

Геоинформатика. 2025. № 2. С. 77–82.
Geoinformatika. 2025;(2):77–82.

Решение практических задач

Научная статья
 УДК 004.67; 519.688
<https://doi.org/10.47148/1609-364X-2025-2-77-82>

Применение алгоритма машинного обучения «Случайный лес» в алмазопрогностике

© 2025 г. — Андрей Викторович Забелин^{а)}, Оксана Васильевна Тарских^{б)}

Акционерная компания «АЛРОСА» (ПАО); Мирный, Республика Саха (Якутия), Россия

^{а)}ZabelinAV@alrosa.ru, ^{б)}TarskikhOV@alrosa.ru

Аннотация: Химический состав индикаторных минералов кимберлитов — один из минералогических критериев, широко применяемых при поисках месторождений алмазов. Приведены результаты использования программных продуктов RandomForestIMK©, RPCA© с реализованным адаптированным алгоритмом «Случайный лес» для идентификации массивов данных химического состава индикаторных минералов кимберлитов относительно известных алмазоносных районов, полей, кимберлитовых тел Якутской алмазоносной провинции. Определены первичные критерии для принятия гипотезы о наличии в области денудации неизвестного коренного источника, принадлежности эталонной выборки району, полю.

Ключевые слова: *RandomForestIMK; RPCA; алмазопрогностика; алгоритм «Случайный лес»; индикаторных минералов кимберлитов; кимберлитовые тела; алмазоносный район; алмазоносный объект.*

Для цитирования: Забелин А.В., Тарских О.В. Применение алгоритма машинного обучения «Случайный лес» в алмазопрогностике // Геоинформатика. — 2025. — № 2. — С. 77–82. <https://doi.org/10.47148/1609-364X-2025-2-77-82>.

Practical application

Original article

Application of the «Random forest» machine learning algorithm in diamond prognostics

© 2025 — Andrey V. Zabelin^{а)}, Oksana V. Tarskikh^{б)}

Public Joint Stock Company «ALROSA»; Mirny, Republic of Sakha (Yakutia), Russia

^{а)}ZabelinAV@alrosa.ru, ^{б)}TarskikhOV@alrosa.ru

Abstract: The chemical composition of kimberlite indicator minerals (IMK) is one of the mineralogical criteria widely used in the search for diamond deposits. The results of using the RandomForestIMK©, RPCA© software products with the implemented adapted Random Forest algorithm for identifying data arrays of the IMC chemical composition regarding known diamondiferous regions, fields, kimberlitic bodies of the Yakut province are given. The primary criteria for accepting the presence of an unknown primary source hypothesis in the denudation area and the belonging of the reference sample to the region, field are defined.

Key words: *RandomForestIMK; RPCA; diamond forecasting; random forest; IMC; igneous formations; geological region; geological object.*

For citation: Zabelin A.V., Tarskikh O.V. Application of the «Random forest» machine learning algorithm in diamond prognostics. *Geoinformatika*. 2025;(2):77–82. <https://doi.org/10.47148/1609-364X-2025-2-77-82>. In Russ.

Введение

Химический состав индикаторных минералов кимберлитов (ИМК) — один из минералогических критериев, широко применяемых при поисках месторождений алмазов [1–4]. Для определения химического состава используется рентгеновский микроанализатор. Химический состав измеряется в мас. % содержания девяти оксидов: SiO₂, TiO₂, Al₂O₃, Cr₂O₃, FeO, MnO, MgO, CaO, Na₂O.

В современных условиях стоит задача определения потенциальной алмазоносности предпола-

гаемого объекта и одним из способов ее решения может служить сравнение проверяемой выборки с эталонной с минимальными трудозатратами. Ознакомление с существующими методами прогноза привело к выбору алгоритма «Случайный лес». Главный плюс алгоритма — точность прогноза, главный минус — принцип классификации остается «за кадром».

