopus, а также 3 публикации в материалах международных и всероссийских научных конференций; получено 3 свидетельства о государственной регистрации программ для ЭВМ.

Личный вклад автора. Все описанные исследования проведены автором лично. Автором составлен необходимый набор данных на основе ежедневных буровых сводок по двум месторождениям и проведена предварительная обработка набора данных для машинного обучения и моделирования целевых свойств. Все этапы разработки и оценки предложенного метода прогнозирования свойств бурового раствора выполнены автором лично. Все результаты достигнуты и интерпретированы автором лично. Концептуализация исследования, постановка цели и задач настоящего исследования осуществлялись автором совместно с научным руководителем.

Структура и объем работы. Настоящее диссертационное исследование включает в себя список таблиц, список рисунков, введение, четыре основные главы, заключение, список номенклатуры и библиографию. Объем диссертационного исследования составляет 182 страниц, включая 52 рисунков и 22 таблицы. Библиография состоит из 165 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертационного исследования, сформулирована цель и поставлены задачи исследования. В данном разделе также освещается научная новизна и потенциальное практическое применение результатов настоящего исследования и формулируются положения, выносимые на защиту.

В **первой главе** «Краткое введение в буровые растворы» приводится краткое введение в технологии буровых растворов и машинного обучения; далее следует комплексный обзор современных подходов машинного обучения, применяемых в основных направлениях нефтегазового дела: разведке, разработке месторождений и бурении. Также в разделе описываются существующие исследования в области прогнозирования свойств буровых растворов, проводится их сравнение, освещаются преимущества и недостатки. Всесторонний обзор литературы свидетельствует о широком применении моделей машинного обучения в различных сегментах нефтегазовой отрасли для решения задач регрессии и классификации. Примечательно, что в предыдущих исследованиях, посвященных прогнозированию реологических свойств, в основном используются модели для конкретных растворов, что ограничивает их применимость в реальном времени при буровых работах. Для таких прогнозов обычно используются искусственные нейронные сети, однако они требуют тщательной настройки гиперпараметров, что сказывается на эффективности прогнозирования.

Что касается прогнозирования водоотдачи, то существующие модели машинного обучения часто опираются на экспериментальные данные по нанофлюидам или требуют редко измеряемых входных параметров, что делает их непригодными для мониторинга водоотдачи в режиме реального времени. Следовательно, существует пробел в моделях, прогнозирующих водоотдачи на основе обычно отслеживаемых параметров бурового раствора ($\rho_{бр}$, ТФ и УВ). В связи с этими ограничениями буровая промышленность ищет более обобщенные модели для точного прогнозирования реологических и фильтрационных свойств различных типов буровых растворов.

Анализ использования алгоритмов LSSVM и MELM в нефтегазовом секторе показывает их перспективность, но в то же время выявляет и недостатки. Оба алгоритма испытывают трудности с оптимизацией гиперпараметров, что является важнейшим фактором, влияющим на точность прогнозирования. Кроме того, MELM требует трудоемких экспериментов для выбора оптимальной архитектуры. Для решения этих проблем в данном исследовании ставится задача интегрировать MELM и LSSVM с передовыми методами оптимизации, позволяющими определить оптимальные структуры моделей и гиперпараметры для повышения точности и обобщаемости.

Во второй главе «Методология, применяемая в новой прогностической системе» подробно рассмотрены теоретические описания алгоритмов, использованных при разработке моделей, описание набора данных и его предварительная обработка, а также метрики оценки точности прогнозирования моделей. Ниже представлена общая схема рабочего процесса разработки и оценки точности прогнозирования целевых параметров буровых растворов.

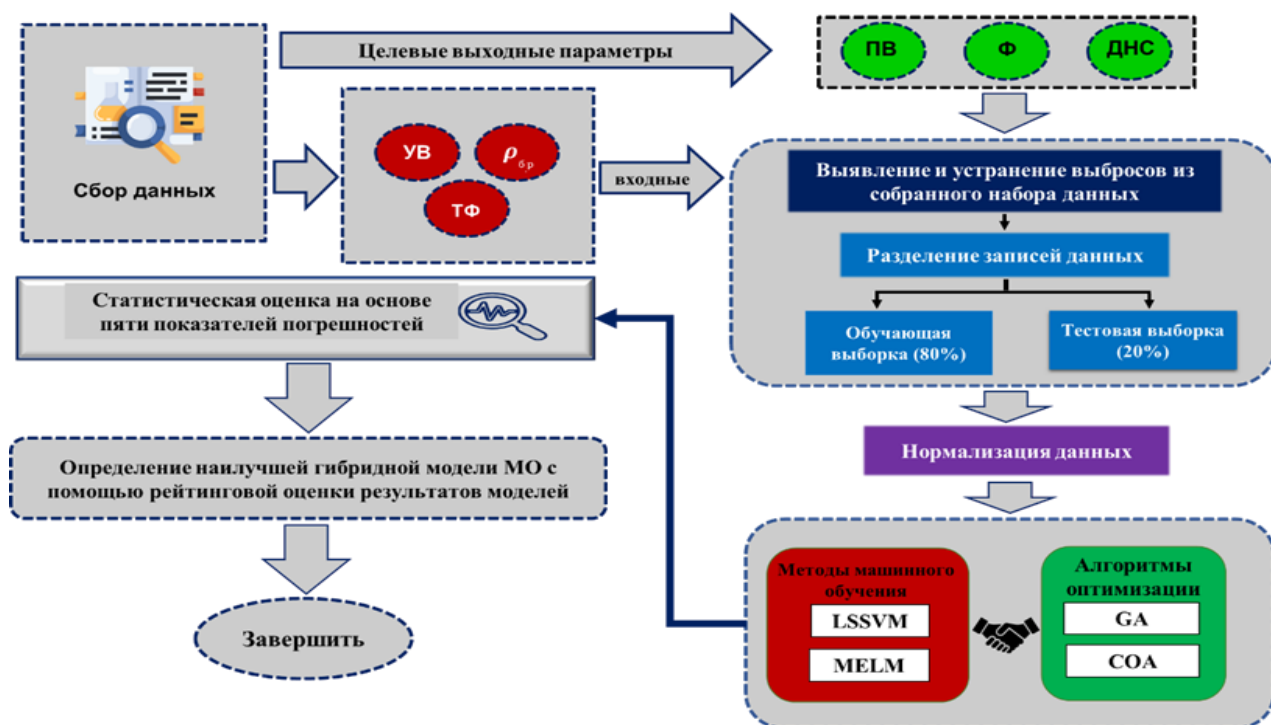


Рисунок 1 – Схема рабочего процесса разработки и оценки точности прогнозирования целевых параметров буровых растворов. GA - генетический алгоритм; COA - алгоритм оптимизации «Cuckoo».

