

Применение методов машинного обучения в обработке данных геофизических исследований скважин отложений викуловской свиты

В.И. Сахнюк*, Е.В. Новиков, А.М. Шарифуллин, В.С. Белохин, А.П. Антонов, М.Ю. Карпушин,
М.А. Большакова, С.А. Афонин, Р.С. Сауткин, А.А. Сулова
Московский государственный университет имени М.В. Ломоносова, Москва, Россия

В настоящее время интерпретация результатов геофизических исследований скважин производится геофизиками-интерпретаторами, которые предварительно обрабатывают данные и нормируют кривые. Процесс подготовки каротажных кривых может занимать большое количество времени особенно в случаях, когда приходится интерпретировать данные по сотням и тысячам скважин. В данной работе исследуется применимость методов машинного обучения в задаче определения литофизических типов по каротажным кривым. В статье рассмотрены три группы алгоритмов: случайный лес, градиентный бустинг и нейронные сети, а также разработана собственная метрика, которая учитывает особенности литофизической типизации исследуемого объекта и основывается на мере близости литофизических типов для фиксированного комплекса методов геофизических исследований скважин.

В результате исследования показано, что алгоритмы машинного обучения способны предсказывать литологию по стандартному набору каротажных диаграмм без нормировки на опорные пласты, что может существенно сократить время на предварительную подготовку кривых.

Ключевые слова: машинное обучение, геофизические исследования скважин, интерпретация каротажа

Для цитирования: Сахнюк В.И., Новиков Е.В., Шарифуллин А.М., Белохин В.С., Антонов А.П., Карпушин М.Ю., Большакова М.А., Афонин С.А., Сауткин Р.С., Сулова А.А. (2022). Применение методов машинного обучения в обработке данных геофизических исследований скважин отложений викуловской свиты. *Георесурсы*, 24(2), с. 230–238. DOI: <https://doi.org/10.18599/grs.2022.2.21>

Введение

Одним из основных источников информации при поиске и разведке на нефть и газ являются методы геофизических исследований в скважинах (ГИС). Интерпретация данных каротажа подразумевает преобразование геофизических параметров, определяемых геофизической аппаратурой в скважине, в «геологические» данные, такие как литологический тип породы, пористость, насыщенность и т.п. Подобное преобразование возможно при соблюдении ряда условий. Во-первых, количество и качество кривых ГИС должно быть достаточным для решения задачи интерпретации. Во-вторых, должен быть разработан алгоритм перехода от физических свойств, измеренных прибором в скважине, к геологическим параметрам. Как правило, это означает наличие установленной связи типа «керна-ГИС».

Процесс интерпретации состоит в тонкой, индивидуальной для каждого пласта или подсчетного объекта настройки зависимостей и критериев, по которым присваивается тип пород, характеризующий литологию в определенном отрезке глубин. Предварительно определяются опорные пласты, в которых снимаются показания кривых для нормирования кривых ГИС и расчета дополнительных

параметров. На геологическую интерпретацию ГИС по одной скважине может уйти от нескольких часов до полного рабочего дня, а если речь идет о месторождении, в стадии полномасштабной разработки с тысячами скважин, то и месяцы.

Для оптимизации временных затрат, учитывая индивидуальность (уникальность) скважинных данных, могут использоваться алгоритмы *машинного обучения*. Явным преимуществом предсказательных моделей машинного обучения является тот факт, что на вход в модель могут идти необработанные каротажные кривые. Этот исходный набор данных называется «признаками» в машинном обучении. Данный подход не требует определения опорных пластов по каждой скважине, калибровки и подсчета новых кривых, таких как aPS (двойной разностный параметр кривой собственной поляризации), dGK (нормализованная кривая гамма-каротажа) и т.д., что существенно снижает объем работ, который необходим для решения этой задачи геофизикам-интерпретаторам.

Рассматриваемая проблема является широко изученной: существующие подходы базируются как на классических моделях машинного обучения (Merembayev et al., 2018; Mohammed et al., 2019; Wu et al., 2018), так и на глубоких нейронных сетях (Peyret et al., 2019; Viggen et al., 2020). В данных работах авторы отталкиваются от имеющегося набора данных, тестируя модели машинного обучения на различных объектах. Вместе с тем следующие важные вопросы – какие данные на конкретном объекте

* Ответственный автор: Владлен Игоревич Сахнюк
e-mail: vladlensakhnyuk@yandex.ru

© 2022 Коллектив авторов

Контент доступен под лицензией Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

представительны, насколько объективно выделение тех или иных классов на основе существующих методов ГИС, какие кривые на исследуемом объекте наиболее информативны – остаются неосвещёнными.

Целью данной работы является построение механизма предсказания литофизических типов породы.

Для достижения цели были поставлены следующие **задачи**:

1. Подготовить набор данных, включающий комплекс ГИС и результаты определения литофизических типов, и провести первичный анализ подготовленных данных;
2. Выбрать алгоритмы машинного обучения;
3. Обучить модели на имеющихся данных;
4. Проанализировать эффективность работы каждого алгоритма на основании метрик качества;
5. Выбрать лучшую модель и методику её дальнейшего применения.

Данная работа посвящена выдающемуся ученому и великому геологу Николаю Брониславовичу Вассоевичу, который внес неоценимый вклад в развитие нефтяной геологии.

