

УДК 550.8.012

## МЕТОД РЕГИОНАЛЬНОГО ПРОГНОЗА НЕФТЕГАЗОНОСНОСТИ ТЕРРИТОРИЙ АЛГОРИТМАМИ МАШИННОГО ОБУЧЕНИЯ НА ПРИМЕРЕ ТЮМЕНСКОЙ СВИТЫ ЗАПАДНОЙ СИБИРИ

Ивлев Дмитрий Александрович,  
dm.ivlev@gmail.com

Бристольский университет,  
Великобритания, BS8 1SE, Клифтон, Бристоль, Куинс-Роуд, Бикон-Хаус.

**Актуальность** исследования обусловлена сокращением фонда структурных ловушек и необходимостью расширения ресурсной базы углеводородов за счет повышения эффективности поиска и разведки месторождений в сложнопостроенных залежах нефти и газа.

**Цель:** на примере прогноза нефтегазоносности территории исследования показать методику прогноза и набор применяемых технологических решений и алгоритмов.

**Объект:** отложения средней юры (тюменская свита) Западной Сибири в пределах района (700×900 км), включающего в себя части Ямало-Ненецкого и Ханты-Мансийского административных округов и Томской области.

**Методы.** С помощью алгоритмов машинного обучения и комплексирования технологического набора методов (геоинформатики, бассейнового моделирования, экспертных оценок) показаны следующие этапы реализации методики прогноза: 1) генерирование признакового пространства изучаемой территории на основе повышения пространственного разрешения структурных построений с применением алгоритмов генеративно-сопоставительной архитектуры нейронных сетей, где в качестве эталонных участков использованы результаты сейсморазведки 3D; 2) отбор признаков статистическим методом и методами машинного обучения; 3) создание подмножества моделей прогноза на основе градиентного бустинга над решающими деревьями; 4) объединение их в метамодель путем стекового обобщения логистической регрессией.

**Результаты.** Формализован и апробирован подход к региональному прогнозу. Сделан прогноз вероятности нефтегазоносности тюменской свиты на изучаемой территории. На его основе и информации по открытым месторождениям оценена методом Монте-Карло ресурсная база УВ. Результаты представлены в виде суммарной таблицы геологических и извлекаемых ресурсов для вероятностей P10, P50, P90 в сравнении с категориями запасов ABC<sub>1</sub> и ABC<sub>1</sub>+C<sub>2</sub> месторождений, числящихся на государственном балансе на территории исследования. В качестве примера приведены графические материалы результатов: работы алгоритма повышения пространственного разрешения; моделирования осадконакопления; моделирования миграции УВ; карта прогноза перспектив нефтегазоносности для северной части Надымского и Пууровского нефтегазовых районов.

### Ключевые слова:

Тюменская свита, средняя юра, машинное обучение, искусственный интеллект, генерация признаков, выбор признаков, градиентный бустинг, генеративно-сопоставительные нейросети, региональный прогноз нефтегазоносности, вероятность геологического успеха, сырьевая база УВ, геоинформатика, бассейновое моделирование, моделирование осадконакопления, моделирование миграции УВ.

### Введение

В традиционных нефтегазовых провинциях РФ на большинстве антиклинальных структур завершены испытанием скважин поисковые работы. Этап открытия крупных месторождений пройден, и основные перспективы по наращиванию ресурсной базы связаны с открытием новых месторождений с мелкими сложнопостроенными залежами УВ.

Относительно высокая степень изученности бурением скважин и сейсморазведочными работами определённой части крупных нефтегазовых регионов и текущий уровень развития технологии анализа данных и машинного обучения позволяют перейти на другой масштаб качественного прогноза нефтегазоносности территорий.

Методика прогноза нефтегазоносности методами машинного обучения имеет богатое прошлое в отечественной геологической науке. Уже в 70-е гг. XX в. с помощью методов машинного обучения (в терминологии того периода – методы распознавания образов) сделаны прогнозы перспективности нефтегазоносности земель для крупных нефтегазовых провинций

[1, 2]. Впоследствии данный подход получил развитие в работах отечественных ученых и был включен в руководства прогнозирования нефтегазоносности как один из математических методов [3].

Предлагаемый в работе подход (рис. 1) является частным случаем в математических методах количественной оценки ресурсов нефти и газа с основой на методах бинарной классификации с учителем. Однако в контексте классификации методов предлагаемый подход скорее является гибридным, так как может использоваться в качестве предикторов любой пространственный признак (атрибут) изучаемой территории, генерируемый самыми разными методами: геологической аналогией, объемно-статистическими, объемно-генетическими, вероятностно-статистическими, экспертными и др.

Цель работы – на примере прогноза нефтегазоносности отложений средней юры (тюменская свита) Западной Сибири в пределах района площадью 700×900 км, расположенного на территориях Ямало-Ненецкого, Ханты-Мансийского административных округов и Томской области, показать методику прогноза и набор применяемых технологических решений.

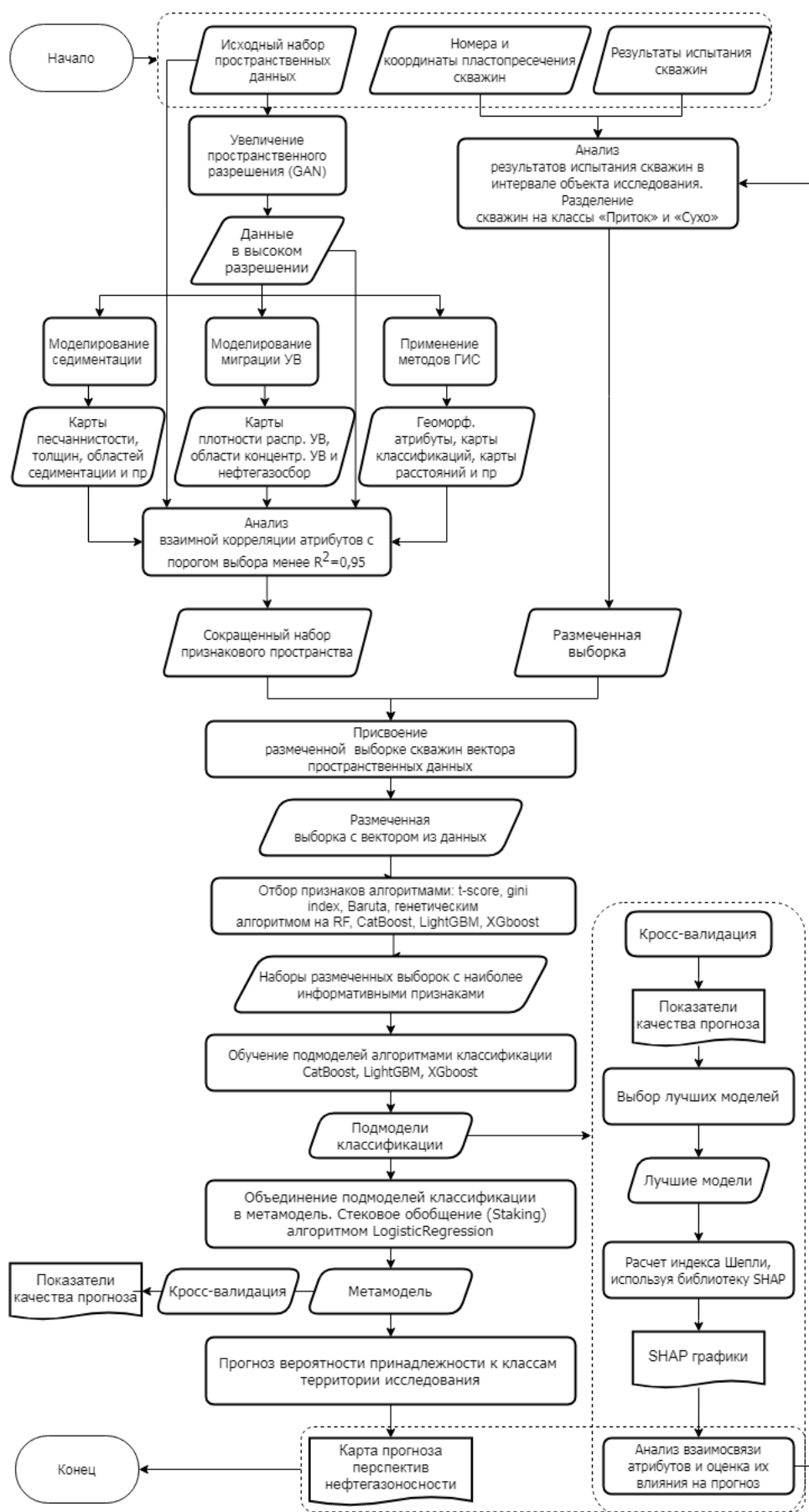


Рис. 1. Алгоритм прогноза перспектив нефтегазоносности территорий  
 Fig. 1. Algorithm for forecasting the prospects for oil and gas potential of territories

Текущая добыча углеводородов из тюменской свиты относительно общей для Западной Сибири не значительна, но в перспективе десятилетия ее объем будет увеличиваться в связи с истощением месторождений в меловых продуктивных горизонтах, а практическая важность поиска и разведки на объектах средней юры, с целью компенсации выпадающих объемов, значительно возрастет.

