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:**  Родионов Алексей Петрович 5203 |
|  |
| **Руководитель выпускной квалификационной работы:**  к.т.н, доцент Смирнова Гульнара Сергеевна |
|  |
|  |

Казань 2025

Оглавление

[ВВЕДЕНИЕ 3](#_Toc188126877)

[ОСНОВНАЯ ЧАСТЬ 4](#_Toc188126878)

[Описание датасета 4](#_Toc188126879)

[Предобработка данных 5](#_Toc188126880)

[Разделение данных 6](#_Toc188126881)

[Обучение модели 6](#_Toc188126882)

[Оценка модели 6](#_Toc188126883)

[Streamlit-приложение 7](#_Toc188126884)

[ЗАКЛЮЧЕНИЕ 11](#_Toc188126885)

[СПИСОК ЛИТЕРАТУРЫ 12](#_Toc188126886)

[Приложение 8](#_Toc188126887)

**ВВЕДЕНИЕ**

Цель данного проекта ؙ— разработать модель машинного обучения, которая предсказывает, произойдёт ли отказ оборудования (Target = 1) или нет (Target = 0), Результаты работы оформить в виде многостраничного Streamlit-приложения с использованием *st.navigation* и *st.Page*.

Уметь предсказывать произойдёт ли отказ оборудования является важной задачей в промышленности. Благодаря этой возможности возможные неисправности обнаруживаются заранее, и предотвращаются потенциальные ошибки. Таким Образом, производственные объекты имеют возможность максимально сократить незапланированные простои.

Ссылка на git-репозиторий: [Репозиторий](https://github.com/Zet992/predictive_maintenance_project)

**ОСНОВНАЯ ЧАСТЬ**

Описание датасета

В данном проекте использовался датасет «AI4I 2020 Predictive Maintenance Dataset». доступен по ссылке [Predictive Maintenance Dataset](https://archive.ics.uci.edu/dataset/601/predictive+maintenance+dataset).

Переменные (Features):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Название переменной** | **Роль** | **Тип данных** | **Описание** | **Единицы измерения** | **Пропущенные значения** |
| UDI | ID | Integer | Уникальный идентификатор записи (от 1 до 10000). | - | Нет |
| Product ID | ID | Categorical | Идентификатор продукта (L, M, H) и серийный номер. | - | Нет |
| Type | Feature | Categorical | Тип продукта (L, M, H). | - | Нет |
| Air temperature [K] | Feature | Continuous | Температура окружающей среды. | Кельвины (К) | Нет |
| Process temperature [K] | Feature | Continuous | Рабочая температура процесса. | Кельвины (К) | Нет |
| Rotational speed [rpm] | Feature | Integer | Скорость вращения. | обороты/мин (rpm) | Нет |
| Torque | Feature | Continuous | Крутящий момент | Ньютон-метры (Nm) | Нет |
| Tool wear [min] | Feature | Integer | Износ инструмента. | минуты (min) | Нет |
| Machine failure | Target | Integer | Бинарная метка: 1 — отказ оборудования, 0 — отказ не произошел. | - | Нет |
| TWF | Target | Integer | Отказ из-за износа инструмента (Tool Wear Failure). | - | Нет |
| HDF | Target | Integer | Отказ из-за недостаточного теплоотвода (Heat Dissipation Failure). | - | Нет |
| PWF | Target | Integer | Отказ из-за недостаточной или избыточной мощности (Power Failure). | - | Нет |
| OSF | Target | Integer | Отказ из-за перегрузки (Overstrain Failure). | - | Нет |
| RNF | Target | Integer | Случайный отказ (Random Failure). | - | Нет |

Пример данных:



Предобработка данных

Из датасета были удалены уникальные идентификаторы, в частности стобицы UDI и Product ID.

Категориальная переменная Type была преобразована в числовой формат при помощи *LabelEncoder* из модуля *sklearn*.

Пропусков не обнаружено, поэтому дополнительная обработка не потребовалась.

