

Оригинальная статья

Применение алгоритмов машинного обучения для прогнозирования золоторудной минерализации Верхнеамгинского щелочного массива, Алдано-Становой щит

П. Л. Чудинов^{1,✉}, В. Ю. Фридовский²

¹АО «Полюс Алдан», п. Нижний Куранах, Российской Федерации

²Институт геологии алмаза и благородных металлов СО РАН, г. Якутск, Российской Федерации

✉Paulenotov@gmail.com

Аннотация

Приводятся результаты применения методов машинного обучения для прогнозирования золоторудной минерализации на поисковой стадии геологоразведочных работ на примере Верхнеамгинского щелочного массива Алдано-Станового щита. Использованы данные анализа 403 штрафных проб методом ICP-AES на 25 химических элементов. Протестированы восемь алгоритмов классификации: Random Forest, Support Vector Machine, Neural Network (Multilayer Perceptron), Boosting (AdaBoost), Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis и Naive Bayes. Наивысшую точность (до 89,6 %) продемонстрировали Random Forest и Support Vector Machine, основанные на выявлении взаимосвязей между рудными элементами (Au, Ag, As, Cu, Sb) и элементами с отрицательной корреляцией (Mg, Ca, Ti). Результаты подтверждены ROC-анализом. При создании модели машинного обучения в качестве целевой переменной приняты значения «рудного» фактора для каждой пробы, использованные в качестве предиктора. С помощью построения аномальных полей значений «рудного» фактора проведено сравнение параметров известных объектов и прогнозируемых площадей. Методы машинного обучения позволяют оперативно и надежно интерпретировать аналитические данные, полученные с использованием спектрометрии или портативных XRF-анализаторов. Для повышения точности прогноза подчеркивается важность комбинации традиционных статистических методов (кластерный, факторный анализ) с современными алгоритмами машинного обучения.

Ключевые слова: машинное обучение, золоторудная минерализация, геохимические данные, поисковые работы, Верхнеамгинский щелочного массив, Алдано-Становой щит

Финансирование. Работа выполнена при финансовой поддержке АО «Полюс Алдан» и по плану НИР ИГАБМ СО РАН (проект FUFG-2024-0006).

Для цитирования: Чудинов П.Л., Фридовский В.Ю. Применение алгоритмов машинного обучения для прогнозирования золоторудной минерализации Верхнеамгинского щелочного массива, Алдано-Становой щит. *Природные ресурсы Арктики и Субарктики*. 2025;30(2):205–219. <https://doi.org/10.31242/2618-9712-2025-30-2-205-219>

Original article

Application of machine learning algorithms to predict gold mineralization in the Verkhneamginsky alkaline massif, Aldan-Stanovoy Shield

Pavel L. Chudinov^{1,✉}, Valery Y. Fridovsky²

¹JSC “Polyus Aldan”, Nizhny Kuranakh, Russian Federation

²Diamond and Precious Metal Geology Institute,
Siberian Branch of the Russian Academy of Sciences, Yakutsk, Russian Federation

✉Paulenotov@gmail.com

Abstract

The study reports on the application of machine learning methods for predicting gold mineralization in the prospecting phase of geological exploration. It focuses on the Verkhneamginsky alkaline massif, situated within the Aldan-Stanovoy Shield, as a case study. The investigation included the analysis of 403 ore samples, which were evaluated through Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) to determine the concentrations of 25 chemi-

cal elements. A total of eight classification algorithms were assessed in this investigation, including Random Forest, Support Vector Machine, Neural Network (Multilayer Perceptron), Boosting (AdaBoost), Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis, and Naive Bayes. The Random Forest and Support Vector Machine algorithms demonstrated the highest accuracy, achieving 89.6%, by identifying the relationships among ore elements (Au, Ag, As, Cu, Sb) and those elements that displayed negative correlations (Mg, Ca, Ti). These results were further validated through Receiver Operating Characteristic (ROC) analysis. In the process of developing the machine learning model, the values corresponding to the “ore” factor for each sample were designated as the target variable, while serving as predictors. To enable a comparative analysis between the parameters of established entities and the predicted regions, anomalous fields of the “ore” factor values were constructed. Additionally, machine learning methods enable the rapid and reliable interpretation of virtually any geochemical analytical data in the field, including data obtained through modern spectrometry methods and portable X-ray fluorescence (XRF) analyzers. The research further underscores the significance of integrating traditional statistical approaches, such as cluster and factor analysis, with contemporary machine learning algorithms to improve the accuracy of predictions.

Keywords: machine learning, gold mineralization, geochemical data, exploration, Verkhneamginsky alkaline massif, Aldan-Stanovoy Shield

Funding. This study was conducted with financial support from JSC “Polyus Aldan” and within the research framework established by the Diamond and Precious Metal Geology Institute, Siberian Branch of the Russian Academy of Sciences (project No. FUFG-2024-0006).

For citation: Chudinov P.L., Fridovsky V.Yu. Application of machine learning algorithms to predict gold mineralization in the Verkhneamginsky alkaline massif, Aldan-Stanovoy Shield. *Arctic and Subarctic Natural Resources*. 2025;30(2):205–219. (In Russ.); <https://doi.org/10.31242/2618-9712-2025-30-2-205-219>

Введение

Для снижения финансовых рисков и повышения точности прогнозирования на разных стадиях геологоразведочных работ применяются высокоэффективные методы обработки аналитических данных [1, 2]. Традиционные статистические подходы обладают ограничениями в интерпретации сложных многомерных зависимостей [3]. Новые возможности открываются при использовании методов машинного обучения за счет автоматизированного выявления скрытых закономерностей в больших массивах геологической информации [3–6].

В данном исследовании показана возможность применения алгоритмов машинного обучения для прогнозирования по результатам штуфного опробования золоторудной минерализации в пределах Верхнеамгинского щелочного массива Алдано-Станового щита. Этот район характеризуется мезозойским щелочным магматизмом и широким развитием метасоматических процессов, связанных с золото-медно-порфировым и золото-порфировым типами оруденения, сходными с месторождениями Рябиновое и Морозкинское [7, 8].

Материалы и методы исследования

Материалом исследования являются результаты анализов штуфных проб, отобранных в пределах Верхнеамгинского щелочного массива в 2021–2024 гг. Штуфные пробы, имеющие видимые метасоматические изменения (гумбейти-

зация, березитизация и фельдшпатизация), отбирались в ходе геологических маршрутов вне сети опробования. Пробы проанализированы пробирным методом с ICP-AES-ом окончанием (на золото) и мультиэлементным (ICP-AES) методом атомно-эмиссионной спектроскопии с индуктивно-связанной плазмой на 35 химических элементов.

Предварительно из общей выборки были исключены пробы, генетически не связанные с горными породами Верхнеамгинского щелочного массива (доломиты, известняки, песчаники и др.). В ходе дальнейшей статистической обработки из выборки удалены химические элементы, для которых количество проб со значениями НПО (нижний порог обнаружения) не является представительным (B, Bi, Ga, Hg, Tl, Th, U, W), а также Al, K, Na.

В сформированной выборке из 25 химических элементов были замены значения НПО на $0,5 \cdot \text{НПО}$, а ВПО (верхний порог обнаружения) на $1,05 \cdot \text{ВПО}$, исключены также ураганные значения. Для дальнейших исследований использованы 403 штуфные пробы, для которых определялся характер распределения элементов (log-нормальный или нормальный). По результатам для всех химических элементов принят log-нормальный закон распределения. Значения логарифмированы и нормализованы методом z-стандартизации для приведения переменных к единому масштабу с нулевым средним и единичной дисперсией.

Статистические исследования являются неотъемлемой частью на любой стадии геологоразведочных работ. Наиболее часто на поисковой стадии применимы методы многомерной статистики (корреляционный, кластерный, факторный (метод главных компонент) анализы), принадлежащие к наиболее ценным и полезным методам для выделения участков, перспективных для выявления любого типа минерализации. Каждый из методов имеет свои ограничения при интерпретации полученных результатов [3, 5].