Объект исследования – средняя юра (тюменская свита), в стратиграфических региональных схемах подразделяется на нижнюю, среднюю и верхнюю подсвиты.

Формирование среднеюрских отложений произошло в условиях континентального и переходного режима седиментации за счет заполнения впадин осадками и размыва древних выступов фундамента, что приводило к выравниванию палеорельефа. В конце средней юры континентальный режим сменился морским. Пласты тюменской свиты (Ю<sub>2-9</sub>) характеризуются резкой фациальной изменчивостью в пространстве и литологической неоднородностью по разрезу [4].

В рамках методики исследования [5] сформирована обучающая выборка из 8791 скважины, которая была разделена на два класса по результатам испытания в интервале тюменской свиты. В класс «приток» отнесено 3079 скважин, в класс «сухо» – 5712.

Атрибуты пространственных данных разделены на исходные и производные. Исходные значения пространственных параметров получены в ходе непосредственного замера, интерпретации, экспертной оценки. Производные сгенерировались на основе исходных данных в процессе конструирования признаков. Исходные значения разделяются на числовые и категориальные. В качестве числовых атрибутов использованы: значения гравитационных и магнитных аномалий; структурные построения по основным стратиграфическим горизонтам; карты изопахит; зартированные крупные тектонические разломы, линейные элементы. Исходные категориальные атрибуты – оцифрованные экспертные оценки территории и палеогеографические реконструкции, деления территории на области и районы нефтегазоносности, фациальное районирование. Последний вид атрибутов использовался в обучении в авторском, неизменном, виде.

Производные параметры генерировались в процессе конструирования признаков (features engineering). Конструирование признаков – это процесс использования предметной области данных для создания признаков, которые применяются для обучения алгоритмов и являются фундаментом для практического приложения машинного обучения [6].

#### Конструирование признаков

В работе применен ряд подходов к конструированию признакового пространства с использованием первичных числовых атрибутов. Используются методы из геоинформатики, геоморфологии, машинного обучения для повышения пространственного разрешения данных, бассейнового моделирования, кластеризации (рис. 1).

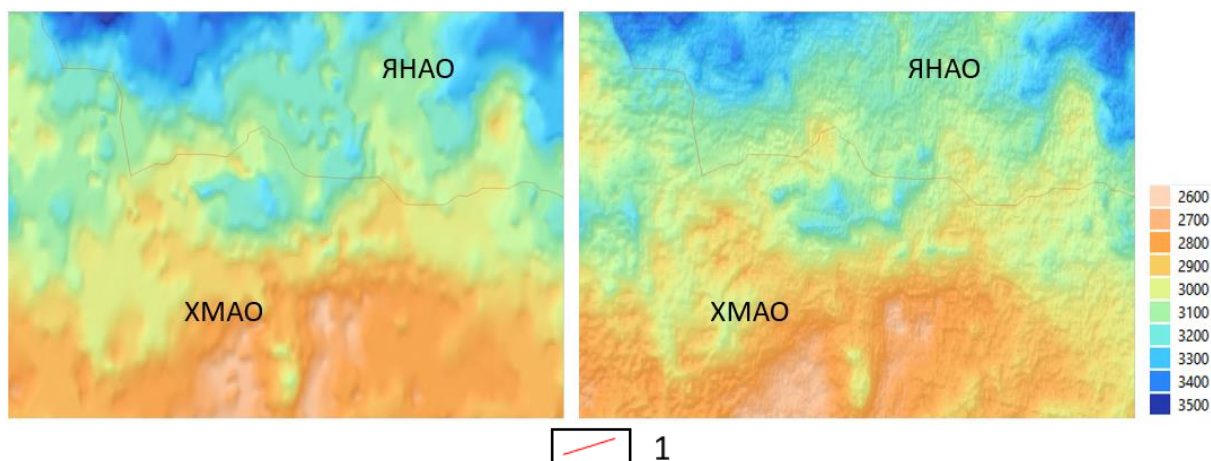
Технология повышения пространственного разрешения данных в предлагаемой методике является основой для дальнейшего конструирования признакового пространства.

Плотность сейсмических профилей и количество скважин на части исследуемой территории изменяется значительно. Выделяются области, покрытые редкой сеткой профилей 2D (участки Приуральского, Шуришкальского, Надымского районов), и территории, детально изученные сейсморазведочными работами 3D (участки Сургутского, Нижневартовского районов). Неравномерно меняется и детализация структурных построений. Для увеличения детализации структурных построений применен подход на основе генеративно-состязательных сетей в его частной реализации: повышение разрешения изображений (Super Resolution – SR).

Генеративно-состязательные сети (Generative adversarial network – GAN) – алгоритм машинного обучения без учителя, построенный на комбинации из двух нейронных сетей, одна из которых (генеративная сеть) генерирует образцы, а другая (дискриминационная сеть) старается отличить правильные («подлинные») образцы от неправильных. Генеративная и дискриминационная нейронные сети имеют противоположные цели – создать образцы и отбраковать образцы, между ними возникает антагонистическая игра [7, 8].

Для реализации модифицированного GAN алгоритма [9] были использованы эталонные участки, покрытые сейсморазведкой 3D. На генеративную нейронную сеть подавались деградированные данные 3D до средней разрешающей детализации сейсморазведки 2D. Задачей дискриминационной нейронной сети было, анализируя образцы эталонные (сейсморазведки 3D) и сгенерированные, дать оценку качества и методом обратного распространения ошибки улучшить качество генерации, изменив набор латентных параметров. В итоге состязания двух нейросетей генеративная сеть научилась справляться с задачей генерации исходного качества 3D из псевдо 2D данных сейсморазведки с учетом специфики оценки и обратного распространения ошибки для реализации алгоритмов SR. В качестве аналогии этого процесса можно представить вместо нейронной сети геолога, который на протяжении многих лет картирует определенную территорию и видел множество раз, как менялись структурные построения от данных сейсморазведки 2D к 3D, и на интуитивном уровне начинает понимать паттерны изменения, к примеру, что пологий структурный нос на карте, созданной по редкой сетке 2D сейсмопрофилей, является серией кулисообразно сочленённых валов. Подобная «интуиция» в настоящем исследовании была сформирована у генеративно-состязательной нейронной сети за 1344 часа непрерывного обучения на графическом процессоре.

По результатам работы алгоритма и объединения лучших экземпляров генеративной нейронной сети получены структурные построения с детализацией псевдо 3D по всей изучаемой территории. На рис. 2 в качестве примера приведены карты до и после работы алгоритма. Дальнейшее конструирование признаков проводилось на новых структурных построениях.

