НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ “МИФИ” (НИЯУ МИФИ)

Инженерно-физический институт биомедицины Кафедра компьютерных медицинских систем

ОТЧЁТ

По производственной практике (научно-исследовательской работе)

«Веб-приложение для диагностики диабета»

Выполнил: студент группы М24-302

Мухин Даниил Дмитриевич

Научный руководитель:

Отченашенко Александр Иванович

Москва 2025

Содержание

[1. Введение 2](#_Toc201532260)

[1.1 Актуальность исследования 2](#_Toc201532261)

[1.2 Анализ текущего состояния проблемы 3](#_Toc201532262)

[1.3. Цель работы 4](#_Toc201532263)

[2. Анализ предметной области 4](#_Toc201532264)

[3. Реализация проекта 6](#_Toc201532265)

[3.1 Набор данных по диабету 6](#_Toc201532266)

[3.2 Обучение модели машинного обучения и сохранение его в виде файла pickle 8](#_Toc201532267)

[3.3 Веб-приложение Streamlit 9](#_Toc201532268)

[3.4 Развертывание приложения 10](#_Toc201532269)

[Заключение 12](#_Toc201532270)

[Список литературы 13](#_Toc201532271)

1. Введение
   1. Актуальность исследования

Диабет — одно из самых распространённых хронических заболеваний в мире. Согласно данным Всемирной организации здравоохранения (ВОЗ), количество больных диабетом стремительно растёт и уже превысило 500 миллионов человек по состоянию на 2023 год. Особую угрозу представляет диабет 2 типа, часто развивающийся бессимптомно, что приводит к поздней диагностике и увеличению риска тяжёлых осложнений — сердечно-сосудистых заболеваний, почечной недостаточности, слепоты и др.

Одной из основных проблем остаётся несвоевременное выявление заболевания. Многие пациенты обращаются за медицинской помощью на поздних стадиях, когда лечение становится менее эффективным. В связи с этим возрастает потребность в доступных, быстрых и точных инструментах предварительной диагностики, которые позволят людям самостоятельно оценить риски развития диабета и своевременно обратиться к специалисту.

Разработка веб-приложения для диагностики диабета отвечает этим вызовам. Такие приложения обладают рядом преимуществ:

* **Доступность** — пользователь может пройти тест из любой точки мира при наличии интернета.
* **Автоматизация** — приложение позволяет быстро обрабатывать входные данные и предоставлять результаты без участия врача.
* **Персонализация** — возможна адаптация рекомендаций в зависимости от возраста, пола, индекса массы тела и других факторов.
* **Образовательный эффект** — пользователь получает информацию о факторах риска и мерах профилактики.
  1. Анализ текущего состояния проблемы

На сегодняшний день проблема ранней диагностики диабета остаётся актуальной во всём мире. Несмотря на развитие медицинских технологий, многие пациенты сталкиваются с поздним выявлением заболевания. По данным Международной диабетической федерации (IDF), значительная часть случаев диабета 2 типа диагностируется уже на стадии осложнений, что обусловлено отсутствием регулярных обследований, низкой осведомлённостью населения и ограниченным доступом к квалифицированной медицинской помощи, особенно в отдалённых регионах.

Современная практика диагностики диабета основывается на лабораторных методах: анализе уровня глюкозы в крови, гликированного гемоглобина (HbA1c), тестах на толерантность к глюкозе. Однако такие методы требуют посещения медицинского учреждения, чего многие избегают по разным причинам — нехватка времени, страха, недоверия к врачам или недоступность лабораторий.

С развитием цифровых технологий появились инструменты, которые частично решают эти проблемы. В настоящее время существует ряд мобильных и веб-приложений, предлагающих пользователям пройти самодиагностику на основе опросников или искусственного интеллекта.

1.3. Цель работы

Цель данной работы состоит в создании веб-приложения с помощью Streamlit-фреймворка для языка программирования Python для прогнозирования диабета с использованием модели машинного обучения.

