



# Модел за предсказване на риска от инсулт

(Stroke Risk Prediction Model)

---

Мария Симеонова



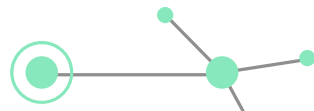
# Съдържание:

**01**    **Цел и обхват на проекта**

**02**    **Описание на данни**

**03**    **Методология и модел**

**04**    **Резултати и заключения**



# 01

## Цел и обхват на проекта

---

модел който да предостави точни  
предсказания за риска от инсулт





## Обхват на проекта

- Анализ на наличните данни и тяхната предварителна обработка.
- Изграждане на модел за предсказване на риска от инсулт.
- Оценка на модела с помощта на метрики като точност, матрица на обърквания и отчет за класификация.
- Създаване на графичен потребителски интерфейс (GUI) за лесно въвеждане на данни и получаване на предсказания.

# 02

## Описание на данни



Наборът от данни е предоставен от Kaggle и съдържа информация за пациенти, използвана за изграждане на модела.

# Атрибути на данните:

- id**: Уникален идентификатор
- gender**: Пол
- age**: Възраст
- hypertension**: Индикатор за наличие на хипертония
- heart\_disease**: Индикатор за наличие на сърдечни заболявания
- ever\_married**: Информация дали пациентът е бил женен
- work\_type**: Тип на работа
- Residence\_type**: Тип на пребиваване
- avg\_glucose\_level**: Средно ниво на глюкоза в кръвта
- bmi**: Индекс на телесна маса.
- smoking\_status**: Статус на пушене
- stroke**: Целевата променлива(инсулт)



## Почистване и обработка на данните

- Зареждане на данните
- Обработка на липсващите стойности
- Почистване на данните
- Кодиране на категориални променливи

# Анализ на корелации



*Играе ли възрастта роля при  
получаването на инсулт?*



*По-податливи ли са  
пушачите към инсулт?*



*По-голям ли е рискът от инсулт при хора със  
сърдечни заболявания?*

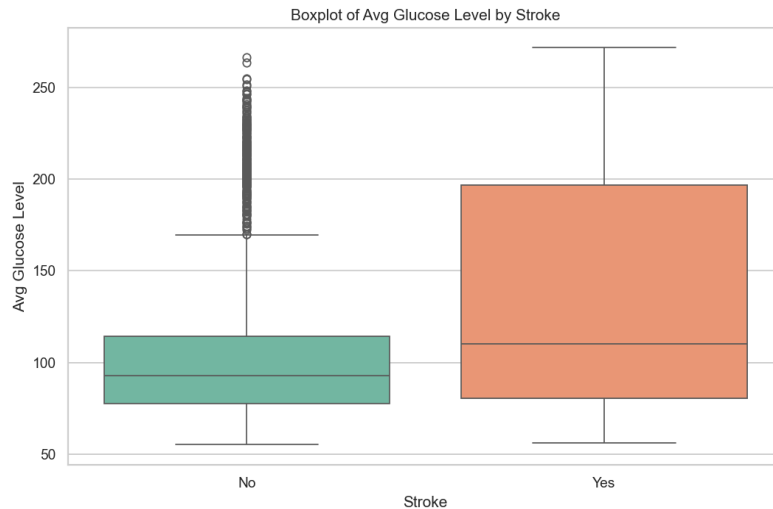
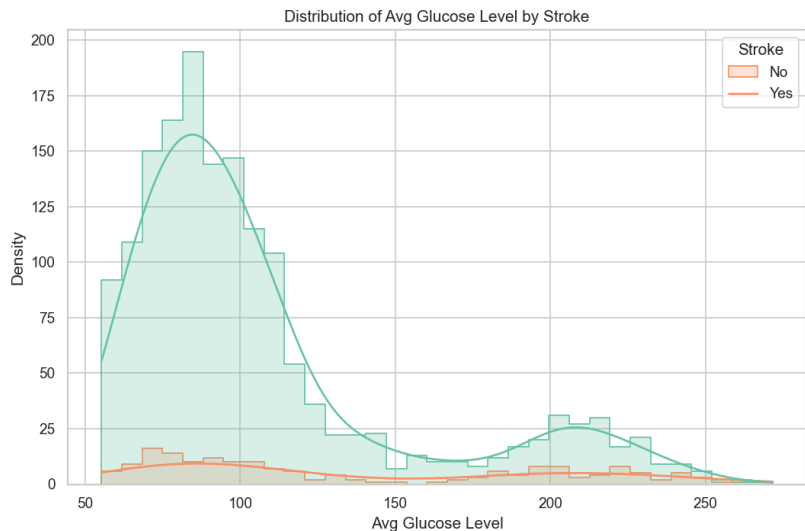


