|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ \_\_\_\_\_ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

КАФЕДРА \_\_\_\_\_СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_Разработка и оценка моделей \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ машинного обучения\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

Студент \_\_ИУ5-6Б\_\_\_\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_Д.Д. Никулин\_\_\_\_\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** **\_\_\_\_\_****Ю.Е. Гапанюк\_\_\_\_\_**

(Подпись, дата) (И.О.Фамилия)

*2024 г.***Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ

Заведующий кафедрой \_\_\_ИУ5\_\_\_\_

(Индекс)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_В.И. Терехов\_\_

(И.О.Фамилия)

«\_\_\_» \_\_\_\_февраля\_\_\_\_\_\_\_ 2024 г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме \_\_\_\_Разработка и оценка моделей методов машинного обучения\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент группы \_ИУ5-61Б\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Никулин Данила Дмитриевич\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ИССЛЕДОВАТЕЛЬСКАЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_КАФЕДРА\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

График выполнения НИР: 25% к \_\_\_\_ нед., 50% к \_\_\_\_ нед., 75% к \_\_\_ нед., 100% к \_\_\_\_ нед.

***Техническое задание*** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на \_\_\_\_\_ листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата выдачи задания «\_\_\_» \_\_\_\_февраля\_\_\_\_\_\_\_ 2024 г.

**Руководитель НИР**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_Ю.Е. Гапанюк\_\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

**Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Оглавление

[Введение 4](#_Toc168493767)

[Основная часть 5](#_Toc168493768)

[Заключение 14](#_Toc168493769)

[Список использованных источников информации 15](#_Toc168493770)

## Введение

В современном мире сердечно-сосудистые заболевания становятся все более значимой проблемой общественного здравоохранения, и понимание факторов риска, влияющих на их возникновение, является ключевым для разработки эффективных стратегий профилактики и лечения. В рамках представленной работы мной было проведено исследование, целью которого было анализировать различные факторы риска, влияющие на вероятность развития сердечно-сосудистых заболеваний.

Данное исследование сосредоточилось на анализе набора данных, включающего информацию о различных клинических показателях, физиологических параметрах и образе жизни участников. Целью работы было разработать и оценить модели машинного обучения для прогнозирования риска сердечно-сосудистых заболеваний на основе представленных в исследовании параметров.

Модели машинного обучения представляют собой мощный инструмент для анализа сложных взаимосвязей между различными признаками и их влиянием на вероятность развития сердечно-сосудистых заболеваний. Представленное исследование включает в себя подробный анализ данных, включая предварительное исследовательское изучение, выбор и кодирование признаков, корреляционный анализ, а также подбор гиперпараметров для моделей машинного обучения.

Результаты моей работы могут послужить основой для дальнейших исследований в области предсказания риска сердечно-сосудистых заболеваний, а также для разработки эффективных методов фракционирования риска и предупреждения данной группы заболеваний. Таким образом, данная работа призвана внести свой вклад в современную медицинскую науку и помочь в борьбе с серьезной медицинской проблемой.

## Основная часть

#### **Постановка задачи**

Целью данной работы является разработка и оценка моделей машинного обучения для прогнозирования риска сердечно-сосудистых заболеваний на основе набора данных, предоставленного Kaggle. В качестве целевой переменной для задачи классификации выбрана переменная о более высокой/низкой вероятности сердечного приступа.

Датасет состоит из файла heart.csv со следующими колонками:

* age: Возраст пациента
* sex: Пол пациента
* cp: Тип боли в груди, 0 = типичная стенокардия, 1 = атипичная стенокардия, 2 = неангинальная боль, 3 = бессимптомная
* trtbps: Артериальное давление в состоянии покоя (в мм рт. ст.)
* chol: Холесторал в мг/дл, полученный с помощью датчика BMI
* fbs: (уровень сахара в крови натощак > 120 мг/дл), 1 = верно, 0 = неверно
* restecg: Результаты электрокардиографии в покое, 0 = норма, 1 = нормальный зубец ST-T, 2 = гипертрофия левого желудочка.
* thalachh: Максимально достигнутая частота сердечных сокращений
* oldpeak: Предыдущий пик
* slp: Склон
* caa: Количество крупных сосудов (0-3)
* thall: Результат стресс-теста с талием ~ (0,3)
* exng: Стенокардия, вызванная физической нагрузкой ~ 1 = Да, 0 = Нет
* output: Целевая переменная, 0 = меньше шансов на сердечный приступ, 1= больше шансов на сердечный приступ

1. **Последовательность действий** 
   1. **Проведение разведочного анализа данных. Построение графиков, необходимых для понимания структуры данных. Анализ и заполнение пропусков в данных**

Список колонок с типами данных:

0 age 303 non-null int64

1 sex 303 non-null int64

2 cp 303 non-null int64

3 trtbps 303 non-null int64

4 chol 303 non-null int64

5 fbs 303 non-null int64

6 restecg 303 non-null int64

7 thalachh 303 non-null int64

8 exng 303 non-null int64

9 oldpeak 303 non-null float64

10 slp 303 non-null int64

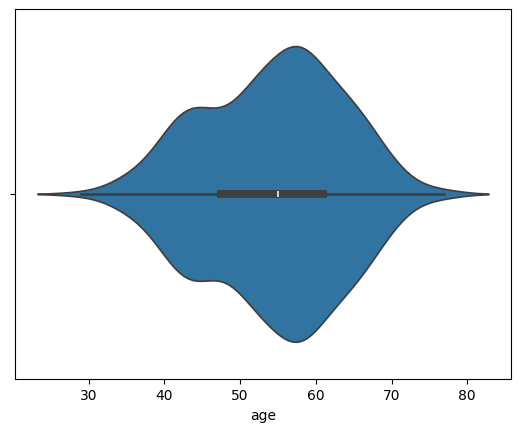
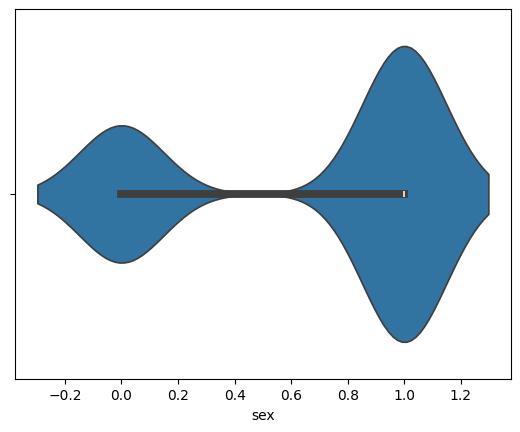
11 caa 303 non-null int64

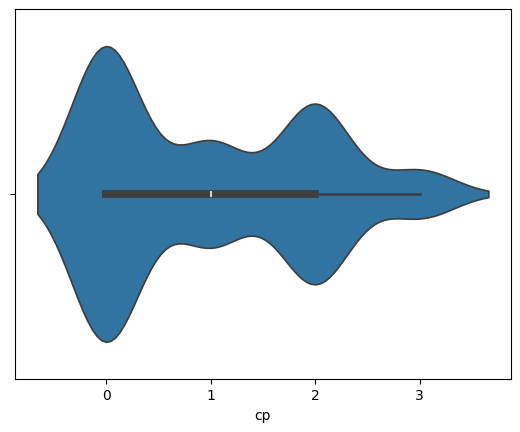
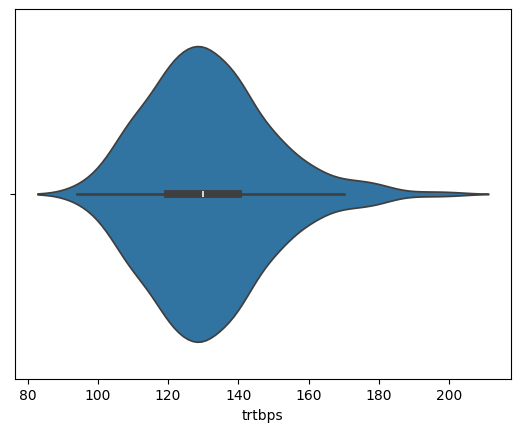
12 thall 303 non-null int64

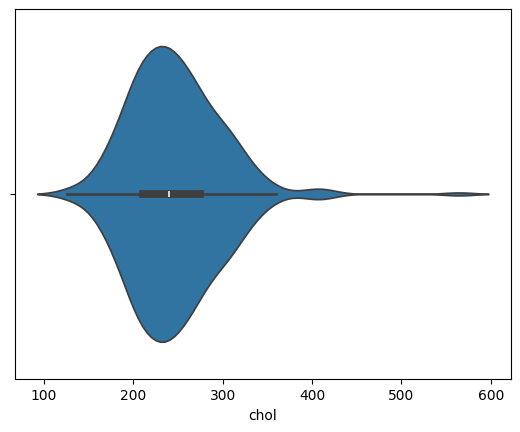
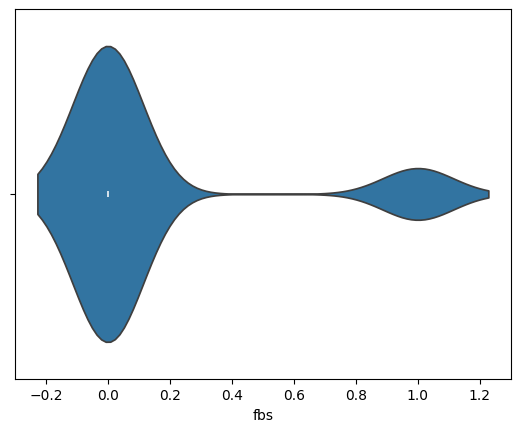
13 output 303 non-null int64