Материалы и методы

Геологическим объектом, на котором отрабатывалась предлагаемая методика, являются породы верхней части викуловской свиты раннемелового возраста (K_1a), распространенные в Западной Сибири (рис. 1) и представленные алевро-песчаными, песчано-алевритовыми,

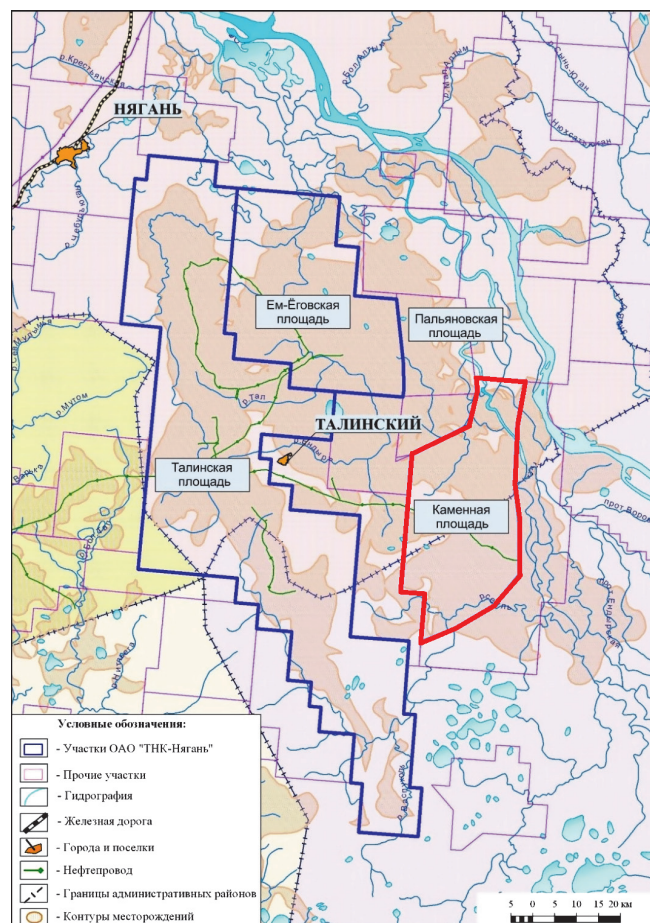


Рис. 1. Обзорная карта исследуемого района (Отчет ТННЦ, 2013 г.)

глинисто-алевролитовыми разностями, неравномерно чередующимися и переслаивающимися по разрезу.

Принятой моделью коллектора является коллектор с микрослоистой текстурной неоднородностью и чередованием слоев алевролитов и песчаников с разной степенью зрелости, при этом глинистая составляющая развита по поровому пространству и цементирующей части, а не в виде отдельных прослоев (Исакова и др., 2020). Породы формировались в прибрежно-морских мелководных и штормовых условиях, при этом по комплексу геофизических данных и сейсмостратиграфическим исследованиям были найдены врезанные речные долины, которые развиты в разных частях Краснотенского свода.

В работе использовались следующие данные по 349 скважинам, вскрывшим викуловские отложения:

1. Результаты проведения геофизических исследований в скважинах, содержащие кривые, представленные в таблице 1;

2. Табличные значения интервалов литологического расчленения разреза и значения литофизических типов (или классов), определенных для каждого интервала.

В результате масштабных лабораторных исследований керна и обработки данных ГИС, группой геофизиков-интерпретаторов была предложена и реализована схема определения литофизических типов интервалов скважин, основанная на значениях двойного разностного параметра кривых гамма-каротажа GK, нейтронного каротажа NK и каротажа собственно поляризации SP (Исакова и др., 2021). К особенностям данного подхода стоит отнести трудоемкость построения двойного разностного параметра, что требует определения в скважинах опорных пластов. А также неудобство использования более полного набора кривых каротажа, что связано с трудностями ручной интерпретации, включающей настройку макросов обработки, содержащих набор условий для различных кривых.

В таблице (табл. 2) представлен набор литофизических типов (классов), используемый в данной работе, а также доля каждого литотипа от общего количества точек литологии. В работе по петрофизической интерпретации

Название кривой	Полное наименование
BK	Кривая бокового каротажа
GR	Кривая гамма-каротажа
SP	Кривая собственных потенциалов
NK	Кривая нейтронного каротажа
GZ 1, 2, 3, 4, 5	Кривые градиент-зондов разной длины
IK	Кривая индукционного каротажа
VIKIZ 1, 2, 3, 4, 5	Кривые высокочастотного индукционного каротажного изопараметрического зондирования
ANK, DGR, ASP, RT	Нормированные на основе опорных пластов кривые нейтронного каротажа ANK, гамма-каротажа DGR, собственных потенциалов ASP и комбинированная кривая сопротивления горной породы RT

Табл. 1. Расшифровка аббревиатур, соответствующих каротажным кривым исследуемого набора данных

Литофизический тип	Доля от общего числа классов литологии, %
Алевро-глинистая порода	24
Алевролит	20
Глина	19
Песчано-алевролитовая порода	11
Песчаник	9
Глинисто-алевролитовая порода	8
Карбонатная порода	5
Алевро-песчаная порода	4

Табл. 2. Литофизические типы (классы) викуловской свиты и их доля от общего количества литотипов

каротажных диаграмм данного района (Исакова и др., 2021) был использован схожий набор литологии. В основном авторы использовали упомянутую статью как пример и работали со своими данными, что и объясняет расхождение в количестве представленных литотипов.

Предварительный анализ данных

На первом этапе анализа данных была построена усредненная матрица корреляции Пирсона для всех методов ГИС (рис. 2), которая отражает степень линейной зависимости между кривыми (Рашка, Мирджалили, 2020). При этом методы сопротивления (БК, ИК, VIKIZ1-5, GZ1-5) были переведены в логарифмический масштаб.

Из данной матрицы видно, что методы сопротивлений сильно коррелируют друг с другом (коэффициент больше 0.7). Линейная зависимость признаков является нежелательным фактором в задаче предсказания литологии, так как может негативно сказываться на разделении классов. Таким образом, для решения задачи определения литофизических типов целесообразно исключить из дальнейшего рассмотрения линейно зависимые методы, поэтому далее авторы решили использовать набор наименее коррелируемых кривых: SP, GR, NK и BK.