Числовые признаки были масштабированы при помощи *StandardScaler* из модуля *sklearn* для улучшения сходимости данных.

Разделение данных

Для разделения данных используется соотношение 80/20 для обучающей и тестовой выборок, поскольку это стандартное соотношение, которое обеспечивает достаточный объем данных для обучения и проверки модели).

Обучение модели

Для проекта были выбраны модели: *Logistic Regression, Random Forest, XGBoost, Support Vector Classification*. *Logistic Regression* — прост интерпретируем, подходит для бинарной классификации. *Random Forest* — Устойчив к переобучению, способен работать с нелинейными данными. *XGBoost* — Мощная модель, которая часто показывает высокую точность на сложных данных. *Support Vector Classification* — хорошо подходит для данных с большим количеством признаков. Модели были обучены на обучающей выборке с использованием стандартных параметров

Оценка модели

Для оценки модели использовались метрики: *Accuracy, Confusion Matrix, ROCAUC*. Результаты оценки: Рис 1. и Табл 1. в приложении.

Модель *XGBClassifier* показала себя лучше всех по метрикам *F1-Score* и *AUC*. Поэтому для проекта выбрана она.

Streamlit-приложение

Streamlit был выбран благодаря своей простоте использования и возможности быстрого создания интерактивных веб-приложений.

Streamlit-приложение поделено на две страницы:

1. Основная страница (Анализ и модель): загрузка данных, обучение модели, визуализация результатов, предсказания. Скриншоты на Рис. 2.1-2.3
2. Страница с презентацией (Презентация): описание проекта. Скриншоты на Рис 3.1-3.2

**ЗАКЛЮЧЕНИЕ**

В результате этой работы удалось сделать модель с удовлетворительной точностью предсказания, выявляющую в среднем 64% неисправностей. В дальнейшем для улучшения предсказывания можно выбрать более мощную модель, научить модель предсказывать тип отказов или же оптимизировать обучение модели.

**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

Ссылка на Git-репозиторий: [Репозиторий](https://github.com/Zet992/predictive_maintenance_project)

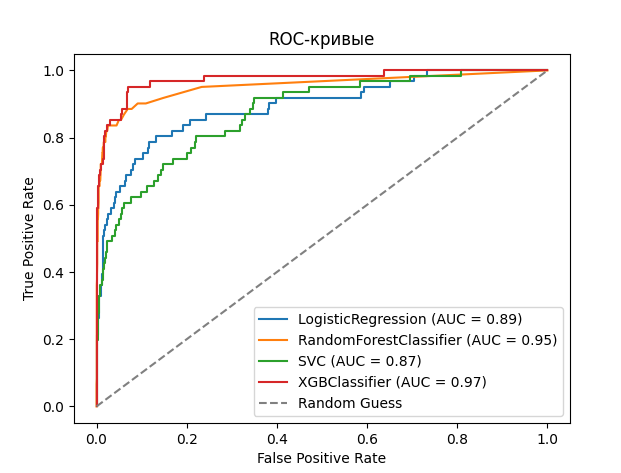


Рис 1. Графики ROC различных моделей.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F1-Score | AUC |
| LogisticRegression | 0,97 | 0,65 | 0,26 | 0,38 | 0,89 |
| RandomForestClassifer | 0,98 | 0,84 | 0,59 | 0,69 | 0,95 |
| SVC | 0,97 | 0 | 0 | 0 | 0,87 |
| XGBClassifier | 0,98 | 0,83 | 0,64 | 0,72 | 0,97 |