Специалистами НИГП АК «АЛРОСА» (ПАО) разработаны программы RandomForestIMK©, RPCA© [5] (зарегистрирована в реестре ЭВМ № 2018666180 от 13.12.2018), позволяющие приме-

нять алгоритм к массивам данных химического состава ИМК. В качестве эталонной выборки используются результаты химического анализа ИМК из кимберлитовых тел Якутской алмазоносной провинции (45250 анализов из 134 тел), проверенные на корректность. Каждый анализ привязан как территориально (алмазоносный район, кимберлитовое поле), так и по средней алмазоносности в кар/т. Эталонная выборка к сожалению не лишена разбалансированности: варьирует как число анализов, так и число изученных индикаторных минералов.

Алгоритм «Случайный лес» оценивает степень сходства проверяемой выборки и эталонной построенной модели-дерева [6]. Прогноз с помощью алгоритма выполняется относительно факторов: геологического положения (алмазоносный район, кимберлитовое поле, кимберлитовое тело), среднему для тела содержанию полезного компонента. Результат представляет собой набор таблиц по фактору. Каждая таблица содержит список названий, включенных в фактор эталона, и число схожих ИМК из проверяемой выборки относительно эталона.

Для минимизации проблемы с разбалансированностью (неравное число результатов для разных кимберлитовых тел) базы эталонов изменен подход к анализу прогноза, полученного с помощью алгоритма «Случайный лес» [7]. Проблему разбалансированности базы удалось минимизировать динамической выборкой, увеличением числа деревьев, выполнением алгоритма «Случайный лес» по одному и тому же эталонному набору данных несколько раз с последующим усреднением результатов и новым подходом к определению «победителя»: главным становится максимальное число минералов, по которым кимберлитовое тело признается «победителем», а затем рассчитывается суммарный процент сходства проверяемой выборки с кимберлитовыми телами из эталонной и таким образом «победитель» определяется окончательно [7].

Результаты исследований

Основными областями применения алгоритма «Случайный лес» в алмазопрогностике являются идентификация коренного источника ИМК в ореолах рассеяния в осадочных коллекторах и проверка гипотезы о существовании на исследуемой территории неизвестных кимберлитовых тел [8, 9]. Для определения достоверности прогноза алгоритм был применен для выборки ИМК нескольких кимберлитовых тел Западной Якутии, входящих в эталонную базу. Условия применения алгоритма — построение 5000 деревьев, повторов применения — 5. Результаты применения алгоритма представлены в табл. 1. Объект прогнозируется однозначно вне зависимости от размера выборки.

Столбец «Суммарное сходство» показывает победителя по наибольшему определенному значению прогноза ИМК. Значения в ячейках столбца

Таб. 1. Результаты применения алгоритма «Случайный лес» [10] для ИМК из кимберлитов трубок Интернациональная (анализы присутствуют в эталонной выборке)

Tabl. 1. The results of applying the "Random Forest" algorithm [10] for IMK (kimberlite minerals indicators) from kimberlite pipes International (analyses are included in the reference sample)

Объект	Суммарное сходство, %
Интернациональная (1413 ан.)	100,00
Летняя-2 (490 ан.)	100,00
ЦНИГРИ (240 ан.)	100,00
Дайковая (99 ан.)	100,00
Синильга (9 ан.)	100,00

«Суммарное сходство» таблиц нормировано, чтобы показать процент деревьев выборки, который был отнесен алгоритмом к эталонному лесу, указанному в строке. Значение 100 % — это однозначная идентификация объекта.

Использование в качестве проверяемой выборки ИМК из двух кимберлитовых тел, входящих в эталонную выборку, также показало однозначную идентификацию объектов (табл. 2). Число анализов по трубке Айхал — 692, 30 лет Айхала — 407 (соотношение 62,97 : 37,03 %).