1. Анализ предметной области

Диабет — это хроническое метаболическое заболевание, характеризующееся устойчиво повышенным уровнем глюкозы в крови. Существует два основных типа заболевания:

* **Сахарный диабет 1 типа** — аутоиммунное заболевание, при котором организм прекращает выработку инсулина. Диагностируется преимущественно в детском или подростковом возрасте.
* **Сахарный диабет 2 типа** — заболевание, при котором снижается чувствительность клеток к инсулину (инсулинорезистентность) при его сохранённой или повышенной выработке. Наиболее распространён, составляет до 90% всех случаев диабета.

Ключевыми факторами риска развития диабета 2 типа являются:

* избыточная масса тела и ожирение,
* малоподвижный образ жизни,
* наследственная предрасположенность,
* повышенное артериальное давление,
* возраст старше 40 лет,
* нарушение толерантности к глюкозе и предиабет.

Для предварительной оценки риска развития диабета применяются различные шкалы и опросники, например:

* **Finnish Diabetes Risk Score (FINDRISC)** — валидированный опросник, учитывающий возраст, ИМТ, физическую активность, питание, наличие гипертензии и родственных случаев диабета.

Эти шкалы могут быть легко адаптированы для использования в цифровых приложениях и являются хорошей основой для автоматической оценки риска.

Современные технологии позволяют создавать веб-приложения с использованием следующих инструментов и подходов:

* **Фронтенд:** HTML, CSS, JavaScript, React, Vue.js — для создания пользовательского интерфейса.
* **Бэкенд:** Python (Django, Flask), Node.js — для обработки логики приложения и хранения данных.
* **Базы данных:** PostgreSQL, MySQL, MongoDB — для хранения пользовательской информации.
* **Интеграция с алгоритмами машинного обучения** (например, логистическая регрессия, деревья решений, нейросети) — для улучшения точности диагностики на основе накопленных данных.
* **Обеспечение безопасности и конфиденциальности** — важный аспект при работе с персональными медицинскими данными (соответствие требованиям GDPR, HIPAA, российского ФЗ-152 «О персональных данных» и др.).

Также важно учитывать принципы UX/UI-дизайна для создания интуитивно понятного интерфейса, доступного для пользователей разных возрастов и уровня компьютерной грамотности.

Таким образом, предметная область объединяет знания из медицины, программирования, анализа данных и проектирования пользовательского опыта. Учитывая мультидисциплинарный характер задачи, успешная разработка веб-приложения требует тесной интеграции клинических подходов с современными веб-технологиями.

1. Реализация проекта
   1. Набор данных по диабету

Набор данных изначально был получен из Национального института диабета, заболеваний пищеварительной системы и почек. Целью является прогнозирование на основе диагностических измерений наличия у пациента диабета.

На выборку этих случаев из более обширной базы данных накладывалось несколько ограничений. В частности, все пациенты здесь — женщины в возрасте не менее 21 года индейского происхождения Пима.

Показатели:

* Беременности: Количество беременностей
* Глюкоза: концентрация глюкозы в плазме через 2 часа в оральном тесте на толерантность к глюкозе
* Кровяное давление: диастолическое артериальное давление (мм рт. ст.)
* SkinThickness: Толщина кожной складки трицепса (мм)
* Инсулин: 2-часовой сывороточный инсулин (мю Ед/мл)
* ИМТ: Индекс массы тела (вес в кг/(рост в м)^2)
* DiabetesPedigreeFunction: Функция родословной диабета
* Возраст: Возраст (лет)
* Результат: переменная класса (0 или 1)

Часть датасета представлена на рисунке 1.

