МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное

учреждение высшего образования

«Казанский национальный исследовательский технический

университет им. А.Н. Туполева – КАИ»

(КНИТУ – КАИ)

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

по дополнительной профессиональной программе профессиональной переподготовки

«Data science. Искусственный интеллект»

по теме: «Бинарная классификация для предиктивного

обслуживания оборудования»

|  |
| --- |
| **Работу выполнил:**  Сидоров Сидор Сидорович |
|  |
| **Руководитель выпускной квалификационной работы:**  к.т.н, доцент Смирнова Гульнара Сергеевна |
|  |
|  |

Казань 2024

Оглавление

[ВВЕДЕНИЕ 3](#_Toc188126877)

[ОСНОВНАЯ ЧАСТЬ 4](#_Toc188126878)

[Описание датасета 4](#_Toc188126879)

[Предобработка данных 4](#_Toc188126880)

[Разделение данных 4](#_Toc188126881)

[Обучение модели 4](#_Toc188126882)

[Оценка модели 5](#_Toc188126883)

[Streamlit-приложение 5](#_Toc188126884)

[ЗАКЛЮЧЕНИЕ 6](#_Toc188126885)

[СПИСОК ЛИТЕРАТУРЫ 7](#_Toc188126886)

[Приложение 8](#_Toc188126887)

**ВВЕДЕНИЕ**

Краткое описание задачи и цели проекта. Обоснование актуальности задачи (например, важность предиктивного обслуживания в промышленности).

Указать ссылку свой git -репозиторий с проектом

**ОСНОВНАЯ ЧАСТЬ**

Описание датасета

В данном разделе должно быть представлено:

Источник данных (ссылка на датасет).

Описание признаков и целевой переменной.

Примеры данных (можно вставить таблицу с несколькими строками

датасета).

Предобработка данных

В данном разделе должно быть представлено: Описание выполненных шагов и их обоснование:

Удаление ненужных столбцов (например, уникальные идентификаторы не несут полезной информации для модели).

Преобразование категориальных переменных (например, Type был преобразован в числовой формат, так как модели машинного обучения работают только с числами).

Проверка на пропущенные значения (пропусков не обнаружено, поэтому дополнительная обработка не потребовалась).

Масштабирование данных (например, числовые признаки были масштабированы для улучшения сходимости модели)

Разделение данных

В данном разделе должно быть представлено: Обоснование выбора соотношения 80/20 для обучающей и тестовой выборок (стандартное соотношение, которое обеспечивает достаточный объем данных для обучения и проверки модели).

Обучение модели

В данном разделе должно быть представлено: Описание выбранных моделей и их обоснование: Logistic Regression: Простая и интерпретируемая модель, подходящая для бинарной классификации. Random Forest: Устойчивая к переобучению модель, способная работать с нелинейными данными. XGBoost: Мощная модель, которая часто показывает высокую точность на сложных данных. 24/25 Краткое описание процесса обучения (например, модели были обучены на обучающей выборке с использованием стандартных параметров).

Оценка модели

В данном разделе должно быть представлено: Описание метрик, используемых для оценки (Accuracy, Confusion Matrix, ROCAUC). Результаты оценки для каждой модели (можно вставить таблицу или графики). Сравнение моделей и выбор наилучшей (например, Random Forest показал наилучшие результаты с Accuracy = 0.95 и ROC-AUC = 0.98).

Streamlit-приложение

В данном разделе должно быть представлено: Описание функционала приложения: Основная страница: загрузка данных, обучение модели, визуализация результатов, предсказания. Страница с презентацией: описание проекта. Скриншоты интерфейса приложения. Обоснование выбора Streamlit (простота использования, возможность быстрого создания интерактивных веб-приложений).

**ЗАКЛЮЧЕНИЕ**

Итоги работы: что удалось сделать, какие результаты получены. Возможные улучшения (например, использование других моделей, более глубокая предобработка данных).

**СПИСОК ЛИТЕРАТУРЫ**

1. Статистика : учебник / под ред. В.В. Глинского. — 5-е изд., перераб. и доп. — Москва : ИНФРА-М, 2024. — 372 с. — (Высшее образование). — DOI 10.12737/1981697. - ISBN 978-5-16-018343-5. - Текст : электронный. - URL: https://znanium.ru/catalog/product/1981697 (дата обращения: 28.04.2024)

2. Загорулько, Ю.А. Искусственный интеллект. Инженерия знаний [Электронный ресурс] : учебное пособие для вузов / Ю.А. Загорулько, Г.Б. Загорулько. Москва : Юрайт, 2022. 93 с. URL: https://www.urait.ru/bcode/494205 (дата обращения: 06.03.2023). Режим доступа: для авториз. пользователей.

3. Официальная страница набора данных на UCI Machine Learning Repository. <https://archive.ics.uci.edu/dataset/601/predictive+maintenance+dataset>

4. Streamlit Reveal Slides Репозиторий с примером использования Streamlit для создания презентаций. <https://github.com/bouzidanas/streamlit-reveal-slides>

5. Scikit-learn Documentation Официальная документация библиотеки Scikit-learn. <https://scikit-learn.org/stable/>

6. Streamlit Documentation Официальная документация библиотеки Streamlit. <https://docs.streamlit.io/> .

.

Приложение

Приложения: Ссылка на репозиторий с полным кодом проекта. Скриншоты интерфейса приложения. Графики и таблицы с результатами.