В последние десятилетия все более широкое распространение получают методы машинного обучения, являющиеся дальнейшим развитием методов многомерной статистики [3, 5]. Машинное обучение (МО) – это область искусственного интеллекта, которая использует алгоритмы для получения информации из данных и создания моделей, способных решать задачи без явного программирования. Искусственный интеллект, в свою очередь, описывает создание систем, способных имитировать человеческую интеллектуальную и творческую деятельность. В этом контексте машинное обучение представляет собой имитацию процесса обучения.

В машинном обучении выделяют четыре основные категории: обучение с учителем (классификация, регрессия), обучение без учителя (ассоциация, кластеризация, уменьшение размерности), полуkontролируемое обучение и активное обучение [5]. В данной работе рассматриваются только классификационные методы машинного обучения с учителем (Boosting (AdaBoost), Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis, Naive Bayes, Neural Network (Multilayer Perceptron), Random Forest и Support Vector Machine). Обучение машины и прогнозирование результата выполняются на основе положительного примера [5]. Основной задачей классификации в машинном обучении с учителем является определение принадлежности объекта к определенному классу или виду [5, 9].

Региональная геологическая позиция

Верхнеамгинский щелочная массив расположен на территории одноименного золотоносного района Алдано-Станового щита. Территория исследуемого района входит в состав Амгинской субмеридиональной зоны тектонического меланжа, отделяющей Нимнырский и Западно-Алданский террейны [10]. В строении Амгинской зоны участвуют архейские комплексы, мета-

морфизованные в амфиболитовой и эпидот-амфиболитовой фациях, раннепротерозойские ортогнейевые и парагнейевые комплексы субгранулитовой–гранулитовой фаций, а также фрагменты архейских и раннепротерозойских зеленокаменных поясов и дифференцированные плутоны ультраосновных, основных и щелочных пород. Верхнеамгинский золотоносный район характеризуется многоярусным строением: нижнекембрийский фундамент, венд-нижнекембрийский осадочный чехол и мезозойские участки активизации (рис. 1) [11]. Архейские структуры фундамента с несогласием перекрыты венд-нижнекембрийским платформенным чехлом морских глинисто-карбонатных осадков. На отдельных участках в северной и восточной частях района на вершинах водоразделов сохранились мезозойские терригенные отложения юрской (юхтинская свита) и кайнозойской систем.

Основной объем магматических образований связан с мезозойской тектономагматической активизацией. Преобладают штоки щелочных сиенитов и монцонитов площадью десятки км². Встречаются отдельные тела и дайки щелочных гранитов, а также силлы и дайки лампрофиров. Дайки лампрофиров (протяженностью до 2 км и мощностью первые десятки метров) имеют преимущественно северо-западное и северо-восточное простирание, соответствующее преобладающим направлениям тектонических разломов. С щелочными интрузиями связаны зоны сульфидизации, окварцевания и скарнирования с золоторудной минерализацией. Специфической особенностью мезозойского магматизма на Алдано-Становом щите является его калиевый уклон, отмечаются породы калий-натриевого ряда [13, 14]. Внедрение основного объема щелочных сиенитов происходило $129,1 \pm 2,5$ млн лет назад, а даек и силлов мезократовых лампрофиров (миннетт) – $117,7 \pm 3,4$ млн лет назад [7].

Металлогеническая специализация Верхнеамгинского золотоносного района определяется карстовыми образованиями (Куранахский тип), скарновой (Лебединский тип) и золото-медно-порфировой (Рябиновый тип) минерализацией [8].

Рудные тела Куранахского типа представлены залежами, приуроченными к карстовым полостям. Это рыхлые, обломочные, супесчано-суглинистые минерализованные образования по метасоматически измененным нижнекембрийским карбонатным отложениям и юрским пес-

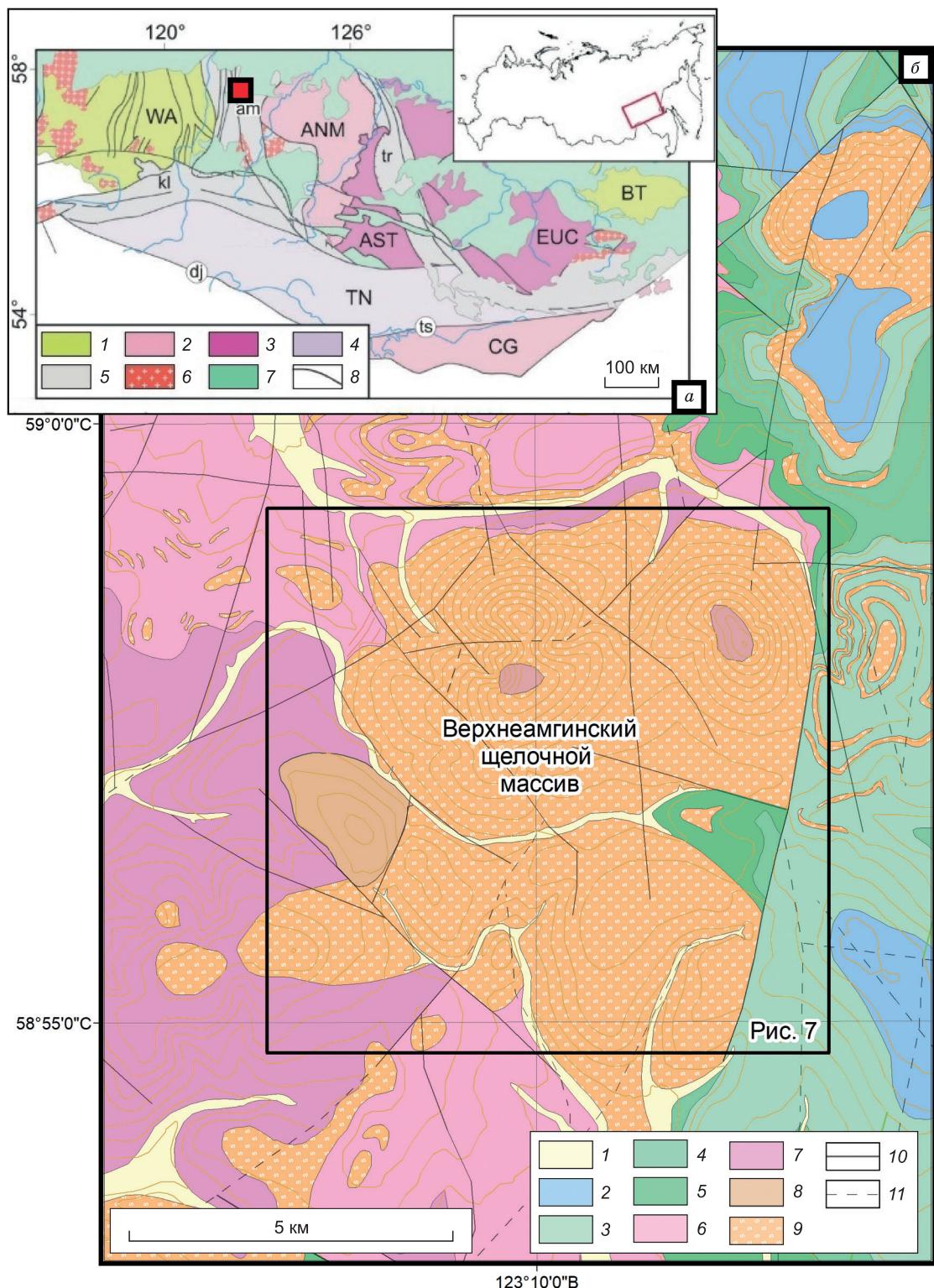


Рис. 1. Тектоническая схема Алдано-Станового щита и положение района работ (по [11] с изменениями) (а) и схема геологического строения Верхнеамгинского щелочного массива (по [8,12] с изменениями) (б).
 а: 1 – гранит-зеленокаменные террейны (WA – Западно-Алданский, BT – Батомгский); 2 – гранулит-ортогнейсовые террейны (ANM – Нимнырский, CG – Чогарский); 3 – гранулит-парагнейсовые террейны (AST – Сутамский, EUC – Учурский); 4 – тоналит-трондемит-гнейсовые террейны (TN – Тындинский); 5 – зоны тектонического меланжа (am – Амгинская, kl –

Каларская, tr – Тыркандинская); 6 – сшивающие раннепротерозойские граниты; 7 – чехол Сибирской платформы; 8 – разломы (dj – Джалтулакский, ts – Таксакандинский).