Как видно на рисунке 1, сначала необходимый набор данных был собран из ежедневных отчетов о бурении 14 скважин, пробуренных на двух различных нефтяных месторождениях, расположенных на юго-западе Ирана. Собранный набор данных содержал 1160 записей по шести входным и выходным переменным ($\rho_{бр}$ – плотности, УВ – условной вязкости, ТФ – содержания твердой фазы, ПВ – пластическая вязкость, Ф – водоотдача, ДНС – динамическое напряжения сдвига). Вначале была проведена предварительная обработка собранного набора данных с использованием расстояния Махаланобиса (MD) для обнаружения и устранения выбросов. Данные о буровом растворе, записанные на месте скважины, могут содержать ошибочные значения. Такие значения могут быть результатом плохо откалиброванного оборудования для мониторинга и/или человеческого фактора. Поэтому необходимо отфильтровать ошибочные записи данных, проведя предварительную обработку набора данных по буровому раствору перед моделированием. Отобранные и сохраненные записи данных были разделены на два поднабора данных: один для обучения модели (80% от общего количества записей данных) и один для тестирования модели (20% от общего количества записей данных). Подмножества для обучения и тестирования были затем отдельно нормализованы для масштабирования всех переменных данных в диапазоне от +1 до -1. После нормализации подмножеств данных контролируемые автономные модели машинного обучения (LSSVM и MELM) и их гибридные формы с методами оптимизации GA или COA (LSSVM-GA, LSSVM-COA, MELM-GA и MELM-COA) были обучены на обучающем подмножестве, для чего было проведено несколько запусков и результаты статистически оценены для определения эффективности прогнозирования параметров бурового раствора Ф, ПВ и ДНС. Затем была проведена независимая проверка обобщаемости обученных автономных и гибридных моделей путем применения обученных моделей к отдельному подмножеству записей данных (тестовое подмножество), ранее не изученных обученными моделями. Затем производительность шести разработанных моделей тщательно оценивалась по пяти статистическим показателям погрешности, а именно: коэффициент детерминации (R^2), среднеквадратическое отклонение, стандартное отклонение, среднее значение относительного отклонения и среднее значение абсолютного относительного отклонения. Затем была определена наиболее эффективная модель для прогнозирования свойств бурового раствора путем применения двух дополнительных оценок эффективности прогнозирования: оценочного анализа и характеристик погрешности регрессии.

Алгоритмы прогнозирования. В настоящей диссертационной работе два алгоритма, модели LSSVM и MELM, применяются для разработки прогнозных моделей для определения трех целевых буровых растворов. Теория алгоритмов кратко описана далее.

Модели LSSVM для задач регрессии определяются следующими уравнениями:

$$\min J(w, \xi) = \min \left(\frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^N \xi_i^2 \right), \quad (1)$$

$$\text{такой, что} \quad y_i = \langle w, \phi(x) \rangle + b + e_i, \quad i = 1, 2, \dots, N \quad (2)$$

где γ - параметр регуляризации, используемый для определения компромисса между сложностью и точностью модели, x представляет собой входные данные, которыми в данном исследовании являются $\rho_{бр}$, ТФ и УВ. ξ_i представляет собой ошибку регрессии между фактическими и прогнозируемыми значениями выходных переменных y_i , которые в данном исследовании являются ПВ, ДНС и Ф, и называется переменной провисания.

Большие значения γ ограничивают значения переменных провисания, что приводит к усложнению модели, в то время как малые значения γ позволяют переменным слабину иметь большие значения, принимая больше ошибок обучения. Следовательно, очень важно определить подходящее значение γ как часть настройки гиперпараметров LSSVM.

Для решения LSSVM требуется функция Лагранжа, выраженная в приведенном ниже уравнении.

$$L_{LSSVM} = \frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N a_i \{ \langle w, \phi(x) \rangle + b + e_i + \gamma_i \}, \quad (3)$$

где a_i ($i = 1, 2, \dots, N$) обозначает множитель Лагранжа. Оптимизация уравнения (3) позволяет выразить решение в виде следующего уравнения.

$$\left\{ \begin{array}{l} \frac{\delta L_{LSSVM}}{\delta w} = 0 \rightarrow w = \sum_{i=1}^N a_i \gamma_i \phi(x_i), \\ \frac{\delta L_{LSSVM}}{\delta b} = 0 \rightarrow \sum_{i=1}^N a_i \gamma_i = 0, \\ \frac{\delta L_{LSSVM}}{\delta \xi_i} = 0 \rightarrow a_i = \gamma \xi_i, \quad i = 1, 2, \dots, N, \\ \frac{\delta L_{LSSVM}}{\delta a_i} = 0 \rightarrow \langle w, \phi(x) \rangle + b + \xi_i + \gamma_i, \quad i = 1, 2, \dots, N. \end{array} \right. \quad (4)$$

В итоге, функция прогнозирования LSSVM будет выражена следующим уравнением:

$$f(x) = \sum_{i=1}^N a_i K(x_i, x_j) + b, \quad (5)$$

где $K(x_i, x_j)$ - нелинейная функция ядра. Тип применяемой функции ядра сильно влияет на эффективность прогнозирования LSSVM и должен быть выбран с осторожностью в рамках процесса настройки модели. В данном исследовании мы протестировали четыре наиболее часто используемые функции ядра, чтобы выбрать наиболее эффективное ядро для модели LSSVM.

Для определения оптимальных гиперпараметров модели LSSVM обычно используются локальные алгоритмы оптимизации: поиск по сетке (GS), градиентный спуск (GD) и метод сопряжённых градиентов (CG). Однако эти методы могут застрять в локальных минимумах. Для преодоления этого ограничения в данном исследовании также используются глобальные алгоритмы оптимизации, а именно генетический алгоритм (GA) и алгоритм оптимизации

кукушки (COA). Таким образом, в этом исследовании оцениваются пять методов оптимизации: три локальных (CG, GD и GS) и два глобальных (GA и COA) метода. Все они направлены на определение наилучших настроек гиперпараметров для модели LSSVM.