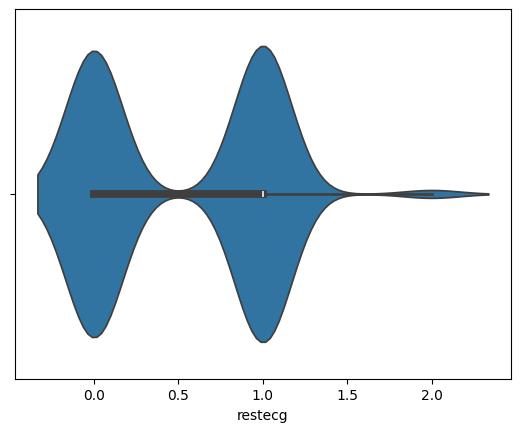
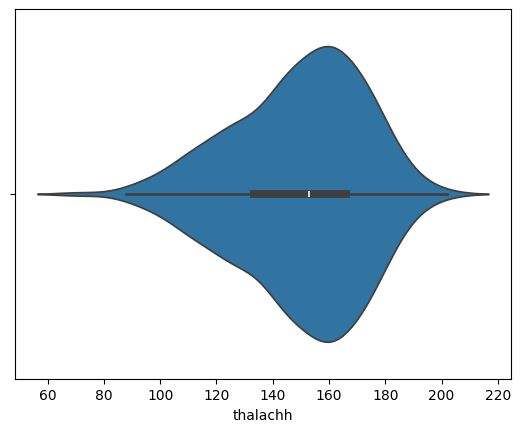
Пустых значений не обнаружено, все категориальные признаки уже были закодированы

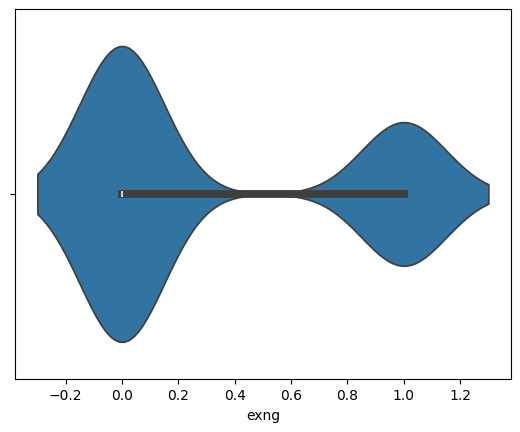
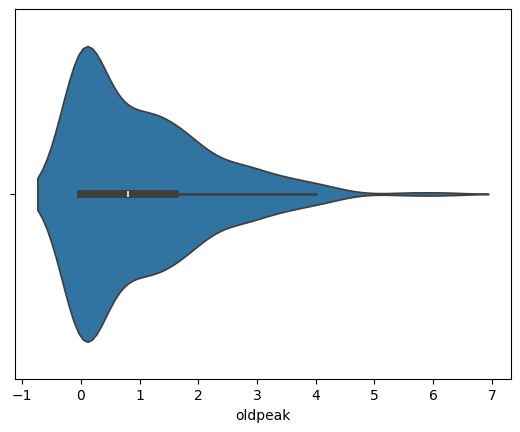
Скрипичные диаграммы:

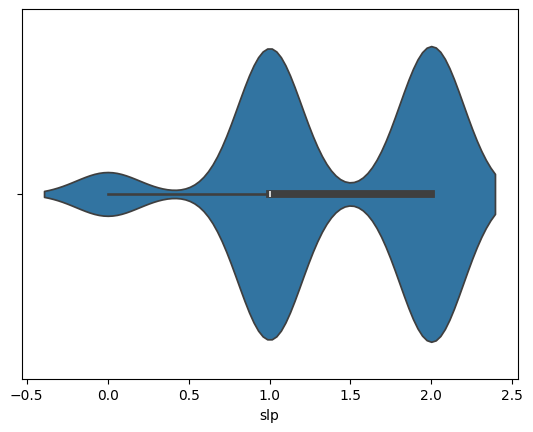
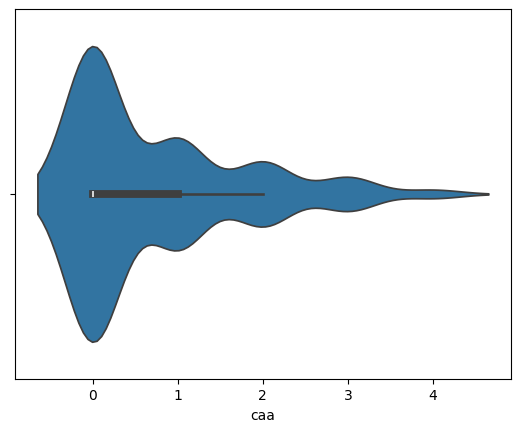
 

