

СОВРЕМЕННЫЕ МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА СЕЙСМИЧЕСКИХ ДАННЫХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

MODERN METHODS OF PROCESSING AND ANALYZING SEISMIC DATA BASED ON MACHINE LEARNING

**A. Aykin
V. Nikishin
P. Igin
S. Remizov**

Summary. The control of seismic events during mining operations is an extremely important aspect of the safety and efficiency of mining. To ensure the control of seismic events during mining operations, various methods are used, such as monitoring of seismic activity, modeling of seismic processes, the use of special technologies and equipment. The most effective model for predicting the occurrence of seismic events is the RandomForestClassifier (Accuracy score 0,98606). This is also confirmed by analytical data, where the result of similar work is similar in value obtained probabilities of occurrence of seismic events.

Keywords: mountain impact, machine learning, seismic events, seismic observation data.

Айкин Андрей Владимирович

Директор, ООО НЦ БИТ, Кемерово
andrey.aykin@yandex.ru

Никишин Виктор Викторович

Ведущий инженер-программист, ООО НЦ БИТ, Кемерово
nikishin.vv@ncbit.ru

Игин Павел Константинович

Ведущий программист, ООО НЦ БИТ, Кемерово
igin.pk@ncbit.ru

Ремизов Станислав Вадимович

Начальник научного отдела,
ООО НПО «АЛЗАМИР», Кемерово
remizov.stanislav@yandex.ru

Аннотация. Контроль сейсмических событий при ведении горных работ является крайне важным аспектом безопасности и эффективности добычи полезных ископаемых. Для обеспечения контроля сейсмических событий при ведении горных работ используются различные методы, такие как мониторинг сейсмической активности, моделирование сейсмических процессов, применение специальных технологий и оборудования. Наиболее эффективной моделью предсказания возникновения сейсмических событий является — RandomForestClassifier (Accuracy score 0,98606). Это так же подтверждается аналитическими данными, где результатом похожих работ являются схожие по значению полученные вероятности возникновения сейсмических событий.

Ключевые слова: горный удар, машинное обучение, сейсмические события, данные сейсмических наблюдений.

Введение

Сейсмические исследования являются одним из основных методов изучения структуры и состава земной коры. Современные методы обработки и анализа сейсмических данных позволяют получить более точные и детальные сведения о геологическом строении подземных образований, что в свою очередь помогает улучшить прогнозирование различных геологических процессов.

Контроль сейсмических событий позволяет предотвратить возможные аварии, уменьшить риск разрушения горных выработок, защитить окружающую среду и обеспечить безопасность работников. Кроме того, контроль сейсмических событий помогает оптимизировать процессы добычи и улучшить качество добытых материалов.

Для обеспечения контроля сейсмических событий при ведении горных работ используются различные ме-

тоды, такие как мониторинг сейсмической активности, моделирование сейсмических процессов, применение специальных технологий и оборудования.

В настоящий момент перспективным направлением исследований является применение алгоритмов машинного обучения для анализа сейсмических данных. Алгоритмы нейронных сетей показали высокую эффективность при классификации сейсмических событий [1, 2]. Так же ведутся исследования по выявлению очагов сейсмических событий методами кластерного анализа [3, 4].

Цель исследования: Разработка модели машинного обучения для предсказания сейсмических ударов.

Задачи исследования:

1. Изучить и описать массив данных о возникновении сейсмических событий и параметров, которые были зарегистрированы в момент их происхождения. Выявить статистические характеристики дата

сета. Выявить ключевые параметры для дальнейшего анализа. Провести обработку данных.

2. Подобрать модели классификации для предсказания высокоэнергетических сейсмических ударов. Подобрать наиболее приемлемые параметры настройки моделей.
3. Оценить производительности моделей, описать результаты исследования.

Объект и методы исследования

Опасности, связанные с подземной добычей угля, многообразны. Однако одной из очень распространенных опасностей является сейсмическая опасность. Сейсмическая опасность — это вероятность того, что в данном географическом районе в течение данного промежутка времени произойдет землетрясение с интенсивностью движения грунта, превышающей определенный порог. Обычной практикой является использование показаний энергии и количества сейсмических событий, зарегистрированных за определенный промежуток времени, для прогнозирования возникновения горного удара. Горный удар — это сейсмическое событие с энергией более 1000 джоулей.

Для проведения исследования был выбран дата сет размещенный на ресурсе «Kaggle» [5]. Данные, представленные в этом наборе данных, взяты из угольной шахты Забже-Бельшовице в Польше.

Дата сет состоит из 19 столбцов, 18 независимых и 1 зависимая переменная. Из 18 независимых переменных 4 имеют тип str. Всего 2584 наблюдения. Не за-

висимые переменные содержат данные сейсмических наблюдений (результат оценки сейсмической опасности сдвига, полученный сейсмоакустическим методом, тип сдвигов, сейсмическая энергия, количество импульсов и т.д.). Зависимая переменная описывает факт возникновения или отсутствия горного удара за период смены (8 часов).

В процессе анализа и предобработки дата сета показано, что зависимая переменная class принимает только два значения 0 и 1. Соответственно для прогнозирования сейсмических событий на данном наборе данных необходимо решить задачу бинарной классификации. Для решения данной задачи были выбраны следующие алгоритмы:

- LogisticRegression;
- RandomForestClassifier;
- Perceptron;
- Support vector machines (svm).

Для анализа работы алгоритмов использованы такие метрики как — Ф — мера, матрица ошибок, accuracy_score.

Результаты исследования

Процедура предобработки данных содержит следующие операции:

- удаление пустых или повторяющихся данных;
- нормализация числовых данных (тип нормализации — z — оценка);
- кодирование вещественных данных;
- разбиение на тренировочную и тестовую выборку.