Важным фактором в алгоритмах оптимизации является определение количества переменных решения. Для комбинированного подхода LSSVM с использованием генетического алгоритма (GA) и/или алгоритма оптимизации COA, учитывая, что ядро LSSVM является радиальной базисной функцией (RBF), при обучении модели оптимизируются два гиперпараметра: ширина ядра (σ) и параметр регуляризации (γ). В результате, количество переменных решения для алгоритма оптимизации составляет два.

Для определения количества переменных решения в алгоритмах оптимизации необходимо установить диапазон изменчивости для каждой переменной. Например, диапазон для значения σ был установлен от 0,001 до 100, в то время как значение γ изменялось в диапазоне от 0,001 до 5000. На начальных этапах алгоритмы оптимизации генерируют значения в пределах установленных диапазонов. Затем эти значения служат входными параметрами для алгоритма LSSVM, позволяя ему построить прогностическую модель. Используя предоставленные гиперпараметры, модель LSSVM прогнозирует целевой параметр. Разница между этими прогнозами и фактическими значениями рассчитывается как ошибка, представляющая собой метрику стоимости, связанную с данными гиперпараметрами. С каждой итерацией алгоритмы оптимизации изменяют гиперпараметры в соответствии со своими внутренними процедурами оптимизации. Этот итеративный механизм продолжается до достижения предопределенного критерия завершения, такого как заранее установленное количество итераций, идентифицируя оптимальные значения гиперпараметров.

Алгоритм MELM. MELM является усовершенствованной версией алгоритма ELM с несколькими скрытыми слоями. Данная модификация позволяет MELM отображать отношения, необходимые для прогнозирования целевых переменных с большей точностью, чем стандартный алгоритм ELM. Отображение отношений переменных проводится путем определения значений, необходимых для весовой матрицы и вектора смещения, связанного со скрытыми слоями сети. Настройка MELM требует семь шагов:

1. Настраивается набор данных входных переменных X_0 и выходных переменных Y с m записями данных $[X_0, Y] = [x_i, y_i]$ ($i = 1, 2, \dots, m$) в матрицы X_0 и Y .

2. Определяется количество скрытых слоев k и узлов для MELM, где каждый скрытый слой в сети включает в себя n узлов с функцией активации.

3. Для каждого слоя за пределами входного слоя определяются случайно инициализированные веса W для узлов, соединяющих j -й и $(j-1)$ -й скрытые слои. Затем каждому из слоев присваиваются случайные смещения B , что приводит к уравнению:

$$W_{j-1,j} = [B_j W_{j-1,j}], \quad X_j = [1 X_{j-1}]^T, \quad (6)$$

где $\mathbf{1}$ обозначает единичный вектор-столбец размера m .

4. Матрица H (вывод скрытого слоя) определяется с помощью уравнения:

$$H = g(W_{j-1,j} X_j). \quad (7)$$

5. Если j меньше k , то инкрементируйте j , установите вход нового слоя в соответствие с выходом предыдущего слоя и переходите к третьему шагу.

$$X_{j-1} = H^T. \quad (8)$$

6. Обеспечивая линейное решение наименьших квадратов, обратный вес Мура-Пенроуза $\text{pinv}()$ применяется для определения весов между выходным слоем и конечным скрытым слоем, в соответствии с следившем уравнением:

$$\beta = \text{pinv}(H^T) \times Y. \quad (9)$$

7. Путем применения, следующего уравнения прогнозируются выходные переменные $\hat{Y}$:

$$\hat{Y} = (H^T \times \beta)^T. \quad (10)$$

Несмотря на описанные выше преимущества, методу MELM присущи два крупных недостатка, требующие пристального внимания: а) точность прогнозирования и обобщаемость MELM в значительной степени зависят от его архитектуры (т.е., соответствующего количества скрытых слоев и достаточного количества нейронов в каждом скрытом слое), б) как и в случае с алгоритмом ELM, стандартной реализации MELM свойственно присвоение случайных весов и смещений узлам и скрытым слоям. Чтобы преодолеть эти ограничения и добиться от модели MELM максимальной эффективности, настоятельно рекомендуется выбрать оптимальное количество скрытых слоев и необходимое количества нейронов в каждом скрытом слое, а также использовать оптимизированные значения в качестве весов и смещений к этим слоям/узлам. В настоящем исследовании применяются два строгих подхода к оптимизации для настройки оптимальной структуры сети MELM и оценки оптимизированных значений смещений и весов для использования в сети MELM. Для создания надежной модели MELM с точными прогнозами необходимо определить её оптимальную конфигурацию. Простое использование метода проб и ошибок для этого является неэффективным. Поэтому интеграция MELM с оптимизационными алгоритмами ускоряет процесс настройки. В этом исследовании применяется двухэтапный подход:

1. Оптимизация структуры:

- Сначала мы оцениваем подходящие диапазоны для количества скрытых слоев и узлов в MELM.

- Используя эти оценки, COA определяет лучшую структуру MELM, рассчитывая RMSE (целевую функцию, стремящуюся к минимизации) для каждой потенциальной структуры.

- Каждая структура тестируется десять раз. Выбирается та, у которой самый низкий средний RMSE из всех десяти тестов.

2. Оптимизация весов и смещений:

- После идентификации оптимальной структуры MELM мы начинаем следующий этап оптимизации, чтобы определить лучшие веса и смещения.
- RMSE служит целевой функцией для оптимизационных алгоритмов GA и СОА. Количество переменных решения в этих алгоритмах соответствует общему числу весов и смещений в оптимальной структуре MELM.
- Каждому члену популяции оптимизации назначаются начальные значения веса и смещения, которые модель MELM оценивает. Полученные значения RMSE направляют оптимизационные алгоритмы.
- На протяжении итеративных оптимизационных раундов веса и смещения корректируются на основе их значений RMSE, двигая модель к минимальной ошибке.
- Этот итеративный процесс останавливается при достижении установленного условия завершения.

Наконец, используя оптимальные веса и смещения, определенные в последней итерации, модель MELM обучается, а затем тестируется на соответствующих наборах данных.

В третьей главе «Эксперименты, проведенные для разработки прогнозных моделей» подробно представлены эксперименты, проведенные для разработки автономной и гибридной моделей машинного обучения.

В качестве одного из первичных этапов моделирования машинного обучения был проведен анализ для определения коэффициента корреляции между входными (УВ, $\rho_{бр}$, ТФ) и выходными параметрами (ПВ, ДНС, Φ) моделей машинного обучения. Тепловая карта показывает самые сильные положительные корреляции красным цветом и самые сильные отрицательные корреляции синим (рисунок 2).