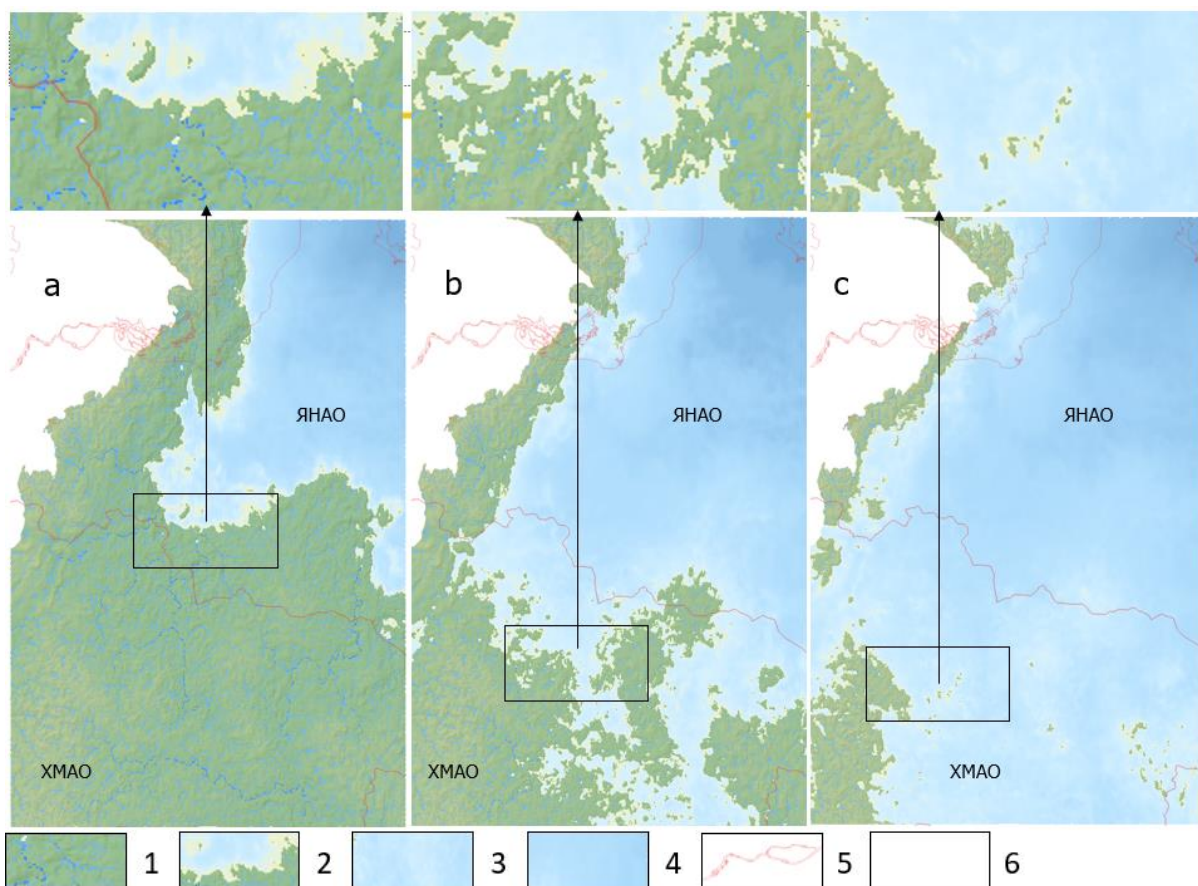


**Рис. 2.** Визуальное сравнение структурных построений по кровле баженовской свиты. Слева показана исходная карта, справа – после работы алгоритма повышения разрешения изображения: 1 – граница административного деления ХМАО, ЯНАО

**Fig. 2.** Visual comparison of structural maps, on the left is the original map, on the right – after the algorithm for increasing the image resolution: 1 – border of the administrative division of the Khanty-Mansi Autonomous Okrug, Yamalo-Nenets Autonomous Okrug

На следующем этапе сгенерированы признаки, базирующиеся на методологии объемно-генетического метода – бассейнового моделирова-

ния: моделирование седиментации осадочного вещества (ОВ) и моделирование миграции углеводородного вещества (УВ).



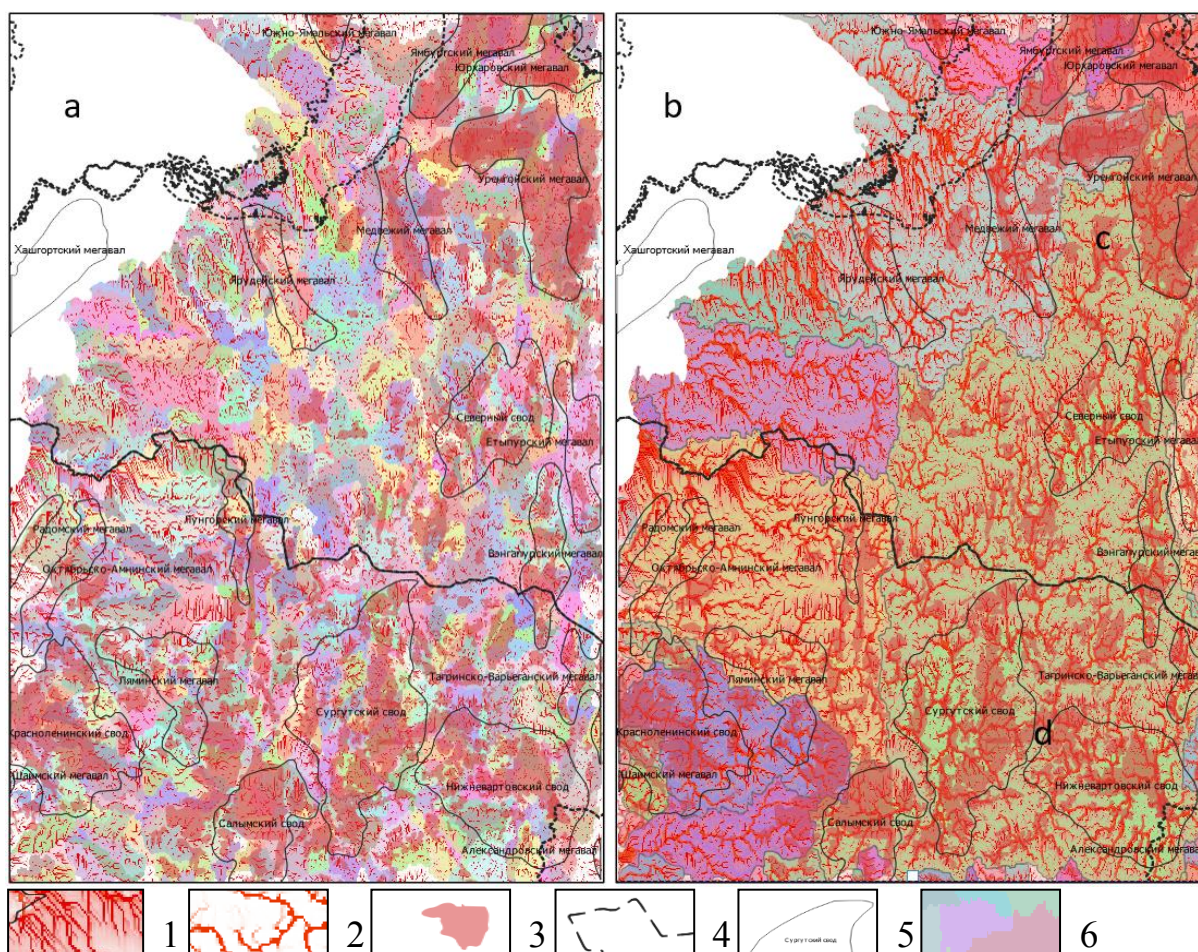
**Рис. 3.** Пример одной из моделей палеорекострукции трансгрессии моря: 1 – суша с гидросетью; 2 – береговая линия; 3 – мелководье; 4 – море; 5 – границы административных областей и современная береговая линия; 6 – отсутствие данных; серия схем последовательной трансгрессии: а) поздний-плисбах; б) баht; в) оксфорд

**Fig. 3.** Example of one of the models of paleoreconstruction of the sea transgression: 1 – land with a hydraulic network; 2 – coastline; 3 – shallow water; 4 – sea; 5 – boundaries of administrative regions and the modern coastline; 6 – lack of data; a series of sequential transgression schemes: a) late-pliesbach; b) baht; c) oxford

Созданы модели седиментации для основных стратиграфических единиц изучаемого объекта. С помощью симуляционных гидрологических моделей [10] воссозданы системы палеорек с учетом палеоландшафта и различного положения морской береговой линии [11]. На основе моделирования осадконакопления и эрозии рельефа, со стохастическим распределением начальных условий седиментации, сгенерированы серии карт песчаных и глинистых тел в заданных граничных условиях палеореконов на базе экспертных гипотез [12, 13]. Настройка моделей проводилась на опорный фонд скважин с известными значениями толщин. В качестве примера реализации моделирования приведены схемы палеореконов части бат-келловейской трансгрессии в Западной Сибири (рис. 3). В целом модели повторяли схемы палеореконов прошлых лет [12, 13], но имели детализированные стохастические свойства распределения фаций. В общей сложности для дальнейшей работы принято

более 250 реализаций карт распределения фаций, созданных за счет вариаций входных параметров моделирования процесса седиментации.

Региональное моделирование миграции УВ осуществлялось только с учетом ее латеральной составляющей на базе полученных моделей седиментации, структурных моделей и карт изопахит, приуроченных к времени формирования коллекторов и миграции УВ. Начальные граничные условия задавались на базе стохастического распределения параметров моделей согласно экспертным оценкам и варьировались в широком диапазоне [14–16]. В качестве примера приведены карты областей нефтегазосбора, а также векторов и плотности потока миграции в зависимости от начальных условий (рис. 4). Получено более 500 вариантов реализации моделей нефтегазонакопления в зависимости от входных параметров. Эти модели можно разделить на три типа по масштабу миграции УВ: локальная, глобальная и миграция промежуточного типа.