б: 1 – четвертичные отложения (современные аллювиальные отложения, пески, галечники, валунники); 2 – юрские отложения (песчаники, гравелиты, конгломераты); 3–6 – венд-нижнекембрийские отложения (известняки, доломиты, мергели): 3 – унгелинская свита; 4 – тумулдурская свита; 5 – пестроцветная свита; 6 – усть-юдомская свита; 7 – архейские образования (кварциты с пачками гнейсов и кристаллических сланцев); 8, 9 – раннемеловые интрузии: 8 – эльконский гипабиссальный комплекс щелочно-сиенитовый; 9 – лебединский плутонический комплекс монцонит-сиенит-гранитовый; 10 – разрывные нарушения достоверные; 11 – разрывные нарушения предполагаемые

Fig. 1. а –Tectonic scheme of the Aldan-Stanovoy Shield and the location of the work area (by [11] with changes); **б** – Geological scheme of the geological structure of the Verkhneamginsky alkaline massif (by [8,12] with changes).

а: 1 – granite-greenstone terranes (WA – West Aldan, BT – Batomga); 2 – granulite-orthogneiss terranes (ANM – Nimnry, CG – Chogar); 3 – granulite-paragneiss terranes (AST – Sutam, EUC – Uchur); 4 – tonalite-trondhjemite-gneiss terrane (TN – Tyndin); 5 – zones of tectonic mélange (am – Amga, kl – Kalar, tr – Tyrkanda); 6 – stitching Early Proterozoic granites; 7 – cover of the Siberian platform; 8 – faults (dj – Dzheltulaksky, ts – Taksakandin).

б: 1 – Quaternary deposits (modern alluvial deposits, sands, pebbles, boulders); 2 – Jurassic deposits (sandstones, gravelstones, conglomerates); 3–6 – Vendian-Cambrian deposits (limestones, dolomites, marls); 3 – Ungelinskaya suite; 4 – Tumuldurskaya suite; 5 – Pestrocvetnaya suite; 6 – Ust-Yudomskaya suite; 7 – Archean formations (quartzites with packs of gneisses and crystalline schists); 8, 9 – Early Cretaceous intrusions: 8 – Elkon hypabyssal alkaline syenite-alkaline granite complex; 9 – Lebedinsky plutonic monzonite-syenite-granite complex; 10 – reliable faults; 11 – inferred faults

чникам. Первичными рудами являются пирит-адуляр-кварцевые метасоматиты [15].

Рудные тела скарнового типа представлены крутопадающими золото-сульфидно-кварцевыми жилами и горизонтальными залежами в низах доломитовой толщи венда, вблизи или на контакте с кристаллическим фундаментом по периферии мезозойских магматитов.

Золото-медно-порфировая минерализация локализована в щелочных массивах и представлена штокверками с золотосодержащими сульфидами и сульфидизированным штоком эпилейцитовых сиенит-порфиров [8].

Одним из важных критериев наличия золото-медно-порфирового и золото-порфирового оруднения является присутствие околоврудных серицит-микроклиновых метасоматитов, гумбейтовых и фельдшпатитовых изменений, приуроченных к зонам разрывных нарушений и локализованных в щелочных магматических комплексах раннемелового возраста [16, 17]. Рудная минерализация в щелочных породах представлена двумя ассоциациями [7,8]. Минералы ранней ассоциации – пирит, халькопирит, галенит, сфалерит, молибденит, буланжерит сформированы из высоконцентрированных 22–44 мас.% NaCl-экв. углекислотно-азотных $\text{CO}_2 \pm \text{N}_2$ флюидов при температуре 330–400 °C и давлении 1150 бар [7]. Формирование минералов поздней ассоциации – алтait, гессит, петцит и самородное золото происходило из низкоконцентрированных 3,3–9,2 мас.% NaCl-экв. углекислотных флюидов при температуре 210–230 °C. Необходимо отметить также наличие в породах Верхнеамгинско-

го щелочного массива редкоземельной (моанцит-Се) и ториевой (торианит) минерализации [7].

Статистическая обработка данных

С целью апробации результатов методов машинного обучения и подтверждения результатов прогнозирования выборка разделена на две части. Одна часть, состоящая из 250 проб, была использована при создании моделей машинного обучения. Другая часть, из 153 проб, была принята как условно новый выделенный объект в пределах площади, не использовалась в процессе обучения модели и являлась дополнительной тестовой выборкой для созданных моделей машинного обучения. К основной выборке из 250 проб применены стандартные методы многомерной статистики, включая расчет коэффициентов кларков концентрации (КК) химических элементов (табл. 1).

Методом корреляционного анализа выявлены основные элементы-спутники золота в штуфных пробах Верхнеамгинского массива ($n = 250$). Результаты корреляционного анализа для золота (элементы ранжированы по убыванию коэффициентов корреляции): Ag 0,55 – As 0,54 – Sb 0,50 (средняя положительная связь); Cu 0,39 – V 0,39 – Fe 0,27 – Pb 0,23 – Zn 0,16 – Sc 0,16 – Ba 0,14 – Mo 0,13 – P 0,13 (слабая положительная связь) и Ca – 0,23 – Ti – 0,27 – Mg – 0,28 (слабая отрицательная связь). Проведен кластерный анализ методом иерархической кластеризации. В качестве правила объединения был использован метод Варда, а мерой расстояния (близости) 1-т Пирсона. В результате была получена дендрограмма

Таблица 1
Кларки концентрации химических элементов

Table 1

Clarks of concentration of chemical elements

Элемент	Кларк
Au	10,87
Cr	5,83
Cu	5,50
Pb	3,87
Mn	3,87
Ag	3,59
Mo	3,40
Cd	1,98
S	1,78
As	1,33
Sb	1,03
V	0,94
Co	0,92
Ni	0,80
Be	0,80
Mg	0,76
Ca	0,55
Sc	0,45
Fe	0,43
P	0,42
Zn	0,31
La	0,24
Sr	0,22
Ti	0,06
Ba	0,06

групп связанных между собой химических элементов (рис. 2).

Кластерный анализ в целом подтверждает результаты корреляционного анализа. Среди элементов на уровне группировки расстояний объединения, равной 0,3–0,8, обособляются шесть основных кластеров: 1) Au–As–Ag–Cu–Sb; 2) Be–S–Pb–Zn; 3) Cd–Cr–Mo; 4) Ba–Fe–P–V–La; 5) Co–Sc–Mn–Ni; 6) Ca–Sr–Mg–Ti.

Среди кластеров отчетливо выделяется основная рудная ассоциация, представленная Au–As–Ag–Cu–Sb. Она соответствует геохимической ассоциации Au–Ag–Cu–Bi, приуроченной к гумбейтизированным породам карбонат-серитит-мусковит-ортоклазовой фации. Ассоциация является индикатором наличия золото-медно-порфирового (Рябиновое месторождение) и золото-порфирового (Морозкинское место-

рождение) оруденения, где выделяются высококонтрастные аномалии Au, Ag, Pb, Cu, Zn, Mo, As [16].

Геохимическая ассоциация Be–S–Pb–Zn может представлять халькофильную полиметаллическую ассоциацию (Pb–Zn–S) в комплексе с Be. Присутствие берилля может быть связано с влиянием плагиоклаза вмещающих пород. Другой возможной причиной может являться мусковитизация пород массива. В целом данная ассоциация характеризует наличие сульфидов свинца и цинка, что является важным поисковым критерием. Геохимическая ассоциация Cd–Cr–Mo имеет сложно объяснимую природу и является переходной между халькофильной Pb–Zn–S полиметаллической ассоциацией (Cd) и сидерофильной Co–Mn–Ni ассоциацией (Cr–Mo). Ассоциация Ba–Fe–P–V–La сложно интерпретируется, представлена двумя группами элементов: основной литофильной Ba–La–V, сидерофильной, включающей только железо, а также фосфором – элементом, обладающим литофильными и сидерофильными свойствами. Сложная конфигурация данного кластера требует дополнительного изучения и интерпретации. Отдельно выделяется кластер сидерофильных элементов группы железа (Transition metals) Co–Mn–Ni и редкоземельных элементов (REE), представленных скандием. Данная минеральная ассоциация отражает наличие в массиве Co–Ni минерализации.