Табл 1. Показатели метрик различных моделей

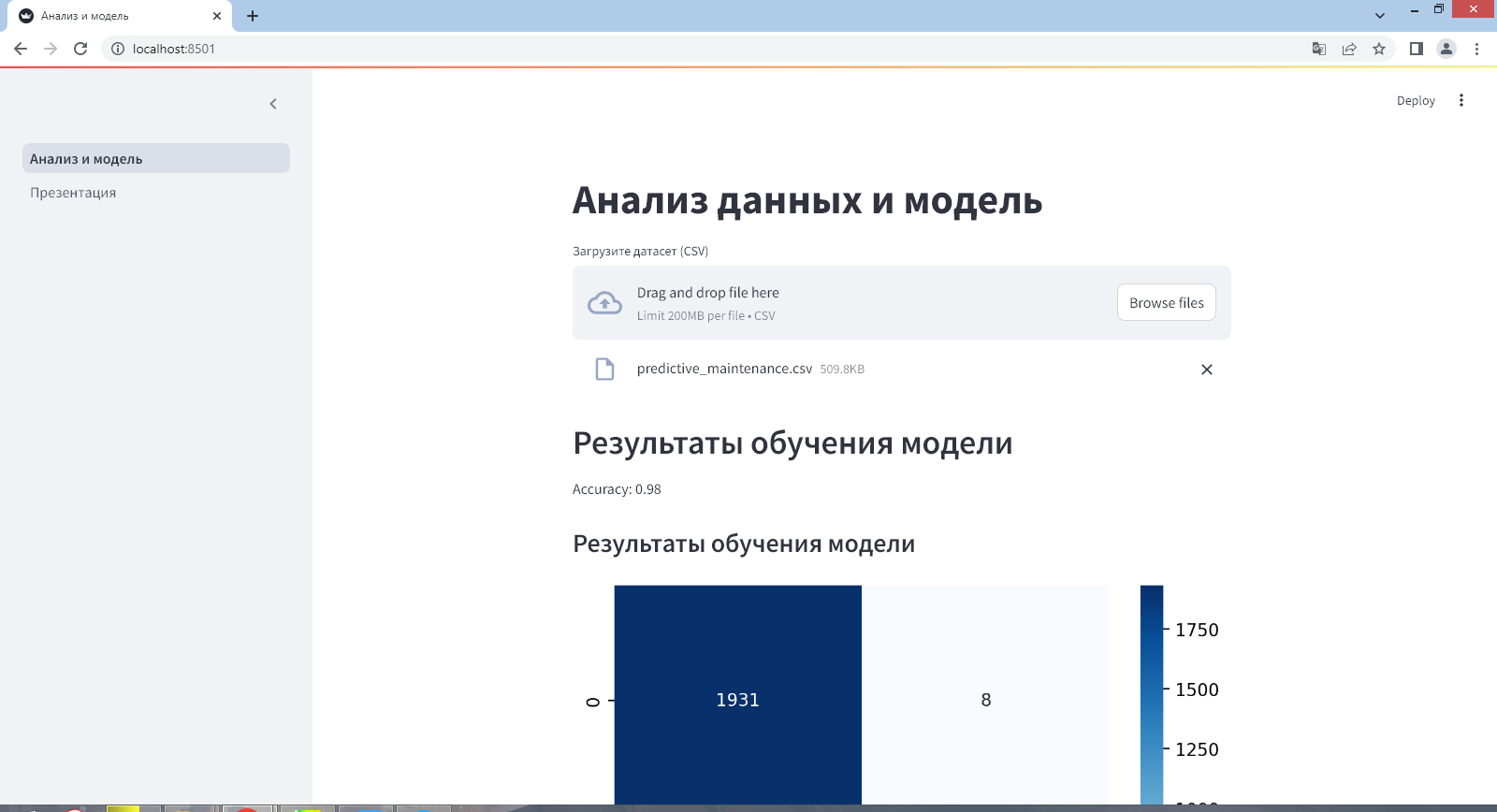


Рис 2.1. Скриншот результатов