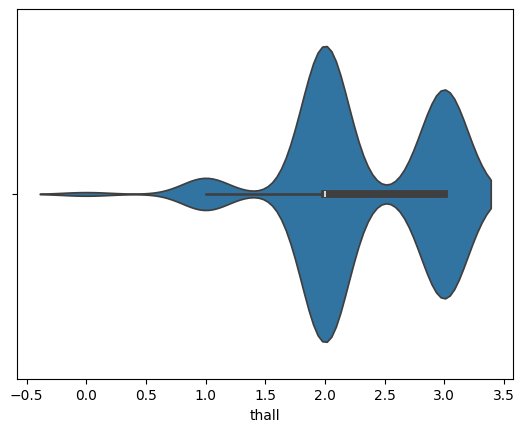
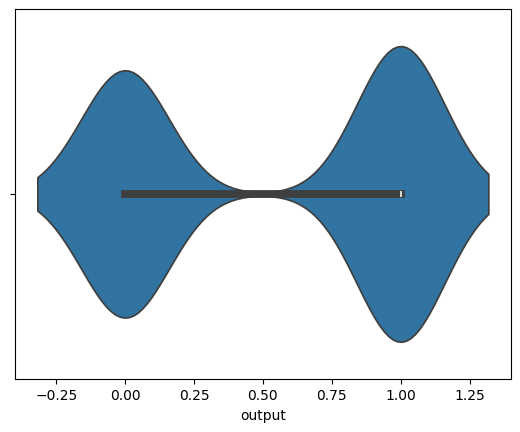
 