Таб. 2. Результаты применения алгоритма «Случайный лес» [10] для ИМК из кимберлитов трубок 30 лет Айхала и Айхал (анализы присутствуют в эталонной выборке)

Tabl. 2. The results of applying the "Random Forest" algorithm [10] for IMK (kimberlite minerals indicators) from kimberlite pipes 30 years Aikhal and Aikhal (analyses are included in the reference sample)

Объект	Суммарное сходство, %
30 лет Айхала	37,03
Айхал	62,97

Следующим шагом проверена идентификация объекта при исключении его из эталонной выборки. Анализ выполнен для всех 134 кимберлитовых тел. В результате установлено, что принадлежность выборки к алмазоносному району (табл. 3) определяется однозначно для тел Далдыно-Алакитского алмазоносного района («суммарное сходство» выше 40 %, тогда как для Малоботуобинского и Среднемархинского значения существенно ниже, что может объясняться небольшим (в пределах 10) числом известных кимберлитовых тел в каждом из этих полей и преобладанием в эталонной выборке анализов ИМК из кимберлитов Далдыно-Алакитского района (32 116 анализов против 4887 для Малоботуобинского и 5048 — для Среднемархинского). При этом фактические районы присутствуют в пуле идентифицированных.

Табл. 3. Результаты применения алгоритма «Случайный лес» [10] — идентификация алмазоносного района (анализы отсутствуют в эталонной выборке)

Tabl. 3. Results of applying the "Random Forest" algorithm [10] — identification of a diamondiferous region (analyses are not included in the reference sample)

Кимберлитовое тело (условная алмазоносность)	Район (фактический)	Район (определен алгоритмом)	Суммарное сходство, %
Айхал (ураганная)	Далдыно-Алакитский	Далдыно-Алакитский	42,77
Юбилейная (высокая)	Далдыно-Алакитский	Далдыно-Алакитский	43,51
Удачная (высокая)	Далдыно-Алакитский	Далдыно-Алакитский	65,53
Академическая (низкая)	Далдыно-Алакитский	Далдыно-Алакитский	88,89
30 лет Айхала (низкая)	Далдыно-Алакитский	Далдыно-Алакитский	75,68
Мир (ураганная)	Малоботубинский	Далдыно-Алакитский	23,02
		Малоботубинский	43,36
Интернациональная (ураганная)	Малоботубинский	Далдыно-Алакитский	27,6
		Малоботубинский	43,81
Амакинская (низкая)	Малоботубинский	Далдыно-Алакитский	25,64
		Малоботубинский	55,59
Ботубинская (ураганная)	Среднемархинский	Далдыно-Алакитский	19,82
		Среднемархинский	47,44
Нюрбинская (ураганная)	Среднемархинский	Далдыно-Алакитский	44,65
		Малоботубинский	20,82
		Среднемархинский	26,24
Озерное (низкая)	Среднемархинский	Далдыно-Алакитский	48,61
		Среднемархинский	31,46

Близкие результаты получены при определении принадлежности анализов к кимберлитовому полю (табл. 4) за исключением трубки Айхал, для кимберлитовых тел Далдыно-Алакитского района принадлежность определяется однозначно.

При этом фактически район и поле присутствуют в числе «победителей» (наибольшее суммарное сходство), что говорит о схожих характеристиках этих районов относительно эталонной выборки и отражает специфику химического состава ИМК из кимберлитов. Адаптированный алгоритм позволяет однозначно идентифицировать либо уменьшить пул подходящих по характеристикам районов/полей/объектов. В спорных случаях можно проводить более точную идентификацию, используя прогноз по алгоритму «Случайный лес» в комплексе с программными продуктами «RN-Test»© [11] (зарегистрирована в реестре ЭВМ № 2020616588 от 04.06.2020), «МСА-Эксперт»© [12] (зарегистрирована в реестре ЭВМ № 2023610345 от 09.01.2023) и др.

Дальнейшие исследования планируются по применению адаптированного алгоритма для выборок ИМК из осадочных образований (россыпи).

Заключение

1. Применение программного продукта RandomForestIMK©, RPCA© с реализованным адаптированным алгоритмом «Случайный лес» расширило инструментарий для обработки результатов химического анализа ИМК с целью идентификации коренного источника ИМК в осадочных коллекторах и проверки гипотезы о существовании на исследуемой территории неизвестных кимберлитовых тел.