обучения модели ч.1



Рис 2.2. Скриншот результатов обучения модели ч.2

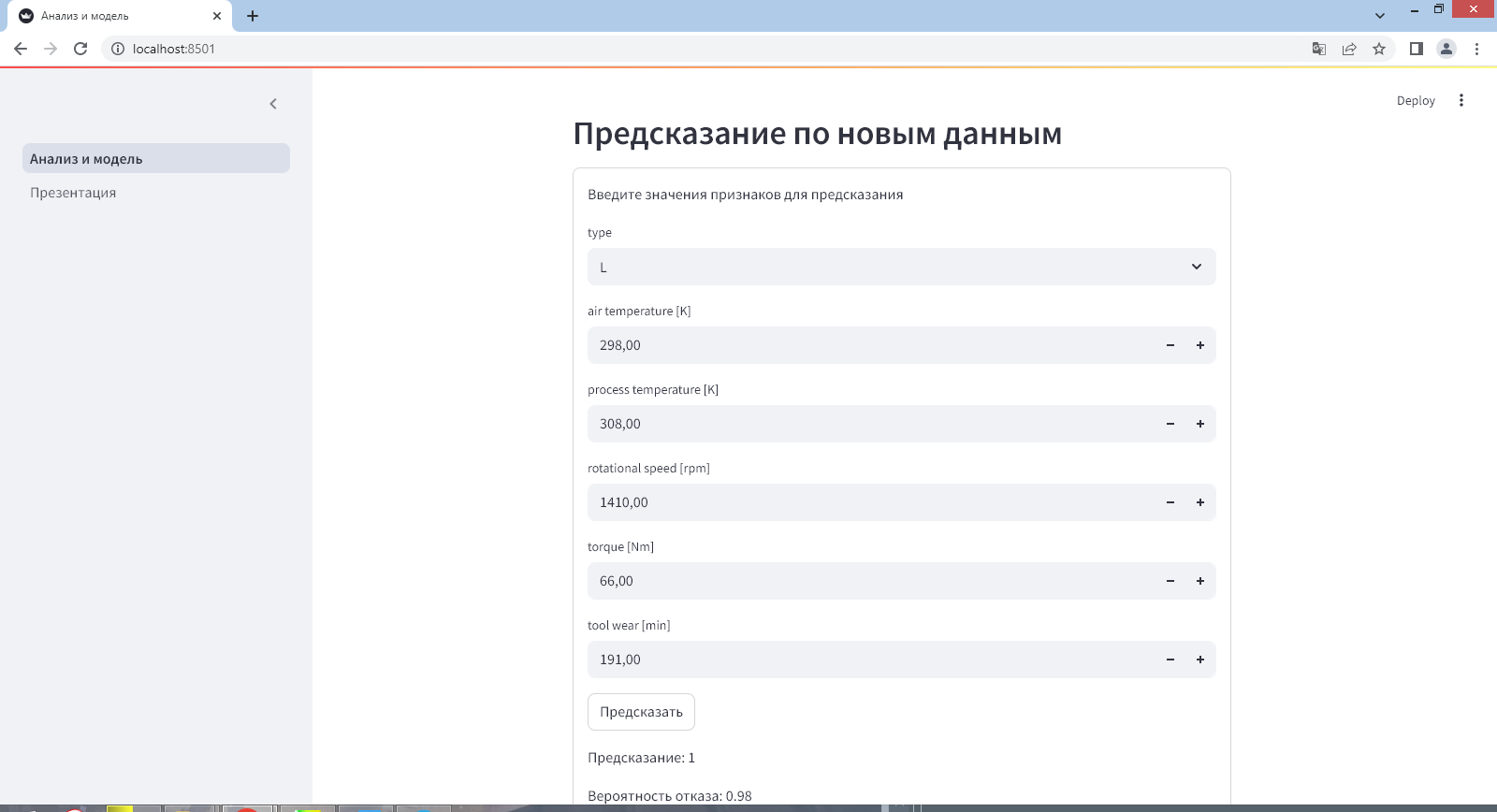


