

Автоматизированный анализ изотопно-геохимических характеристик битумоидов и нефтей с применением методов машинного обучения

Научный руководитель – Краснова Елизавета Андреевна

Лекомцева Анастасия Валерьевна

Студент (магистр)

Московский государственный университет имени М.В.Ломоносова, Геологический факультет, Кафедра геологии и геохимии горючих ископаемых, Москва, Россия

E-mail: anasta.lekomceva@gmail.com

Изотопные данные широко применяются для геохимической типизации нефтей, битумоидов и керогена. Анализ изотопного состава фракций помогает выявить вторичные изменения углеводородов (биodeградацию, окисление и др.) [1]. В свою очередь, общий изотопный состав углерода служит индикатором для определения источника органического вещества, палеогеографических условий седиментации и идентификации связей в системе «нефть—материнская порода» [1].

В рамках работы разработан инструмент для автоматизированной изотопно-геохимической характеристики битумоидов и нефтей. В его основу легло обучение моделей на данных лабораторных исследований изотопно-фракционного анализа более 800 образцов. Для решения задачи применялся алгоритм машинного обучения CatBoostClassifier. Обучающая выборка характеризовалась наличием выбросов и дисбаланса классов. Для обработки пропусков данных использовались методы восстановления значений IterativeImputer и KNNImputer, а для балансировки классов применялся алгоритм SMOTE. После предобработки данные проходили дополнительную проверку на геологическую корректность с помощью сравнения графиков исходной и обработанной выборок. Сгенерированные SMOTE синтетические образцы и восстановленные пропуски повторяют биологический тренд и пригодны для обучения моделей.

Всего было обучено пять моделей: (1) определения типа органического вещества нефтей и битумоидов, (2) прогнозирования термического воздействия, (3) окисления, (4) биodeградации и (5) миграции углеводородов. В качестве метрики оценки качества использовалась Macro F1-score, поскольку важным являлось равномерное качество классификации по всем классам. По результатам тестирования на независимой выборке были получены значения Macro F1 в диапазоне от 0.89 до 0.94. Полученные результаты (Macro F1 > 0.85) свидетельствуют о высокой обобщающей способности моделей и их надежности при классификации образцов.

Наилучшие результаты были достигнуты для моделей, решающих задачи определения источника органического вещества и диагностики термического воздействия. Это объясняется тем, что данные процессы оставляют наиболее контрастный и однозначный изотопно-геохимический след в углеводородах.

Разработанный инструмент на основе методов машинного обучения демонстрирует высокую эффективность и может использоваться для ускорения и стандартизации геохимического анализа при решении задач нефтегазовой геологии.

Источники и литература

- 1) Галимов Э. М. Природа биологического фракционирования изотопов / Э. М. Галимов. – Москва : Наука, 1981. – 247 с.