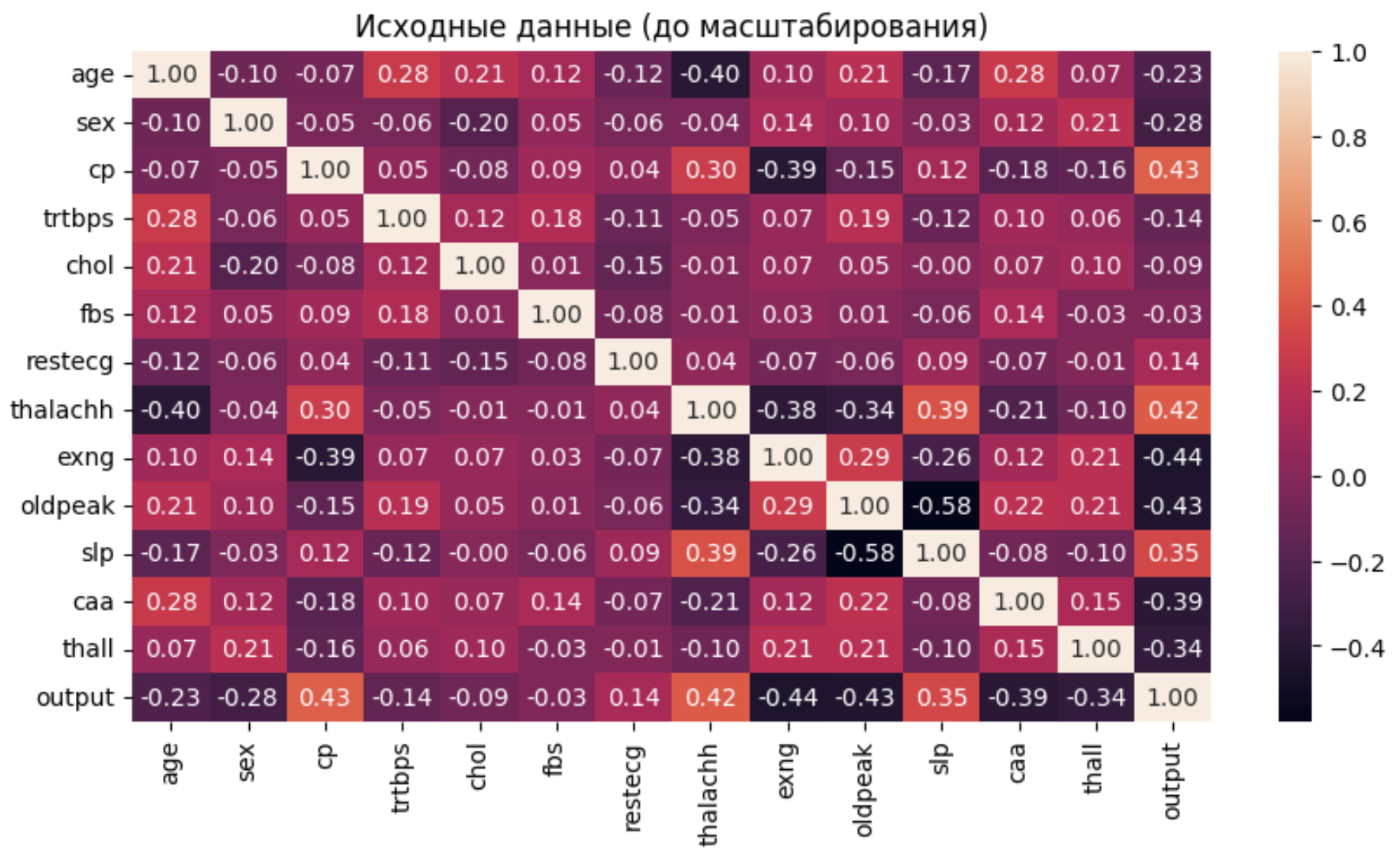
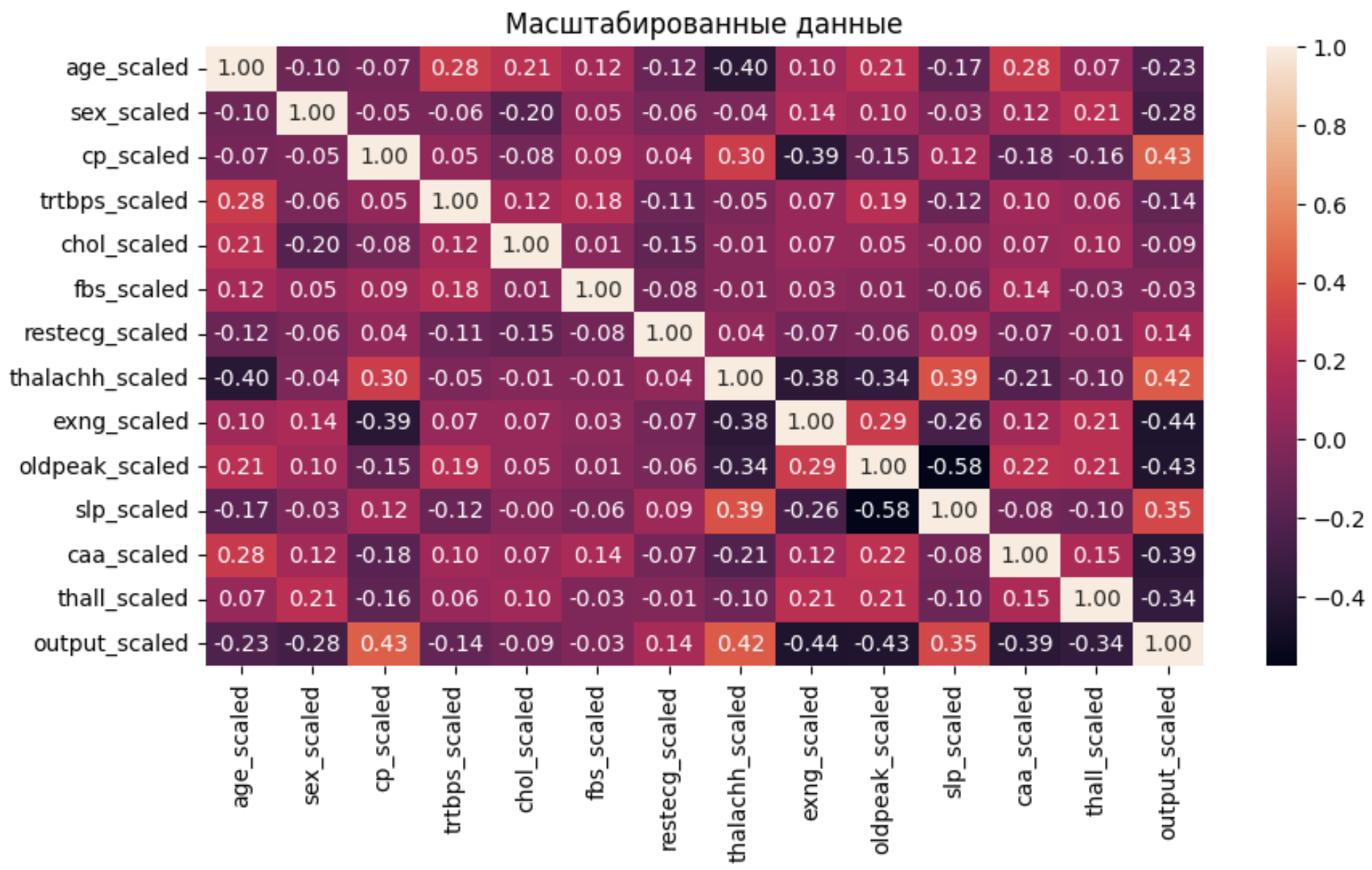
* 1. **Подготовка Данных: В процессе подготовки данных были выполнены следующие действия:**

Масштабирование числовых признаков, чтобы обеспечить равномерное распространение значений и улучшить стабильность обучения моделей.