Рис 2.3. Скриншот формы предсказания по новым данным

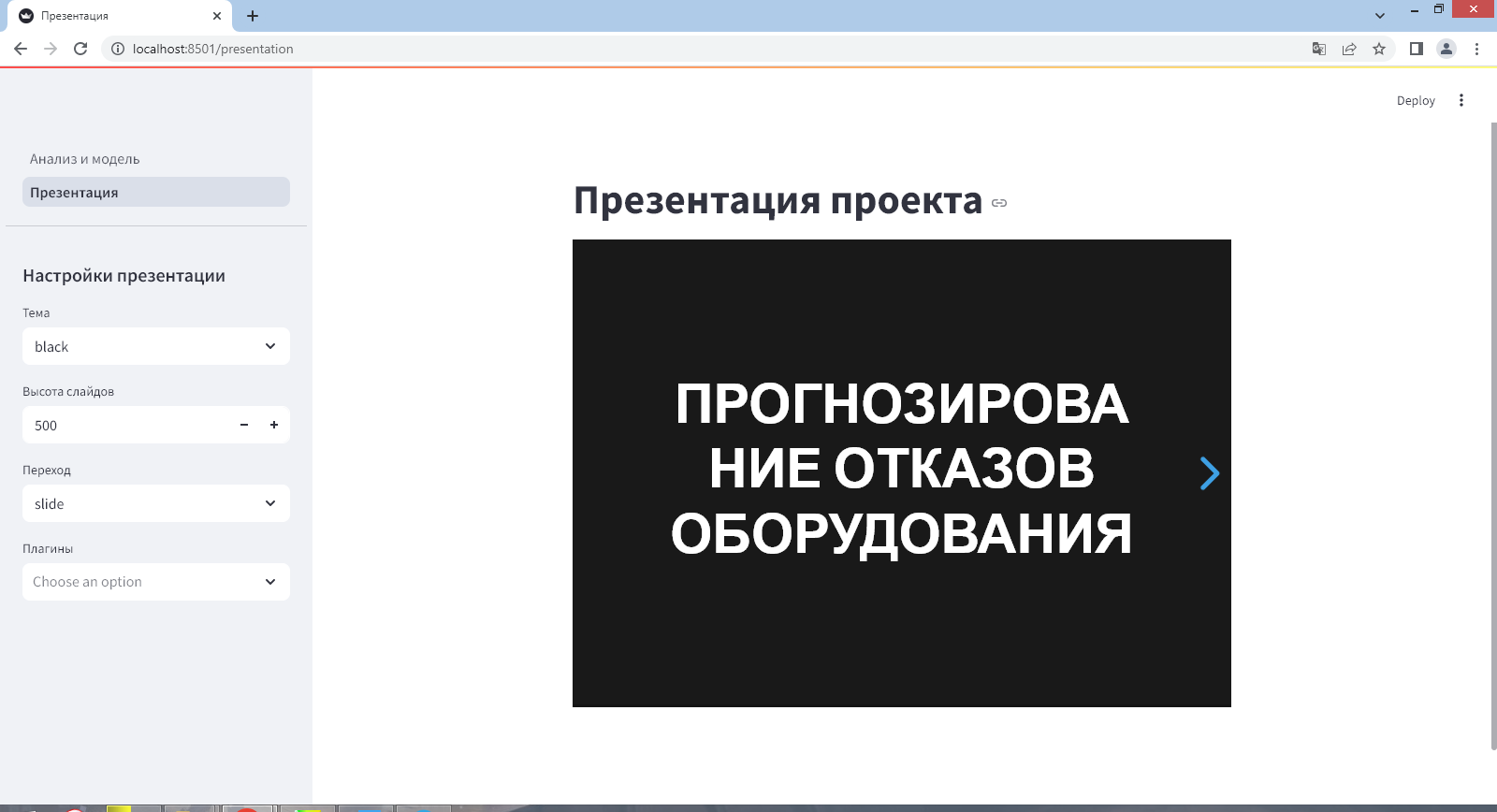


Рис 3.1. Скриншот