**Рис. 4.** Пример результатов моделирования миграции углеводородов: 1 – вектора миграции в области концентрации углеводородов; 2 – крупные потоки углеводородов по типу «русла»; 3 – области нефтегазосбора и месторождения; виды миграции: а) местная; б) глобальная; положение крупных потоков углеводородов в областях нефтегазосбора: с) исток; д) место концентрации; 4 – административные границы; 5 – границы крупных тектонических элементов; 6 – области нефтегазосбора

**Fig. 4.** Example of the results of modeling of hydrocarbon migration: 1 – migration vectors in the area of hydrocarbon concentration; 2 – large hydrocarbon flows by the «channel» type; 3 – oil and gas catchment areas and deposits; types of migration: a) local; b) global; position of large hydrocarbon flows in the oil and gas catchment areas: c) source; d) place of concentration; 4 – administrative boundaries; 5 – boundaries of large tectonic elements; 6 – oil and gas catchment areas

Наибольшее количество реализаций моделей было связано с локальной миграцией УВ, однако при определенных начальных заданных параметрах миграция приобретала промежуточный или глобальный масштаб (рис. 4, *b*), что приводило к формированию потоков с характерными «руслowymi» признаками концентрации и миграции УВ. Так, для одной из моделей (рис. 4, *b*) истоком «русла» служила область в районе юга Уренгойского мегавала (рис. 4, *c*), а местом впадения – области глобальной концентрации в районе Сургутского свода и Нижневартовского свода (рис. 4, *d*).

Стоит отметить, что особенностью предлагаемого подхода (рис. 1) является отсутствие необходимости выбора «правильного» признака (модели, карты и пр.) для дальнейшего анализа: каждый признак, несущий информацию об изучаемом объекте исследования, может быть использован для разделения обучающей выборки на классы в многомерном признаковом пространстве, его значение для прогноза определяется эмпирически – по степени его эффективности для задачи классификации. Ограничение для генерации признаков связано с последующей интерпретируемостью моделей.

С помощью методов геоинформатики (LS-Factor, Downslope distance gradient, Slope, Aspect, Profile & Plan Curvature, Based Landform Classification и пр.), широко применяющихся в задачах анализа цифровых моделей рельефа дневной поверхности, снято 1511 числовых и категориальных атрибутов с основных структурных планов горизонтов, карт изопахит, магнитных и гравитационных аномалий. Построены карты расстояний до ближайших линейных элементов, крупных тектонических разломов и специфических геоморфометрических точек изучаемого пространства (ближайшие максимальные и минимальные отметки в заданных окнах, границы разделов и др.). На основе карт с числовыми признаками были созданы категориальные, с помощью алгоритма K-mean, с классификацией пространства на следующее количество кластеров: 3, 5, 8, 13, 21.

#### Отбор признаков

После процедуры конструирования признаков пространства проведена процедура отбора (features selection). Целью отбора является редукция гиперпространства признаков для более эффективного его разделения классификаторами машинного обучения (снижение «проклятия размерности») [17]. Отбор осуществлялся алгоритмами на основе статистики и машинного обучения с учителем. Статистический метод был применен ко всему набору пространственных данных. Из признакового пространства были отсеяны признаки с квадратичным коэффициентом корреляции Пирсона более 0,95. По итогам этой процедуры каждую точку исследуемого пространства характеризовал вектор из 5242 признаков.

Для отбора признаков с использованием методов обучения с учителем полученная векторная характеристика пространства была присвоена каждой скважине согласно ее пространственному положению. Далее эта размеченная выборка подавалась на вход следующим

алгоритмам отбора: T-Score, Gini Index, Gain Ratio, Baruta, генетическая селекция на RandomForest, CatBoost, LightGBM, XGboost. Все модели отбора признаков проверялись на чувствительность к «шуму». К набору данных (вектору) из категориальных и числовых признаков пространства добавлялись случайные значения. В данной работе использованы параметры, снятые с наложенной на территорию с различным смещением карты температуры реликтового излучения по результатам работы телескопа Планка. Так, если бы модель при отборе признаков с заданным генератором псевдослучайных чисел (seed=137) произвела выбор «шумящих» признаков как значимых, вся выборка признаков этой модели была бы исключена из дальнейшего исследования. Все предложенные алгоритмы отбора с учителем «проигнорировали» «шумные» признаки. Итогом работы алгоритмов стал набор отличающихся между собой выборок с наиболее информативными, по «мнению» алгоритмов, признаками для дальнейшей классификации.

В качестве алгоритма машинного обучения с целью прогноза перспектив нефтегазоносности был выбран градиентный бустинг над решающими деревьями в его частных реализациях CatBoost, LightGBM и XGboost. На текущий момент этот класс алгоритмов является зарекомендовавшим себя на практике стандартом для обучения классификаторов на разнородном наборе числовых и категориальных признаков.

Каждый набор отобранных признаков разделялся на обучающую и тестовую выборку в разных пропорциях. Для контроля обучения использовалась метрика F1-score на тестовой выборке. F1-score достигает максимума при полноте и точности прогноза, равного единице, и близка к нулю, если один из аргументов силы моделей велась методом кросс-валидации на 8 подвыборках (folds), со стратификацией выборок по соотношению классов.

С учетом разной настройки гиперпараметров для выбранных алгоритмов и входных наборов данных, сформированных на этапе отбора признаков, обучено 68 моделей со следующими лучшими показателями качества, достигнутыми при обучении (табл. 1).

**Таблица 1.** Оценка качества моделей с разными метриками

**Table 1.** Assessment of the quality of models with different metrics

Метрика/Модель Metric/Model	CatBoost	LightGBM	XGBoost
F1	0,87332	0,87013	0,8697
AUC	0,94626	0,94401	0,94041
CA	0,87549	0,86985	0,86898

#### Анализ и интерпретация результатов машинного обучения

Лучшие по метрикам модели обучения могут быть выбраны для интерпретации. В качестве примера показаны возможности аналитики с использованием библиотеки SHAP. Для каждого признака рассчитывался вектор Шепли на основе теории игр, с оценкой важности для прогноза, путем включения и исключения его из обучающей выборки [19].

По итогам обучения лучшей модели в качестве примера показаны следующие средства визуализации, входящие в библиотеку SHAP: сводный график важности признаков, график взаимосвязей признаков, график влияния признаков на конкретный случай (рис. 5–7).

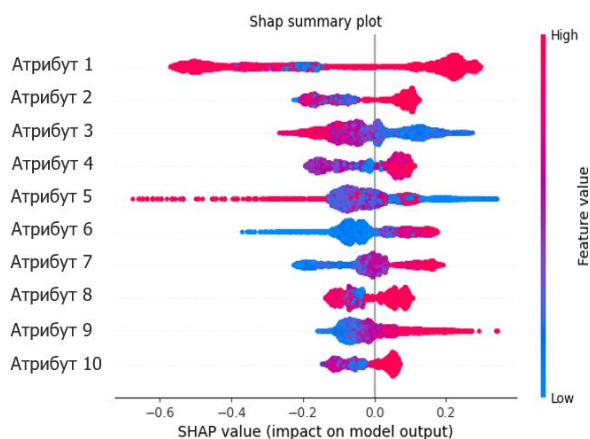


Рис. 5. Сводный график важности признаков для классификации

Fig. 5. Summary graph of the importance of features for classification

Сводный график важности признаков создает представление о том, какие признаки наиболее важны, и диапазон их влияния на модель (рис. 5). Чем выше находится признак в вертикальном списке, тем больше его интегральное влияние на модель. Слева от базовой линии, проведенной через ноль на нижней шкале влияния, располагаются значения от нормированного на ноль индекса Шепли – признака в виде точек, влияющих на категорию класса «сухо» – левая часть графика, и «приток» – правая его часть. Концентрация точек показывает количество значений с этим параметром. Цветом закодировано изменение

числового значения признака: чем краснее точки, тем выше значение. На графике интерес представляют признаки с наибольшим интегральным влиянием на модель и четким разделением по цвету базовой линии. Важными признаками для всех обученных моделей стали следующие характеристики пространства: принадлежность к нефтегазовому району, принадлежность к фациальной зоне, прогнозные температуры на кровле баженовской свиты, расстояние до ближайшего крупного тектонического элемента, связанная песчанность тюменской свиты, производные от аномалий гравитационного поля.