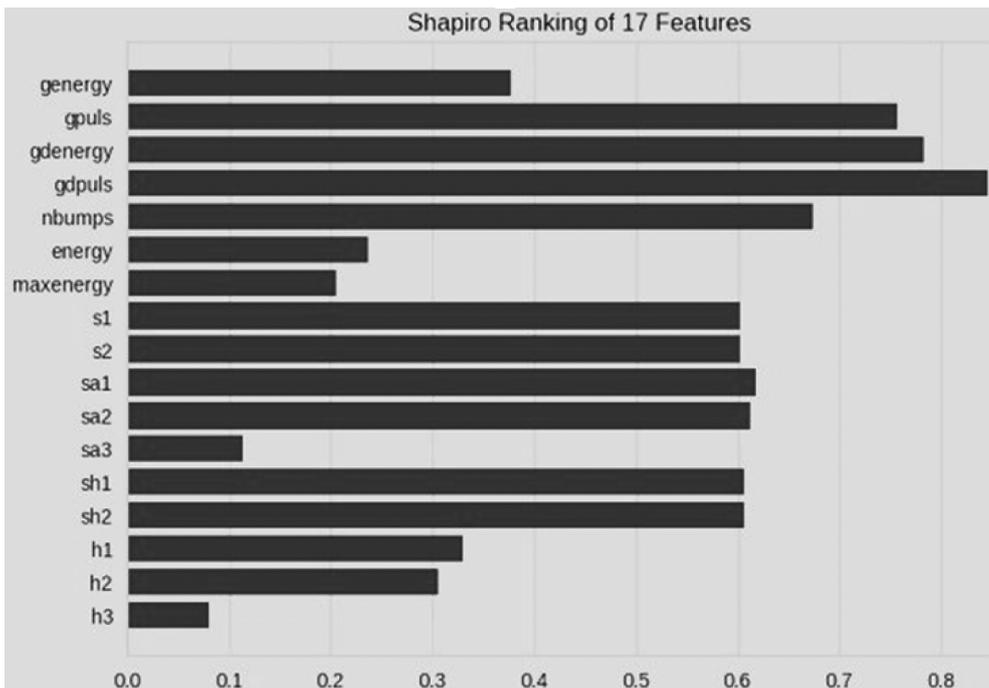


Рис. 1. Ранжирование признаков дата сета

Далее проведено ранжирование признаков (рисунок 1).

После обучения моделей были проведены предсказания, и сравнение точности предсказания используя метрику качества (таблица 1).

Таблица 1.

Результаты предсказания

Модель	Accuracy score	Матрица ошибок		Φ — мера
Logistic Regression	0,93266	2405	9	0,965088
		165	5	
RF	0,98606	2411	3	0,99259
		33	137	
Perceptron	0,93382	2413	1	0,96577
		170	0	
Svm	0,94543	2408	6	0,97155
		135	35	

Согласно метрикам качества максимально эффективно предсказывать возникновение сейсмических событий можно с применением алгоритма Random Forest.

Заключение

Использование алгоритмов машинного обучения позволяет с высокой точностью предсказывать возможные

сейсмические удары, что может быть полезно для принятия мер по предотвращению чрезвычайных ситуаций и защите населения.

Наиболее эффективной моделью предсказания возникновения сейсмических событий является — RandomForestClassifier (Accuracy score 0,98606). Это так же подтверждается аналитическими данными, где результатом похожих работ являются схожие по значению полученные вероятности возникновения сейсмических событий.

Дальнейшее развитие данной работы состоит в улучшении качества данных, использование более сложных алгоритмов машинного обучения и расширение набора признаков для более точного прогнозирования сейсмических ударов. Также возможно внедрение данной модели в реальные системы мониторинга и предупреждения о сейсмических ударах.

Решении вопроса прогнозирования возникновения сейсмических ударов является актуальной и перспективной задачей, которая может значительно улучшить безопасность и защиту населения от природных катастроф.

ЛИТЕРАТУРА

1. Сергунин М.П., Еременко В.А. Обучение нейронной сети предсказывать параметры сдвижения горных пород налегающей толщи на основании данных о трещиноватости массива на примере рудника «Заполярный» // Горный информационно-аналитический бюллетень. 2019. № 10. С. 106–116. DOI: 10.25018/0236-1493-2019-10-0-106-116.
2. Liang WZ, Sari YA, Zhao GY, McKinnon S, Wu H. Probability Estimates of Short-Term Rockburst Risk with Ensemble Classifiers // Rock Mechanics and Rock Engineering. 2021. 54. P. 1799–1814. doi.org/10.1007/s00603-021-02369-3.
3. А.П., Ревин И.Е., Морозов К.В. Композитная модель анализа данных сейсмического мониторинга при ведении горных работ на примере Кукисвумчоррского месторождения АО «Апатит» // Записки Горного института. 2023. Т. 262. С. 571–580. DOI: 10.31897/PMI.2023.9
4. Захаров Л.А., Мартюшев Д.А., Пономарева И.Н. Прогнозирование динамического пластового давления методами искусственного интеллекта // Записки Горного института. 2022. Т. 253. С. 23–32. DOI: 10.31897/PMI.2022.11
5. Seismic Bumps Data Set [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/pranabroy94/seismic-bumps-data-set/data> (Дата обращения: 20.03.2024).

© Айкин Андрей Владимирович (andrey.aykin@yandex.ru); Никишин Виктор Викторович (nikishin.vv@ncbit.ru); Игин Павел Константинович (igin.pk@ncbit.ru); Ремизов Станислав Вадимович (remizov.stanislav@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»