первого слайда презентации

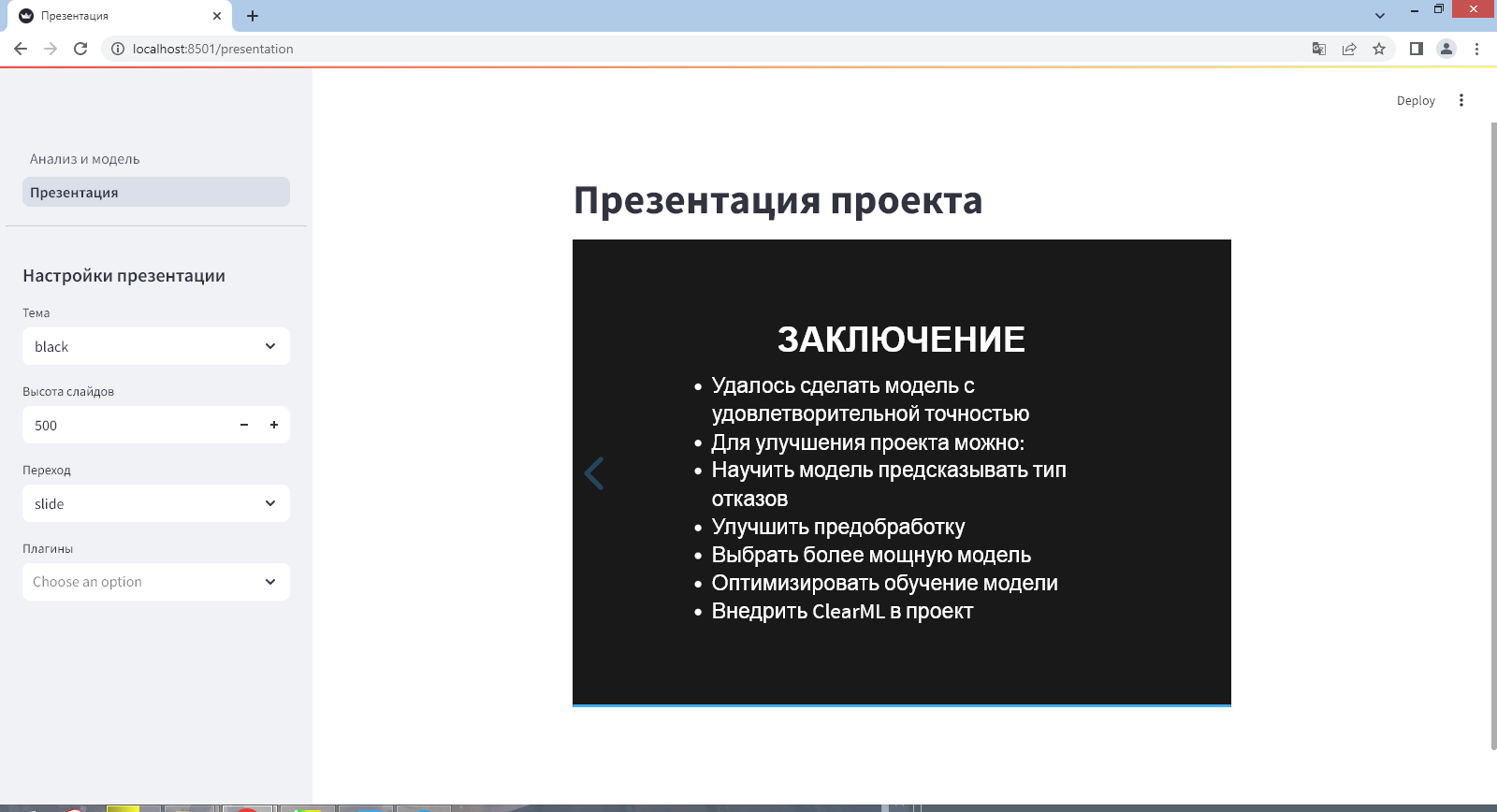


Рис 3.2. Скриншот последнего слайда презентации