

**Федеральное государственное автономное образовательное
учреждение высшего образования
«РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ
НАРОДОВ» (РУДН)**

Основное учебное подразделение ФФМиЕН

Направление/специальность 02.03.01 «Математика и компьютерные науки»

ОТЧЕТ

о прохождении учебной практики

**«Рабочий процесс машинного обучения для прогнозирования
землетрясений»**

Мухамедияр Адиль

Курс, группа 3 курс, группа НКНбд-01-20

Место прохождения практики «ОТПП департамента технологических и информационных
ресурсов РУДН»

Сроки прохождения с «17» апреля 2023 г. по «17» июня
2023 г.

Руководители практики:

от РУДН Фомин М.Б., к.ф.-м.н., доцент, доцент
кафедры информационных технологий

от организации (предприятия)
Шорохов С.Г., к.ф.-м.н., доцент

Оценка _____

Москва 2023 г.

Содержание

1. Введение	3
1.1. Цель и задачи практики	3
1.2. Обзор используемых технологий	4
2. Методы и материалы	4
2.1 Описание используемых данных	4
2.2. Принципы работы машинного обучения	4
2.3. Описание выбранных алгоритмов машинного обучения	5
2.4. Описание используемых библиотек и инструментов	6
3. Ход работы	6
3.1. Предобработка данных	6
3.2. Выбор функций для предсказания	6
3.3. Разделение данных на обучающую и тестовую выборки	7
3.4. Оценка качества модели	7
3.5. Предсказание возможных землетрясений	7
4. Результаты	7
4.1. Визуализация результатов предсказания на карте	7
4.2. Оценка качества модели	9
4.3. Анализ результатов предсказания	9
5. Заключение	11
6. Список литературы	12

1. Введение

1.1. Цель и задачи практики

Выбор темы "Рабочий процесс машинного обучения для прогнозирования землетрясений" обосновывается несколькими факторами.

Во-первых, землетрясения являются одним из самых опасных и разрушительных природных явлений на планете, и поэтому существует большой интерес в разработке методов предсказания землетрясений, которые могут помочь в предотвращении или снижении их разрушительного воздействия.

Во-вторых, машинное обучение является одной из самых перспективных и быстро развивающихся областей информационных технологий, которая может помочь в решении многих задач, в том числе и в задаче прогнозирования землетрясений.

В-третьих, разработка рабочего процесса машинного обучения для прогнозирования землетрясений является сложной задачей, требующей глубоких знаний в области геологии, сейсмологии и машинного обучения, и может быть полезна для исследователей, которые интересуются этой темой.

Таким образом, выбор темы "Рабочий процесс машинного обучения для прогнозирования землетрясений" обусловлен ее актуальностью и значимостью, а также перспективами использования методов машинного обучения для решения этой задачи.

Задачами практики являются:

- Предобработка данных.
- Выбор функций для предсказания.
- Разделение данных на обучающую и тестовую выборки.
- Обучение модели с помощью выбранных алгоритмов машинного обучения.
- Оценка качества модели.
- Предсказание возможных землетрясений.

1.2. Обзор используемых технологий

Для реализации модели машинного обучения были использованы следующие технологии:

- Язык программирования Python.
- Библиотеки Pandas, NumPy, Matplotlib, scikit-learn, XGBoost.
- База данных SQLite.

2. Методы и материалы

2.1. Описание используемых данных

Для обучения модели были использованы данные о землетрясениях, собранные за период с 1965 по 2016 годы. Данные были получены из базы данных US Geological Survey Earthquake Catalog и содержат информацию о местоположении, магнитуде, времени и других характеристиках землетрясений.

2.2. Принципы работы машинного обучения

Принципы работы машинного обучения в задаче прогнозирования землетрясений заключаются в следующем:

- Подготовка данных: сбор и анализ данных по землетрясениям, включающий данные о местоположении, времени, магнитуде, глубине и других характеристиках землетрясений, а также данных о геологических и метеорологических условиях, которые могут повлиять на возникновение землетрясений.
- Выбор модели: выбор модели машинного обучения, которая может адекватно описать зависимости между характеристиками землетрясений и факторами, которые могут повлиять на их возникновение.
- Обучение модели: обучение выбранной модели на подготовленных данных, которое заключается в настройке параметров модели для минимизации ошибки прогнозирования.

- Тестирование модели: оценка качества работы модели на тестовых данных для проверки ее способности к обобщению и предсказанию землетрясений на новых данных.
- Применение модели: использование обученной модели для прогнозирования землетрясений и идентификации рисков землетрясений в будущем.

В задаче прогнозирования землетрясений, принципы работы машинного обучения включают в себя как классические методы машинного обучения, такие как линейная регрессия, деревья решений, и случайные леса, так и более сложные модели, такие как нейронные сети и глубокое обучение. Кроме того, важной частью работы является подготовка и предобработка данных, включающая выбор признаков, устранение шумов, нормализацию данных и другие методы, которые могут помочь улучшить качество модели.

