ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ

**Программа для обнаружения спам-сообщений**

Руководство пользователя к заданию по дисциплине «Проектный семинар «Python в науке о данных» студентов образовательной программы «Информатика и вычислительная техника» учебной группы БИВ248:

* Погоняев Сергей Сергеевич, [sspogoniaev@edu.hse.ru](mailto:sspogoniaev@edu.hse.ru), 89234102924
* Шишелякин Матвей Романович, [mrshisheliakin@edu.hse.ru](mailto:mrshisheliakin@edu.hse.ru), 89142355181

Москва 2025

# **Оглавление**

[**Оглавление** 2](#_Toc200280632)

[**Общая информация. Функционал приложения** 3](#_Toc200280633)

[**Установка и запуск приложения** 4](#_Toc200280634)

[**Установка приложения** 4](#_Toc200280635)

[**Запуск приложения** 4](#_Toc200280636)

[**Интерфейс приложения** 5](#_Toc200280637)

[**Главная страница.** 5](#_Toc200280638)

[**Добавление новой записи.** 6](#_Toc200280639)

[**Скачивание отчетов.** 6](#_Toc200280640)

[**Настройки.** 7](#_Toc200280641)

[**Описание отчетов** 8](#_Toc200280642)

[**Графический отчет.** 8](#_Toc200280643)

[**Отчет по модели.** 8](#_Toc200280644)

[**Текстовый отчет.** 9](#_Toc200280645)

# **Общая информация. Функционал приложения**

Приложение разработано с целью определения спам-сообщений среди SMS-сообщений. Предсказание строится на основе модели, обученной на начальном датасете. Помимо непосредственно категоризации сообщения в приложении предусмотрен следующий функционал:

* Возможность добавления новых записей в датасет с пометкой (spam/ham – спам/не спам);
* Возможность переобучить модель (применительно в случае добавления новых записей в датасет);
* Возможность скачать отчеты следующих типов: отчет по обучению модели, текстовый отчет о текущем датасете, графический отчет о текущем датасете;
* Возможность настройки пользовательского интерфейса.

# **Установка и запуск приложения**

## **Установка приложения**

Для установки приложения требуется скачать архив «python\_project.zip», распаковать его на своем компьютере. Также убедитесь, что на вашем компьютере установлен дистрибутив Anaconda, а в переменной Path переменных сред на вашем компьютере добавлены «conda/Scripts» и «conda/Library/bin».

## **Запуск приложения**

Для запуска приложения вам нужно двойным кликом запустить файл runner.bat. Приложение работает на локальном сервере (порт 8000), поэтому после создания и активации среды в консоли появится ссылка на главную страницу:



Для начала работы с приложением просто перейдите по этой ссылке.

# **Интерфейс приложения**

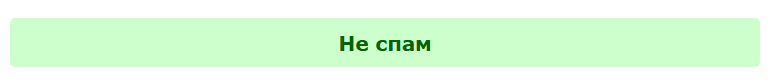
## **Главная страница.**

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Содержимое, созданное искусственным интеллектом, может быть неверным.

Здесь пользователю предлагается проверить его сообщение на спам. Для этого сообщение вводится в соответствующее поле, после чего нужно нажать кнопку «Проверить» (!!!ВАЖНО: для наиболее корректной работы приложения рекомендуется вводить сообщения именно на английском языке, поскольку модель была обучена на датасете с сообщениями из Великобритании). Далее внизу появляется поле с вердиктом:

1. Не спам:



1. Спам:



Также на странице представлены кнопки:

* «Добавить запись» - для перехода на страницу добавления записи в датасет;
* «Скачать отчеты» - для перехода на страницу скачивания отчетов;
* «Настройки» - для перехода на страницу кастомизации интерфейса.

## **Добавление новой записи.**

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Содержимое, созданное искусственным интеллектом, может быть неверным.

В соответствующее поле вводится сообщение, после чего выбирается маркер (спам/не спам). Поле ввода и маркер являются обязательными для заполнения полями.

Если вы хотите переобучить модель с учетом добавленных записей, то нажмите кнопку «Переобучить модель». При этом автоматически обновятся все отчеты.

Вернуться на главную страницу – кнопка «На главную».

## **Скачивание отчетов.**

Изображение выглядит как текст, снимок экрана, Шрифт, Цвет электрик

Содержимое, созданное искусственным интеллектом, может быть неверным.

Существует 3 вида отчетов:

1. Графический отчет (столбчатая диаграмма, отображающая общую информацию о группах в датасете, их количестве и соотношении);
2. Отчет по модели (текстовый отчет по обучению модели, отображающий матрицу ошибок, классификационных отчет, общую точность модели и необработанные моделью сообщения)
3. Текстовый отчет (текстовый отчет о датасете, содержащий общую информацию о сообщениях, а также информацию по каждой группе в отдельности).

Выберите желаемые для скачивания отчеты и нажмите кнопку «Скачать». Если был выбран 1 отчет, то он скачается, как одиночный файл, иначе скачается архив «reports.zip», в котором содержатся выбранные отчеты.

Вернуться на главную страницу – кнопка «На главную».

## **Настройки.**

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Содержимое, созданное искусственным интеллектом, может быть неверным.

Для настройки пользовательского интерфейса доступны следующие изменения:

* Цвет фона (можно задать любой цвет);
* Шрифт (Arial, Verdana, Courier New, Georgia);
* Размер текста (от 12px до 24px);
* Стиль кнопок (Плоские, Скругленные).

Для сохранения изменений нажмите на кнопку «Сохранить изменения»

# **Описание отчетов**

## **Графический отчет.**

Показывает соотношение SMS-сообщений из групп «Спам» и «Не спам» в столбчатом виде:

Изображение выглядит как текст, снимок экрана, диаграмма, Прямоугольник

Содержимое, созданное искусственным интеллектом, может быть неверным.

## **Отчет по модели.**

Текстовый отчет по модели содержит следующую информацию о результатах обучения модели на датасете:

1. Матрица ошибок в формате

|  |  |  |
| --- | --- | --- |
|  | Не спам (предсказано) | Спам (предсказано) |
| Не спам (маркер) | Правильное предсказание | Неправильное предсказания |
| Спам (маркер) | Правильное предсказание | Неправильное предсказание |

1. Классификационный отчет;
2. Значение общей точности модели;
3. Список необработанных сообщений (сообщения, которые помечены как «Спам», но модель определила их как «Не спам»).

## **Текстовый отчет.**

Содержит следующую информацию:

1. Общая информация:
   1. Общее количество сообщений;
   2. Средняя длина сообщения;
   3. Среднее количество слов в сообщении;
   4. Процентное соотношение спам-сообщений к обычным сообщениям (spam/ham).
2. Информация по каждой группе:
   1. Количество сообщений из группы;
   2. Средняя длина сообщения из группы;
   3. Среднее количество слов в сообщении из группы.

Изображение выглядит как текст, снимок экрана, Шрифт

Содержимое, созданное искусственным интеллектом, может быть неверным.