Затем все кривые были стандартизованы следующим образом: из всех значений каждого метода ГИС вычиталось среднее значение, и проводилась нормировка

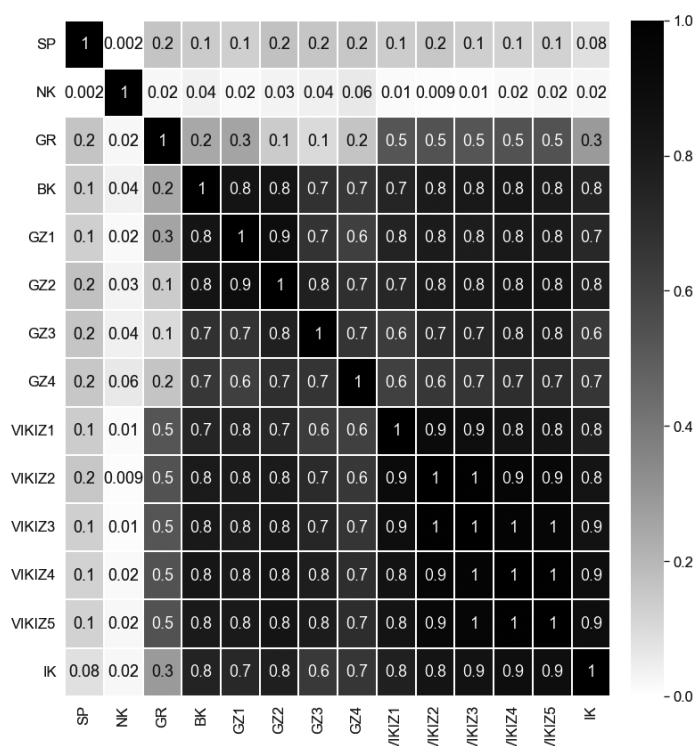


Рис. 2. Матрица корреляции Пирсона для методов ГИС

на стандартное отклонение этого метода, взятое в интервале исследования. Такая процедура позволяет привести все методы к единому масштабу значений, что с одной стороны позволяет учесть различия в калибровках геофизических методов каротажа, а с другой проводить сравнение и математические операции между различными методами ГИС.

Далее для стандартизованных кривых были вычислены средние значения для каждого литофизического типа (рис. 3).

Визуальный анализ графика показал, что классы в основном разделимы, т.е. характеризуются различными значениями признаков, но при этом наблюдается явная «близость» одних классов и «удаленность» других. Так, наиболее удалены друг от друга карбонатная порода и

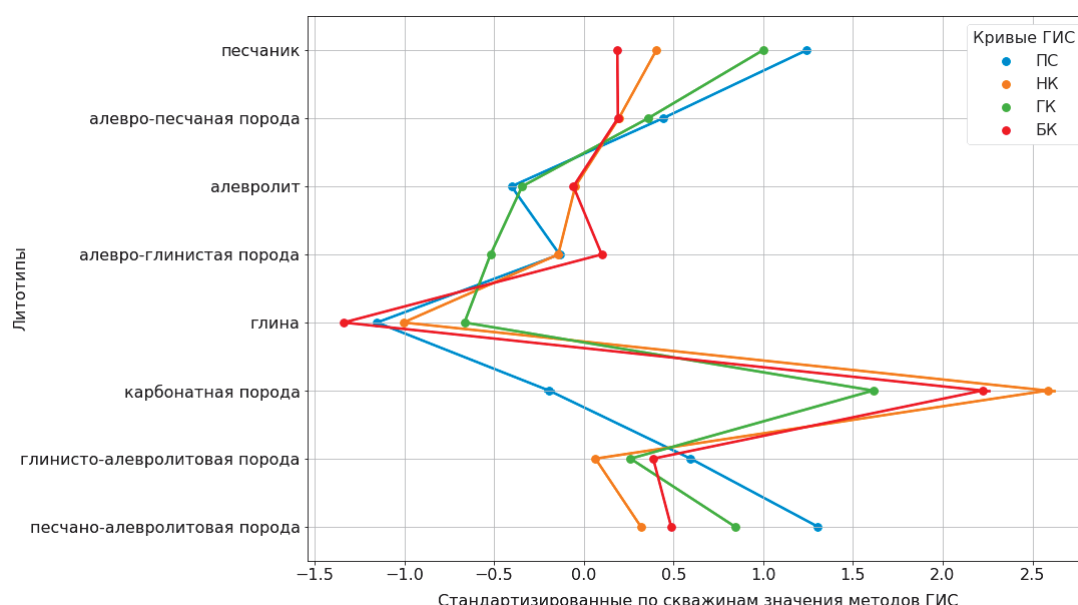


Рис. 3. Средние по классам стандартизованные в пределах скважин значения методов ГИС

глина. А например, песчаник и песчано-алевролитовая порода напротив близки.

Для численной оценки расстояния между классами были рассчитаны коэффициенты «близости», представляющие собой евклидово расстояние в 4-мерном пространстве средних значений методов ГИС для каждого класса. Другими словами, каждый класс охарактеризован 4 средними значениями методов ГИС, и эти значения определяют вектор в четырехмерном пространстве, базисом в котором служат методы ГИС. А расстояние между векторами, как раз и характеризует меру близости литофизических типов (рис. 4).

Значения, изображенные на рисунке 4, тем больше, чем больше разница между классами на основе исследуемых данных, что в целом соответствует гипотезе о том, что выбранный подход позволяет учесть «близость» рассматриваемых классов. Согласно рассчитанному параметру наиболее близкими оказались алевро-глинистая порода и алевролит (значение 0.15), дальше всего – глина и карбонатная порода (значение 3.6).

Полученные результаты далее использовались в специально разработанной для задачи предсказания литологии метрике качества, где учитывалась «близость» рассматриваемых классов исходя из значений кривых ГИС с использованием рассчитанных весов. Для визуализации эмпирически было установлено граничное значение, с которым сравнивался каждый вес (1.5). В случае, если вес больше граничного, предсказанная литология помечалась флагом «разделимый класс», в ином случае – «неразделимый» (рис. 5).

Методика

Для решения задачи предсказания литологии по каротажным кривым были использованы алгоритмы

машинного обучения. В данной работе использовалось 3 группы моделей: случайный лес (реализация при помощи открытой библиотеки scikit-learn), градиентный бустинг (реализация Catboost) и нейронные сети (собственной архитектуры).