* 1. **Выбор Признаков:**

Был проведен анализ корреляции между признаками и целевой переменной "output" для определения наиболее значимых признаков, которые будут использоваться для построения моделей.

Корреляционные матрицы для исходных и масштабированных данных совпадают. Категориальные признаки отсутствуют, их кодирования не требуется. Вспомогательные признаки для улучшения качества моделей в данном примере мы строить не будем

* 1. **Выбор метрик для последующей оценки качества моделей:**
* Accuracy - Общая точность модели, показывает долю правильных ответов.
* Precision - Точность положительных прогнозов, хорошо подходит, когда важно минимизировать ложные тревоги.
* Recall - Полнота положительных прогнозов, важно, когда критично не упустить положительные случаи.
* F1-score - Гармоническое среднее precision и recall, баланс между ними.
* ROC-AUC - Площадь под ROC-кривой для оценки качества бинарных моделей классификации.
  1. **Выбор наиболее подходящих моделей для решения задачи классификации.**
* Логистическая регрессия
* Решающее дерево
* Случайный лес
* Градиентный бустинг
* Адаптивный бустинг
* Опорно-векторная машина
  1. **Формирование обучающей и тестовой выборок на основе исходного набора данных.**

X\_train.shape, y\_train.shape, X\_test.shape, y\_test.shape  
Размер выборок 80% и 20%

* 1. **Построение базового решения (baseline) для выбранных моделей без подбора гиперпараметров.**

Logistic Regression Decision Tree Random Forest

accuracy 0.852459 0.819672 0.852459

precision 0.870968 0.862069 0.848485

recall 0.843750 0.781250 0.875000

f1 0.857143 0.819672 0.861538

roc\_auc 0.852909 0.821659 0.851293

Gradient Boosting AdaBoost SVM

accuracy 0.770492 0.803279 0.704918

precision 0.800000 0.857143 0.666667

recall 0.750000 0.750000 0.875000

f1 0.774194 0.800000 0.756757

roc\_auc 0.771552 0.806034 0.696121

* 1. **Подбор гиперпараметров для выбранных моделей. Анализ Результатов**

Best parameters for Random Forest: {'bootstrap': True, 'ccp\_alpha': 0.0, 'class\_weight': None, 'criterion': 'gini', 'max\_depth': None, 'max\_features': 'auto', 'max\_leaf\_nodes': None, 'max\_samples': None, 'min\_impurity\_decrease': 0.0, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'min\_weight\_fraction\_leaf': 0.0, 'n\_estimators': 100, 'n\_jobs': None, 'oob\_score': False, 'random\_state': None, 'verbose': 0, 'warm\_start': False}

Best parameters for Gradient Boosting: {'ccp\_alpha': 0.0, 'criterion': 'friedman\_mse', 'init': None, 'learning\_rate': 0.1, 'loss': 'log\_loss', 'max\_depth': 3, 'max\_features': None, 'max\_leaf\_nodes': None, 'min\_impurity\_decrease': 0.0, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'min\_weight\_fraction\_leaf': 0.0, 'n\_estimators': 100, 'n\_iter\_no\_change': None, 'random\_state': None, 'subsample': 1.0, 'tol': 0.0001, 'validation\_fraction': 0.1, 'verbose': 0, 'warm\_start': False}

Best parameters for AdaBoost: {'algorithm': 'SAMME.R', 'base\_estimator': 'deprecated', 'estimator': None, 'learning\_rate': 0.1, 'n\_estimators': 50, 'random\_state': None}

Best parameters for SVM: {'C': 10, 'break\_ties': False, 'cache\_size': 200, 'class\_weight': None, 'coef0': 0.0, 'decision\_function\_shape': 'ovr', 'degree': 3, 'gamma': 'scale', 'kernel': 'linear', 'max\_iter': -1, 'probability': True, 'random\_state': None, 'shrinking': True, 'tol': 0.001, 'verbose': False}

Best parameters for Logistic Regression: {'C': 1.0, 'class\_weight': None, 'dual': False, 'fit\_intercept': True, 'intercept\_scaling': 1, 'l1\_ratio': None, 'max\_iter': 1000, 'multi\_class': 'auto', 'n\_jobs': None, 'penalty': 'l2', 'random\_state': None, 'solver': 'lbfgs', 'tol': 0.0001, 'verbose': 0, 'warm\_start': False}

Best parameters for Decision Tree: {'ccp\_alpha': 0.0, 'class\_weight': None, 'criterion': 'gini', 'max\_depth': None, 'max\_features': None, 'max\_leaf\_nodes': None, 'min\_impurity\_decrease': 0.0, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'min\_weight\_fraction\_leaf': 0.0, 'random\_state': None, 'splitter': 'best'}

* 1. **Сравнение значений базового и с гиперпараметром**

Baseline results vs Tuned results:

Baseline \

Logistic Regression Decision Tree Random Forest Gradient Boosting

accuracy 0.852459 0.819672 0.852459 0.770492

precision 0.870968 0.862069 0.848485 0.800000

recall 0.843750 0.781250 0.875000 0.750000

f1 0.857143 0.819672 0.861538 0.774194

roc\_auc 0.852909 0.821659 0.851293 0.771552

Tuned \

AdaBoost SVM Random Forest Gradient Boosting AdaBoost

accuracy 0.803279 0.704918 0.852459 0.770492 0.852459

precision 0.857143 0.666667 0.848485 0.800000 0.870968

recall 0.750000 0.875000 0.875000 0.750000 0.843750

f1 0.800000 0.756757 0.861538 0.774194 0.857143

roc\_auc 0.806034 0.696121 0.851293 0.771552 0.852909

SVM Logistic Regression Decision Tree

accuracy 0.868852 0.852459 0.803279

precision 0.875000 0.870968 0.916667

recall 0.875000 0.843750 0.687500

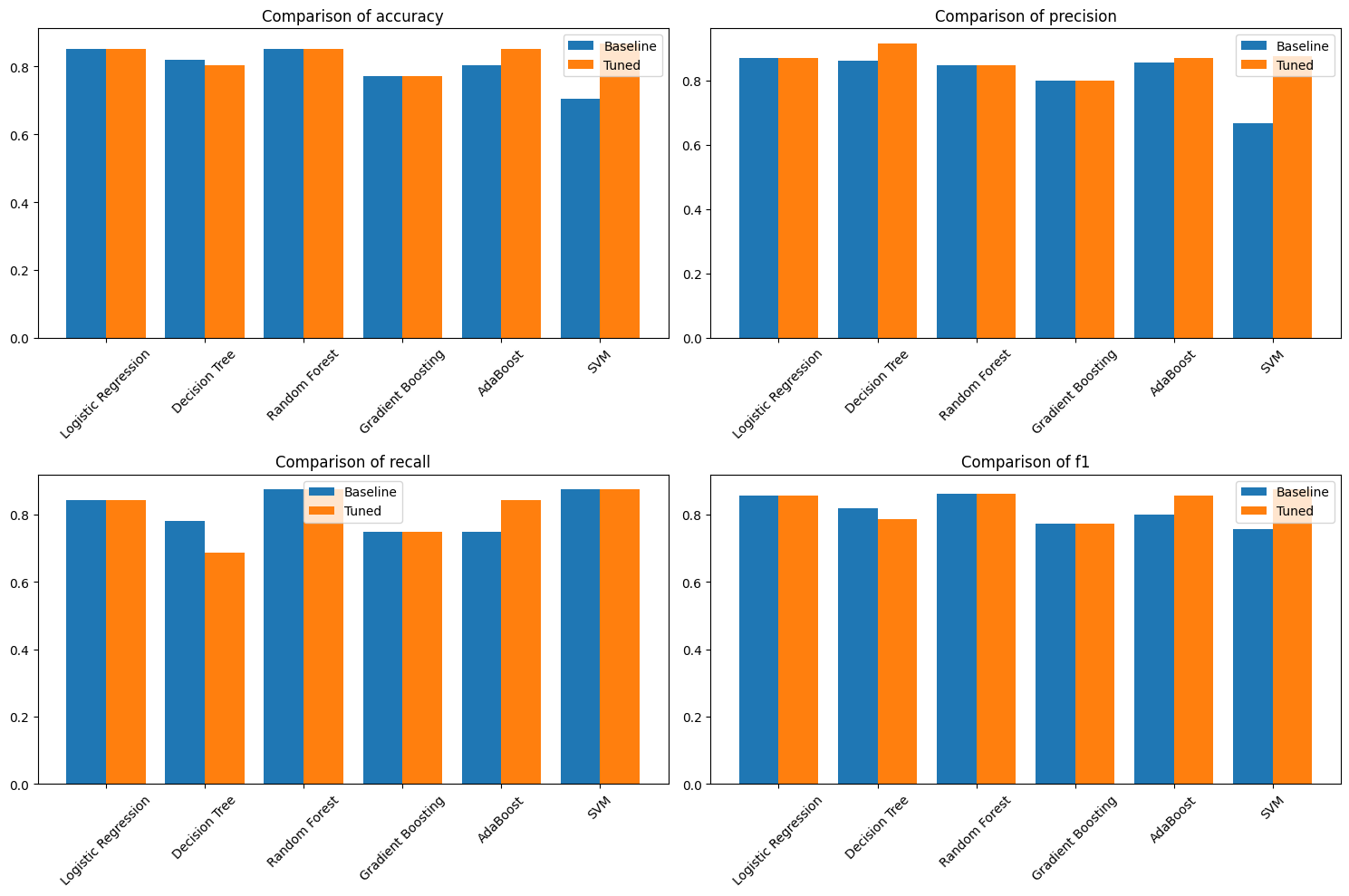
f1 0.875000 0.857143 0.785714

roc\_auc 0.868534 0.852909 0.809267

Итог: гиперпараметры помогли незначительно улучшить результаты метрик

* 1. **Формулирование Выводов.**

На основе проведенного анализа были сформулированы выводы о том, какие факторы оказывают наибольшее влияние на вероятность развития сердечно-сосудистых заболеваний и какая модель машинного обучения наиболее эффективна для прогнозирования.