В качестве примера на рис. 5 переведен результат расчета индекса Шепли для одной из 68 моделей с лучшими показателями качества (табл. 1). Из 2111 признаков пространства показаны 10, оказывающих максимальное влияние на прогноз, – это атрибуты под номерами:

- 1) принадлежности к нефтегазовому району;
- 2) фациальная принадлежность района согласно палеорекострукции Батского времени;
- 3) связанная песчанность тюменской свиты;
- 4) принадлежность к бассейну нефтегазоносности при средних настройках миграции УВ;
- 5) плотность потока миграции углеводородов;
- 6) экспозиция значений гравитационных аномалий в редукции Буге;
- 7) прогнозная температура в кровле баженовской свиты;
- 8) принадлежность к кластеру (алгоритмом K-mean выделен 21 кластер на структурной поверхности по кровле баженовской свиты);
- 9) расстояние до ближайшего крупного тектонического элемента;
- 10) вертикальное расстояние до базового уровня сети каналов по структурной поверхности кровли фундамента.

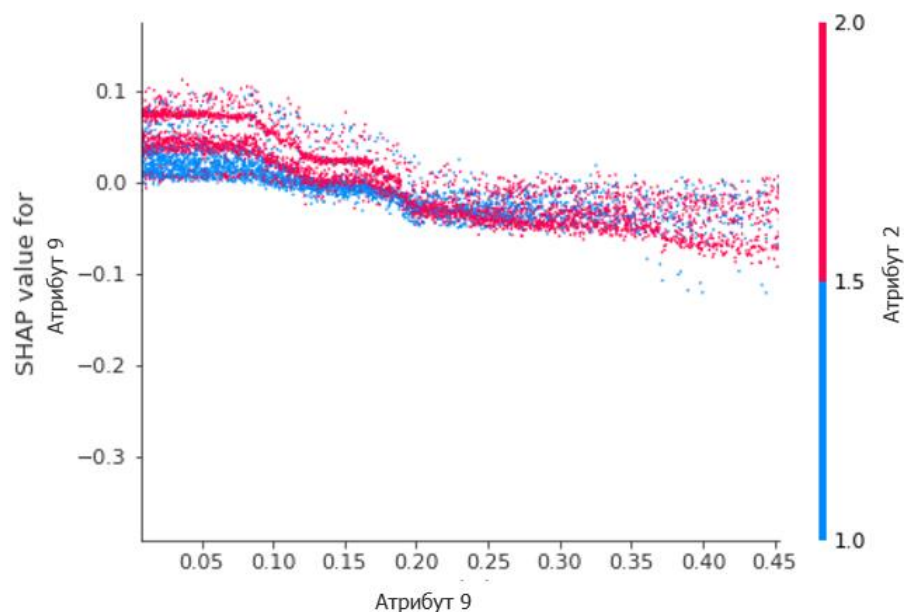


Рис. 6. График влияния на предсказание отдельного признака

Fig. 6. Graph of influence on prediction of a single feature

Влияние отдельного признака на модель визуализируется на графике взаимосвязью признаков (рис. 6). По оси абсцисс откладывается значение признака, по оси ординат – индекс Шепли. Чем выше индекс, тем

более «уверенно» модель прогнозирует класс «приток». Цветом кодируются значения другого признака (атрибут 2), с которым основной отображаемый признак имеет наибольшее взаимодействие при прогнозе.

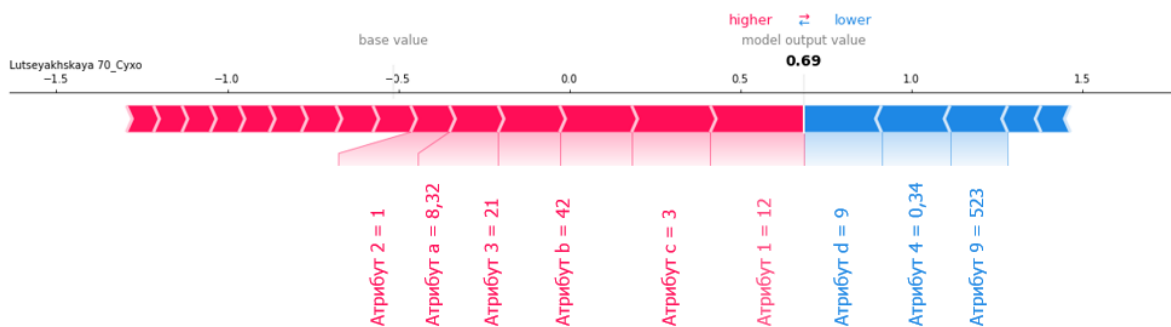


Рис. 7. График влияния признаков на конкретный случай  
Fig. 7. Graph of influence of features on a specific case

В примере (рис. 6) чем меньше в градусах расстояние до ближайшего крупного тектонического элемента (атрибут 9), тем выше индекс Шепли и больше вероятность нефтегазоносности территории, в то же время, если территория находилась в батское время в мелководной части с индексом 2 (красные точки), эта тенденция усиливается. При увеличении расстояния происходит инверсия, и уже более глубоководная область с индексом 1 (синие точки) увеличивает индекс Шепли. Однако в целом тенденция сохраняется: чем дальше от крупного тектонического элемента, тем меньше индекс Шепли.

График влияния признаков для конкретного случая может быть построен для каждой классифицируемой точки пространства, но в данном исследовании имеет смысл визуализировать прогноз для точек местонахождения скважин с известным исходом испытания, для сравнения факта и предсказания (рис. 7). Степень отклонения от порогового значения индекса Шепли, рассчитанного для скважины, показывает «уверенность» модели в классификации – чем правее и, соответственно, больше индекс относительно базового, тем модель более «уверена» в принадлежности ее к классу «приток». Длинной и близостью к порогу, а также цветом показана степень влияния признака на данную классификацию. Красным выделены признаки, «склоняющие» модель к классу «приток», синим – к классу «сухо». Иногда интерес представляют скважины, на которых модель ошиблась в классификации с высокой степенью «уверенности», эти случаи желательно рассматривать более детально, изучая дело скважины, анализируя нюансы испытания, показания станции ГТИ при бурении, описание керна, возможно, потребуется уточнение интерпретации каротажного материала. После детального анализа в ряде случаев выявлена некорректная исходная классификация скважин по результатам испытания как «сухих». Продуктивный объект был пропущен в ходе апробации продуктивных интервалов, или само испытание было некондиционное, в то время как ГИС и высокие газопоказания ГТС свидетельствовали о наличии продуктивного интервала.

В качестве примера приведен полученный результат на той же модели для скважины 70Р Луцьянского месторождения нефти. Помимо влияния атрибутов из первых десяти под номерами 2, 3, 4, 9, на прогноз оказывают влияния следующие атрибуты: а) мультиразрешающий индекс плоскостности дна долины, снятый с карты общей толщины юрских отложений; б) стандартное отклонение от осреднённой поверхности общей толщины юрских отложений с окном снятия 40 км<sup>2</sup>; в) фациальная принадлежность согласно палеорекострукции Мальшевского горизонта.

Значение атрибутов под индексами 1, с, b, 3, a, 2 увеличивают вероятность принадлежности скважины к классу «приток», в то же время значения атрибут под индексами d, 4, 9 снижают эту вероятность. Интегральный показатель индекса Шепли (вклада) каждого атрибута – 0,69, что выше порогового значения – 0,5. Таким образом, согласно модели, в данной скважине с высокой вероятностью тюменская свита нефтегазоносная. Однако по результату испытания притока получено не было, и скважина была отнесена в обучающей выборке к классу «сухо».

Анализ дела скважины и актов испытания показал значительное поглощение бурового раствора до 420 м<sup>3</sup> при бурении в интервале тюменской свиты. После спуска хвостовика 127 мм и перфорации интервалов Ю<sub>8</sub>, Ю<sub>7</sub>, Ю<sub>4</sub> и Ю<sub>3</sub> общим погонным метражом 38 м зарядами ЗПРК-42С-01 после промывки (80 м<sup>3</sup>) проведено испытание продуктивных горизонтов единым объектом, методом снижения уровня технической жидкости в колонне с помощью сваба за несколько рейсов до глубины 1870 м при глубине верхних дыр перфорации 3516 м. За период испытания откачено 40 м<sup>3</sup> технической воды без признаков УВ, общее время освоения составило 28 часов. Согласно заключению по результатам геофизических исследований пористость в перфорированных залежах – от 0,08 до 0,16 д.е. В шламе при проходке средней юры отмечались признаки УВ. Учитывая совокупность факторов, скважина не была освоена, испытание не было кондиционным, данный интервал рекомендуется к переиспытанию.