2. Принадлежность выборки к алмазоносному району, полю определяется однозначно для тел Далдыно-Алакитского алмазоносного района. Исключением является трубка Айхал.

3. Критерий «суммарное сходство» не менее 20,81 % эталонной выборки» может свидетельство-

Табл. 4. Результаты применения алгоритма «Случайный лес» [10] — идентификация кимберлитового поля (анализы отсутствуют в эталонной выборке)

Tab. 4. Results of applying the "Random Forest" algorithm [10] — identification of a kimberlite field (analyses are not included in the reference sample)

Кимберлитовое тело	Поле (фактическое)	Поле (определено алгоритмом)	Суммарное сходство, %
Айхал	Алакит-Мархинское	Далдынское	21,97
		Алакит-Мархинское	20,81
Юбилейная	Алакит-Мархинское	Алакит-Мархинское	26,76
30 лет Айхала	Алакит-Мархинское	Алакит-Мархинское	64,62
Удачная	Далдынское	Далдынское	42,28
Академическая	Далдынское	Далдынское	52,38
Мир	Мирнинское	Алакит-Мархинское	13,70
		Мирнинское	43,36
Интернациональная	Мирнинское	Накынское	19,32
		Мирнинское	43,81
Амакинская	Мирнинское	Алакит-Мархинское	19,93
		Мирнинское	55,59
Ботубобинская	Накынское	Алакит-Мархинское	14,48
		Накынское	47,44
Нюрбинская	Накынское	Алакит-Мархинское	33,47
		Малоботубобинский	20,82
		Мирнинское	26,24
Озерное	Накынское	Алакит-Мархинское	30,61
		Накынское	31,46

вать о принадлежности эталонной выборки геологическому району, полю.

4. Критерий «суммарное сходство» менее 50 % выборки из осадочного коллектора с локальным эталоном может свидетельствовать о наличии в области денудации неизвестного коренного источника [7].

Благодарности

Авторы выражают глубочайшую признательность руководителям центров ВГРЭ АК АЛРОСА (ПАО) С.С. Ратькову и к.г.-м.н. Е.В. Проценко за всемерную поддержку.

Список источников

1. *Соболев Н.В.* Глубинные включения в кимберлитах и проблема состава верхней мантии. — Новосибирск: Наука, 1974. — 263 с.
2. *Харькив А.Д., Зинчук Н.Н., Крючков А.И.* Геолого-генетические основы шлихо-минералогического метода поисков алмазных месторождений. — М.: Недра, 1995. — 348 с.
3. *Афанасьев В.П., Зинчук Н.Н., Похиленко Н.П.* Поисковая минералогия алмаза. — Новосибирск: Гео, 2010. — 650 с.
4. *Афанасьев В.П., Зинчук Н.Н., Похиленко Н.П.* Морфология и морфогенез индикаторных минералов кимберлитов. — Новосибирск: Гео, 2001. — 276 с.

5. Применение программных инструментов RandomForestIMK®, RPCA®, MCA-Эксперт® в комплексе позволяет не только оценить вероятность наличия кимберлитового тела на площади, но и его потенциальную продуктивность [9].