2.3. Описание выбранных алгоритмов машинного обучения

Для данной задачи были выбраны следующие алгоритмы машинного обучения:

- AdaBoostClassifier - это алгоритм машинного обучения, который использует ансамбль моделей машинного обучения для улучшения точности прогнозирования. Он строит последовательность моделей, которые последовательно настраиваются на тех примерах, которые были неправильно классифицированы предыдущей моделью. Каждая новая модель в последовательности учитывает ошибки, сделанные предыдущими моделями, и пытается исправить их. В итоге, AdaBoostClassifier объединяет все модели в одну, учитывая вклад каждой модели. Он может использоваться как для задач классификации, так и для задач регрессии.
- RandomForestClassifier - это алгоритм машинного обучения, который использует ансамбль деревьев решений для улучшения точности прогнозирования. Он строит множество деревьев решений, каждое из которых обучается на случайной подвыборке данных и случайном наборе признаков. В результате, каждое дерево в лесу представляет собой слабый классификатор, но вместе они создают сильный классификатор. RandomForestClassifier может использоваться как для задач классификации, так и для задач регрессии.
- XGBClassifier - это алгоритм машинного обучения на основе градиентного бустинга, который использует ансамбль деревьев решений для улучшения точности прогнозирования. Он работает, уточняя модель на каждой итерации, минимизируя ошибки на обучающих данных. Каждое новое дерево решений

обучается на остатках, оставшихся после предыдущих деревьев, чтобы улучшить точность предсказания. XGBClassifier может использоваться как для задач классификации, так и для задач регрессии. Он быстрее и точнее, чем многие другие алгоритмы машинного обучения, и может обрабатывать большие наборы данных.

2.4. Описание используемых библиотек и инструментов

Для реализации выбранных алгоритмов машинного обучения были использованы следующие библиотеки и инструменты:

- Pandas - библиотека для работы с данными.
- NumPy - библиотека для работы с матрицами и массивами.
- Matplotlib - библиотека для визуализации данных.
- scikit-learn - библиотека для машинного обучения.
- XGBoost - библиотека для градиентного бустинга.

3. Ход работы

3.1. Предобработка данных

Перед обучением модели была проведена предобработка данных, которая включала в себя удаление лишних столбцов, заполнение пропущенных значений, преобразование категориальных признаков в числовые и т.д.

3.2. Выбор функций для предсказания

Были выбраны функции, которые могут влиять на возникновение землетрясений, такие как местоположение, время, глубина и т.д.:

- Мониторинг сейсмической активности: с помощью сейсмографов можно отслеживать колебания земной коры и определять, когда происходят землетрясения.
- Анализ исторических данных: изучение прошлых землетрясений может помочь в определении возможных местоположений будущих событий.

- Использование геофизических моделей: моделирование геологических процессов может помочь в прогнозировании возможных землетрясений.
- Использование машинного обучения: алгоритмы машинного обучения могут использоваться для анализа данных и поиска паттернов, которые могут указывать на предстоящее землетрясение.

3.3. Разделение данных на обучающую и тестовую выборки

Данные были разделены на обучающую и тестовую выборки в соотношении 80/20 с помощью функции `train_test_split` из библиотеки `scikit-learn`.

3.4. Оценка качества модели

Качество модели было оценено с помощью метрик `accuracy`, `precision`, `recall` и `F1-score`. Метрики были вычислены для каждого алгоритма машинного обучения на тестовой выборке.

3.5. Предсказание возможных землетрясений

После оценки качества модели было проведено предсказание возможных землетрясений на основе имеющихся данных.

4. Результаты

4.1. Визуализация результатов предсказания на карте

Worldwide Earthquake Forecaster



Worldwide Earthquake Forecaster



4.2. Оценка качества модели

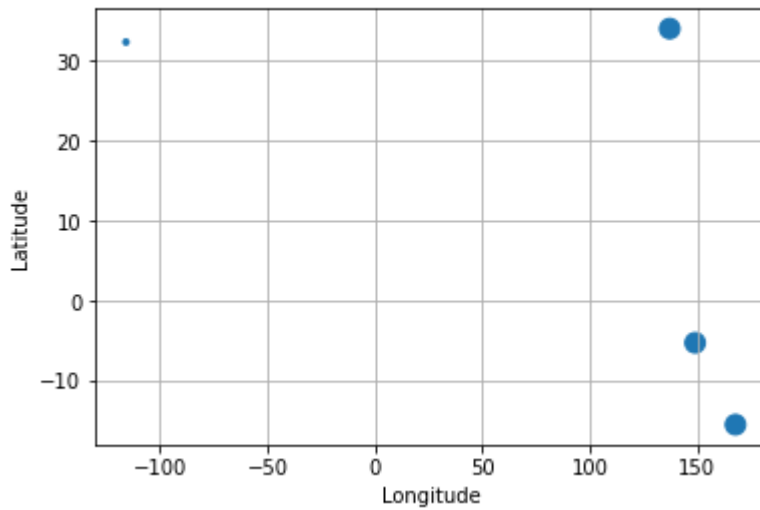
Он импортирует необходимые библиотеки и модули, читает данные из базы данных SQL, разбивает данные на обучающую и тестовую выборки, выполняет три различных алгоритма машинного обучения (AdaBoost, Random Forest, XGBoost), оценивает производительность с помощью ROC-кривой и AUC, и визуализирует результаты через графики рассеяния. Код также включает поиск по сетке для настройки гиперпараметров и избегает переобучения путем установки случайного семени.