Особого внимания заслуживают литофильные элементы Ca–Sr–Mg–Ti с явной отрицательной связью с минеральной ассоциацией Au–As–Ag–Cu–Sb, что отражает особенности рудообразования. Вынос Mg и Ca сопровождает процесс окварцевания и серититизации. Снижение концентраций Ti может указывать на растворение титансодержащих минералов (ильменит и сфен) в метасоматическом процессе [18].

По результатам факторного анализа основной выборки получены шесть основных факторов. Для них рассчитаны значения факторных нагрузок каждого химического элемента. Рассчитанная модель объясняет 73 % общей дисперсии, что является вполне удовлетворительным результатом (табл. 2).

Первый фактор с максимальным вкладом в общую изменчивость 28,3 % отражает изменение большинства исследуемых элементов, таких

Дендрограмма для 25 перемен.

Метод Варда

1-г Пирсон

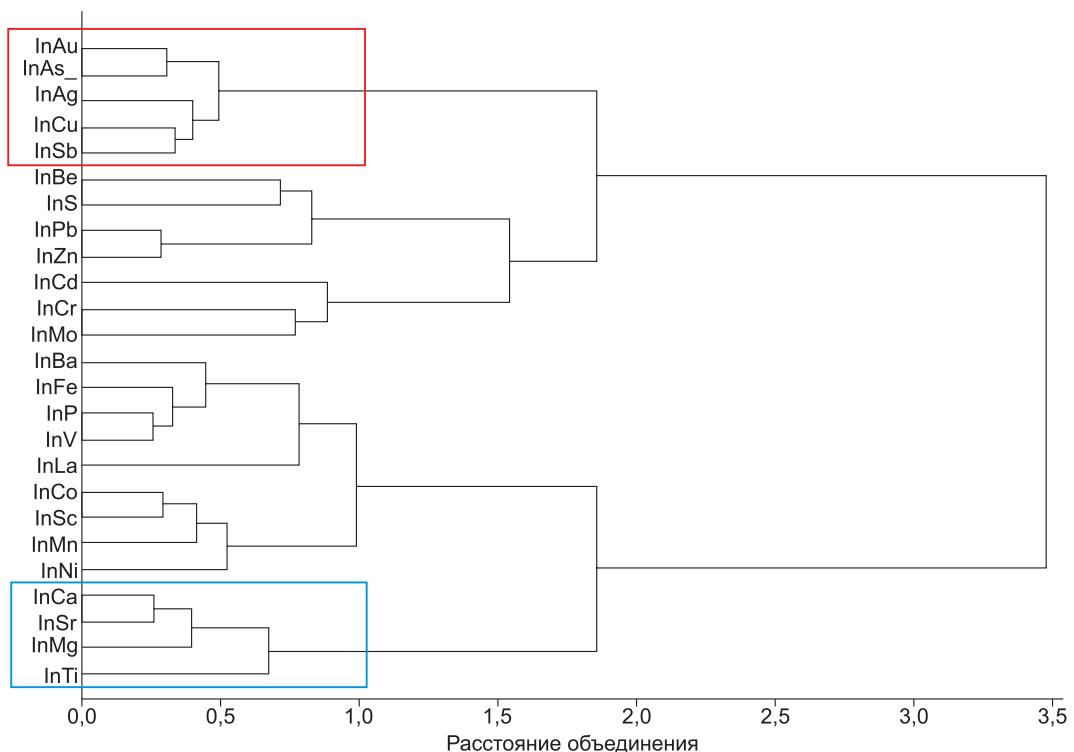


Рис. 2. Результаты кластерного анализа по основной выборке (n = 250)

Fig. 2. Results of cluster analysis for the main sample (n = 250)

как Ba, Co, Fe, Mn, P, Sc, V, Zn, а также Cu, La, и Pb, значения которых близки к существенным.

Второй фактор с весом 19,5 % и высокими положительными нагрузками Au, Ag, As, Sb, а также близкими значениям к значимым факторным нагрузкам Cu, с одной стороны, отражает накопление рудного комплекса элементов (Au, As, Ag, Sb, Cu), а с другой, отрицательными значениями Ca, Mg, Ti показывает вынос данных элементов в процессе рудообразования. Вынос Ca и Mg может быть связан с раскислением плагиоклаза, замещением амфибола (роговой обманки) и биотита карбонат-серийцитовой минеральной ассоциацией, развитием щелочного метасоматизма (гумбейтизация, фельдшпатизация) с привносом Na и K. Этот фактор полностью соответствует результатам кластерного анализа и подтверждает наличие устойчивых геохимических связей. В дальнейшем, в ходе машинного обучения модели, «рудный» фактор будет использован как целевой. С этим метасоматозом сопряжена Zn–Pb минерализация, что подтверждается результатами кластерного и факторного

анализов. Фактор принят как условно «рудный», является продуктивным и перспективным показателем для выявления золотого оруденения. Достаточно высокий фактический вес (17,5 %) отражает высокую значимость и широкое развитие данного процесса в пределах Верхнеамгинского массива. В «рудном» факторе выделяются две ассоциации химических элементов: положительная, $As_{74}Ag_{71}Sb_{69}Au_{64}Cu_{57}$ и отрицательная, $Mg_{74}Ca_{68}Ti_{65}$ (рис. 3).

Третий фактор с весом 8,5 % обусловлен высокой положительной нагрузкой Cr и Ni. Четвертый фактор с весом 8 % определяется высокими отрицательными нагрузками Cd и Mo и существенными значениями Cr, что хорошо соотносится с результатами кластерного анализа. Пятый и шестой факторы имеют незначительные доли общей дисперсии, составляющие 5 и 4,4 % соответственно, малозначимые нагрузки для химических элементов.

Оценка достоверности данных, приведенных в выборке, проведена с использованием критерия Кайзера–Майера–Олкина (КМО). Средний

Значения факторных нагрузок по основной выборке (n = 250)

Таблица 2

Factor loading values for the main sample (n = 250)

Table 2

Элемент	Fac1	Fac2	Fac3	Fac4	Fac5	Fac6
Au	-0,32	0,64	0,01	0,03	0,17	0,42
Ag	-0,36	0,71	-0,05	-0,07	0,09	0,11
As	-0,48	0,74	-0,02	-0,03	-0,02	0,20
Ba	-0,73	-0,06	-0,10	0,31	0,03	-0,21
Be	-0,44	-0,14	-0,34	-0,30	-0,28	0,00
Ca	-0,22	-0,68	-0,40	-0,21	0,20	0,34
Cd	-0,11	-0,12	-0,19	-0,61	-0,10	0,00
Co	-0,75	-0,33	0,35	0,17	-0,18	0,13
Cr	-0,02	-0,08	0,73	-0,44	-0,18	0,03
Cu	-0,55	0,57	0,18	-0,08	0,07	0,09
Fe	-0,76	0,15	0,28	0,26	-0,06	-0,28
La	-0,50	-0,27	0,17	-0,24	0,39	-0,26
Mg	-0,34	-0,74	-0,06	-0,34	-0,01	0,21
Mn	-0,64	-0,41	-0,04	0,18	-0,29	0,10
Mo	-0,05	0,24	0,11	-0,64	0,21	-0,47
Ni	-0,45	-0,39	0,60	-0,15	-0,04	0,17
P	-0,82	-0,06	-0,03	0,27	0,17	-0,15
Pb	-0,58	0,40	-0,33	-0,18	-0,05	-0,20
S	-0,22	0,31	-0,37	-0,16	-0,59	-0,10
Sb	-0,38	0,69	0,01	-0,13	0,27	0,18
Sc	-0,74	-0,18	0,27	-0,08	-0,18	0,17
Sr	-0,49	-0,49	-0,48	-0,06	0,32	0,12
Ti	-0,31	-0,65	0,02	0,18	0,16	-0,24
V	-0,86	0,08	-0,05	-0,03	0,22	0,02
Zn	-0,77	0,09	-0,25	0,02	-0,27	-0,14
Общая дисперсия	7,08	4,86	2,12	1,74	1,26	1,10
Доля объясненной дисперсии	0,28	0,19	0,08	0,07	0,05	0,04

индекс КМО по всей выборке равен 0,84, что означает высокие достоверность и адекватность этих данных. Для значений Cu, P, Sc и V показатель КМО максимальный, а для Cr, Mo и S – минимальный (табл. 3).

