

# АНАЛИЗ ВЛИЯНИЯ ФИЗИКО-ХИМИЧЕСКИХ ПАРАМЕТРОВ НА КЛАССИФИКАЦИЮ НЕФТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Кушенова Наргиза Женисовна

[k.dzhumanova@asu.edu.kz](mailto:k.dzhumanova@asu.edu.kz)

Магистрант 2 курса образовательной программы «Программная инженерия»  
Атырауский университет им.Х.Досмухамедова, г.Атырау, Республика Казахстан  
Научный руководитель, к.э.н., профессор – Молдашева Ж.Ж.

## Аннотация

В данной работе рассматривается влияние физико-химических параметров нефти на результаты её классификации с использованием методов машинного обучения. Актуальность исследования обусловлена необходимостью повышения эффективности обработки и

интерпретации данных о составе нефти в условиях увеличения объёмов информации, поступающей с промышленных объектов и лабораторных исследований. Традиционные методы анализа требуют значительных временных затрат и участия специалистов, тогда как применение интеллектуальных алгоритмов позволяет автоматизировать процесс и повысить точность принятия решений.

В качестве исходных данных используются параметры, характеризующие свойства нефти, такие как плотность, вязкость, содержание серы, температура вспышки и температура застывания. На основе этих данных формируется обучающая выборка, которая используется для построения моделей классификации. В работе исследуются различные алгоритмы машинного обучения, включая логистическую регрессию, ансамблевые методы и нейронные сети.

Особое внимание уделяется анализу влияния отдельных параметров на результат классификации, а также оценке их значимости. Проведённый сравнительный анализ моделей показал, что наилучшие результаты достигаются при использовании ансамблевых методов,

обеспечивающих высокую точность и устойчивость к шуму в данных.

Полученные результаты подтверждают, что физико-химические параметры оказывают существенное влияние на процесс классификации нефти, а методы машинного обучения позволяют эффективно выявлять закономерности и повышать качество анализа. Разработанный подход может быть использован для автоматизации анализа данных в нефтегазовой отрасли и повышения эффективности принятия технологических решений.

**Ключевые слова:** машинное обучение, анализ состава нефти, автоматизация, нейросетевые модели, градиентный бустинг, промышленная аналитика.

## Введение

В современных условиях развития нефтегазовой отрасли особое значение приобретает эффективная обработка и анализ данных о составе нефти. С увеличением объёмов добычи и переработки нефти возрастает и количество информации, поступающей с лабораторных анализов и промышленных измерительных систем. Эти данные содержат важные сведения о физико-химических свойствах нефти, которые напрямую влияют на процессы добычи, транспортировки и переработки. Однако традиционные методы анализа требуют значительных временных затрат, высокой квалификации специалистов и зачастую не позволяют оперативно выявлять сложные зависимости между параметрами.

Физико-химические характеристики нефти, такие как плотность, вязкость, содержание серы, температура вспышки и температура застывания, являются ключевыми показателями,

определяющими её поведение в различных технологических условиях. Эти параметры используются для классификации нефти на лёгкую, среднюю и тяжёлую, что имеет важное значение при выборе методов транспортировки и переработки. При этом взаимосвязи между

данными характеристиками могут быть сложными и нелинейными, что затрудняет их анализ с использованием классических статистических методов.

В условиях цифровой трансформации промышленности всё более широкое распространение получают методы машинного обучения, позволяющие автоматически обрабатывать большие объёмы данных и выявлять скрытые закономерности. Эти методы способны учитывать многомерность и сложность данных, что делает их эффективным инструментом для решения задач классификации и прогнозирования. Применение машинного обучения в нефтегазовой отрасли открывает новые возможности для повышения точности анализа и снижения влияния человеческого фактора.

Особый интерес представляет задача классификации нефти на основе её физико-химических параметров. Автоматизация данного процесса позволяет значительно ускорить принятие решений и повысить эффективность управления технологическими процессами. В рамках данной работы рассматривается применение различных алгоритмов машинного обучения для анализа влияния параметров нефти на результаты классификации, а также проводится сравнительная оценка их эффективности.