*По-податливи ли са мъжете на  
инсулти?*



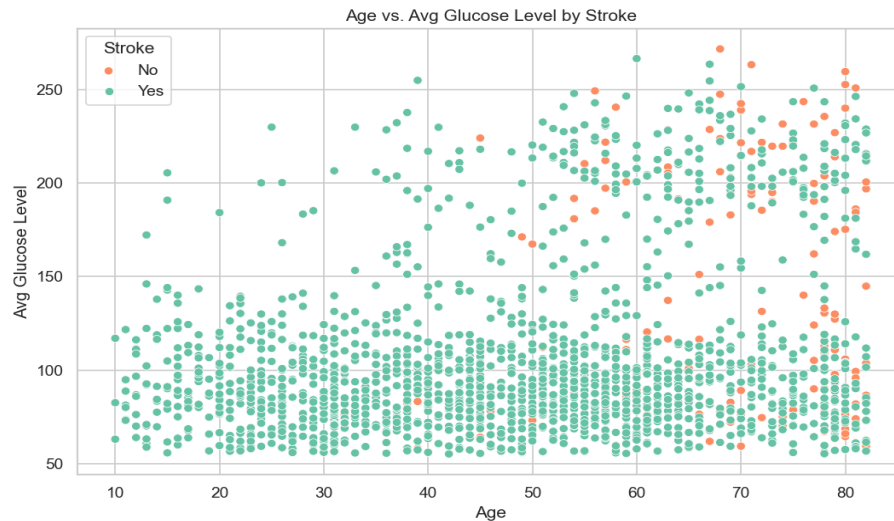
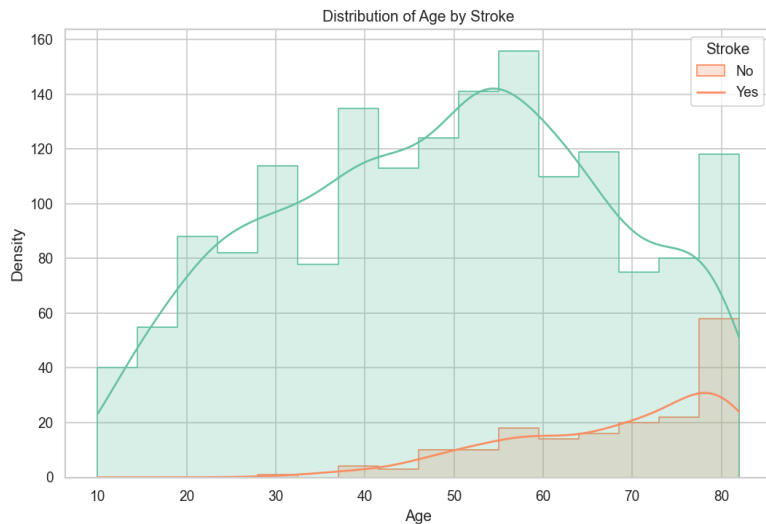