Случайный лес (Random Forest) (Breiman, 2001). Главная идея метода заключается в построении множества независимых деревьев решений, обучении их на тренировочной выборке и получении предсказания литологии. Алгоритм использует метод выборки с возвращением: из исходного набора данных извлекается некоторое их количество, на котором обучается простой алгоритм решающих деревьев, далее процесс повторяется. Эта методика применяется для предотвращения переобучения математической модели, то есть чтобы модель предсказывала приблизительно одинаково любые тестовые данные. Итоговое предсказание основано на усреднении предсказаний простых алгоритмов.

Градиентный бустинг (Gradient Boosting) (Friedman, 2001). В настоящее время данная группа алгоритмов пользуется популярностью в силу своей надежности и предсказательной способности. Метод также основан на решающих деревьях: простые алгоритмы комбинируются последовательно, причем следующий учится на основе ошибок предыдущего. В данной работе используется модель CatBoost (Prokhorenkova et al., 2019), которая, по сравнению с другими реализациями, является наиболее эффективной и показывает себя отлично в задачах классификации и регрессии.

Нейронные сети (Artificial Neural Networks) (Haykin, 1994; Schmidhuber, 2015). Нейронные сети являются алгоритмическим аналогом простейшей модели работы мозга, они состоят из нескольких связанных друг с другом слоев, которые позволяют постепенно находить сложные