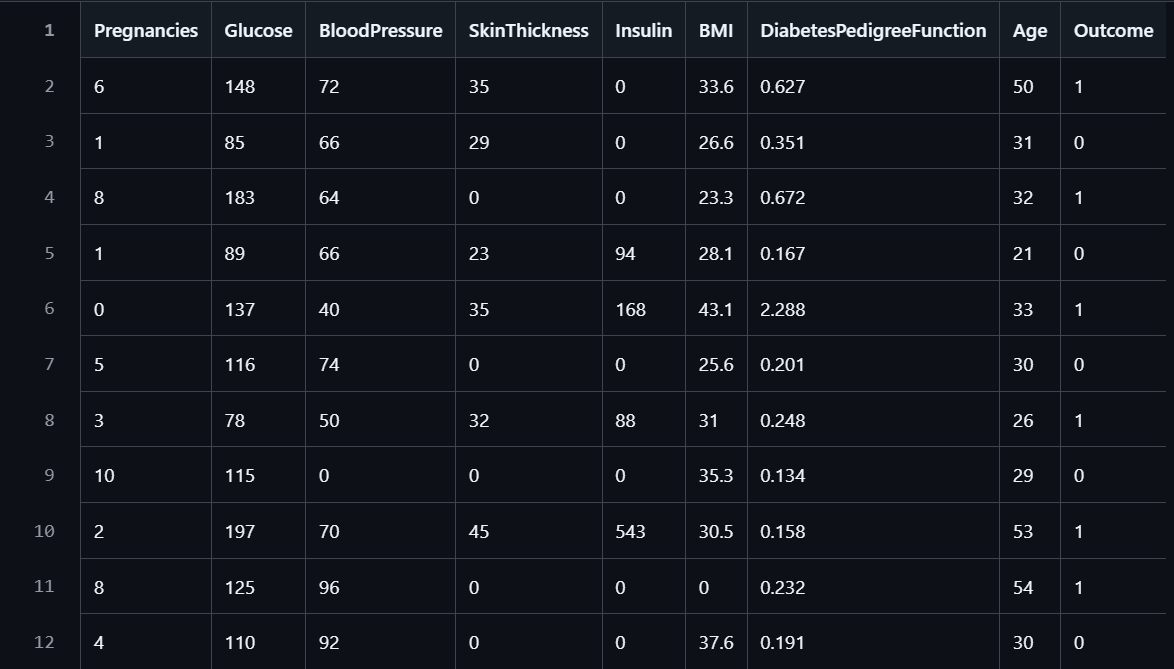


Рисунок 1-Набор данных по диабету

* 1. Обучение модели машинного обучения и сохранение его в виде файла pickle

В качестве метода машинного обучения используется LDA(Linear Discriminant Analysis)- это алгоритм классификации, который помогает отличить объекты разных классов, найти, чем один класс отличается от другого, предсказать, к какому классу относится новый объект.

LDA ищет направления, где классы( например больной/здоровый) отличаются друг от друга. В контексте диабета он получает данные пациента(глюкоза, возраст, давление и.т.д.) и учится распознавать, болен человек диабетом или нет. Затем можно подать новые данные- и LDA покажет, в какую группу попадает человек.

На рисунке 2 показано обучение модели LDA и инициализация.

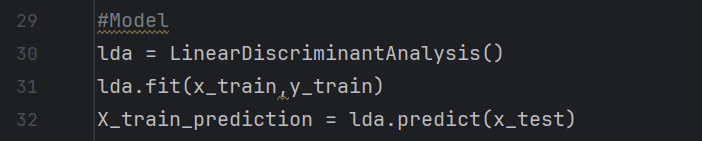


Рисунок 2-Инициализация и обучение модели LDA

**Также в процессе работы  периодически возникает необходимость выгрузить обрабатываемую в программе информацию, для её хранения или дальнейшей передачи. При этом важно не утратить информацию или её часть, и быть уверенным, что тот, кому файл будет передаваться, сможет с ним работать. И тогда на помощь приходит сериализация данных.**

**Сериализация данных-это преобразование данных, обрабатываемых в программе в форматы, которые можно хранить и передавать.**

**В данном проекте используется стандартная библиотека pickle. Библиотека работает с двоичными потоками данных, как в файл, так и по сети. Pickle позволяет сериализовать большое количество разнообразных объектов, используемых в Python. Сохранение модели в бинарные файлы представлено на рисунке 3.**