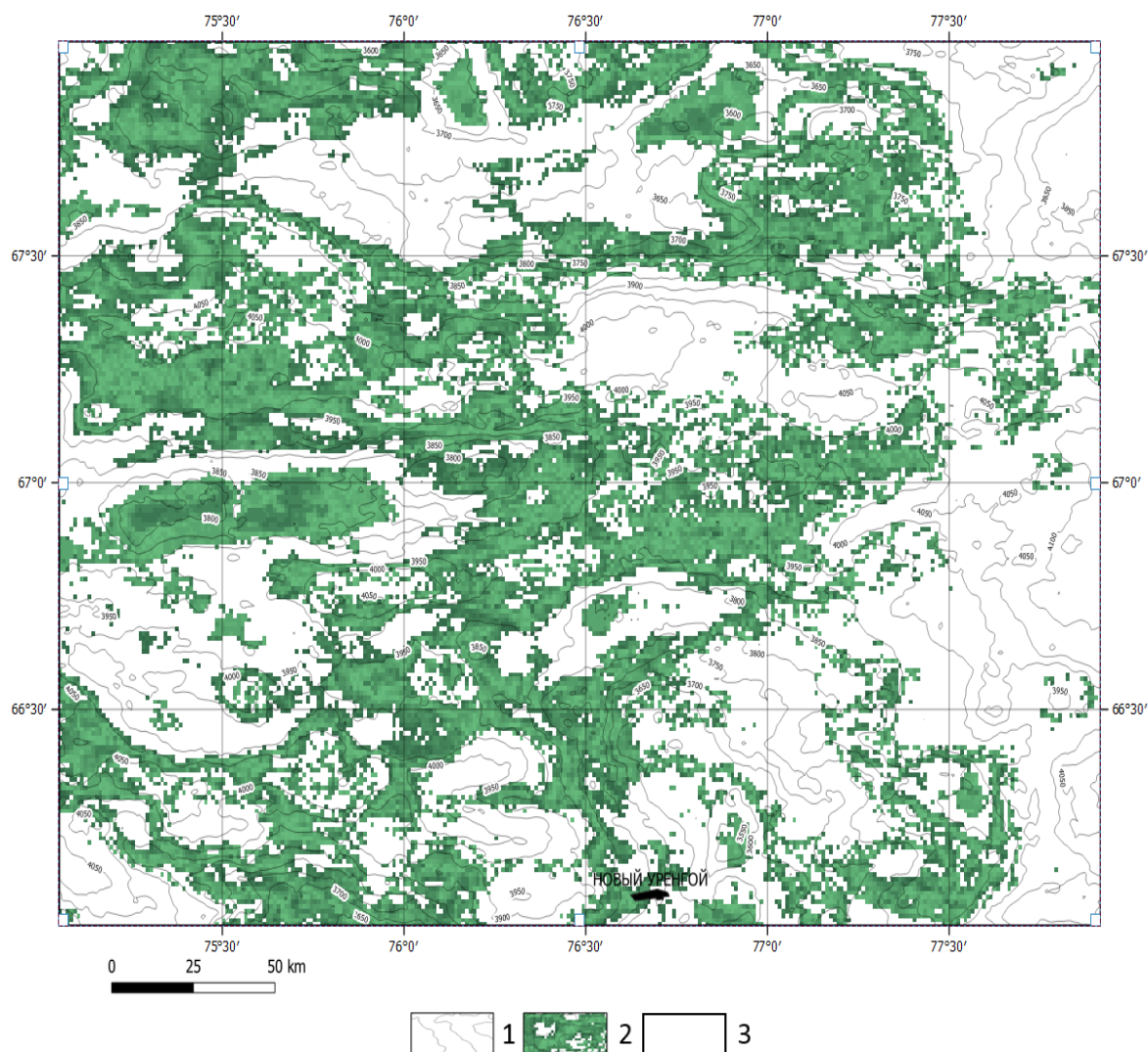


Проблема исходной классификации скважин по результатам испытания является существенной для всех методов прогноза нефтегазоносности, однако в данном методе ошибки модели на основе глобальных и локальных выявленных закономерностей могут быть использованы для последующей ее донстройки или корректировки разметки обучающей выборки.

Аналитическая работа с графиками библиотеки SHAP позволяет раскрыть полученную модель взаимосвязи между пространственными признаками и их влиянием как на прогноз в целом для изучаемой территории, так и для конкретной скважины.

### Прогноз нефтегазоносности

Модели были объединены в метамодель (ensemble) со стековым обобщением (stacking) алгоритмом логистической регрессии. Стековое обобщение использует предсказание каждой отдельной «слабой» модели и, комбинируя сильные предсказательные свойства каждой из них, делает потенциально более качественный прогноз. Метрики качества итоговой метамодели на кросс-валидации составили: F1 – 0,91212, AUC – 0,95761, Accurasy – 0,9112. Метамодель показала лучшие результаты, чем алгоритмы по отдельности. Качество модели, согласно метрикам, хорошее.



**Рис. 8.** Схема прогноза нефтегазоносности тюменской свиты в пределах северной части Надымского и Пуловского НГР: 1 – изолинии по абсолютным отметкам кровли баженовской свиты; 2 – перспективная территория; 3 – неперспективная территория для поисков углеводородов

**Fig. 8.** Scheme of forecasting the oil and gas content of the Tyumen suite within the northern part of the Nadym and Purovsky oil and gas regions: 1 – isolines at the absolute elevations of the top of the Bazhenov formation; 2 – promising territory; 3 – unpromising territory for hydrocarbon exploration

На основе обученной метамодели сделан прогноз вероятности принадлежности к классам в интервале тюменской свиты на изучаемой территории.

В зависимости от выбранного порога значений вероятности принадлежности к классу «приток» по

критерию степени важности для прогноза ошибки первого (False Positive Rate – FPR) или второго (False Negative Rate – FNR) рода проводится классификация территории. Ошибка первого рода при поисках на УВ обычно является более значимой, так как ее цена –

это затраты на бурение «сухой» скважины в область, классифицируемую как «приток».

В качестве примера прогноза приведена карта классификации территории на перспективные и неперспективные земли для поисков УВ для северной части Надымского и Пуровского НГР (рис. 8).

На карте каждый пиксель с пространственным разрешением 400×400 м ассоциирован с вероятностью нефтегазоносности. Вероятность нефтегазоносности, или степень «уверенности» модели к принадлежности территории к классу «приток», варьируется от нуля до единицы; там, где значения не достигают выбранного порогового значения, территория относится к классу «сухо», где равны или более – к классу «приток». На рис. 8 чем интенсивнее зеленый цвет, тем значения ближе к единице, а белым отмечены неперспективные земли. Особенностью полученного прогноза стали области, выделенные как перспективные вне привязки к структурному фактору, т. е. находящиеся на текущий момент вне поисково-разведочной парадигмы большинства нефтегазовых компаний.

Дополнительная проверка объективности полученных значений вероятности перспектив нефтегазоносности территории проведена на данных, по которым информации в обучающей выборке не было – это фактические результаты поисковой деятельности компаний за 2018 и 2019 гг. в регионе исследования. Для вновь открытых месторождений им. Ю.Е. Батурина вероятность открытия составляла 0,93, а для Восточно-Унлорского месторождения – 0,89. Прогнозируемые высокие коэффициенты вероятности открытия месторождений УВ в тюменской свите подтверждают эффективность выявленных моделями связей внутри признакового пространства и сделанных обобщений.

Информация по каждой вновь пробуренной скважине, законченной испытанием целевого объекта, позволяет уточнить существующую модель прогноза, а новые данные сейсморазведки 3D – увеличить точность региональных структурных построений. Созданные модели, при появлении новых данных, можно дообучать, создавая таким образом динамично меняющийся прогноз в зависимости от поступления новых данных о территории исследования.