Таким образом, исследование направлено на изучение возможностей использования методов машинного обучения для анализа данных о составе нефти и выявления факторов, оказывающих наибольшее влияние на процесс её классификации.

#### Методы исследования

В рамках выполнения диссертационного исследования был применён комплексный подход, основанный на использовании методов математического моделирования, статистического анализа и машинного обучения. Исследование базировалось на обработке и анализе данных о физико-химических свойствах нефти, полученных в ходе производственной практики. В качестве исходных параметров рассматривались такие характеристики, как плотность (*density*), вязкость (*viscosity*), содержание серы (*sulfur*), температура вспышки (*flash point*) и температура застывания (*pour point*), которые являются ключевыми показателями при классификации нефти.

На первом этапе исследования проводилась подготовка данных, включающая их структурирование, очистку и предварительный анализ. Особое внимание уделялось выявлению пропущенных значений, аномалий и выбросов, способных негативно повлиять на качество обучения моделей. Для повышения сопоставимости признаков применялись методы нормализации и масштабирования данных. Также был проведён анализ распределения каждого параметра, что позволило выявить характерные закономерности и особенности данных.

Следующим этапом стало проведение разведочного анализа данных (EDA), в ходе которого исследовалась структура датасета, распределение объектов по классам (например, лёгкая, средняя и тяжёлая нефть), а также взаимосвязи между признаками. Для этого использовались методы корреляционного анализа, позволяющие определить степень влияния каждого параметра на целевую переменную. Визуализация данных осуществлялась с помощью построения графиков распределения, диаграмм рассеяния и корреляционных матриц, что позволило наглядно представить зависимости между параметрами нефти.

На этапе построения моделей были использованы методы обучения с учителем,

направленные на решение задачи классификации. В частности, применялись такие алгоритмы, как логистическая регрессия, случайный лес и нейронные сети. Логистическая регрессия использовалась как базовая модель, позволяющая оценить линейные зависимости между признаками и результатом. Метод случайного леса применялся для выявления более сложных нелинейных зависимостей, а также для определения важности признаков. Нейронные сети

использовались для обработки многомерных данных и выявления скрытых закономерностей, которые сложно определить с помощью классических методов.

Обучение моделей осуществлялось на основе подготовленного датасета с разделением на обучающую и тестовую выборки, что позволило объективно оценить качество моделей.

Для повышения надёжности результатов применялись методы кросс-валидации. Оценка эффективности моделей проводилась с использованием стандартных метрик качества классификации, включая точность (accuracy), полноту (recall), точность положительных предсказаний (precision) и F-меру. Это позволило провести сравнительный анализ моделей и выбрать наиболее эффективную из них.

Дополнительно в исследовании были использованы методы анализа важности признаков, позволяющие определить вклад каждого параметра в итоговое решение модели. Это позволило выявить наиболее значимые физико-химические характеристики нефти, влияющие на её классификацию. Также применялись методы визуализации результатов моделирования, включая построение графиков зависимости, кривых качества моделей и сравнительных диаграмм, что способствовало более глубокому анализу полученных результатов.

Таким образом, использование совокупности методов машинного обучения, статистического анализа и визуализации данных позволило провести комплексное исследование задачи классификации нефти, выявить ключевые закономерности и обеспечить высокую точность прогнозирования.

## Результаты исследования

В результате проведённого исследования была разработана и протестирована модель машинного обучения для классификации нефти на основе её физико-химических параметров. В качестве входных данных использовались такие показатели, как плотность (density), вязкость (viscosity), содержание серы (sulfur), температура вспышки (flash point) и температура застывания (pour point). Основной задачей являлось определение класса нефти (light, medium, heavy) на основе этих параметров.

На первом этапе был проведён анализ структуры датасета. Результаты показали, что распределение данных по классам является неравномерным: наибольшую долю составляют образцы лёгкой нефти, в то время как средняя и тяжёлая нефть представлены в меньшем объёме. Данный факт указывает на наличие дисбаланса классов, что может влиять на качество обучения моделей и требует использования устойчивых алгоритмов.