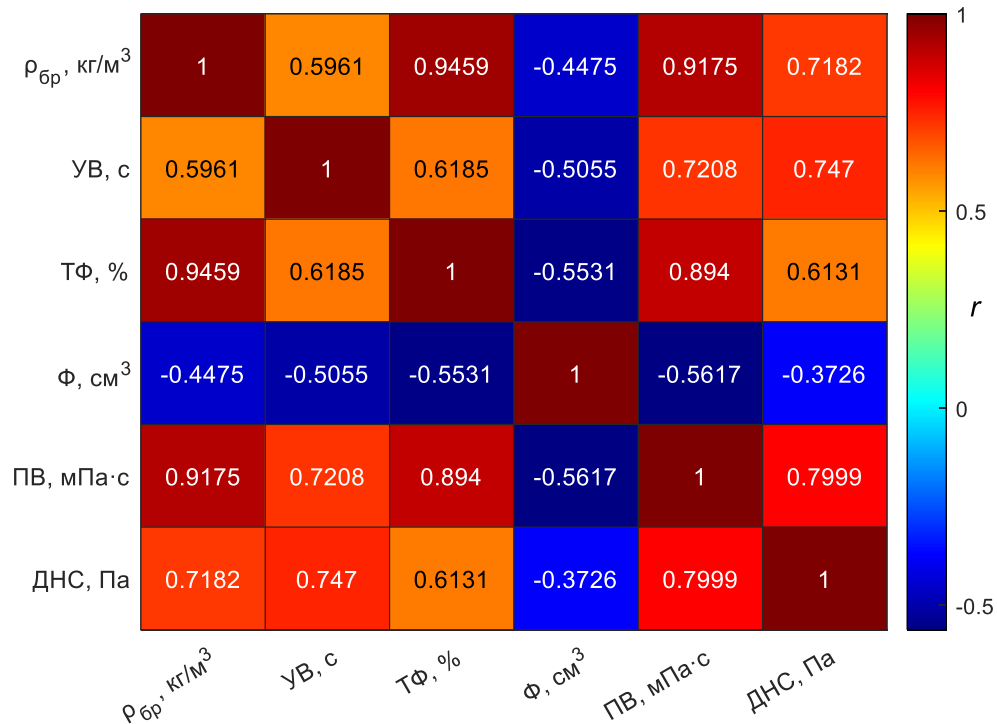


Рисунок 2 – График тепловой карты коэффициентов корреляции между изучаемыми свойствами бурового раствора.

Как видно из рисунка 2, ПВ и ДНС имеют положительные корреляции со всеми тремя входными параметрами; это означает, что по мере увеличения значений переменных УВ, ρ_{br} и ТФ увеличиваются и значения ДНС и ПВ. Демонстрируя существенную прямую положительную корреляцию друг с другом, ТФ и ρ_{br} аналогично влияют и на реологические свойства буровых растворов (ПВ и ДНС). ρ_{br} показывает самые сильные положительные корреляции с ПВ и ДНС. Напротив, три входные переменные имеют обратные корреляции с Ф, где самые слабые и самые сильные корреляции с Ф наблюдаются с ρ_{br} и УВ соответственно.

Данные, записанные на месторождении в лаборатории буровых жидкостей, обычно содержат некоторые внутренние ошибки, связанные с неправильной калибровкой оборудования, используемого при проведении испытаний, и/или человеческими ошибками при выполнении и записи этих испытаний. Следовательно, перед тем как приступить к моделированию машинного обучения, собранный набор данных о буровых жидкостях требует предварительной обработки для обнаружения выбросов и исключения неточных записей. Для этой цели была применена фильтрация данных с использованием расстояния Махаланобиса (MD) в сочетании с результатами прогноза алгоритма гауссовского процесса регрессии. Этот метод позволил выявить 42, 44 и 50 записей данных как выбросы в показателях фильтрации, пластической вязкости и динамического касательного напряжения соответственно.

В данном исследовании оптимальное соотношение обучающих и тестовых данных было определено с использованием метода проб и ошибок. Данные были случайным образом разделены на различные пропорции 60/40, 70/30 и 80/20 для обучения и тестирования. Результаты, полученные методом проб и ошибок, показали, что, ошибка на этапе обучения также увеличивается с увеличением объема обучающих данных. В то же время ошибка на этапе тестирования уменьшается при уменьшении объема тестовых данных. Учитывая, что увеличение ошибки на этапе обучения менее значимо по сравнению с уменьшением ошибки на этапе тестирования, и учитывая также высокую важность способности обобщения модели на невидимых данных, пропорция 80/20 обучающих и тестовых данных, по сравнению с двумя другими пропорциями, является наиболее подходящей для разделения данных в данном исследовании. Аналогичные результаты были достигнуты и для двух других целевых параметров (ПВ и ДНС).

Первым этапом разработки гибридной модели MELM является определение оптимальной архитектуры MELM. Для определения оптимальной структуры MELM необходимо провести анализ методом проб и ошибок для различных структур архитектур MELM, первоначально рассматриваемых для прогнозирования Ф, ПВ и ДНС. Результаты показывают, что MELM с 3-7 скрытыми слоями (HL) и с каждым HL, содержащим от 2 до 10 нейронов, обеспечивают наилучшие прогностические показатели. На рисунке 3а представлена контурная диаграмма изменений RMSE для модели MELM с различными архитектурами, разработанными для прогнозирования водоотдачи. Эти предварительные диапазоны затем используются в качестве верхнего и

нижнего ограничений для алгоритма оптимизации COA, применяемого для достижения оптимальной структуры MELM. На рисунке 3б представлена картина сходимости COA за 50 итераций при определении оптимальной структуры MELM с использованием обучающих данных.

График показывает снижение и последующую стабилизацию RMSE на протяжении итераций для прогнозируемых водоотдачи. В начале значения RMSE для Ф, ПВ и ДНС повышены, что указывает на неоптимальную конфигурацию модели MELM. Однако уже в первых 15 итерациях COA заметное снижение RMSE наблюдается для всех трех выходных параметров, что отражает эффективность алгоритма в настройке структуры MELM. К 30-ой итерации RMSE стабилизируется для всех переменных, что свидетельствует о достижении оптимального решения с минимальной ошибкой.