В полученном «рудном» факторе наиболее существенное положительное влияние на вероятность выявления комплексной полиэлементной рудной ассоциации оказывают элементы: Au, Ag, As, Cu, Sb, где наблюдается прямая положительная корреляция (рис. 4). Остальные элементы выборки, хоть и имеют влияние на фактор, но значительно меньшее. Отмеченное позволяет использовать методы машинного обучения для оперативного прогнозирования участков, перспек-

тивных на выявление Au, Ag, As, Cu, Sb рудной геохимической ассоциации.

В ходе дальнейшей статистической обработки значений «рудного» фактора был принят закон log-нормального распределения, выполнен стандартный набор операций по расчету минимально аномального, фонового содержаний, коэффициента вариации и других показателей, отражающих статистические особенности распределения значений. В соответствии с этим выделены и ранжированы уровни аномальных значений «рудного» фактора: No – фоновое, Low – низкое аномальное, Med – среднее аномальное, High – высокое аномальное. В процессе машинного обучения моделей данная многоклассовая клас-

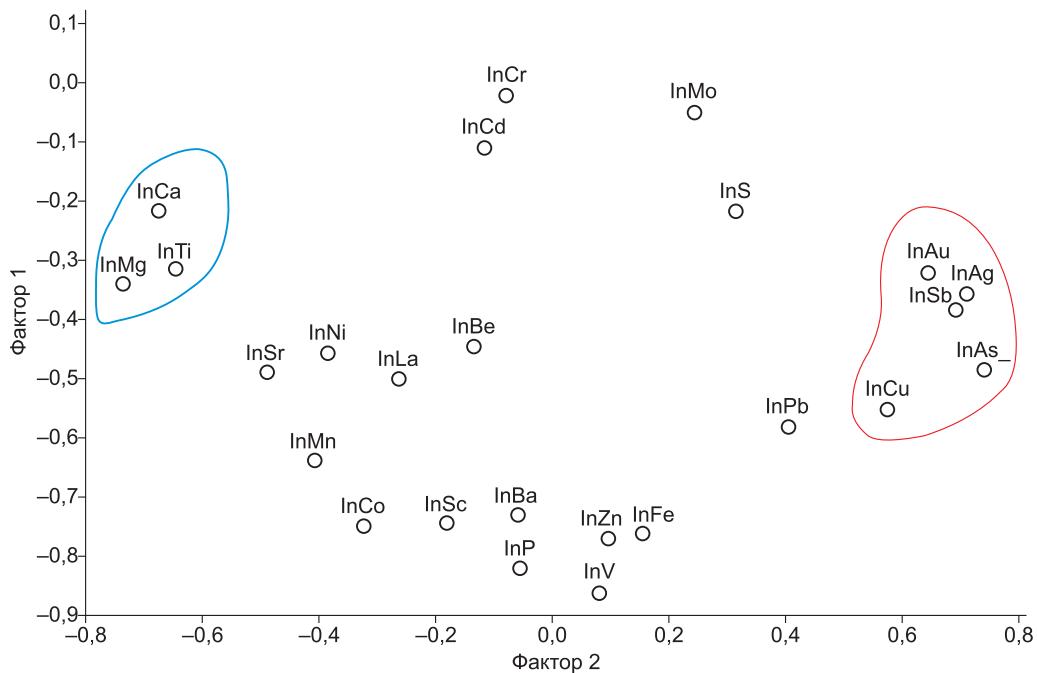


Рис. 3. График отношений химических элементов, отраженный в «рудном» факторе

Fig. 3. Graph of chemical element relationships reflected in the “ore” factor

сификация будет являться целевой, а значения «сырых» значений химических элементов будут переменными, представляющими информацию о целевой переменной.

Применение алгоритмов машинного обучения

С целью выявления наиболее эффективного метода прогнозирования были использованы восемь основных методов классификации: Boosting (AdaBoost), Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis, Naive Bayes, Neural Network (Multilayer Perceptron), Random Forest и Support Vector Machine [4, 6, 19]. Широкий выбор методов обоснован разными алгоритмами обучения для каждого из них. В процессе обучения были сформированы три локальные выборки: тренировочная, для настройки модели и проверки ее производительности (160 проб); валидационная, для подбора оптимального набора гиперпараметров (40 проб); тестовая, для итоговой оценки модели (50 проб). Распределение проб в таких пропорциях позволяет избежать недообучения или переобучения модели и дать объективную оценку работе модели [4].

На основе созданной по 250 пробам модели были прогнозированы аномальные значения в пробах тестовой выборки (153 пробы). Резуль-

таты прогноза приведены в табл. 4. По приведенным данным наиболее точными оказались три метода Random Forest, Support Vector Machine и Neural Network (Multilayer Perceptron), точность прогнозирования которых превысила 85 %. При этом стоит отметить, что все алгоритмы показали самые худшие результаты по предсказанию средних аномальных значений (Med). Высокий показатель неточности предсказаний средних аномальных значений, предположительно, связан с малым количеством проб данного кластера (менее 30). С одной стороны, можно предположить наличие определенных проблем при обработке малых массивов информации, но с другой стороны, по данному кластеру относительно неплохие результаты показали методы машинного обучения Support Vector Machine и Naive Bayes, что характеризует их способность корректно работать с малыми массивами информации. Несмотря на достаточно низкие показатели по данному кластеру, при дальнейшей тренировке моделей не представляется возможным увеличить точность предсказания. Отдельно стоит отметить, что наилучшие результаты каждый из методов показал по фоновым (No) и высоким аномальным значениям (High).

В целом все методы показали удовлетворительные результаты и могут быть использова-

Таблица 3
**Значения критериев
 Кайзера–Мейера–Олкина (KMO)
 для химических элементов**

Table 3

**Kaiser–Meyer–Olkin (KMO)
 criteria values for chemical elements**

Элемент	KMO
Средний	0,84
Au	0,82
Ag	0,88
As	0,83
Ba	0,87
Be	0,83
Ca	0,75
Cd	0,71
Co	0,85
Cr	0,55
Cu	0,90
Fe	0,83
La	0,89
Mg	0,86
Mn	0,88
Mo	0,56
Ni	0,74
P	0,91
Pb	0,82
S	0,61
Sb	0,83
Sc	0,92
Sr	0,80
Ti	0,82
V	0,90
Zn	0,84

ны при прогнозировании, за исключением Naive Bayes, где общая точность предсказания составила всего лишь 9,2 %. Возможной причиной низкого результата может являться корреляция предикторов между собой, что противоречит предположению о независимости признаков в алгоритме Naive Bayes.

Наиболее точными алгоритмами предсказаний являются Support Vector Machine. Они обладают высокой общей точностью 89,2 % и одним из самых высоких уровней предсказания средних аномальных значений. Алгоритмы Random Forest и Neural Network (Multilayer Perceptron) имеют близкие результаты, различные в точно-

сти определения фоновых и низких аномальных значений. Данные алгоритмы отличают достаточно высокие показатели предсказания, спрогнозированные относительно всех других классов для каждой из выборок, что отражают ROC-кривые (рис. 5). В данной выборке по выбранным трем алгоритмам графики показывают зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров. В целом ROC-анализ подтверждает фактически спрогнозированные результаты.

С целью контроля качества построенной модели и измерения производительности для каждого элемента был построен график среднего уменьшения точности (Mean Decrease in Accuracy), в котором наиболее высокие значения указывают на степень участия и важность в прогнозировании данного химического значения (рис. 6). На графике отображается высокая значимость как рудных элементов: As–Ag–Sb–Au–Cu, так и элементов с отрицательной корреляцией: Mg–Ca–Ti. Дополнительно отмечается повышенное влияние на прогнозирование V–Fe–Pb элементов, ранее не выделяемых как значимые, что должно учитываться при интерпретации прогнозных значений и может привести к изменению самого алгоритма обучения.