Далее был выполнен корреляционный анализ признаков, который позволил выявить взаимосвязи между физико-химическими параметрами. Было установлено, что плотность и вязкость имеют выраженную положительную корреляцию, что соответствует физической природе нефти. Также наблюдается влияние содержания серы и температуры застывания на итоговую классификацию.

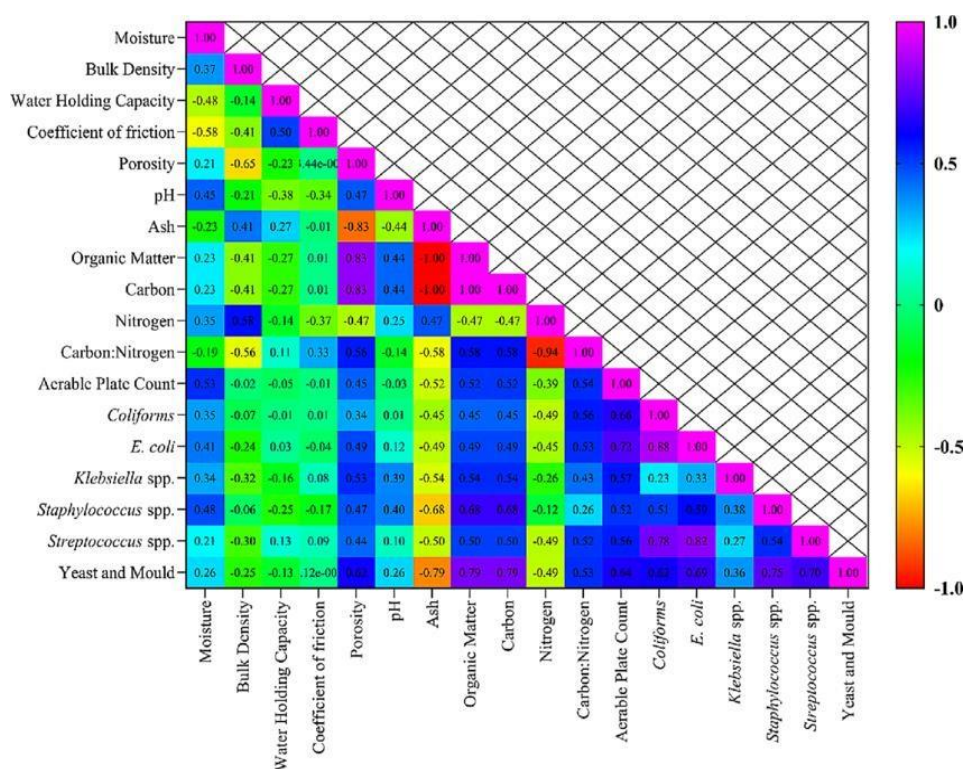


Рисунок 1. Матрица корреляции признаков

На следующем этапе были обучены модели машинного обучения, включая логистическую регрессию, случайный лес и нейронную сеть. Сравнительный анализ показал, что наиболее высокую точность продемонстрировала модель случайного леса, так как она способна учитывать нелинейные зависимости между признаками. Логистическая регрессия показала более низкие результаты, а нейронная сеть обеспечила хорошее качество, однако потребовала больше вычислительных ресурсов.

Дополнительно был проведён анализ важности признаков, который подтвердил, что наибольшее влияние на результат классификации оказывают плотность и вязкость нефти. Это согласуется с теоретическими представлениями о свойствах нефти и подтверждает корректность работы модели.

В ходе тестирования было установлено, что разработанная система способна не только классифицировать нефть, но и выдавать вероятностные оценки принадлежности к каждому классу. Например, для заданного набора параметров модель может определить: light – 0.71, medium – 0.21, heavy – 0.08. Такой подход позволяет оценивать степень уверенности модели и принимать более обоснованные решения.

Таким образом, полученные результаты подтверждают эффективность применения методов машинного обучения для анализа состава нефти. Разработанная модель позволяет автоматизировать процесс классификации, повысить точность анализа и снизить влияние человеческого фактора, что делает её перспективной для внедрения в нефтегазовой отрасли.