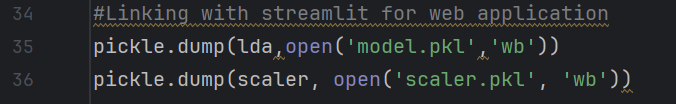
****

Рисунок 3-Сохранение модели

* 1. Веб-приложение Streamlit

Streamlit- это библиотека для создания веб-приложении на Python. Она помогает перенести модель машинного обучения в веб и добавить интерфейс, кнопки, визуализацию. Для запуска приложения в терминале требуется прописать команду <streamlit run web.py>. В скрипте, содержащем код для веб-приложения, описываются функции масштабирования пользовательского ввода и предсказания наличия диабета, а также интерфейс и пользовательский ввод. На рисунке 4 представлено описание кнопки для предсказания диабета.

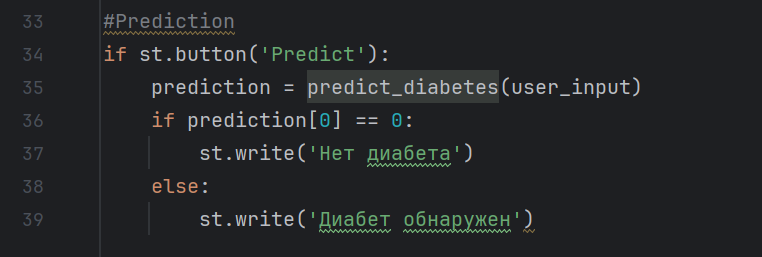


Рисунок 4-Кнопка для предсказания

Если пользователь нажимает кнопку «Predict», то вызывается функция predict\_diabetes. Если результат 0, появляется сообщение «Нет диабета»