Результаты и обсуждение

Сравнение и интерпретация результатов, полученных с использованием разных алгоритмов, позволяют оценить их точность относительно друг друга и выбрать наиболее подходящий и эффективный метод для решения задач по прогнозированию. Для данной выборки наилучший результат был показан при использовании алгоритма Support Vector Machine.

Алгоритм Support Vector Machine основан на разделении геохимических и других данных через гиперплоскость. Например, разделение проб на «рудные» и «безрудные» через оптимальную гиперплоскость и определение трех областей поиска целей по обе стороны от данной гиперплоскости. Также используется генетический алгоритм для оптимизации гиперпараметров Support Vector Machine, чтобы уменьшить их влияние на результаты прогнозирования [6, 19].

Для сравнения предсказанных результатов по всей выборке построены схемы настоящих аномальных содержаний «рудного» фактора и карта аномальных предсказанных значений, получен-

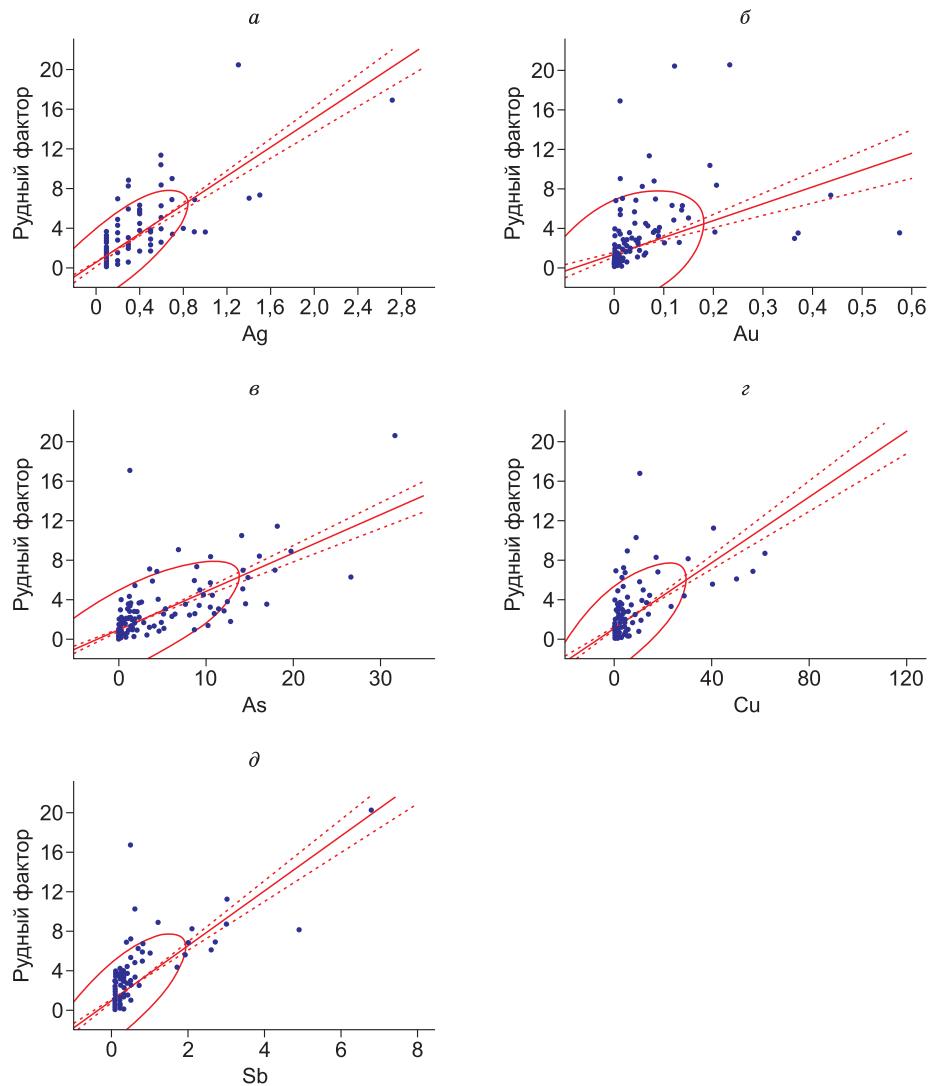


Рис. 4. Диаграммы рассеяния для «рудного» фактора и содержаний Ag (*a*), Au (*b*), As (*c*), Cu (*e*), Sb (*d*)

Fig. 4. Scatter plots for the “ore” factor and the contents of Ag (*a*), Au (*b*), As (*c*), Cu (*e*), Sb (*d*)

ных методом машинного обучения Support Vector Machine (рис. 7). На схемах отражены относительные значения в условных единицах «рудного» и прогнозного факторов. Участки разделены красной линией на две ранее использованные при обучении модели выборки: обучающую ($n = 250$) и тестовую ($n = 153$). При детальном изучении и сопоставлении двух карт вполне видны основные участки минерализации, направления минерализации, и в целом карты отличаются схожим обликом.

Полученные результаты можно считать удовлетворительными и использовать при планировании поисковых работ. В целом прогноз отражает основные участки минерализации, даже несмотря на крайне неравномерное опробова-

ние в ходе геологических маршрутов и качественные показатели (No, Low, Med, High), полученные при прогнозировании, что подтверждает его универсальность и возможность применения на поисковой стадии.

Заключение

Применение алгоритмов машинного обучения является эффективным инструментом для анализа данных и построения прогнозов. Машинное обучение позволяет автоматически обрабатывать большие объемы данных и выделять закономерности локализации перспективных участков. Для успешного применения методов машинного обучения необходимо правильно выбирать модели и алгоритмы прогнозирования,

Таблица 4

**Фактическое количество корректно предсказанных значений
для каждого уровня аномальных значений и среднее по всему алгоритму, %**

Table 4

**The actual number of accurately predicted values
for each level of outliers and the average for the entire algorithm, %**

№ п/п	Метод обучения	No	Low	Med	High	Сред.
1	Random Forest	99,2	81,3	47,4	93,3	89,6
2	Support Vector Machine	90,2	89,1	63,2	97,8	89,2
3	Neural Network (MLP)	89,3	89,1	36,8	95,6	86,4
4	K-Nearest Neighbors	84,4	81,3	36,8	86,7	80,4
5	Boosting (AdaBoost)	96,7	48,4	0	91,1	76
6	Decision Tree	94,3	53,1	26,3	66,7	73,6
7	Linear Discriminant Analysis	59,8	73,4	36,8	77,8	64,8
8	Naive Bayes	1,6	14,1	68,4	0	9,2

обучать их на соответствующих геологических данных и проводить качественную проверку результатов.

На примере Верхнеамгинского щелочного массива были построены модели МО и оценены результаты прогнозирования золотого оруденения восемью алгоритмами машинного обучения: Boosting (AdaBoost), Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis, Naive Bayes, Neural Network (Multilayer Perceptron), Random Forest и Support Vector Machine. Среди них выбраны и применены наиболее точные для прогнозирования оруденения.

Методы машинного обучения отличаются универсальностью и могут быть применимы при поисках на разные виды твердых полезных ископаемых. Ввиду комплексной и многофакторной обработки геохимических данных, особое зна-

чение методы МО могут иметь в первую очередь при поисках благородных металлов и полиметаллов, а также при прогнозировании скрытого оруденения по геохимическим данным, характеризующим надрудный уровень.

При поисковых работах методы МО могут эффективно применяться в комплексе с XRF-анализатором, непосредственно в полевых условиях, когда в предварительно подготовленной модели используются массивы «сырых» данных и оперативно определяется степень их перспективности, с последующим построением схем аномальных значений.