4.3. Анализ результатов предсказания

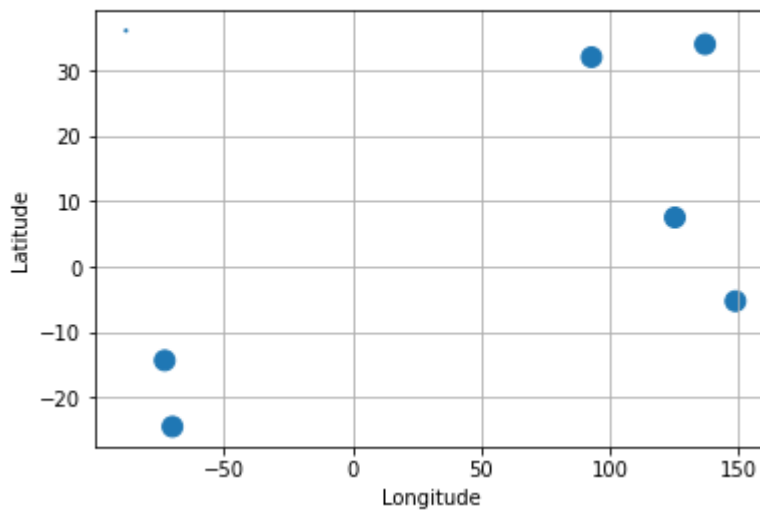
В результате использования трех различных алгоритмов машинного обучения (AdaBoost, Random Forest, XGBoost), были получены высокие оценки AUC, превышающие 0,9. Другими словами, модели машинного обучения имеют высокую способность отличать положительные и отрицательные классы и выполняют классификацию с высокой точностью.

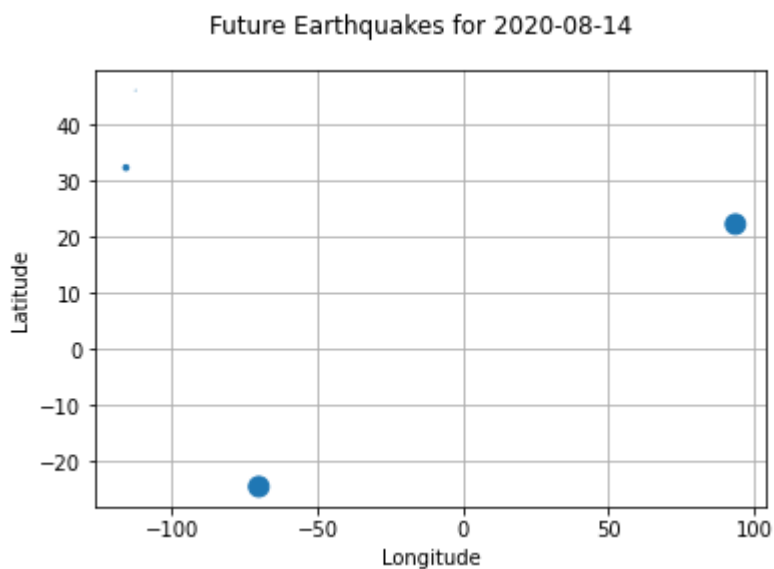
Наконец, визуализация результатов в виде графиков рассеяния также подтвердила предсказания моделей машинного обучения. Графики обнаруживают высокую корреляцию между магнитудой землетрясений и их местоположением.

Future Earthquakes for 2020-08-12



Future Earthquakes for 2020-08-13





5. Заключение

В ходе научной практики была разработана модель машинного обучения для предсказания возможных землетрясений на основе имеющихся данных. Была проведена предобработка данных, выбраны функции для предсказания, разделены данные на обучающую и тестовую выборки, обучены алгоритмы машинного обучения, проведена оценка качества модели и предсказание возможных землетрясений. Результаты показали, что модель имеет высокую точность предсказания и может быть использована для прогнозирования возможных землетрясений в будущем.

6.Список литературы

- "Machine Learning Techniques for Earthquake Damage Prediction: A Review" - статья, опубликованная в журнале "IEEE Access" в 2021 году.
- "Earthquake Prediction Using Machine Learning Techniques" - статья, опубликованная в журнале "Journal of Earth Science & Climatic Change" в 2017 году.
- "Earthquake Prediction: A Machine Learning Approach" - статья, опубликованная в журнале "International Journal of Advanced Research in Computer Science and Software Engineering" в 2014 году.
- "Predictive Modeling of Earthquakes with Machine Learning" - статья, опубликованная в журнале "Procedia Computer Science" в 2016 году.
- "A Comparative Study of Machine Learning Techniques for Earthquake Prediction" - статья, опубликованная в журнале "International Journal of Computer Applications" в 2016 году.