#### Оценка ресурсной базы УВ

Оценка ресурсной базы УВ тюменской свиты проведена вероятностным объемным методом через расчет плотности запасов, приведенных к единице площади, со стохастическим моделированием методом Монте-Карло [20]. По 281 месторождению, стоящему на государственном балансе на 2019 г. с залежами УВ в тюменской свите, определены плотности вероятностного распределения основных подсчетных параметров объемного метода. Рассчитана плотность вероятностного распределения запасов категорий  $ABC_1+C_2$  на единицу площади для открытых месторождений в тоннах нефтяного эквивалента (т.н.э.). Полученное распределение умножалось на значения, взятые с карты вероятной нефтегазоносности терри-

тории в каждой точке пространства. Таким образом, на выходе генерировалась карта ресурсной базы по региону с учетом вероятности нефтегазоносности. Для получения интегральных показателей значения с карты просуммированы. Извлекаемые ресурсы оценивались путем перемножения полученных значений ресурсов на плотность вероятностного распределения КИН для залежей тюменской свиты. Вероятностная оценка геологической и извлекаемой ресурсной базы и суммарные значения этих параметров по месторождениям, числящимся на государственном балансе на территории исследования, приведены в табл. 2.

**Таблица 2.** Оценка ресурсной базы углеводородов тюменских отложений

**Table 2.** Assessment of the resource hydrocarbons of the Tyumen deposits

Ресурсная база УВ/ Оценка в вероятности, гос. баланс Resource HC/Assessment in probability, State Register	P90	P50	P10	Гос. баланс 2018 State Register 2018	
				ABC <sub>1</sub>	ABC <sub>1</sub> +C <sub>2</sub>
Геологическая, млрд т.н.э Geological, billion toe	6,3	23,8	178,5	5,4	9,3
Извлекаемая, млрд т.н.э Recoverable, billion toe	1,4	5,3	46,7	1,3	2,1

На изучаемой территории геологическая ресурсная база УВ для тюменских отложений с вероятностью более 90 % равна или превышает 6,3 млрд т.н.э., что на 0,9 млрд т.н.э. больше поставленных на государственный баланс запасов по категории ABC<sub>1</sub>. С вероятностью 50 % ресурсная база объекта исследования равна или более 23,8 млрд т.н.э., что на 14,5 млрд т.н.э. больше суммарных запасов по категориям ABC<sub>1</sub>+C<sub>2</sub>, поставленных на государственный баланс. В самом оптимистичном сценарии, с вероятностью P10, извлекаемая ресурсная база УВ равна или более 46,7 млрд т.н.э., что сопоставимо с потенциальной ресурсной базой мелового нефтегазоносного комплекса Западной Сибири.

#### Заключение

Предложенная методика прогноза нефтегазоносности с помощью методов машинного обучения и комплексирования технологического стека методов: геоинформатики, бассейнового моделирования, экспертных оценок, показала высокое качество предсказания вероятности нефтегазоносности территорий как на кросс-валидации, так и по результатам сделанных открытий прошлых лет.

Методика отличается от других методов прогноза эмпирическим подходом, при котором на основе фактических результатов испытания отдельных скважин методами машинного обучения ведется поиск взаимосвязей среди пространственных атрибутов территории и их многомерное обобщение. Пространственные атрибуты могут быть сгенерированы различными методами, способными привнести даже косвенную информацию об изучаемом объекте. Лучшие результаты возможны в тандеме между искусственным интеллектом (ИИ), частью которого является машинное обучение, и геологом. Используя возможности ИИ

искать связи и делать обобщения в признаковом гиперпространстве, геолог извлекает знания, создавая и совершенствуя гипотезы. При получении новых данных по результатам поисково-разведочных работ существующие модели могут быть доучены, совершенствуя прогноз, создавая динамический цикл с положительной обратной связью между прогнозом и поисковой деятельностью.

Результатом построения региональной карты вероятности нефтегазоносности территории и карты ресурсной базы УВ с учетом геологического риска может быть: определение и ранжирование поисково-разведочных объектов; оценка ожидаемой стоимости (EMV) проектов; оценка целесообразности приобретения или развития активов; понимание будущего

вектора развития конкурентов по итогам поисковых работ; изменение стратегии самой компании в связи с потенциальной возможностью вовлечения в разработку значительной традиционной ресурсной базы УВ с синергетическим экономическим эффектом от существующей и будущей инфраструктуры.

На основании оценки методом Монте-Карло ресурсной базы УВ в тюменской свите с учетом прогноза вероятности нефтегазоносности территории исследования показан ее значительный потенциал для поисковой деятельности. Так, в равновероятной оценке P50 ресурсный потенциал свиты в регионе исследования превышает в два с лишним раза поставленные на государственный баланс геологические запасы месторождений по сумме категорий АВС<sub>1</sub> и С<sub>2</sub>.

#### СПИСОК ЛИТЕРАТУРЫ

1. Распознавание образов гигантских нефтяных месторождений / А.А. Трофимук, В.С. Вышемирский, А.Д. Дмитриев, Д.П. Дробот, П.Н. Карагодин, И.Н. Сулимов // Проблемы нефтеносности Сибири. – 1971. – Т. 1. – С. 34–50.
2. Конторович А.Э. Геология нефти и газа: избранные труды. Методы прогноза нефтегазоносности. Планирование геолого-разведочных работ. – Новосибирск: СНИИГГиМС, 2008. – Т. 3. – 331 с.
3. Методическое руководство по количественной и экономической оценке ресурсов нефти, газа и конденсата России / под ред. К.А. Клещева – М.: ВНИГНИ, 2000. – 189 с.
4. Геологическое строение и нефтегазоносность нижней-средней юры Западно-Сибирской провинции / Ф.Г. Гурари, В.П. Девятков, В.И. Демин, А.Е. Еханин, А.М. Казаков, Г.В. Касаткина, Н.И. Курушин, Н.К. Могучева, В.В. Сапьяник, О.В. Серебренникова, Л.В. Смирнов, Л.Г. Смирнова, В.С. Сурков, Г.Г. Сысолова, О.В. Шиганова. – Новосибирск: Наука, 2005. – 156 с.
5. Ивлев Д.А. Региональный прогноз областей притока нефти из баженовско-абалакского комплекса на территории ХМАО-Югры методом машинного обучения // Нефтяное Хозяйство. – 2016. – № 6. – С. 90–93.
6. Zheng A., Casari A. Feature engineering for machine learning: principles and techniques for data scientists. – Boston: O'Reilly, 2018. – 218 p.
7. Ganguly Kuntal. Learning generative adversarial networks. – Birmingham: Packt Publ., 2017. – 180 p.
8. Langr J., Bok V. GANs in action: deep learning with generative adversarial networks. – New York: Manning Publishing, 2019. – 240 p.
9. High-throughput, high-resolution registration-free generated adversarial network microscopy / Hao Zhang, Xinlin Xie, Chunyu Fang, Yicong Yang, Di Jin, Peng Fei // arXiv.org. – 2018. – 21 p. URL: <https://arxiv.org/ftp/arxiv/papers/1801/1801.07330.pdf> (дата обращения: 08.08.2020).
10. Path sampling method for modeling overland water flow, sediment transport, and short term terrain evolution in Open Source GIS / H. Mitasova, C. Thaxton, J. Hofierka, A. Moore, L. Mitas // Developments in Water Science. – 2004. – V. 55. – № 2. – P. 1479–1490.
11. Hearn C.J. The dynamics of coastal models. – Cambridge: Cambridge University Press, 2008. – 503 p.
12. Палеогеография Западно-Сибирского осадочного бассейна в юрском периоде / А.Э. Конторович, В.А. Конторович, С.В. Рыжкова, Б.Н. Шурыгин, Л.Г. Вакуленко, Е.А. Гайдебурова, В.П. Данилова, В.А. Казаненков, Н.С. Ким, Е.А. Костырева, В.И. Москвин, П.А. Ян // Геология и геофизика. – 2013. – Т. 54. – № 8. – С. 972–1012.
13. Литолого-палеогеографические реконструкции юрского периода севера Западно-Сибирского осадочного бассейна / Г.Г. Шемин, В.А. Врениковский, В.И. Москвин, Л.Г. Вакуленко, Е.В. Деев, Н.В. Первухина // Геология нефти и газа. – 2018. – № 6. – С. 35–61.
14. Шиманский В.В., Танинская Н.В., Раевская Е.Г. Выявление структурно-литологических ловушек в юрских и нижнемеловых отложениях Западной Сибири на основе палеогеографических реконструкций // Геология нефти и газа. – 2019. – № 3. – С. 39–46.
15. Неручев С.Г., Смирнов С.В. Оценка потенциальных ресурсов углеводородов на основе моделирования процессов их генерации и формирования месторождений нефти и газа // Нефтегазовая геология. Теория и практика. – 2007. – № 2. – С. 33–45.
16. Verweij J.M. Hydrocarbon migration systems analysis. – Delft: Elsevier Science Publishers, 1993. – 289 p.
17. Zheng A., Casari A. Feature engineering for machine learning. – Sebastopol: O'Reilly Media, 2018. – 216 p.
18. Opitz J., Burst S. Macro F1 and Macro F1 // arXiv.org. – 2019. – 12 p. URL: <https://arxiv.org/pdf/1911.03347.pdf> (дата обращения: 08.08.2020).
19. SHAP // github.com. URL: <https://github.com/slundberg/shap> (дата обращения: 08.08.2020).
20. Rose P.R. Risk analysis and management of petroleum exploration ventures. – Tulsa: American Association of Petroleum Geologists, 2001. – 164 p.