Алгоритмы машинного обучения применимы при камеральной обработке информации, интерпретации результатов разных видов работ поисковой стадии: штрафное опробование, опробование по вторичным и первичным ореолам рассеяния

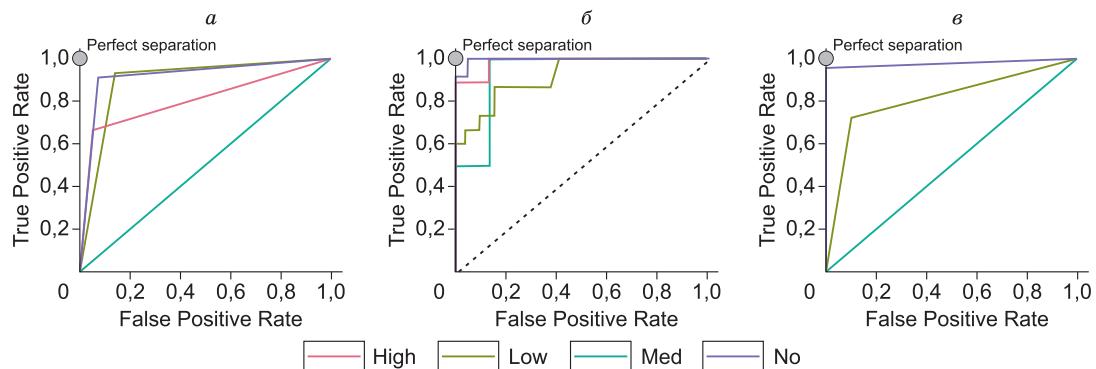


Рис. 5. ROC-кривая моделей: Support Vector Machine (a), Random Forest (b), Neural Network (c)

Fig. 5. ROC-curve of the models: Support Vector Machine (a), Random Forest (b), Neural Network (c)

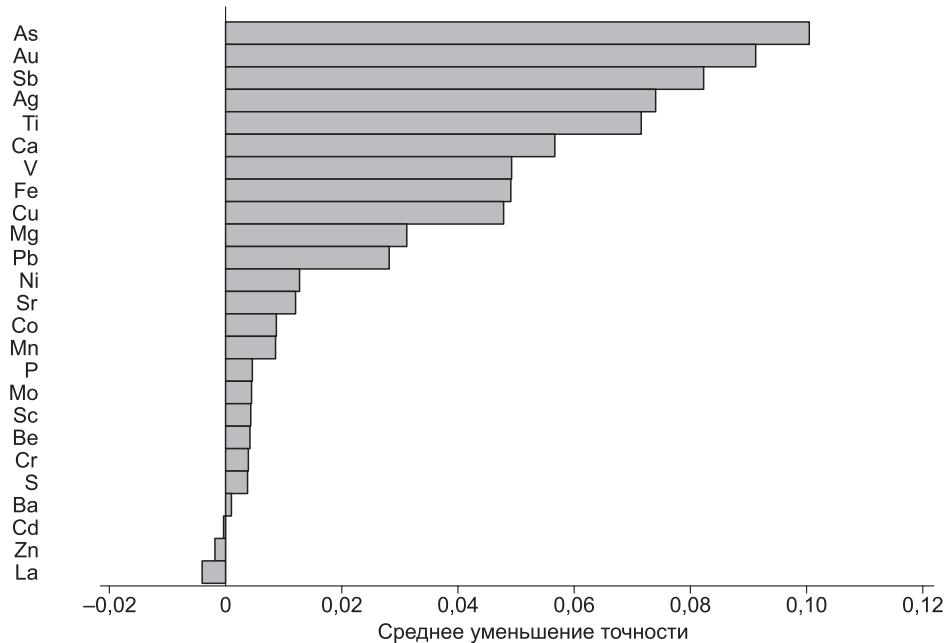


Рис. 6. График среднего уменьшения точности

Fig. 6. Plot of mean decrease in accuracy

и др. Достоинством МО является возможность актуализации модели, внесения изменений в свои алгоритмы в процессе получения первичной геологической информации. Свойство обучаемости позволяет прогнозируемым данным максимально соответствовать всей базе данных геологической информации, накопленной в процессе проведения геологоразведочных работ. При корректной обработке первичной информации и интерпретации получаемых результатов представляется возможным прогнозировать объекты-эталоны даже по небольшому количеству отобранных проб ($n < 100$).

Список литературы / References

1. Lindi O.T., Aladejare A.E., Ozoji T.M., et al. Uncertainty quantification in mineral resource estimation. *Natural Resources Research*. 2024;33:2503–2526. <https://doi.org/10.1007/s11053-024-10394-6>
2. Daya Sagar B.S., Cheng Q., Agterberg F. *Handbook of Geosciences*. Springer Cham; 2018. 366 p. <https://doi.org/10.1007/978-3-319-78999-6>
3. Mahboob M.A., Celik T., Genc B. Review of machine learning-based mineral resource estimation. *The Journal of the Southern African Institute of Mining and Metallurgy*. 2022;122(11):665–664. <https://dx.doi.org/10.17159/2411-9717/1250/2022>
4. Kanevski M., Foresti L., Kaiser C., et al. *Machine learning models for geospatial data. Handbook of theoretical and quantitative geography*. Lausanne: University of Lausanne; 2009, pp. 175–227.
5. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. Amsterdam: Elsevier; 2011. 432 p.
6. Akpan U.I., Starkey A. Review of classification algorithms with changing inter-class distances. *Machine Learning with Applications*. 2021;4:100031. <https://doi.org/10.1016/j.mlwa.2021.100031>
7. Прокопьев И.Р., Дорошкевич А.Г., Редина А.А. и др. Петрография щелочных пород, минералогия и условия образования рудной минерализации Верхнеамгинского золоторудного района (Алданский щит, Южная Якутия). В кн.: Эрнст Р.Э. (ред.) *Петрология магматических и метаморфических комплексов: материалы IX Всероссийской конференции с международным*

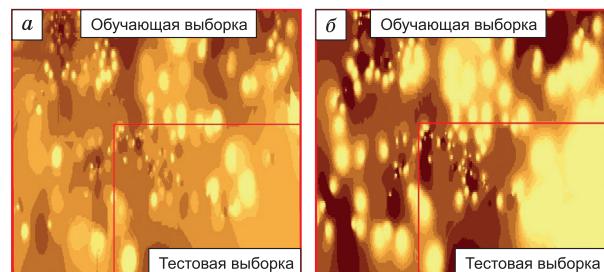


Рис. 7. Схемы настоящих «рудных» аномальных значений (a) и карты прогнозируемых аномальных значений (б) Верхнеамгинского щелочного массива Алдано-Станового щита. Положение рисунка 7 смотрите рисунок 1.

Fig. 7. Maps of real “ore” anomalous values (a) and a map of predicted anomalous values (б) of the Verkhneamginsky alkaline massif of the Aldan-Stanovoy shield. Position of figure 7 see figure 1.

участием. Выпуск 9, г. Томск, 28 ноября – 2 декабря 2017 года. Томск: Изд-во ЦНТИ; 2017. С. 360–366.

Prokopyev I.R., Doroshkevich A.G., Redina A.A., et al. Petrography of alkaline rocks, mineralogy and conditions of formation of ore mineralization of the Verkhneamginsky gold ore region (Aldan shield, Southern Yakutia). In: Ernst R.E. (ed.) *Petrology of igneous and metamorphic complexes: Proceedings of the 9th All-Russian Conference with International Participation. Issue 9, Tomsk, November 28 – December 2, 2017*. Tomsk: Izd-vo TsNTI; 2017, pp. 360–366. (In Russ.)

8. Анисимова Г.С., Кондратьева Л.А., Соколов Е.П. и др. Золотое оруденение лебединского и куранахского типов в Верхнеамгинском районе (Южная Якутия). *Отечественная геология*. 2018;(5):3–13. <https://doi.org/10.24411/0869-7175-2018-10010>

Anisimova G.S., Kondratyeva L.A., Sokolov E.P., et al. Gold mineralization of the Lebedinsky and Kuranakh types in the Verkhneamginsky region (Southern Yakutia). *Otechestvennaya Geologiya* = [Domestic geology]. 2018;(5):3–13. (In Russ.) <https://doi.org/10.24411/0869-7175-2018-10010>

9. Wang Q. Reservoir permeability prediction based on machine learning. *International Journal of Natural Resources and Environmental Studies*. 2024;3(1):129–138. <https://doi.org/10.62051/ijnres.v3n1.17>

10. Смелов А.П., Зедгенизов А.Н., Тимофеев В.Ф. Фундамент Северо-Азиатского кратона. Алдано-Становой щит. В кн.: *Тектоника, геодинамика и металлогенез Республики Саха (Якутия)*. М.: МАИК «Нauка/Интерпериодика»; 2001. С. 81–100.