# Разпределение на средното ниво на глюкоза според инсулт



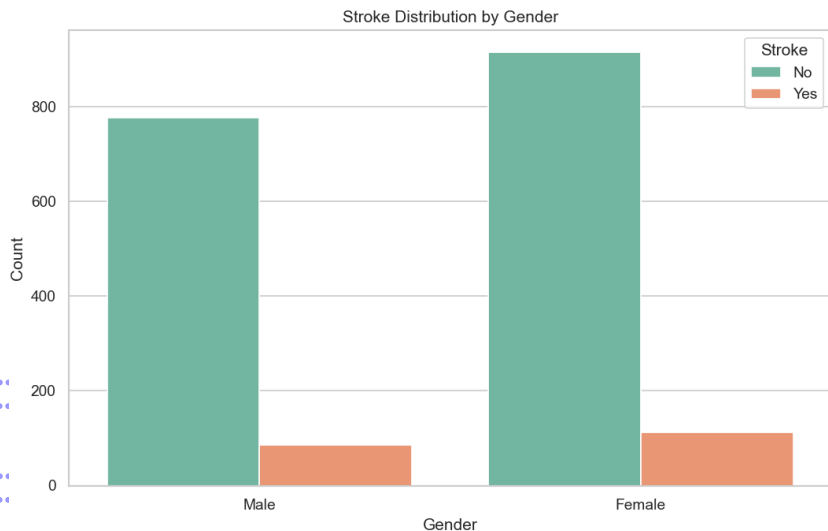
Лицата с по-високо средно ниво на глюкоза са по-податливи на инсулт.

# Разпределение на възрастта според инсулт

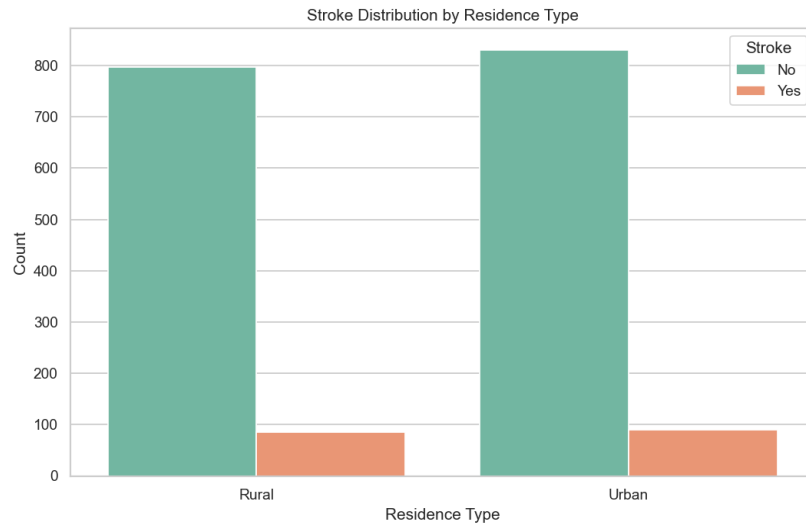


*Положителна връзка между възрастта и риска от инсулт, като по-възрастните индивиди са по-изложени на риск.*

# Разпределение на пола и местожителството според инсулт



Полът сам по себе си *не е определящ* фактор за риска от инсулт.



Няма значителна връзка между местожителството и риска от инсулт.