Accuracy Comparison

Logistic Regression: Baseline = 0.8525, Tuned = 0.8525

Decision Tree: Baseline = 0.8197, Tuned = 0.8033

Random Forest: Baseline = 0.8525, Tuned = 0.8525

Gradient Boosting: Baseline = 0.7705, Tuned = 0.7705

AdaBoost: Baseline = 0.8033, Tuned = 0.8525

SVM: Baseline = 0.7049, Tuned = 0.8689

Precision Comparison

Logistic Regression: Baseline = 0.8710, Tuned = 0.8710

Decision Tree: Baseline = 0.8621, Tuned = 0.9167

Random Forest: Baseline = 0.8485, Tuned = 0.8485

Gradient Boosting: Baseline = 0.8000, Tuned = 0.8000

AdaBoost: Baseline = 0.8571, Tuned = 0.8710

SVM: Baseline = 0.6667, Tuned = 0.8750

Recall Comparison

Logistic Regression: Baseline = 0.8438, Tuned = 0.8438

Decision Tree: Baseline = 0.7812, Tuned = 0.6875

Random Forest: Baseline = 0.8750, Tuned = 0.8750

Gradient Boosting: Baseline = 0.7500, Tuned = 0.7500

AdaBoost: Baseline = 0.7500, Tuned = 0.8438

SVM: Baseline = 0.8750, Tuned = 0.8750

F1 Comparison

Logistic Regression: Baseline = 0.8571, Tuned = 0.8571

Decision Tree: Baseline = 0.8197, Tuned = 0.7857

Random Forest: Baseline = 0.8615, Tuned = 0.8615

Gradient Boosting: Baseline = 0.7742, Tuned = 0.7742

AdaBoost: Baseline = 0.8000, Tuned = 0.8571

SVM: Baseline = 0.7568, Tuned = 0.8750

На основании итоговых метрик, метод опорных векторов (SVM) показал лучший результат после оптимизации гиперпараметров:

* Accuracy: 86.89%
* Precision: 87.50%
* Recall: 87.50%
* F1 score: 87.50%

Обоснование:

SVM демонстрирует высокую точность и F1-метрику, что подчеркивает высокое качество предсказаний. Улучшение метрик после настройки гиперпараметров означает, что модель лучше адаптировалась к данным. Высокие показатели точности (Precision) и полноты (Recall) говорят о хорошей сбалансированности между ложноположительными и ложноотрицательными предсказаниями. Таким образом, для данной задачи можно рекомендовать метод опорных векторов (SVM) с настроенными гиперпараметрами как лучшую модель с точки зрения общей производительности.

## Заключение

В ходе данной работы было проведено комплексное исследование, направленное на прогнозирование риска сердечно-сосудистых заболеваний на основе представленных в исследовании параметров на основе набора данных, предоставленного Kaggle. Были разработаны и оценены различные модели машинного обучения, включая базовые и оптимизированные с использованием подбора гиперпараметров.

Результаты показали, что подбор гиперпараметров не существенно повышает качество прогнозирования, делая модели более точными и надежными. Особенно эффективной показалась модель линейной регрессии.

Таким образом, разработанные модели машинного обучения и проведенный анализ позволяют не только прогнозировать риск сердечно-сосудистых заболеваний, но и выявлять ключевые факторы. Результаты исследования могут быть использованы как основа для дальнейших научных и прикладных работ в области медицины, биоинформатики и общественного здравоохранения. Они могут помочь в разработке более точных и эффективных стратегий предупреждения сердечно-сосудистых заболеваний, что в свою очередь способствует повышению качества жизни и продления жизненного потенциала населения.

В будущем, расширение анализа на дополнительные признаки и использование более продвинутых методов машинного обучения может еще больше улучшить точность прогнозирования и дать возможность более глубокого понимания механизмов принятия решений

## Список использованных источников информации

1. Kaggle. (2024). Heart Attack Analysis & Prediction Dataset

URL:<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/data>

1. Скриптовый язык программирования Python. (2023). Documentation.

URL: <https://docs.python.org/3/>

1. Scikit-learn. (2023). Machine Learning in Python.

URL: <https://scikit-learn.org/stable/>

1. Scikit-learn. (2023). Machine Learning in Python.

URL: <https://github.com/ugapanyuk/courses_current/wiki/COURSE_TMO_SPRING_2024>