Smelov A.P., Zedgenizov A.N., Timofeev V.F. Foundation of the North Asian Craton. Aldano-Stanovoy shield. In: *Tectonics, Geodynamics and Metallogeny of the Republic of Sakha (Yakutia)*. Moscow: MAIK “Nauka/Interperiodica”; 2001, pp. 81–100. (In Russ.)

11. Парфенов Л.М., Кузьмин М.И. *Тектоника, геодинамика и металлогенез территории Республики Саха (Якутия)*. М.: МАИК «Нauка/Интерпериодика»; 2001. 571 с.

Parfenov L.M., Kuzmin M.I. *Tectonics, Geodynamics and Metallogeny of the Republic Sakha (Yakutia)*. Moscow: MAIK “Nauka/Interperiodica”; 2001. 571 p. (In Russ.)

12. Соколов Е.П. Рудное золото Верхнеамгинского золотоносного района. В кн.: Биллер А.Я. (ред.) *Геология и минерально-сырьевые ресурсы Северо-Востока России: материалы всероссийской научно-практической конференции, г. Якутск, 31 марта – 2 апреля 2015 г.* Якутск: Издательский дом СВФУ; 2015. С. 458–462.

Sokolov E.P. Ore gold of the Verkhneamginsky gold-bearing region. In: Biller A.Ya. (ed.) *Geology and mineral resources of the northeast of Russia: materials of the All-Russian scientific and practical conference, Yakutsk, March 31 – April 2, 2015*. Yakutsk: NEFU Publishing House; 2015, pp. 458–462. (In Russ.)

13. Жумадилова Д.В., Прокопьев И.Р., Дорошкевич А.Г. и др. Золотоносность Верхнеамгинского рудного района (Алданский щит, Якутия). В кн. *Научно-методические основы прогноза, поисков, оценки месторождений алмазов, благородных и цветных металлов*. М.: ФГБУ «ЦНИГРИ»; 2019. С. 25–28.

Zhumadilova D.V., Prokopyev I.R., Doroshkevich A.G., et al. Gold-bearing capacity of the Verkhneamginsky ore region (Aldan shield, Yakutia). In: *Scientific and methodological foundations for forecasting, prospecting, and evaluating deposits of diamonds, precious and non-ferrous metals*. Moscow: Central Research Institute of Geology of Non-Ferrous Metals; 2019, pp. 25–28. (In Russ.)

14. Пономарчук А.В., Прокопьев И.Р., Дорошкевич А.Г., и др. $^{40}\text{Ar} / ^{39}\text{Ar}$ возраст щелочных пород Верхнеамгинского массива (Алданский щит, Южная Якутия). *Известия Томского политехнического университета. Инженеринг ресурсов*. 2019;330(3):28–39.

Ponomarchuk A.V., Prokopyev I.R., Doroshkevich A.G., et al. $^{40}\text{Ar} / ^{39}\text{Ar}$ age of alkaline rocks of the Verkhneamginsky massif (Aldan shield, South Yakutia). *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*. 2019;330(3):28–39. (In Russ.)

15. Никифорова З.С., Каженкина А.Г. Типоморфизм россыпного золота Хатырхайского рудно-россыпного узла (Верхнеамгинская площадь). *Природные ресурсы Арктики и Субарктики*. 2018;23(2):39–48. <https://doi.org/10.31242/2618-9712-2018-24-2-39-48>

Nikiforova Z.S., Kazhenkina A.G. Typomorphism of placer gold of the Khatyrkhay ore-placer cluster (Upper-Amga area). *Arctic and Subarctic Natural Resources*. 2018;23(2):39–48. (In Russ.) <https://doi.org/10.31242/2618-9712-2018-24-2-39-48>

16. Шатова Н.В., Шатов В.В., Молчанов А.В. и др. Геохимия и петрография гидротермально измененных пород Рябинового рудного поля (Южная Якутия) как основа прогноза золото-медно-порфирового оруденения. *Региональная геология и металлогенез*. 2020; (84):71–96.

Shatova N.V., Shatov V.V., Molchanov A.V., et al. Geochemistry and petrography of hydrothermally altered rocks of the Ryabinovoye ore field (South Yakutia) as a basis for prediction gold-copper-porphyry ore mineralization. *Regional Geology and Metallogeny*. 2020;(84):71–96. (In Russ.)

17. Дворник Г.П. Поисковые критерии и признаки золотого оруденения в калиевых щелочных массивах (на примере месторождений и рудопроявлений Алданского щита). *Литосфера*. 2017;17(6):118–135. <https://doi.org/10.24930/1681-9004-2017-6-118-135>

Dvornik G.P. Searching criterions and indications of gold mineralization in potash alkaline massifs (on example deposits and displays ore Aldanskiy Shield). *LITHOSPHERE (Russia)*. 2017;17(6):118–135. (In Russ.) <https://doi.org/10.24930/1681-9004-2017-6-118-135>

18. Barnes H.L. *Geochemistry of Hydrothermal Ore Deposits*. New York: Wiley & Sons. 1979. 798 p.

19. Dwarampudi T.M., Alugolu A. Review of machine learning classification algorithms. *International Journal of Progressive Research in Engineering Management and Science (IJPREMS)*. 2024;4(11):1104–1106
<https://www.doi.org/10.58257/IJPREMS36877>

Об авторах

ЧУДИНОВ Павел Леонидович, старший геолог, <https://orcid.org/0009-0003-3951-8413>, e-mail: Paulenotov@gmail.com
ФРИДОВСКИЙ Валерий Юрьевич, член-корреспондент РАН, доктор геолого-минералогических наук, директор, <https://orcid.org/0000-0002-4994-2141>, Scopus Author ID: 6505824025, SPIN: 3383-0398, e-mail: fridovsky@diamond.ysn.ru

Вклад авторов

Чудинов П.Л. – разработка концепции, верификация данных, проведение статистического анализа, проведение исследования, ресурсное обеспечение исследования, администрирование данных, создание черновика рукописи, визуализация и представление данных; **Фридовский В.Ю.** – верификация данных, администрирование данных, редактирование рукописи, визуализация данных, руководство исследованием и администрирование проекта

Конфликт интересов

Один из авторов – член-корреспондент РАН, доктор геолого-минералогических наук Фридовский Валерий Юрьевич является заместителем главного редактора журнала «Природные ресурсы Арктики и Субарктики». Авторам неизвестно о каком-либо другом потенциальном конфликте интересов, связанном с этой рукописью.

About the authors

CHUDINOV, Pavel Leonidovich, Senior Geologist, <https://orcid.org/0009-0003-3951-8413>, e-mail: Paulenotov@gmail.com
FRIDOVSKY, Valery Yurievich, Corresponding Member of the Russian Academy of Sciences, Dr. Sci. (Geol. and Mineral.), Director, <https://orcid.org/0000-0002-4994-2141>, Scopus Author ID: 6505824025, SPIN: 3383-0398, e-mail: fridovsky@diamond.ysn.ru

Authors' contribution

Chudinov P.L. – conceptualization, validation, formal analysis, investigation, resources, data curation, writing – original draft, visualization; **Fridovsky V.Yu.** – validation, data curation, writing – review & editing, visualization, supervision, project administration

Conflict of interest

One of the authors – Valery Yu. Fridovsky, Corresponding Member of the Russian Academy of Sciences, Dr. Sci. (Geol. and Mineral.), is the Deputy Editor-in-Chief for the journal “Arctic and Subarctic Natural Resources”. The authors are not aware of any other potential conflict of interest relating to this article.

Поступила в редакцию / Submitted 11.05.2025

Поступила после рецензирования / Revised 22.05.2025

Принята к публикации / Accepted 28.05.2025