# Разделяне на данните



80%

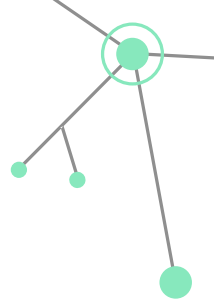
Обучаващи

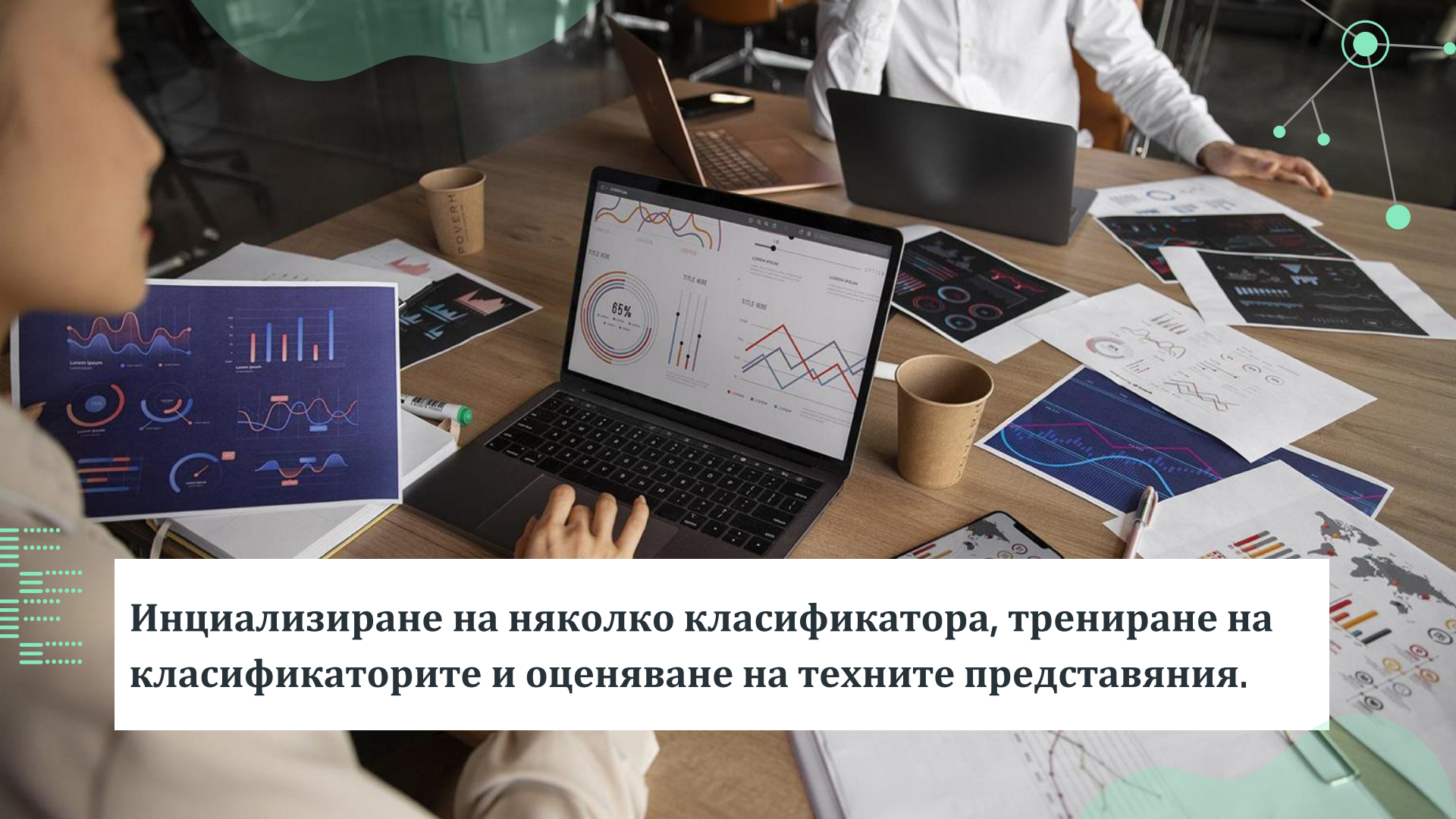


20%

Тестови

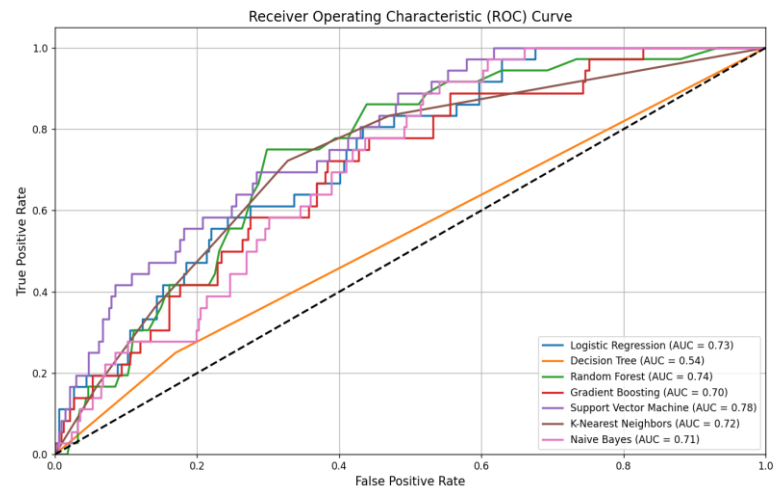
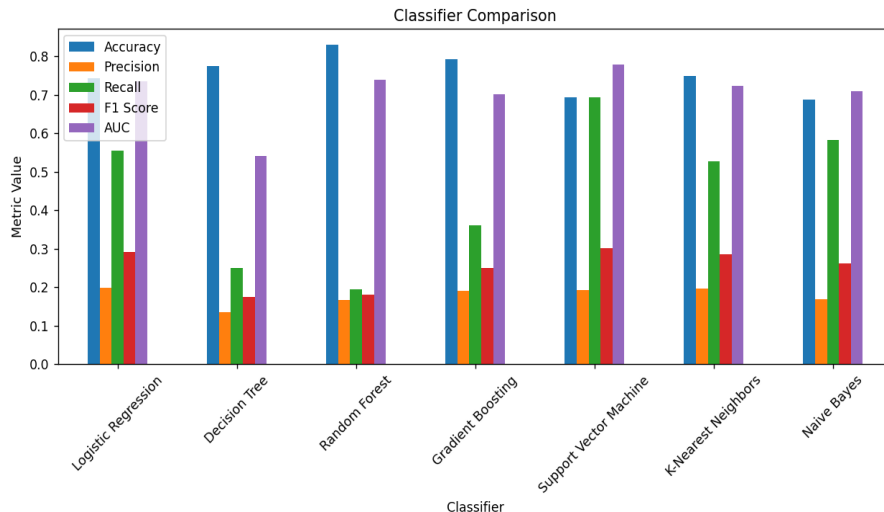
Прилагане на **SMOTE** (Synthetic Minority Over-sampling Technique)  
за справяне с дисбаланса между класовете.





**Инциализиране на няколко класификатора, трениране на класификаторите и оценяване на техните представяния.**

# Оценка на Класификаторите



**Най-добър Класификатор: Случайната гора (Random Forest)**  
*показва най-висока точност (0.828)*

## Random Forest Classifier:

Този модел е избран заради неговата гъвкавост и способност да се справя с неравномерно разпределени данни и сложни взаимодействия между функциите. Моделът е обучен с помощта на обработените данни и е

оценен с точност от **87.81%.**



# Използване на потребителския интерфейс

Stroke Risk Prediction

Gender

Age

Hypertension

Heart Disease

Average Glucose Level

BMI

Ever Married

Residence Type

Work Type

Smoking Status

Stroke Risk Prediction

Gender

Age

Hypertension

Heart Disease

Average Glucose Level

BMI

Ever Married

Residence Type

Work Type

Smoking Status

**Low risk of having a stroke.**

Stroke Risk Prediction

Gender

Age

Hypertension

Heart Disease

Average Glucose Level

BMI

Ever Married

Residence Type

Work Type

Smoking Status

**High risk of having a stroke.**



# Заключение

Въпреки че моделът показва висока обща точност от 87.81% при тестовите данни, резултатите показват значителен дисбаланс в предсказанията.

Моделът има **висока точност** за отрицателните случаи (без инсулт), но значително **по-ниска точност** при идентифициране на положителни случаи (с инсулт), което може да се дължи на дисбаланса в данните.

Необходимо е да се разгледат допълнителни методи за подобряване на способността за предсказване на положителните случаи.



# Благодаря Ви за вниманието!

*Въпроси?*