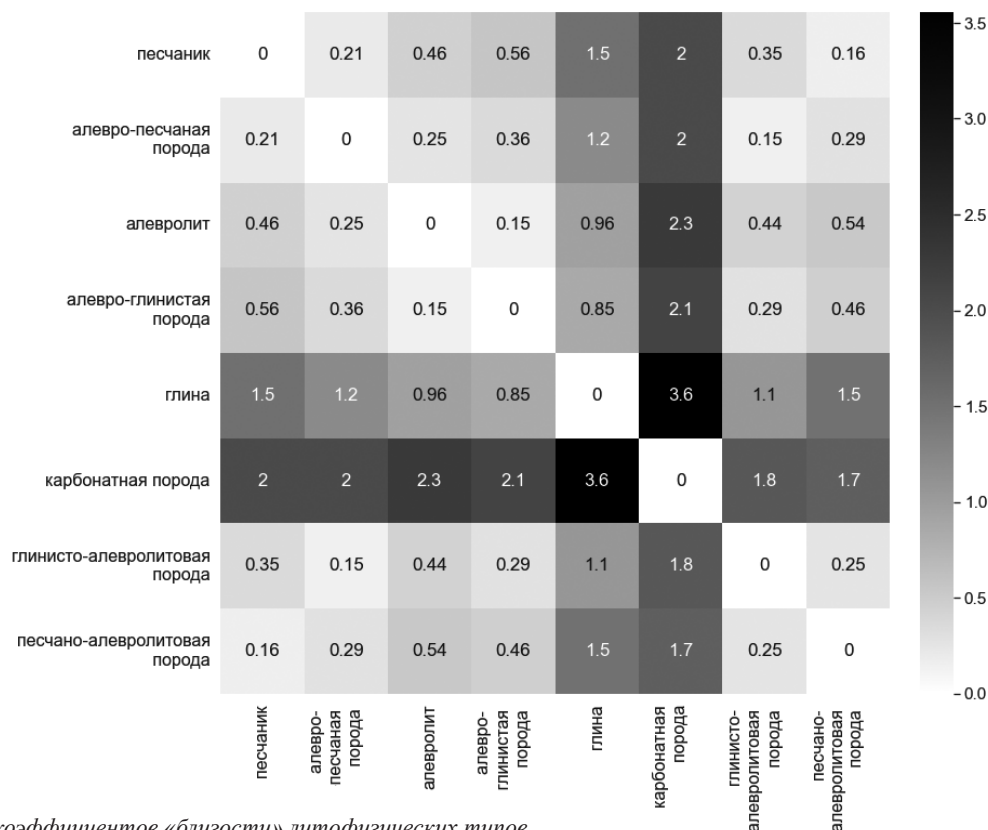


Рис. 4. Значения коэффициентов «близости» литофизических типов

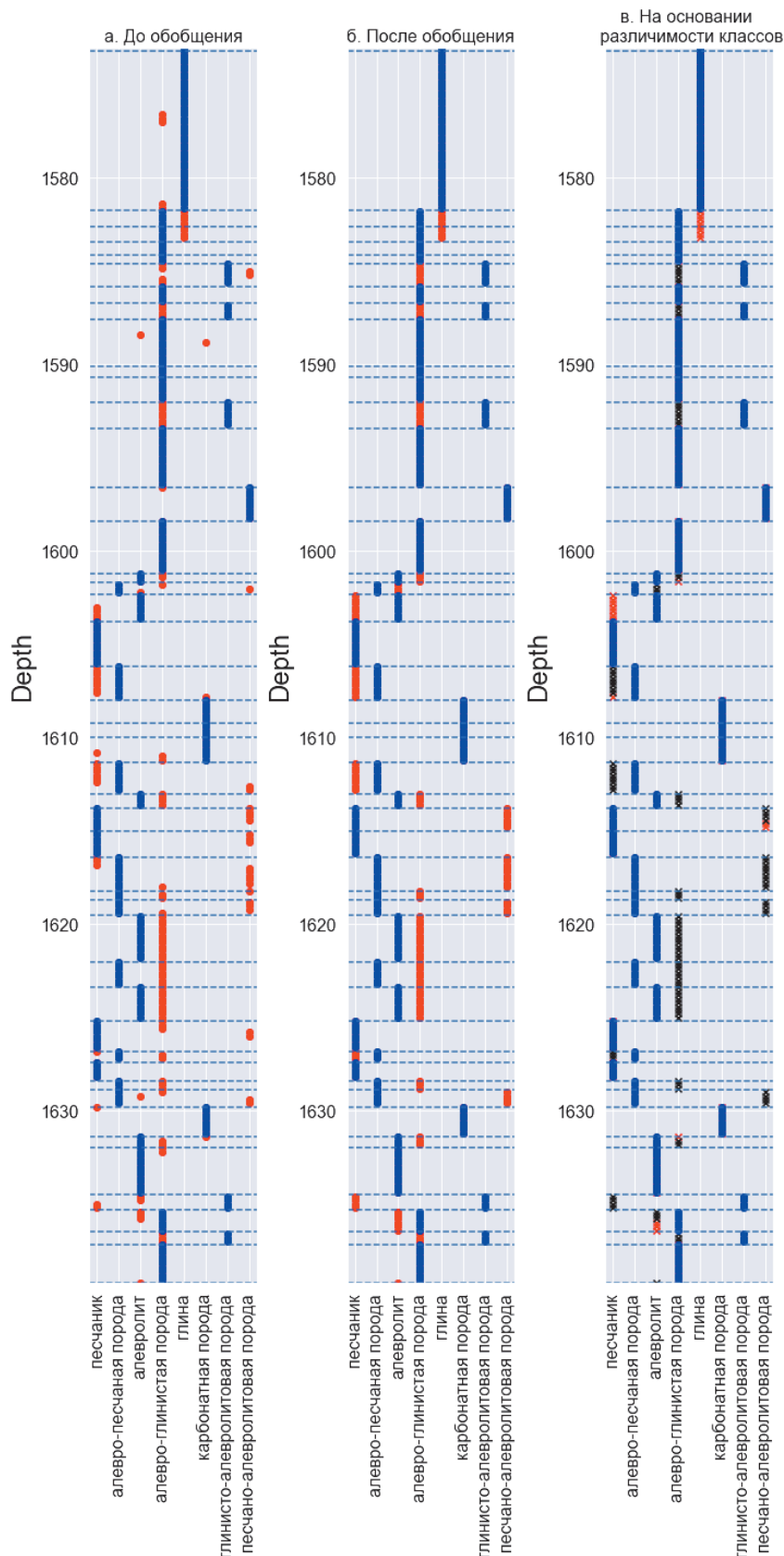


Рис. 5. Результаты работы алгоритма предсказания литологии (Random Forest): а) без учета разделимости классов до обобщения предсказаний литологии (красные точки – неверные предсказания, синие – верные); б) без учета разделимости классов после обобщения (красные точки – неверные предсказания, синие – верные); в) на основании разделимости классов после обобщения (черные точки – предсказание неверное, но «близкое» (мера «близости» менее 1.5), красные точки – предсказание неверное, классы обладают большим значением меры «близости» (более 1.5), синие – верные).

зависимости. В данной работе для классификации используется многослойный перцептрон (*Multilayer Perceptron – MLP*) с архитектурой, включающей 3 последовательно соединенных полносвязных слоя нейронов.

В качестве метрик, по которым можно оценить работу алгоритмов, были использованы доля верных ответов (accuracy) и F1-мера:

$$accuracy = \frac{\sum_{i=1}^n [y_i = \bar{y}_i]}{n}$$

где y_i – действительное значение, $\bar{y}_i$ – предсказанное на основе модели значение, n – количество измерений;

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

где *precision* – точность модели (доля действительных значений литологии среди всех предсказанных значений данного класса), *recall* – её полнота (доля найденных классификатором литотипов, принадлежащих к классу, относительно всех литотипов этого класса).

Помимо этого, дополнительно было принято решение ввести адаптивную метрику, которая учитывает геологические особенности исследуемого объекта. В разделе «Предварительный анализ данных» была произведена оценка сходства литологии (рис. 4), в результате чего разработана матрица весов (как ранее уже говорилось, представляющих собой оценку меры «близости» рассматриваемых литотипов на основании наблюдаемых физических значений) для разных литотипов. Данная метрика штрафует предсказания алгоритма тем больше, чем больше значение оценки меры «близости» литофизических типов. Так, если используемый алгоритм предсказал значение, которое статистически разделимо с действительным типом литологии (к примеру, модель предсказала карбонат, тогда как на деле это песчаник), то метрика штрафует такое предсказание. В ином случае (алевритистый песчаник и песчаник) метрика реагирует слабо. Финальное значение варьируется от 0 до 1 и интерпретируется как близость предсказанной литологии к истинной, где 0 – модель не может предсказывать даже близкие литотипы, 1 – модель точно предсказывает всю литологию, включая неразделимые классы:

$$accuracy' = 1 - \frac{\sum_{i=1}^n (e^{ij} / \max(e^i))}{n}$$

где $accuracy'$ – суммарный итоговый результат оценки, e^{ij} – значение из матрицы весов e на пересечении i -того истинного класса и j -того предсказанного,

e^i – значение из матрицы весов e на пересечении i -того истинного класса и класса, являющегося самым далеким от i -го истинного и, тем самым, дающего максимальную ошибку в суммму.

Результаты и обсуждение

В данном разделе приведены результаты, а также сравнительный анализ работы алгоритмов машинного обучения с ненормированными кривыми и нормированными в задаче предсказания литотипов.

Важно отметить, что геофизик-интерпретатор выполняет интерпретацию ГИС попластово. Алгоритм машинного обучения же выдает предсказание для каждой точки кривых ГИС, не учитывая их принадлежность к пласту. Поэтому было принято решение обобщить предсказания для одного пласта по принципу наиболее часто встречающегося класса литологии (таким образом, если в пласте предсказаны 10 точек песчаника, 2 точки аргиллита и 5 точек алевролита, то обобщенным будет предсказание песчаника).

Помимо этого, для оценки качества применялась специально разработанная адаптивная метрика (раздел «Методы»), учитывающая разделимость классов.

Пример предсказания литологии и визуализации применяемых метрик представлен на рис. 5.

Тест проводился на 61 скважине (22561 точка) с полным набором кривых (ненормированных – SP, GR, NK, BK, преобразованных (нормированных + кривая RT) – ANK, DGR, ASP, RT), а также глубины пластов и метки литологии для каждого пласта. Метрики качества, полученные по каждой скважине, усреднялись. Обучение производилось на 242 скважинах (86191 точка).