5. Программа для ЭВМ «RPCA»: свидетельство о гос. регистрации № 2018666180 Рос. Федерация / Ф.К. Зайцевский, А.В. Забелин, О.В. Тарских; правообладатель АК «АЛРОСА» (ПАО). – № 2018663378; дата поступления 26.11.2018; дата гос. регистрации 13.12.2018.
6. Шитиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. – Тольятти: Creative Commons, 2017. – 351 с.
7. Забелин А.В., Тарских О.В., Гибшер А.А., Мальковец В.Г. Применение адаптированного алгоритма «случайный лес» к массивам результатов химического анализа индикаторных минералов кимберлитов для алмазопромышленных работ на примере участка Сохолох-Верховье, Алакит-Мархинское кимберлитовое поле (западная Якутия) // Геология и минерально-сырьевые ресурсы северо-востока России: материалы XII Всероссийской научно-практической конференции, посвященной 65-летию ИГАБМ СО РАН. – Якутск: СВФУ им. М.К. Аммосова, 2022. – С. 292–297. DOI: 10.52994/9785751332846_2022_60.
8. Тарских О.В., Забелин А.В., Зайцевский Ф.К., Старостин П.В. Опыт применения алгоритма «случайный лес» в прогнозно-поисковых целях на примере Алакит-Мархинского кимберлитового поля (Якутия) // Геология и минерально-сырьевые ресурсы Северо-Востока России: материалы X Всероссийской научно-практической конференции с международным участием. – Якутск: СВФУ им. М.К. Аммосова, 2020. – С. 409–413.
9. Тарских О.В., Гибшер А.А., Забелин А.В., Овчинников И.М. Опыт применения комплексирования программ и технологий при обработке результатов химического анализа для прогнозирования и поисков коренных месторождений алмазов на примере участка Тегюрюк, Алакит-Мархинское кимберлитовое поле, западная Якутия // Геология и минерально-сырьевые ресурсы Северо-Востока России: материалы XI Всероссийской научно-практической конференции. – Якутск: СВФУ им. М.К. Аммосова, 2021. – С. 332–335. DOI: 10.52994/9785751331399_2021_88.
10. Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – № 1. – pp. 5–32. DOI: 10.1023/A:1010933404324.
11. Программа для ЭВМ «RN-Test»: свидетельство о гос. регистрации № 2020616588 Рос. Федерация / О.Е. Ковальчук, А.В. Забелин, Ф.К. Зайцевский; правообладатель АК «АЛРОСА» (ПАО). – № 2020615611; дата поступления 04.06.2020; дата гос. регистрации 18.06.2020.
12. Программа для ЭВМ «MCA-ЭКСПЕРТ»: свидетельство о гос. регистрации № 2023611824 Рос. Федерация / В.К. Гаранин, А.В. Забелин, А.А. Гибшер, В.Г. Мальковец, А.Ю. Никифорова, Ф.К. Зайцевский; правообладатель АК «АЛРОСА» (ПАО). – № 2023610345; дата поступления 09.01.2023; дата гос. регистрации 25.01.2023.