Поступила 25.11.2020 г.

#### Информация об авторах

**Ивлев Д.А.**, соискатель ученой степени в области геологических наук Бристольского университета.

UDC 550.8.012

## METHOD FOR REGIONAL FORECAST OF OIL AND GAS POTENTIAL TERRITORIES BY MACHINE LEARNING ALGORITHMS ON THE EXAMPLE OF THE TYUMEN FORMATION OF WESTERN SIBERIA

Dmitry A. Ivlev,  
dm.ivlev@gmail.com

University of Bristol,  
Beacon House, Queens Road, Bristol, Clifton, BS8 1SE, UK.

**The relevance** of the research is caused by the reduction in the fund of structural traps and the need to expand the resource base of hydrocarbons by increasing the efficiency of prospecting and exploration of fields in complex oil and gas deposits.

**The main aim** of the research is to show the forecasting methodology and the set of applied technological solutions and algorithms using the example of forecasting the oil and gas content of the study area.

**Object:** Middle Jurassic deposits (Tyumen Formation) of Western Siberia within the region (700×900 km), which includes parts of the Yamalo-Nenets and Khanty-Mansiysk administrative districts and the Tomsk region.

**Methods.** Using the machine-learning algorithms and integrating a technological set of methods: geoinformatics, basin modeling, and expert assessments, the following stages of the forecast method implementation: 1) generation of the feature space of the studied area based on increasing the spatial resolution of structural constructions using algorithms of generative-adversarial architecture of neural networks, where the results of 3D seismic survey are used as reference areas; 2) selection of features by statistical method and machine learning methods; 3) creation of a subset of forecast models based on gradient boosting over decision trees; 4) combining them into a metamodel by stacking generalization by logistic regression, are shown.

**Results.** An approach to regional forecasting has been formalized and tested. A forecast of the probability of oil and gas content of the Tyumen suite in the study area was made. On its basis and information on discovered fields, the hydrocarbon resource base was estimated by the Monte Carlo method. The results are presented in the form of a summary table of geological and recoverable resources for probabilities P10, P50, P90 in comparison with the categories of reserves ABC1 and ABC<sub>1</sub>+C<sub>2</sub> of the fields listed on the state balance sheet in the study area. As an example, the graphic materials of the results are given: the work of the algorithm for increasing the spatial resolution; sedimentation modeling; modeling of hydrocarbon migration; hydrocarbon potential forecast map for the northern part of the Nadym and Purovsky oil and gas regions.

### Key words:

Tyumen formation, Middle Jurassic, machine learning, artificial intelligence, feature generation, feature selection, gradient boosting, generative adversarial neural networks, regional oil and gas forecast, probability of geological success, hydrocarbon resource, geoinformatic, basin modeling, sedimentation modeling, modeling HC migration.

### REFERENCES

1. Trofimuk A.A., Vyshemirskiy V.S., Dmitriyev A.D., Drobot D.P., Karagodin P.N., Sulimov I.N. Recognition of images of giant oil fields. *Problems of oil-bearing capacity of Siberia*, 1970, vol. 1, pp. 34–50. In Rus.
2. Kontorovich A.E. *Geologiya nefii i gaza: izbrabnye Trudy. Metody prognoza neftegazonosnosti. Planirovanie geologorazvedochnykh rabot* [Geology of Oil and Gas: Selected Works. Methods for forecasting oil and gas content. Planning of geological exploration works]. Novosibirsk, SNIIGiMS Publ., 2008. Vol. 3. 331 p.
3. *Metodicheskoe rukovodstvo po kolichestvennoy i ekonomicheskoy otsenke resursov nefii, gaza i kondensata Rossii* [Methodological guide for the quantitative and economic assessment of oil, gas and condensate resources in Russia]. Ed. by K.A. Kleshchev. Moscow, VNIGNI Publ., 2000. 189 p.
4. Gurari F.G., Devyatov V.P., Demin V.I., Ekhanin A.E., Kazakov A.M., Kasatkina G.V., Kurushin N.I., Mogucheva N.K., Sapyanik V.V., Serebrennikova O.V., Smirnov L.V., Smirnova L.G., Surkov V.S., Sysolova G.G., Shiganova O.V. *Geologicheskoe stroenie i neftegazonosnost nizhney-sredney yury Zapodno-Sibirskoy provintsii* [Geological structure and oil and gas content of the Lower-Middle Jurassic of the West Siberian province]. Novosibirsk, Nauka Publ., 2005. 156 p.
5. Ivlev D.A. Regional forecast of areas of oil inflow from the Bazhenov-Abalak complex on the territory of the Khanty-Mansi Autonomous Okrug-Yugra by the machine learning method. *Oil Industry*, 2016, no. 6, pp. 90–93. In Rus.
6. Zheng A., Casari A. *Feature engineering for machine learning: principles and techniques for data scientists*. Boston, O'Reilly, 2018. 218 p.
7. Kuntal G. *Learning generative adversarial networks*. Birmingham, Packt Publ., 2017. 180 p.
8. Langr J., Bok V. *GANs in action: deep learning with generative adversarial networks*. New York, Manning Publ., 2019. 240 p.
9. Hao Zhang, Xinlin Xie, Chunyu Fang, Yicong Yang, Di Jin, Peng Fei. High-throughput, high-resolution registration-free generated adversarial network microscopy. *arXiv.org*, 2018, 21 p. Available at: <https://arxiv.org/ftp/arxiv/papers/1801/1801.07330.pdf> (accessed 08 August 2020).
10. Mitasova H., Thaxton C., Hofierka J., Moore A., Mitas L. Path sampling method for modeling overland water flow, sediment transport, and short term terrain evolution in Open Source GIS. *Developments in Water Science*, 2004, vol. 55, no. 2, pp. 1479–1490.
11. Hearn C.J. *The dynamics of coastal models*. Cambridge, Cambridge University Press, 2008. 503 p.
12. Kontorovich A.E., Kontorovich V.A., Ryzhkova S.V., Shurygin B.N., Vakulenko L.G., Gaideburova E.A., Danilova V.P., Kazanenkov V.A., Kim N.S., Kostyreva E.A., Moskvina V.I., Yan P.A. Paleogeography of the West Siberian sedimentary basin in the Jurassic period. *Geology and Geophysics*, 2013, vol. 54, no. 8, pp. 972–1012. In Rus.
13. Shemin G.G., Vrenikovskiy V.A., Moskvina V.I., Vakulenko L.G., Deev E.V., Pervukhina N.V. Lithological-paleogeographic reconstructions of the Jurassic period in the north of the West Siberian sedimentary basin. *Geology of oil and gas*, 2018, no. 6, pp. 35–61. In Rus.
14. Shimansky V.V., Taninskaya N.V., Raevskaya E.G. Identification of structural and lithological traps in the Jurassic and Lower Cretaceous sediments of Western Siberia based on paleogeographic reconstructions. *Geology of Oil and Gas*, 2019, no. 3, pp. 39–46. In Rus.

15. Neruchev S.G., Smirnov S.V. Assessment of potential hydrocarbon resources based on modeling the processes of their generation and formation of oil and gas fields. *Neftegazovaya Geologia*, Theory and Practice, 2007, no. 2, pp. 33–45. In Rus.
16. Verweij J.M. *Hydrocarbon migration systems analysis*. Delft, Elsevier Science Publishers, 1993. 289 p.
17. Zheng A., Casari A. *Feature engineering for machine learning*. Sebastopol, O'Reilly Media, 2018. 216 p.
18. Opitz J., Burst S. Macro F1 and Macro F1. *arXiv.org*, 2019, 12 p. Available at: <https://arxiv.org/pdf/1911.03347.pdf> (accessed 08 August 2020).
19. Shap. *github.com*. Available at: <https://github.com/slundberg/shap> (accessed 08 August 2020).
20. Rose P.R. *Risk analysis and management of petroleum exploration ventures*. Tulsa, American Association of Petroleum Geologists, 2001. 164 p.

*Received: 25 November 2020.*

#### **Information about the authors**

**Dmitry A. Ivlev**, PhD applicant in Geological Sciences, University of Bristol.