Результаты работы классификационных алгоритмов представлены в таблице 3.

Исходя из этого можно сделать выводы, что во всех алгоритмах предсказания на ненормированных кривых метрики качества относительно равны метрикам при предсказании литологии на нормированных. Разница предсказания не столь велика, при этом время, потраченное на нормирование кривых, существенно. Следовательно, алгоритмы машинного обучения способны при минимальной потере качества работы повысить эффективность за счет сокращения времени на работы интерпретатора. Также подход с использованием машинного обучения является

помощником для интерпретатора при оперативном просмотре качества данных.

На крайних правых графиках рис. 6 представлены литологические колонки предсказаний для одной скважины разными методами. Надо отметить, что от скважины к скважине показатели точности варьируются в среднем от 0.7 до 0.9. Это объясняется тем, что в наборе данных есть скважины, в которых некоторые типы литологии не представлены, однако алгоритм научен выделять данные типы (поскольку в тренировочных данных эти типы есть).

Важной особенностью алгоритмов случайного леса и градиентного бустинга является возможность определить вклад каждой кривой в предсказательную способность моделей при помощи встроенного атрибута feature importances (рис. 7). Величины значимости каждой кривой на рисунках показывают, на какие кривые алгоритмы акцентируют свое внимание в первую очередь при предсказании литологии.

Малые показатели значимости кривой свидетельствуют о том, что точность алгоритма незначительно уменьшится при исключении данной кривой из списка входных параметров. К примеру, для градиентного бустинга метрика (доля ошибок после попластового округления) без кривой PS будет составлять 0.61, что всего лишь на 3 % меньше предсказания со всеми четырьмя кривыми.

Заключение

По итогам проведенных исследований получены следующие результаты.

По данным корреляционного анализа представленных методов ГИС, установлено, что линейно независимыми, а значит, пригодными для математического моделирования, являются кривые GR, BK, SP, NK.

Все рассматриваемые алгоритмы машинного обучения (случайный лес, градиентный бустинг и многослойный перцептрон) в целом способны улавливать основные закономерности и пригодны для задачи предсказания литотипов. В дальнейшем рекомендуется настройка параметров выбранных алгоритмов, что гарантированно повлияет на качество работы алгоритмов.

Предсказание литологии на ненормированных кривых (GR, SP, NK, BK) не сильно отличается от предсказания на нормированных (DGR, ASP, ANK, RT).

Модель (метрика)	Нормированные кривые, в д.е.	Ненормированные (“сырые”) кривые, в д.е.
CatBoost (accuracy)	0.65	0.60
CatBoost (accuracy после попластового округления)	0.69	0.64
CatBoost (адаптивная метрика)	0.93	0.91
CatBoost (F1)	0.45	0.52
Random Forest (accuracy)	0.63	0.58
Random Forest (accuracy попластового округления)	0.69	0.63
Random Forest (адаптивная метрика)	0.93	0.90
Random Forest (F1)	0.43	0.49
MLP (accuracy)	0.60	0.58
MLP (accuracy после попластового округления)	0.62	0.62
MLP (адаптивная метрика)	0.91	0.91
MLP (F1)	0.37	0.43

Табл. 3. Результаты работы алгоритмов классификации на двух наборах кривых (преобразованных (ANK, DGR, ASP, RT) и ненормированных (SP, NK, BK, GR))

