

Федеральное государственное бюджетное образовательное учреждение высшего образования  
"Красноярский государственный медицинский университет имени профессора В.Ф. Войно-  
Ясенецкого" Министерства здравоохранения Российской Федерации

ВСЕРОССИЙСКАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ «ЗДРАВООХРАНЕНИЕ  
БУДУЩЕГО: ОРИЕНТИР НА КАЧЕСТВО» 8-9 декабря 2023 г.

# Сравнительный анализ прогностических моделей в риске возникновения инсульта

Студенческая наука "Здравоохранение будущего"

Выполнила: Сирунян Кристина Саргисовна  
Научный руководитель: Васильева Мария Равильевна

# Актуальность и цель исследования

| Проблема                                   | Суть                                                            | Востребованность                                                                                                                                                                   | Пути решения                                                                 |
|--------------------------------------------|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Острые нарушения мозгового кровообращения. | Высокая доля в структуре заболеваемости и смертности населения. | Разработка математической модели, включающей прогностический комплекс из факторов риска, клинических проявлений, показателей лабораторных и инструментальных методов обследования. | С помощью сравнительного анализа выбор наилучшей модели с высокой точностью. |

Цель: сравнительный анализ сконструированных классификационных моделей риска возникновения инсульта.

# Методы



**Метод опорных векторов (SVM)**



**Метод k-Ближайших соседей (kNN)**



**"Случайный лес" (Randomforestclassifier)**



**Деревья решений( DecisionTreeClassifier)**



**Логистическая регрессия(LogisticRegression)**



**Стэкинг(Stacking)**

# Материалы

► Источник базы данных: **kaggle**

Исходная база содержит следующие параметры:

- Категориальные: пол, гипертензия, болезни сердца, семейное положение, тип работы, место жительства, статус курения, диагноз
- Количественные: возраст, уровень глюкозы, ИМТ

## Пол:

Female: 2897  
Male: 2011  
Other: 1

## Гипертензия:

0(нет): 4458  
1(есть): 451

## Сердечная болезнь:

0(нет): 4666  
1(есть): 243

## Диагноз инсульт:

0(нет) : 4700  
1(есть): 209

## Семейное положение:

В браке: 3204  
Не в браке: 1705

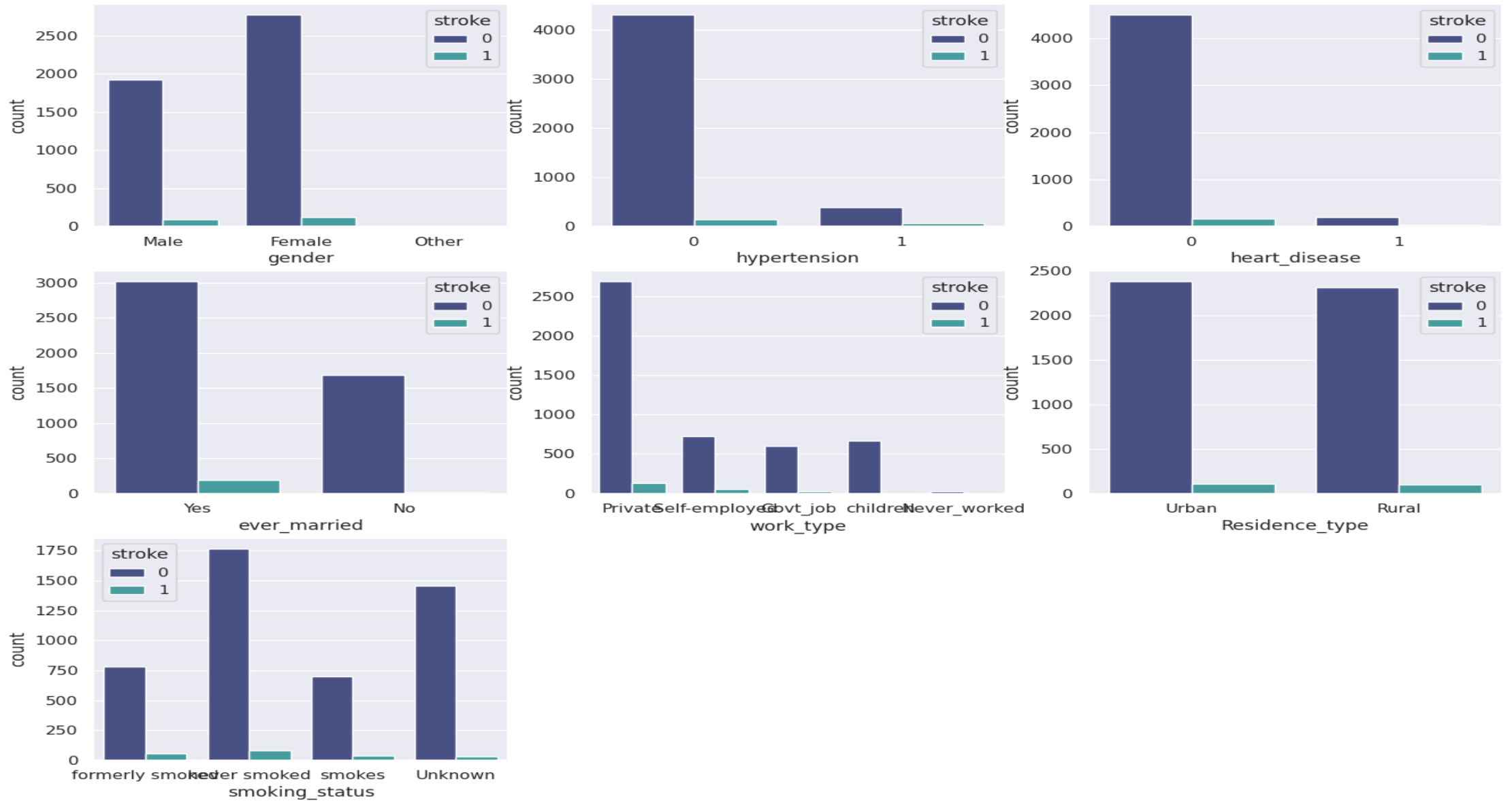
## Тип работы:

Скрыто: 2811  
Самозанятый: 775  
В декрете: 671  
Государственная работа: 630  
Никогда не работали: 22

## Статус курения:

Никогда не курили: 1852  
Скрыто: 1483  
Ранее курили: 837  
Курят: 737

# Проблема дисбаланса данных



# Сравнение классификационных моделей

| Модель                 | Точность |
|------------------------|----------|
| SVC                    | 80.78%   |
| KNeighborsClassifier   | 92.51%   |
| RandomForestClassifier | 99.13%   |
| DecisionTreeClassifier | 97.05%   |
| LogisticRegression     | 76.37%   |
| Stacking Classifier    | 99,82%   |

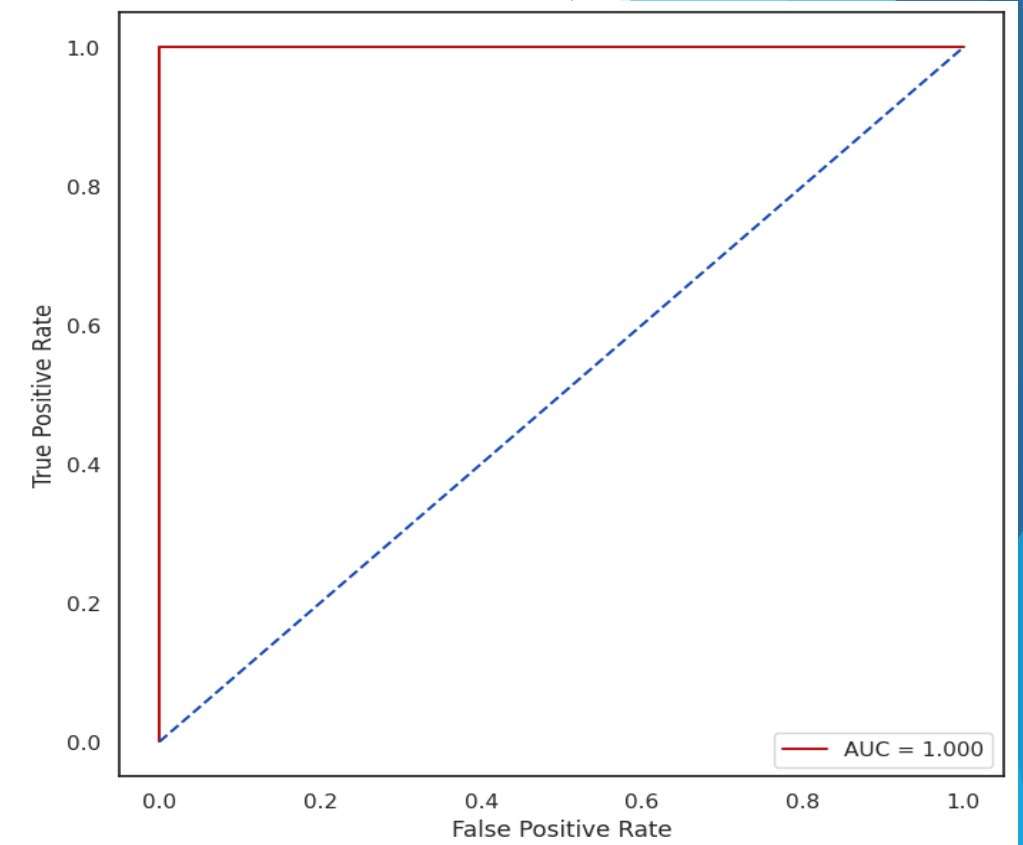
Точность(accuracy)-доля правильных классификаций