References

1. Sobolev N.V. Glubinnye vklucheniya v kimberlitakh i problema sostava verkhnei mantii [Deep inclusions in kimberlites and the problem of the composition of the upper mantle]. Novosibirsk: Nauka; 1974. 263 p.
2. Khar'kiv A.D., Zinchuk N.N., Kryuchkov A.I. Geologo-geneticheskie osnovy shlikho-mineralogicheskogo metoda poiskov almaznykh mestorozhdenii [Geological and genetic foundations of the heavy concentrate mineralogical method of prospecting for diamond deposits]. Moscow: Nedra; 1995. 348 p.
3. Afanas'ev V.P., Zinchuk N.N., Pokhilenko N.P. Poiskovaya mineralogiya almaza [Diamond prospecting mineralogy]. Novosibirsk: Geo; 2010. 650 p.
4. Afanas'ev V.P., Zinchuk N.N., Pokhilenko N.P. Morfologiya i morfogenez indikatornykh mineralov kimberlitov [Morphology and morphogenesis of kimberlite indicator minerals]. Novosibirsk: Geo; 2001. 276 p.
5. Computer program “RPCA”: certificate of state registration No. 2018666180 Rus. Federation. F.K. Zaitsevsky, A.V. Zabelin, O.V. Tarskikh; copyright holder PJSC ALROSA. «RPCA». No. 2018663378; date of receipt 26.11.2018; date of state registration 13.12.2018.
6. Shitikov V.K., Mastitskii S.Eh. Klassifikatsiya, regressiya i drugie algoritmy Data Mining s ispol'zovaniem R [Classification, regression and other algorithms Data Mining with using R]. Tolyatti: Creative Commons; 2017. 351 p.
7. Zabelin A.V., Tarskikh O.V., Gibsher A.A., Malkovets V.G. Adapted algorithm “random forest” application on chemical composition of kimberlite indicator minerals data arrays for predictive search on the example of Sokhsolokh-Verhov'ye area, Alakit-Markha kimberlite field (western Yakutia). In: Geologiya i mineral'no-syr'evye resursy severo-vostoka Rossii: materialy XII Vserossiiskoi nauchno-prakticheskoi konferentsii, posvyashchennoi 65-letiyu IGABM SO RAN. Yakutsk: SVFU im. M.K. Ammosova, 2022. pp. 292–297. DOI: 10.52994/9785751332846_2022_60.
8. Tarskikh O.V., Zabelin A.V., Zaitsevskiy F.K., Starostin P.V. Results of the algorithm Random Forest application to predictive search on the example of Alakit-Markha kimberlite field (Yakutia). In: Geologiya i mineral'no-syr'evye resursy Severo-Vostoka Rossii: materialy X Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem. Yakutsk: SVFU im. M.K. Ammosova; 2020. pp. 409–413.
9. Tarskikh O.V., Gibsher A.A., Zabelin A.V., Ovchinnikov I.M. Results of the program and technology complex application to predictive search on the example of Teguryuk area, Alakit-Markha kimberlite field (Yakutia). In: Geologiya i mineral'no-syr'evye resursy Severo-Vostoka Rossii: materialy XI Vserossiiskoi nauchno-prakticheskoi konferentsii. Yakutsk: SVFU im. M.K. Ammosova; 2021. pp. 332–335.
10. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5–32. DOI: 10.1023/A:1010933404324.
11. Computer program “RN-Test”: certificate of state registration No. 2020616588 Rus. Federation. O.E. Kovalchuk, A.V. Zabelin, F.K. Zaitsevsky; copyright holder PJSC ALROSA. No. 2020615611; date of receipt 04.06.2020; date of state registration 18.06.2020.
12. Computer program “KIM-EXPERT”: certificate of state registration No. 2023611824 Rus. Federation. V.K. Garanin, A.V. Zabelin, A.A. Gibsher, V.G. Malkovets, A.Y. Nikiforova, F.K. Zaitsevsky; copyright holder PJSC ALROSA. No. 2023610345; date of receipt 09.01.2023; date of state registration 25.01.2023.

Статья поступила в редакцию 06.02.2025 г., одобрена после рецензирования 09.03.2025 г., принята к публикации 01.06.2025 г.
The article was submitted 06.02.2025; approved after reviewing 09.03.2025; accepted for publication 01.06.2025.

Информация об авторах

Забелин Андрей Викторович

Эксперт центра геоинформационных технологий
Акционерная компания «АЛРОСА» (ПАО),
Вилуйская геологоразведочная экспедиция
678174 Мирный, Республика Саха (Якутия),
ул. Вилуйская, д. 7Б
e-mail: ZabelinAV@alrosa.ru
SPIN-код: 7772-0301
AuthorID: 998861

Тарских Оксана Васильевна

Кандидат геолого-минералогических наук,
эксперт экспертного геологического центра
Акционерная компания «АЛРОСА» (ПАО),
Вилуйская геологоразведочная экспедиция
678174 Мирный, Республика Саха (Якутия),
ул. Вилуйская, д. 7Б
e-mail: TarskikhOV@alrosa.ru
ORCID: 0000-0002-1004-0546
SPIN-код: 1052-2422
AuthorID: 68083

Information about authors

Andrei V. Zabelin

Expert (Vilyuyskaya Geological Exploration Expedition)
PJSC ALROSA
7, build. B, Vilyujskaya Str., Mirny, 678174,
Republic of Sakha (Yakutia), Russia
e-mail: ZabelinAV@alrosa.ru
SPIN-код: 7772-0301
AuthorID: 998861

Oksana V. Tarskikh

Candidate of Geological and Mineralogical Sciences
Expert (Vilyuyskaya Geological Exploration Expedition)
PJSC ALROSA
7, build. B, Vilyujskaya Str., Mirny, 678174,
Republic of Sakha (Yakutia), Russia
e-mail: TarskikhOV@alrosa.ru
ORCID: 0000-0002-1004-0546
SPIN-код: 1052-2422
AuthorID: 68083