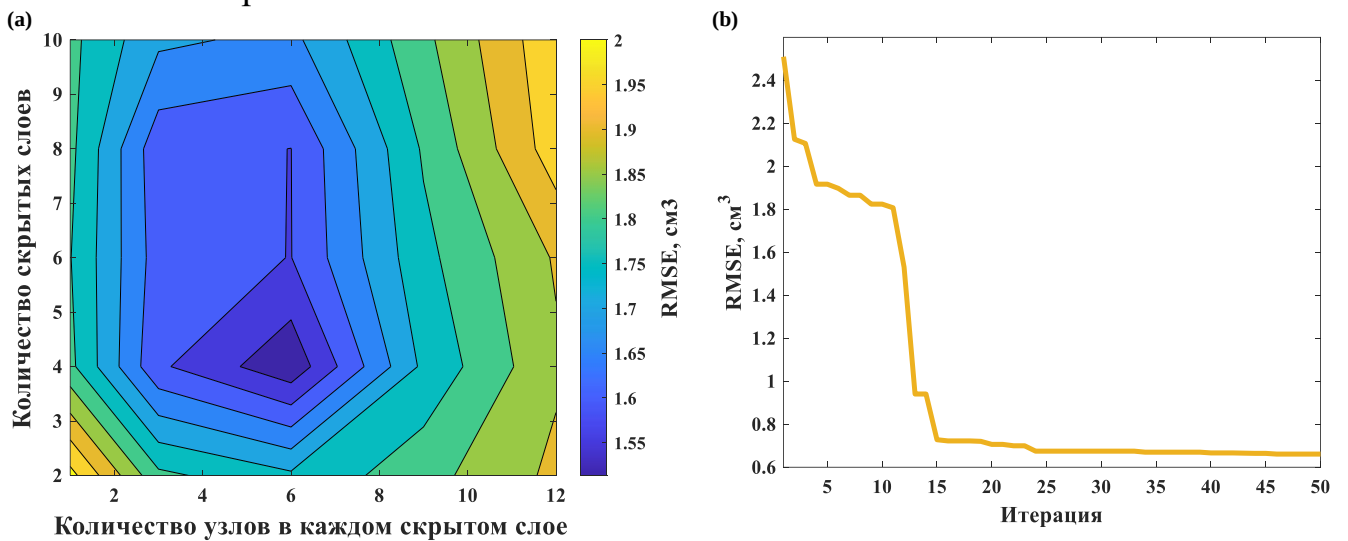


Рисунок 3 – Изменение среднеквадратического отклонения а) в моделях MELM с разным количеством скрытых слоев и узлов для прогнозирования водоотдачи, и б) различным количеством итераций COA для оптимизации архитектуры MELM.

Последующий анализ результатов COA выявил, что оптимальной структурой MELM для прогнозирования всех трех параметров является композиция из четырех скрытых слоев, содержащих 4, 3, 2 и 2 узла в каждом слое соответственно. Эта конфигурация будет использоваться для моделирования Ф, ПВ и ДНС в MELM.

При разработке прогнозных моделей для параметров Ф, ПВ и ДНС с использованием LSSVM, функция ядра является ключевым параметром, который должен быть выбран. Для определения оптимальной функции ядра для данного применения, в данном исследовании был также использован подход метода проб и ошибок. Конкретно, модели LSSVM были разработаны и оценены с использованием четырех различных функций ядра - линейная, полиномиальная, многослойный перцептрон, радиальная базисная функция (RBF). Проведенный эксперимент над рассмотренными функциями ядра подтвердил, что функция ядра RBF выдала наименьшее значение RMSE, RMSE, 0,5186 cm^3 для водоотдачи. Необходимо отметить, что аналогичные результаты были достигнуты и для двух других целевых параметров (ПВ и ДНС). Исходя из этих результатов, функция ядра RBF была выбрана для построения моделей LSSVM для прогнозирования

параметров Φ , PB и $ДНС$ в данном исследовании. Оптимизационные алгоритмы для LSSVM должны лишь найти лучшие значения для двух гиперпараметров: сигмы и гаммы, основанные на выбранной функции ядра. После того как оптимальная структура модели MELM и правильная функция ядра для LSSVM установлены, настало время использовать оптимизаторы (GA и COA) для определения оптимальных гиперпараметров для моделей MELM и LSSVM.

На рисунке 4 представлен процесс повышения эффективности прогнозирования водоотдачи каждой из моделей HML за 200 итераций. Как видно на рисунке 4, для моделей MELM, гибридных с COA и GA, оба глобальных оптимизатора быстро сходятся к лучшему решению за первые 20 итераций. Однако после 20-той итерации снижение RMSE резко замедляется. Кроме того, анализ рисунка позволяет сделать вывод, что алгоритм COA достигает оптимального решения быстрее и предоставляет более точное решение по сравнению с методом GA. Модель MELM-COA сходится наиболее быстро и эффективно к решениям с наименьшей ошибкой в водоотдаче ($RMSE < 0,23 \text{ см}^3$), по сравнению с другими тремя гибридными моделями машинного обучения. Необходимо отметить, что, аналогичные результаты были достигнуты и для двух других целевых параметров (PB и $ДНС$).