# Результат

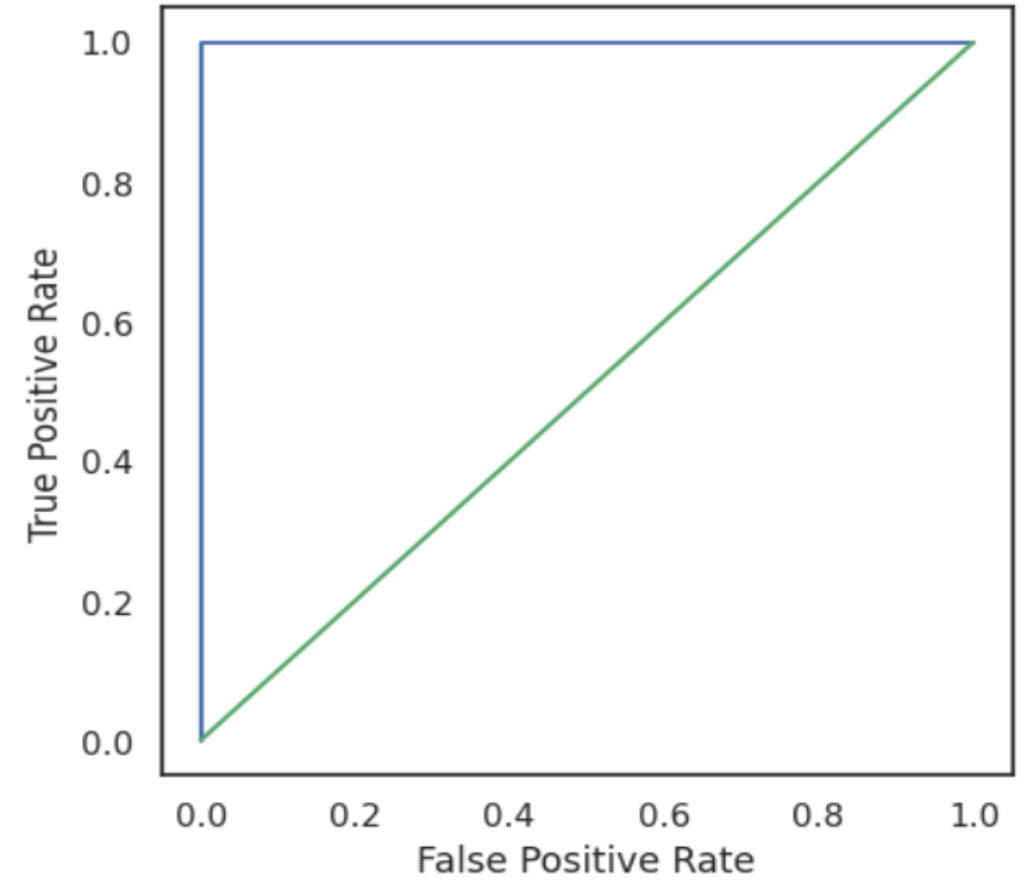
## Random Forest Classifier

|           |                     |                  |
|-----------|---------------------|------------------|
| No stroke | 1389                | 21               |
| Stroke    | 0                   | 1410             |
|           | Predicted no stroke | Predicted stroke |



# Stacking Classifier

|           |                     |                  |
|-----------|---------------------|------------------|
| No stroke | 1406                | 4                |
| Stroke    | 0                   | 1410             |
|           | Predicted no stroke | Predicted stroke |



# Выводы

- ▶ Проведена предобработка данных.
- ▶ Построены 5 моделей на основе следующих методов:
  - RandomForestClassifier("случайный лес"),
  - DecisionTreeClassifier( деревья решений),
  - K-nearest neighbors Classifier (метод k ближайших соседей)
  - SVM( метод опорных векторов),
  - Logistic Regression (логистическая регрессия).
- ▶ Наилучшие модели построенные на основе RandomForestClassifier и Stacking Classifier показали результаты 99,13% и 99,82% соответственно.