* 1. Развертывание приложения

Скриншот приложения показан на рисунке 5

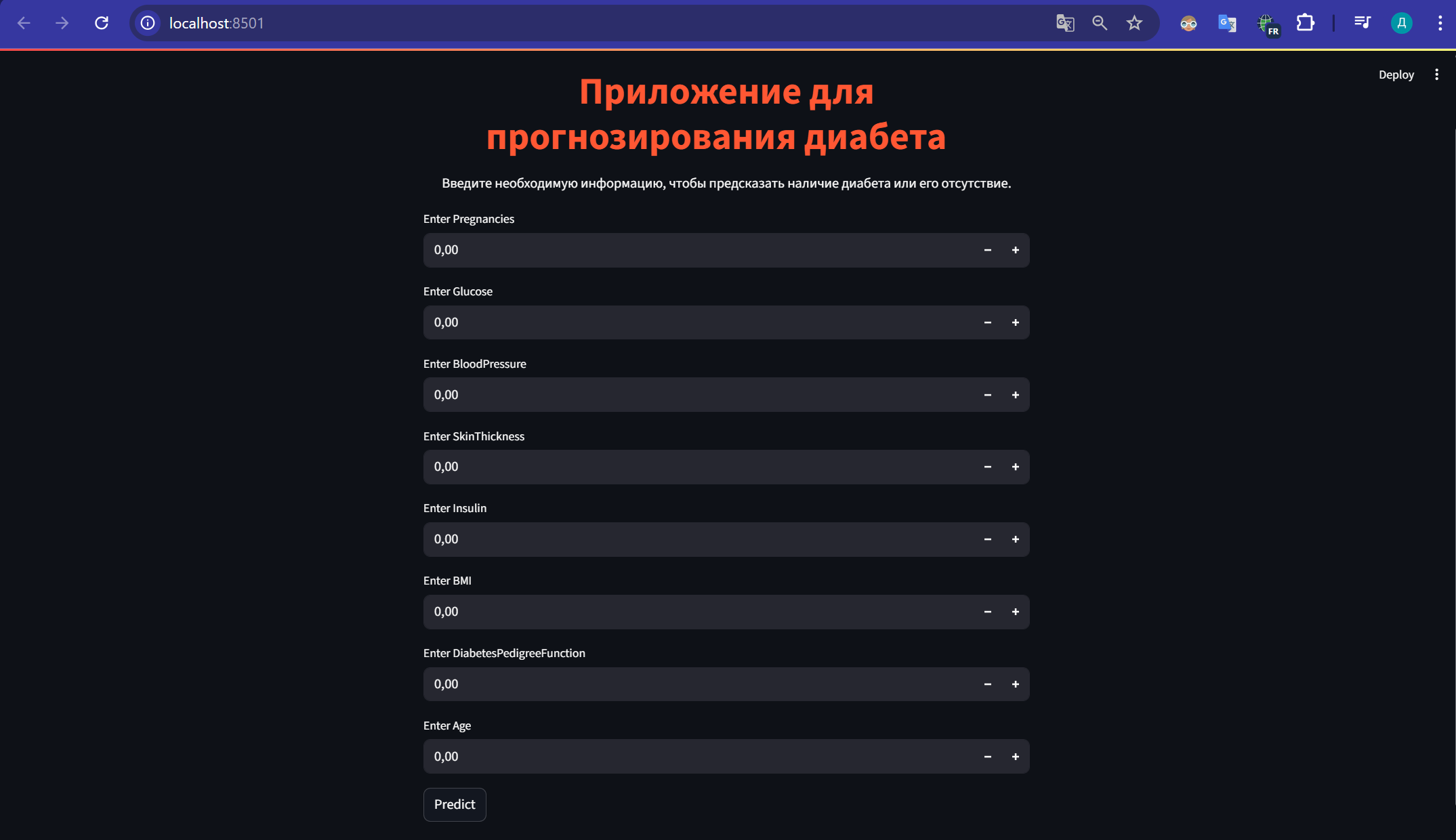


Рисунок 5-приложение для прогнозирования диабета

Для использования вводится необходимая информация, такая как уровень инсулина, возраст, глюкоза и.т.д. в предоставленную форму и нажимается кнопка «predict» для получения прогноза диабета.

Гитхаб проекта представлен по ссылке <https://github.com/villareal593/Diabetes-Prediction-App.git>

Также возможны будушие улучшения: интеграция дополнительных показателей здоровья для улучшения прогноза, расширение выбора моделей с помощью более продвинутых алгоритмов, улучшение пользовательского интерфейса.

# Заключение

Разработка веб-приложения для диагностики диабета представляет собой важный шаг в сторону доступной и быстрой предварительной медицинской оценки. С помощью машинного обучения и таких инструментов, как Streamlit, удалось создать простую в использовании систему, которая на основе медицинских показателей (уровень глюкозы, давление, возраст и др.) позволяет предсказать вероятность наличия диабета.

Приложение не заменяет профессиональную медицинскую диагностику, но может быть полезным инструментом для **раннего выявления риска заболевания** и **повышения осведомлённости** пользователей о состоянии своего здоровья. Благодаря использованию модели Linear Discriminant Analysis и предварительной стандартизации данных, достигается высокая точность предсказаний, а пользовательский интерфейс делает приложение доступным даже для людей без технической подготовки.

В будущем такое приложение может быть дополнено:

* визуализацией результатов и рисков,
* хранением истории пользователя,
* интеграцией с медицинскими базами данных.

Таким образом, веб-приложение является примером практического применения искусственного интеллекта в области здравоохранения, способствуя улучшению качества жизни и профилактике хронических заболеваний.

# Список литературы

**1. Сравниваем модуль pickle и альтернативные способы сериализации // Хабр URL: https://habr.com/ru/articles/766904/ (дата обращения: 23.06.2025).**

**2.** Линейный дискриминантный анализ (LDA). Принцип работы и реализация с нуля на Python // Хабр URL: https://habr.com/ru/articles/802511/ (дата обращения: 23.06.2025).

3. Streamlit // Skillfactory URL: https://blog.skillfactory.ru/glossary/streamlit/ (дата обращения: 23.06.2025).

4. P. T. Siva Gurunathan; Raghuraj S; Roahit S; Nithishkumar S Web Application-based Diabetes Prediction using Machine Learning // 7th International Conference on Computing Methodologies and Communication (ICCMC). – 2023

5. Mujumdar, A., Vaidehi, V.: Diabetes prediction using machine learning algorithms. Procedia Comput. Sci. **165**, 292–299 (2019)