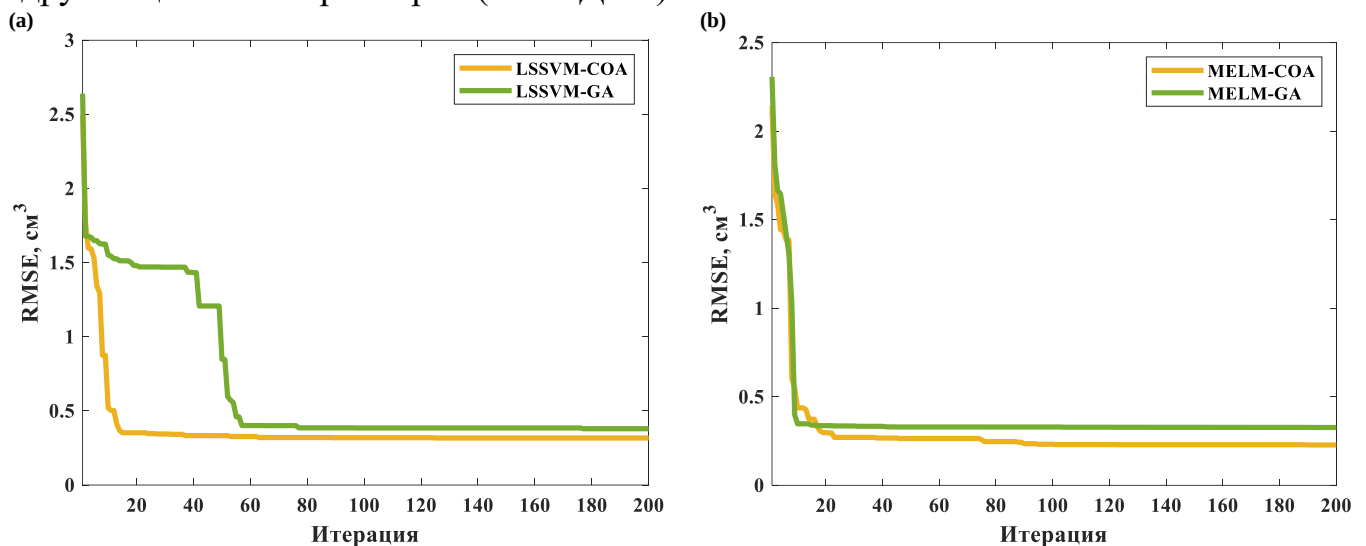


Рисунок 4 – Среднеквадратическое отклонение прогнозирования расхода воды на каждой итерации гибридных моделей машинного обучения: а) LSSVM-GA и LSSVM-COA; б) MELM-GA и MELM-COA.

Четвёртая глава «Результаты прогнозирования гибридных моделей машинного обучения» данного исследования описывает результаты, полученные в результате использования моделей прогнозирования, связанных с целевыми характеристиками буровых растворов. Эти прогнозы осуществляются с использованием как автономно разработанной модели прогнозирования, так и гибридной модели прогнозирования. Полученные результаты тщательно анализируются и подвергаются сравнительному анализу для определения наиболее эффективной модели.

На рисунке 5 сравниваются значения $RMSE$ и R^2 , связанные с прогнозными значениями водоотдачи, полученными каждой из шести обученных моделей для

тестового подмножества данных. Эффективность прогнозирования водоотдачи обученными гибридными моделями МО для тестового подмножества данных может быть ранжирована в порядке возрастания RMSE: MELM-COA > LSSVM-COA > MELM-GA > LSSVM-GA.

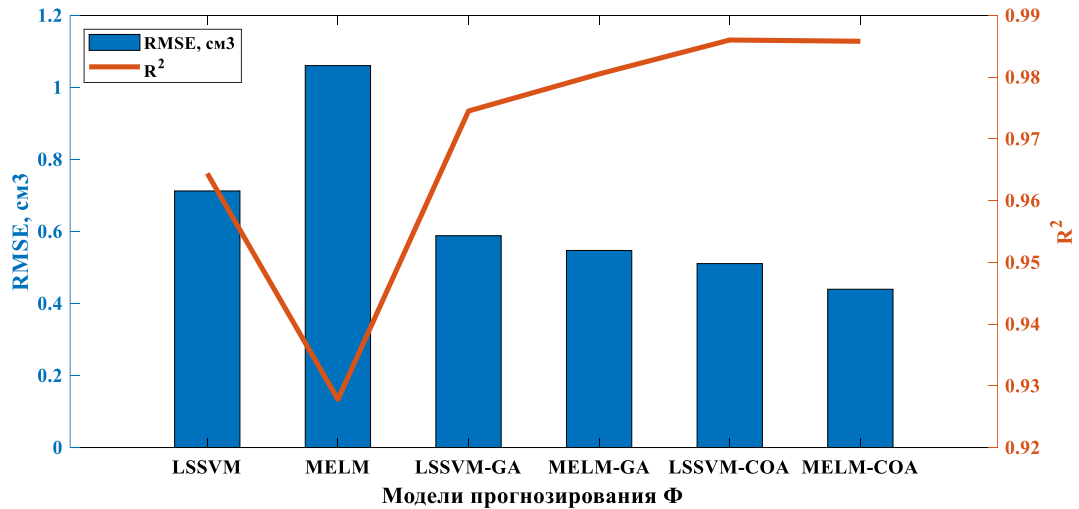


Рисунок 5 – Среднеквадратическое отклонение (столбцы) и коэффициент детерминации (линия) для прогнозирования водоотдачи предложенными моделями.

На рисунке 6 сравниваются значения RMSE и R^2 , связанные с прогнозами пластической вязкости для тестового подмножества данных, сгенерированными каждой из шести обученных моделей. Все шесть оцениваемых моделей прогнозируют пластическую вязкость со значениями R^2 выше 0,97 и значениями RMSE ниже 0,99 мПа·с. Данные результаты подтверждают, что все шесть обученных моделей обеспечивают высокую достоверность прогнозирования пластической вязкости для тестового подмножества данных.

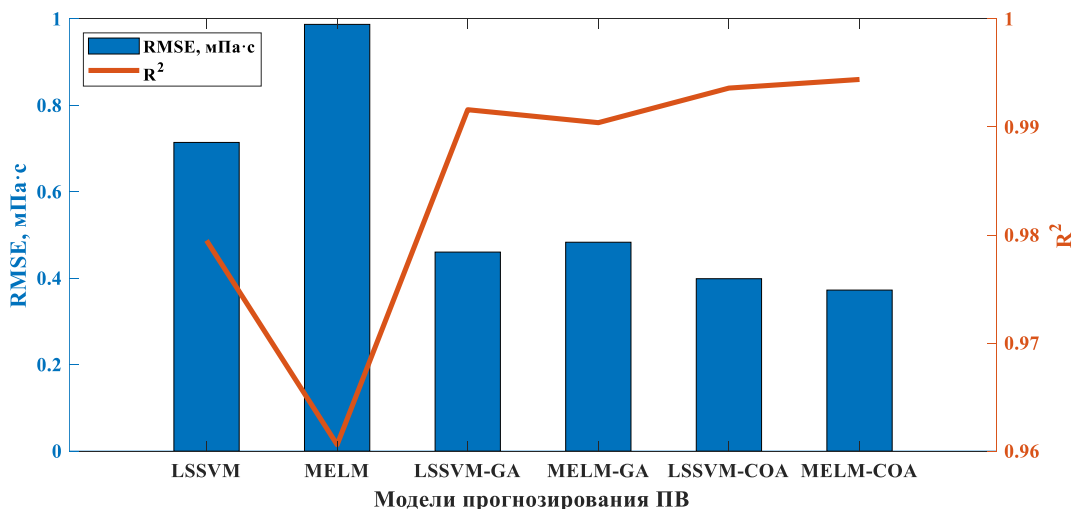


Рисунок 6 – Среднеквадратическое отклонение (столбцы) и коэффициент детерминации (линия) для прогнозирования пластической вязкости предложенными моделями.