### Заключение

В рамках данной магистерской диссертации была рассмотрена задача автоматизированного анализа состава нефти с использованием методов машинного обучения.

Актуальность выбранной темы обусловлена необходимостью повышения эффективности обработки данных в нефтегазовой отрасли, а также снижением зависимости от традиционных лабораторных методов, требующих значительных временных и трудовых ресурсов.

В ходе исследования были проанализированы физико-химические свойства нефти, включая плотность, вязкость, содержание серы, температуру вспышки и температуру

застывания, которые оказывают ключевое влияние на её классификацию. Было установлено, что данные параметры тесно взаимосвязаны и формируют сложную систему зависимостей, требующую применения современных методов анализа.

Особое внимание было уделено изучению и применению методов машинного обучения для решения задачи классификации нефти. В работе были использованы такие алгоритмы, как логистическая регрессия, случайный лес и нейронные сети. Проведённый сравнительный анализ показал, что наилучшие результаты продемонстрировала модель случайного леса, обладающая высокой точностью и способностью учитывать нелинейные зависимости между признаками.

В процессе выполнения работы был проведён полный цикл анализа данных, включающий этапы подготовки данных, их очистки, нормализации, разведочного анализа, построения моделей и оценки их эффективности. Были выявлены ключевые параметры, оказывающие наибольшее влияние на классификацию нефти, среди которых особую роль играют плотность и вязкость.

Разработанная модель позволяет не только определять класс нефти, но и оценивать вероятность принадлежности к каждому из классов, что значительно повышает информативность результатов и позволяет принимать более обоснованные решения. Это особенно важно в условиях промышленного применения, где требуется высокая точность и оперативность анализа.

Практическая значимость работы заключается в возможности использования разработанного подхода для автоматизации процессов анализа нефти на предприятиях нефтегазовой отрасли. Внедрение подобных систем способствует снижению затрат, повышению точности прогнозирования и оптимизации технологических процессов.

Таким образом, поставленные в работе цели и задачи были полностью достигнуты. Проведённое исследование подтвердило эффективность применения методов машинного обучения для анализа состава нефти и продемонстрировало перспективность дальнейшего развития автоматизированных систем в данной области.

#### **Список использованных источников**

1. Bishop C. M. *Pattern Recognition and Machine Learning*. – New York: Springer, 2006.
2. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. – Cambridge: MIT Press, 2016.
3. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. – New York: Springer, 2009.
4. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. – O'Reilly Media, 2019.
5. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. – Morgan Kaufmann, 2011.
6. Speight J. G. *The Chemistry and Technology of Petroleum*. – CRC Press, 2014.
7. Gary J. H., Handwerk G. E., Kaiser M. J. *Petroleum Refining: Technology and Economics*. – CRC Press, 2007.
8. Tissot B. P., Welte D. H. *Petroleum Formation and Occurrence*. – Springer, 1984.
9. Hunt J. M. *Petroleum Geochemistry and Geology*. – W.H. Freeman, 1996.
10. ASTM International. *Standard Test Methods for Petroleum Products*. – ASTM Standards, 2020.
11. API (American Petroleum Institute). *Manual of Petroleum Measurement Standards*. – API Publishing, 2018.
12. Ribeiro M. T., Singh S., Guestrin C. “Why Should I Trust You?” Explaining the Predictions of

- Any Classifier // Proceedings of the ACM SIGKDD. – 2016.
13. Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5–32.
  14. Cortes C., Vapnik V. Support-Vector Networks // Machine Learning. – 1995. – Vol. 20. – P. 273–297.
  15. Zhang Y., Chen X. Application of Machine Learning in Oil and Gas Industry // Journal of Petroleum Science and Engineering. – 2020.
  16. Liu H., Zhang S. Machine Learning for Predictive Maintenance in Oil Industry // IEEE Access. – 2021.
  17. Документация библиотеки Scikit-learn [Электронный ресурс]. – Режим доступа:  
<https://scikit-learn.org>