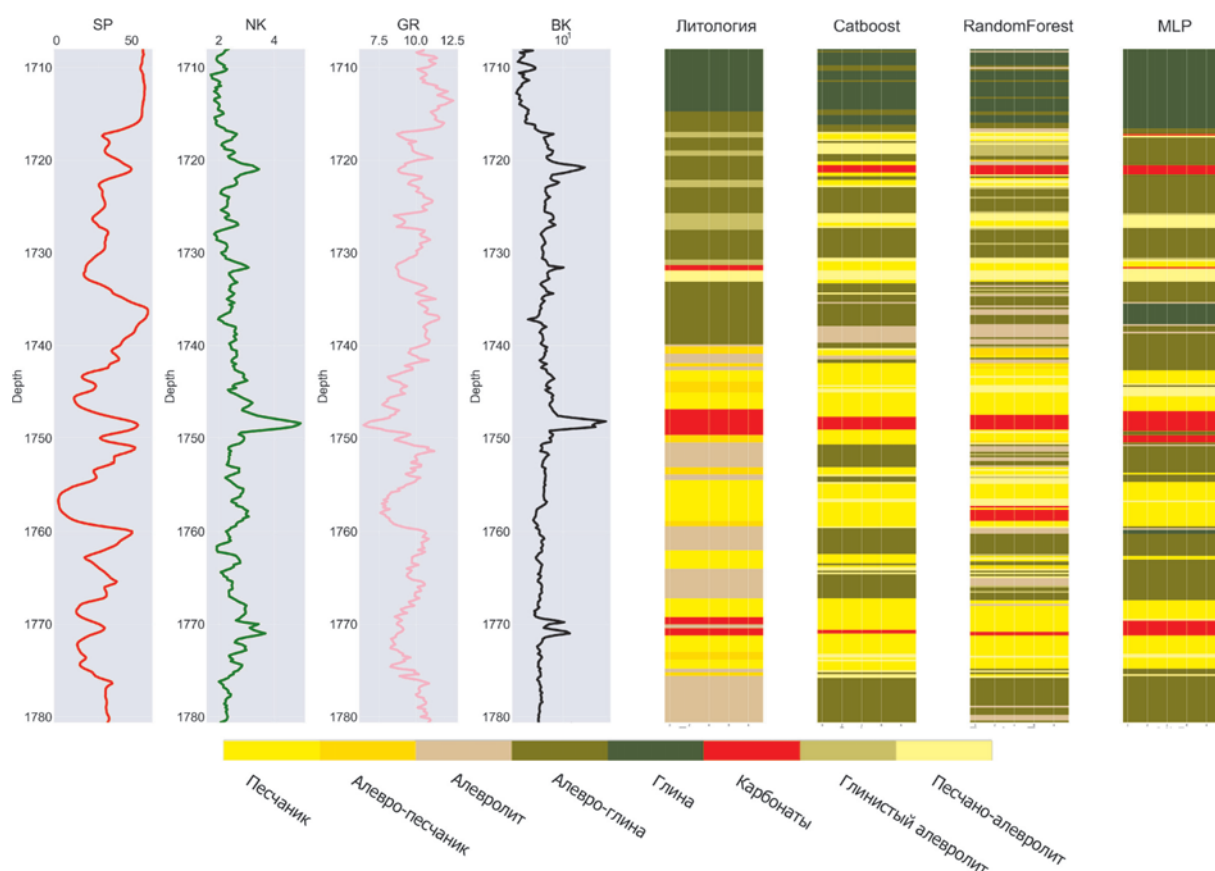


Рис. 6. Результат работы алгоритмов: справа – колонки литологии для каждого метода, слева – каротажные диаграммы 4-х ненормированных кривых

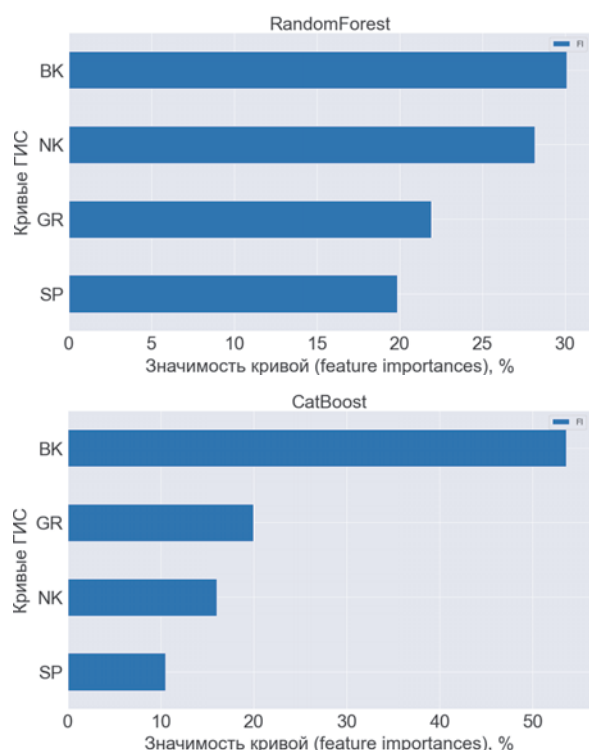


Рис. 7. Значения важности кривых для алгоритмов случайного леса RandomForest и градиентного бустинга CatBoost.

Для оценки работы алгоритмов разумно использовать метрику, которая учитывает степень различия литотипов и штрафует за предсказание литотипов, которые далеки друг от друга по значениям кривых.

Финансирование/Благодарности

Исследование выполнено при финансовой поддержке ПАО «НК «Роснефть» в рамках научного проекта – грант на тему «Применение методов машинного обучения в обработке данных геофизических исследований скважин».

Авторы статьи признательны рецензенту за потраченное время и рецензию.

Литература

- Исакова Т.Г., Дьяконова Т.Ф., Носикова А.Д., Калмыков Г.А., Акиншин А.В., Яценко В.М. (2021). Прогнозная оценка фильтрационной способности тонкослоистых коллекторов викуловской свиты по результатам исследования керна и ГИС. *Георесурсы*, 23(2), с. 170–178. <https://doi.org/10.18599/grs.2021.2.17>
- Рашка С., Мирджалили В. (2020). Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2. 3-е изд. СПб: ООО «Диалектика», 848 с.
- Breiman L. (2001). Random Forests. *Machine Learning* 45, p. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Friedman J. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, 29(5), pp. 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Haykin S. (1994). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River.
- Merembayev T. Yunussov R. and Amirgaliyev Y. Machine learning algorithms for classification geology data from well logging. *14th International Conference on Electronics Computer and Computation (ICECCO)*, pp. 206–212. <https://doi.org/10.1109/ICECCO.2018.8634775>
- Mohamed I.M., Mohamed S., Mazher I. et al. (2019). Formation lithology classification: insights into machine learning methods. *SPE Annual Technical Conference*. <https://doi.org/10.2118/196096-MS>
- Peyret A.P., Ambía J., Torres-Verdin C. et al. (2019). Automatic Interpretation of Well Logs with Lithology-Specific Deep-Learning Methods. *SPWLA 60th Annual Logging Symposium*. https://doi.org/10.30632/T60ALS-2019_SSSS

Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. (2019). CatBoost: unbiased boosting with categorical features. <https://doi.org/10.48550/arXiv.1706.09516>

Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, pp. 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

Viggen E.M., Merciu I.A., Løvstakken L. et al. (2020). Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *Journal of Petroleum Science and Engineering*, 195. <https://doi.org/10.1016/j.petrol.2020.107539>

Wu P., Jain V., Kulkarni M.S. et al. (2018). Machine learning-based method for automated well log processing and interpretation. *SEG Technical Program Expanded Abstracts*. <https://doi.org/10.1190/segam2018-2996973.1>

Сведения об авторах

Владлен Игоревич Сахнюк – магистрант кафедры геологии и геохимии горючих ископаемых

Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Евгений Владимирович Новиков – магистрант кафедры геологии и геохимии горючих ископаемых

Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Александр Маратович Шарифуллин – магистрант кафедры геологии и геохимии горючих ископаемых

Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Василий Сергеевич Белохин – кандидат физ.-мат. наук, научный сотрудник кафедры геологии и геохимии горючих ископаемых

Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Алексей Петрович Антонов – кандидат физ.-мат. наук, доцент кафедры математического анализа, руководитель НОЦ ПАО «НК «Роснефть» по цифровым технологиям в нефтегазовой отрасли на базе кафедры геологии и геохимии горючих ископаемых, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Михаил Юрьевич Карпушин – геолог, сотрудник кафедры геологии и геохимии горючих ископаемых, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Мария Александровна Большакова – кандидат геол.-мин. наук, старший научный сотрудник кафедры геологии и геохимии горючих ископаемых, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Сергей Александрович Афонин – кандидат физ.-мат. наук, доцент кафедры вычислительной математики, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Роман Сергеевич Сауткин – кандидат геол.-мин. наук, старший научный сотрудник кафедры геологии и геохимии горючих ископаемых, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Анна Анатольевна Суслова – кандидат геол.-мин. наук, ведущий научный сотрудник кафедры геологии и геохимии горючих ископаемых, Московский государственный университет имени М.В. Ломоносова

Россия, 119234, Москва, ул. Ленинские горы, д. 1

Статья поступила в редакцию 28.02.2022;

Принята к публикации 19.04.2022; Опубликована 16.05.2022

IN ENGLISH

ORIGINAL ARTICLE

Machine learning applications for well-logging interpretation of the Vikulov Formation

V.I. Sakhnyuk, E.V. Novickov, A.M. Sharifullin, V.S. Belokhin, A.P. Antonov, M. U. Karpushin, M.A. Bolshakova, S.A. Afonin, R.S. Sautkin, A.A. Suslova*

Lomonosov Moscow State University, Moscow, Russian Federation

**Corresponding author: Vladlen I. Sakhnyuk, email: vladlensakhnyuk@yandex.ru*

Abstract. Nowadays well logging curves are interpreted by geologists who preprocess the data and normalize the curves for this purpose. The preparation process can take a long time, especially when hundreds and thousands of wells are involved. This paper explores the applicability of Machine Learning methods to geology tasks, in particular the problem of lithology interpretation using well-logs, and also reveals the issue of the quality of such predictions in comparison with the interpretation of specialists. The authors of the article deployed three groups of Machine Learning algorithms: Random Forests, Gradient Boosting and Neural Networks, and also developed its own metric that takes into account the

geological features of the study area and statistical proximity of lithotypes based on log curves values.

As a result, it was proved that Machine Learning algorithms are able to predict lithology from a standard set of well logs without calibration on reference layers, which significantly saves time spent on preliminary preparation of curves.

Keywords: machine learning, well logging, logging interpretation

Acknowledgements

The study was carried out with the financial support of Rosneft Oil Company as part of the scientific project – a grant

on the topic “Application of machine learning methods in the processing of well logging data”.

Recommended citation: Sakhnyuk V.I., Novickov E.V., Sharifullin A.M., Belokhin V.S., Antonov A.P., Karpushin M. U., Bolshakova M.A., Afonin S.A., Sautkin R.S., Suslova A.A. (2022). Machine learning applications for well-logging interpretation of the Vikulov Formation. *Georesury = Georesources*, 24(2), pp. 230–238. DOI: <https://doi.org/10.18599/grs.2022.2.21>

References

- Breiman L. (2001). Random Forests. *Machine Learning* 45, p. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Friedman J. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, 29(5), pp. 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Haykin S. (1994). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River.
- Isakova T.G., Dyakonova T.F., Nosikova A.D., Kalmykov G.A., Akinshin A.V., Yatsenko V.M. (2021). Predictive assessment of the fluid loss properties of thin-layer reservoirs of Vikulovskaya series based on the results of core and well logs. *Georesury = Georesources*, 23(2), pp. 170–178. <https://doi.org/10.18599/grs.2021.2.17>
- Merembayev T. Yunussov R. and Amirgaliyev Y. Machine learning algorithms for classification geology data from well logging. *14th International Conference on Electronics Computer and Computation (ICECCO)*, pp. 206–212. <https://doi.org/10.1109/ICECCO.2018.8634775>
- Mohamed I.M., Mohamed S., Mazher I. et al. (2019). Formation lithology classification: insights into machine learning methods. *SPE Annual Technical Conference*. <https://doi.org/10.2118/196096-MS>
- Peyret A.P., Ambía J., Torres-Verdín C. et al. (2019). Automatic Interpretation of Well Logs with Lithology-Specific Deep-Learning Methods. *SPWLA 60th Annual Logging Symposium*. https://doi.org/10.30632/T60ALS-2019_SSSS
- Raschka S., Mirjalili V. (2019). *Python machine learning. Machine Learning and Deep Learning with Python, scikit-learn and TensorFlow 2*. Birmingham: Packt Publishing Ltd, 741 p.
- Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. (2019). CatBoost: unbiased boosting with categorical features. <https://doi.org/10.48550/arXiv.1706.09516>
- Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, pp. 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Viggen E.M., Merciu I.A., Løvstakken L. et al. (2020). Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *Journal of Petroleum Science and Engineering*, 195. <https://doi.org/10.1016/j.petrol.2020.107539>
- Wu P., Jain V., Kulkarni M.S. et al. (2018). Machine learning-based method for automated well log processing and interpretation. *SEG Technical Program Expanded Abstracts*. <https://doi.org/10.1190/segam2018-2996973.1>

About the Authors

Vladlen I. Sakhnyuk – Graduate student, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Evgeniy V. Novikov – Graduate student, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Alexander M. Sharifullin – Graduate student, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Vasily S. Belokhin – PhD (Physics and Mathematics), Researcher, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Alexey P. Antonov – PhD (Physics and Mathematics), Associate Professor of Mathematical Analysis Department, Head of Rosneft Research Center, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Mikhail U. Karpushin – Geologist, Researcher, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Maria A. Bolshakova – PhD (Geology and Mineralogy), Senior Researcher, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Sergey A. Afonin – PhD (Physics and Mathematics), Associate Professor, Department of Computational Mathematics, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Roman S. Sautkin – PhD (Geology and Mineralogy), Senior Researcher, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

Anna A. Suslova – PhD (Geology and Mineralogy), Leading Researcher, Petroleum Geology Department, Lomonosov Moscow State University 1, Leninskie gory, Moscow, 119234, Russian Federation

*Manuscript received 28 February 2022;
Accepted 19 April 2022; Published 16 May 2022*