Эффективность прогнозирования пластической вязкости обученными гибридными моделями МО для тестового подмножества данных может быть

ранжирована в порядке возрастания RMSE: MELM-COA < LSSVM-COA < LSSVM-GA < MELM-GA.

На рисунке 7 сравниваются значения RMSE и R^2 , связанные с прогнозами динамического напряжения сдвига для тестового подмножества данных, сгенерированными каждой из шести обученных моделей. Все шесть оцениваемых моделей прогнозируют динамическое напряжение сдвига со значениями R^2 выше 0,90 и значениями RMSE ниже 0,32 Па. Наиболее эффективная модель MELM-COA достигает значений $R^2 = 0,9598$, и $RMSE = 0,1970$ Па для тестового подмножества данных. Эффективность прогнозирования динамического напряжения сдвига обученными гибридными моделями МО для тестового подмножества данных может быть ранжирована в порядке возрастания RMSE: MELM-COA < LSSVM-COA < MELM-GA < LSSVM-GA.

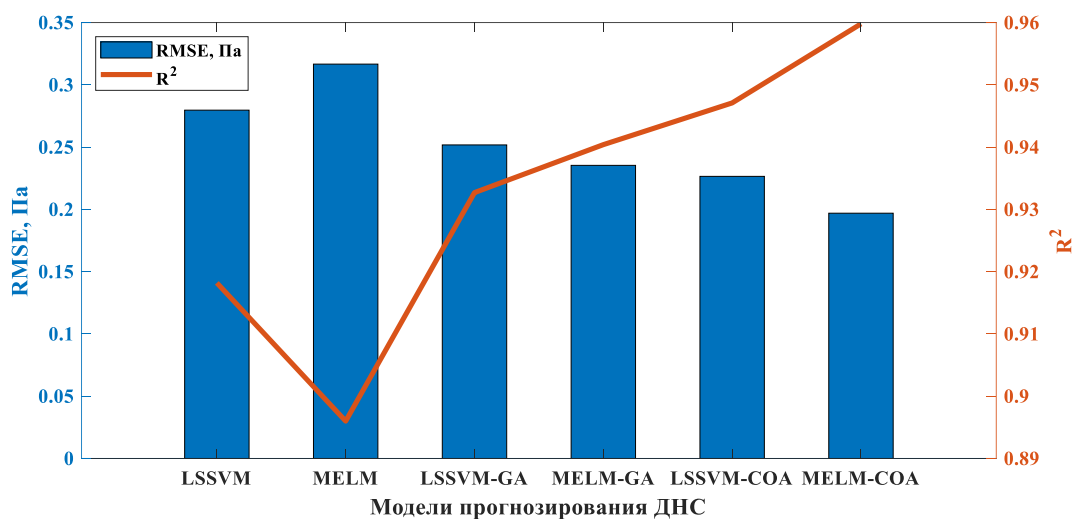
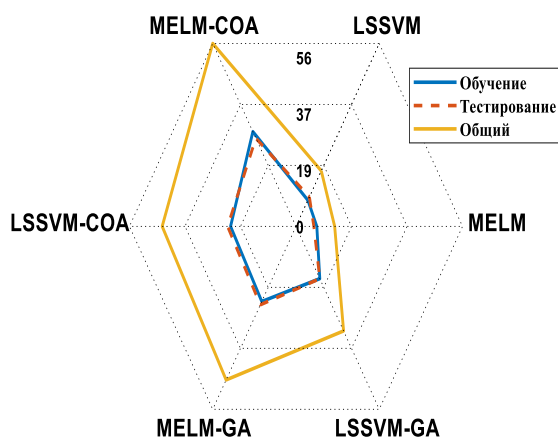


Рисунок 7 – Среднеквадратическое отклонение (столбцы) и коэффициент детерминации (линия) для прогнозирования динамического напряжения сдвига предложенными моделями.



Для дальнейшей оценки эффективности прогнозирования предложенных моделей по трем целевым параметрам буровых растворов были проведены Score Analysis (SA) для прогнозов, полученных с помощью автономных и гибридных моделей. На рисунке 8 графически сравниваются оценки эффективности прогнозирования оцениваемых автономных моделей и гибридных моделей МО.

Рисунок 8 – Оценки эффективности прогнозирования для обучающего и тестового подмножеств данных, а также суммарные оценки, полученные автономными и гибридными моделями МО.

Как видно из рисунка 8, модели, оптимизированные с помощью методов глобальной оптимизации (GA и COA), показывают значительно более высокие оценки эффективности прогнозирования, чем автономные модели. Наибольшую

оценку при прогнозировании водоотдачи получила модель MELM, оптимизированная с помощью оптимизатора COA. Полученные результаты свидетельствуют о том, что модель MELM-COA обладает наилучшей точностью прогнозирования среди всех шести моделей, разработанных для прогнозирования параметров буровых растворов в данном исследовании. Таким образом, мы приходим к выводу, что значение водоотдачи, пластической вязкости, динамического напряжения сдвига, можно достоверно прогнозировать с помощью гибридных моделей МО, использующих в качестве входных параметров только три часто измеряемые входные переменные (плотность, условная вязкость, содержание твердой фазы), тем самым избегая необходимости проводить многочисленные трудоемкие испытания бурового раствора в полевых условиях для измерения целевых параметров. Наличие надежных моделей для оценки водоотдачи, пластической вязкости и динамического напряжения сдвига буровых растворов может повысить производительность буровых работ и минимизировать риск повреждения пласта. Это позволит оперативно корректировать кондиционирование бурового раствора в соответствии с изменениями условий бурения и состава бурового раствора в ходе буровых работ.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

- Предложена, обоснована и экспериментально проверена прогнозирующая модель, основанная на многослойном экстремальном обучении MELM, обеспечивающая повышение точности прогнозирования свойств бурового раствора с помощью применением оптимальной архитектура MELM. Разработанная прогнозирующая модель обеспечивает повышение точности прогнозирования свойств бурового раствора на 35-62 % в зависимости от целевого параметра по сравнению с традиционным подходом.
- Предложена, обоснована и экспериментально проверена прогнозирующая модель, основанная на методе опорных векторов LSSVM, обеспечивающая повышение точности прогнозирования свойств бурового раствора за счет определения подходящей функции ядра и организации поиска гиперпараметров, при которых достигается глобальный минимум среднеквадратического отклонения. Разработанная прогнозирующая модель обеспечивает повышение точности прогнозирования свойств бурового раствора на 19-45 % в зависимости от целевого параметра по сравнению с традиционным подходом.
- Разработанные гибридные модели машинного обучения позволяют прогнозировать водоотдачу, пластическую вязкость и динамическое напряжение сдвига буровых растворов с погрешностями (2,95 – 4,27) %, (1,27 – 1,73) %, и (1,83 – 2,42) % соответственно.
- Результаты настоящего диссертационного исследования внедрены в учебный процесс Инженерной школы природных ресурсов ТПУ, а также в производственный процесс компании ООО "Инновационные Технологии". Компанией были протестированы разработанные в ходе диссертационного исследования модели машинного обучения на реальных данных по буровому

раствору ранее пробуренных скважин. Высокая точность прогноза $R^2 \approx 90-92\%$ дала компании основание для использования данных моделей при мониторинге бурения скважин с целью повышения точности и оперативности определения целевых параметров бурового раствора. Акты внедрения приложены к диссертационной работе.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК

1. Davoodi, S. Hybridized machine-learning for prompt prediction of rheology and filtration properties of water-based drilling fluids / S. Davoodi, M. Mehrad, D.A. Wood, H. Ghorbani, V.S. Rukavishnikov // *Engineering Applications of Artificial Intelligence*. – 2023. – Vol. 123. – P. 106459. (Q1, WoS, Scopus).
2. Davoodi, S. Machine-learning models to predict hydrogen uptake of porous carbon materials from influential variables / S. Davoodi, H.V. Thanh, D.A. Wood, M. Mehrad, M. Al-Shargabi, V.S. Rukavishnikov // *Separation and Purification Technology*. – 2023. – Vol. 316. – P. 123807. (Q1, WoS, Scopus).
3. Davoodi, S. Machine-learning predictions of solubility and residual trapping indexes of carbon dioxide from global geological storage sites / S. Davoodi, H.V. Thanh, D.A. Wood, M. Mehrad, V.S. Rukavishnikov, Z. Dai // *Expert Systems with Applications*. – 2023. – Vol. 222. – P. 119796. (Q1, WoS, Scopus).
4. Davoodi, S. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms / S. Rashidi, M. Mehrad, H. Ghorbani, D.A. Wood, N. Mohamadian, J. Moghadasi // *Journal of Petroleum Science and Engineering*. – 2021. – Vol. 202. – P. 108425. (Q1, WoS, Scopus).
5. Davoodi, S. Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields / A.R. Abad, H. Ghorbani, N. Mohamadian, S. Davoodi, M. Mehrad, S.K. Aghdam, H.R. Nasriani // *Fuel*. – 2022. – Vol. 308. – P. 121872. (Q1, WoS, Scopus).
6. Davoodi, S. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs / A.R. Behesht Abad, S. Mousavi, N. Mohamadian, D.A. Wood, H. Ghorbani, S. Davoodi, M.A. Alvar, K. Shahbazi // *Journal of Natural Gas Science and Engineering*. – 2021. – Vol. 95. – P. 104210. (Q1, WoS, Scopus).
7. Davoodi, S. Prediction of permeability of highly heterogeneous hydrocarbon reservoir from conventional petrophysical logs using optimized data-driven algorithms / A. Sheykhinasab, A.A. Mohseni, A. Barahooie Bahari, E. Naruei, Davoodi, S., A. Aghaz, M. Mehrad // *Journal of Petroleum Exploration and Production Technology*. – 2023. – Vol. 13. – P. 66-689. (Q1, WoS, Scopus).
8. Davoodi, S. Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms / M. Farsi, N. Mohamadian, H. Ghorbani, D.A. Wood, S. Davoodi, J. Moghadasi, M. Ahmadi Alvar // *Natural Resources Research*. – 2021. – Vol. 30. – P. 3455-3481. (Q1, WoS, Scopus).

9. Davoodi, S. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning / N. Mohamadian, H. Ghorbani, D.A. Wood, M. Mehrad, S. Davoodi, S. Rashidi, A. Soleimanian, A.K. Shahvand // Journal of Petroleum Science and Engineering. – 2021. – Vol. 196. – P. 107811. (Q1, WoS, Scopus).

Свидетельства о государственной регистрации

10. Свидетельство о регистрации программы для ЭВМ № 2022685136 (RU); Заявка № 2022684195 от 05.12.2022, дата рег. 21.12.2022; Бюл. № 1 // Давуди Шадфар. Программа для расчета фильтрационных свойств бурового раствора на основе гибридной модели машинного обучения MELM-COA.
11. Заявка № 2022685740 от 16.12.2022, дата рег. 17.01.2023; Бюл. № 1 // Давуди Шадфар. Программа для расчета пластической вязкости бурового раствора на основе гибридной модели машинного обучения MELM-COA.
12. Свидетельство о регистрации программы для ЭВМ № 2023610692 (RU); Заявка № 2022685760 от 16.12.2022, дата рег. 12.01.2023; Бюл. № 1 // Давуди Шадфар. Программа для расчета динамического напряжения сдвига бурового раствора на основе модели машинного обучения MELM-COA.

Статьи в других изданиях

13. Davoodi, S. Accurate determination of shear wave velocity using LSSVM-GA algorithm based on petrophysical log / H. Ghorbani, A. Davarpanah, S. Davoodi // Conference Proceedings, Third EAGE Eastern Mediterranean Workshop, 1-3 December 2021, Larnaca, Cyprus, – Vol. 201, – P. 1-3. (Scopus).
14. Davoodi, S. Application of an eco-friendly green additive for drilling fluids / S. Davoodi, M. Al-Shargabi, M.K. Minaev // Труды XXVI Международного научного симпозиума студентов и молодых ученых имени академика М.А. Усова «Проблемы геологии и освоения недр», 4-8 апреля 2022 г., г. Томск. – Томск: Изд-во ТПУ. – 2022. – Т. 2. – С. 341-344.
15. Davoodi, S. Application Machine-Learning Predictions Co2 Solubility and Residual Trapping Indexes / S. Davoodi, M. Al-Shargabi // Труды XXVII Международного научного симпозиума студентов и молодых ученых имени академика М.А. Усова «Проблемы геологии и освоения недр», 3-7 апреля 2023 г., г. Томск. – Томск: Изд-во ТПУ. – 2023. – Т. 2. – С. 95-96.