

KNN, K=8 14,56

Таблиця 1 - Частоти помилки моделей

Отже, найкращими моделями є логістична регресія та лінійний дискримінантний аналіз з помилками в 14,34%.

Визначимо коефіцієнти моделей та побудуємо їх.

Модель логістичної регресії:

$$y = 1.762 * \text{FastingBS} + 2.461 * \text{cptASY} + 1.889 * \text{sex} - 2.96 * \text{stUp} - 1.493$$

Модель лінійного дискримінантного аналізу:

$$y = 0.765 * \text{FastingBS} + 1.418 * \text{cptASY} + 0.881 * \text{sex} - 1.798 * \text{stUp} - 0.8622$$

Список літератури:

1. Серцева недостатність: симптоми та методи лікування [Електронний ресурс] . – Режим доступу до ресурсу: <https://oxford-med.com.ua/ua/media-center/publikacii/serdechnaya-nedostatochnost/>
2. Heart Failure Prediction Dataset [Електронний ресурс] . – Режим доступу до ресурсу: <https://www.kaggle.com/fedesoriano/heart-failure-prediction>
3. Джеймс Г., Уиттон А., Хасті Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R Изд. Второе, испр. Пер с англ. С.Э. Мاستицкого –М. ДМК Пресс, 2017. -456с

УДК 519.2:004.8:004.62

*Чернега В.М., студентка 3 курсу
спеціальності 122 «Комп'ютерні науки»
Нескородева Т.В., к.т.н., доцент, доцент
кафедри інформаційних технологій*

АНАЛІЗ ФАКТОРІВ, ЩО ПРИЗВОДЯТЬ ДО СЕРЦЕВО-СУДИННИХ ЗАХВОРЮВАНЬ

Донецький національний університет імені Василя Стуса, м. Вінниця

На сьогоднішній день серцево-судинні захворювання – одна з найбільш частих причин смертності людей у світі. Вони пов'язані із патологіями в серцево-

судинній системі, що виникають на фоні різних факторів. В залежності від патології серцево-судинні захворювання поділяють на різні типи, наприклад аритмія, хвороби кровоносних судин, ішемічна хвороба серця, кардіоміопатія, міокардит, аневризма аорти тощо.

На даному етапі розвитку медицини відкрито багато нових методів, що допомагають лікувати та попереджати серцево-судинні захворювання. Актуальність теми полягає в тому, що не дивлячись на широкий діапазон методів боротьби із даним типом хвороб, велика кількість людей зіштовхуються з ними сьогодні. Тому існує необхідність в аналізі та виявленні чинників, що призводять до захворювань. За останніми дослідженнями виявлено фактори, що впливають на ймовірність виникнення хвороби серця у людини:

1. Високий артеріальний тиск
2. Погане харчування
3. Рівень холестеролу
4. Куріння (тютюновий дим)
5. Індекс маси тіла
6. Рівень глюкози
7. Забруднення повітря
8. Порушення функцій роботи нирок
9. Надмірне вживання алкоголю
10. Низька фізична активність

Метою роботи є виявлення та аналіз чинників, що впливають на ймовірність розвитку серцево-судинних захворювань у людей різного віку та статі.

В якості інструменту аналізу використовується мова програмування R, середовище розробки RStudio та дані з датасету Heart.

Опис даних в датасеті:

1. age: вік пацієнта
2. sex: стать пацієнта
3. cp: тип болю у грудях (chest pain):
 - 0 = Typical Angina,
 - 1 = Atypical Angina,
 - 2 = Non-anginal Pain,
 - 3 = Asymptomatic
4. trtbps: кров'яний тиск у спокої в mm/Hg (resting blood pressure)
5. chol: холестерин в mg/dl виміряний через BMI сенсор.
6. fbs: рівень цукру в крові (fasting blood sugar) якщо значення більше 120 mg/dl: 1 = True, 0 = False
7. restecg: електрокардіограма в стані спокою (resting electrocardiographic results):
 - 0 = Normal,
 - 1 = ST-T wave normality,
 - 2 = Left ventricular hypertrophy
8. thalachh: максимальний пульс (maximum heart rate achieved)

9. oldpeak: попередній максимальний пульс (previous peak)
- 10.slp: спуск (slope)
- 11.caa: Кількість великих судин (number of major vessels)
- 12.thall: Талієвий стрес тест: притік крові в серце через корональні судини (thallium Stress Test result ~ (0,3))
- 13.exng: стенокардія, викликана фізичним навантаженням (exercise induced angina). Результати: 1 = Так, 0 = Ні.
- 14.output: Цільовий показник (target variable):
 - 0 = низька ймовірність виникнення захворювання серця,
 - 1 = висока ймовірність виникнення захворювання серця.

Побудова графіків розподілу кожного атрибуту за допомогою функції `ggplot()` з пакету `tidyverse`:

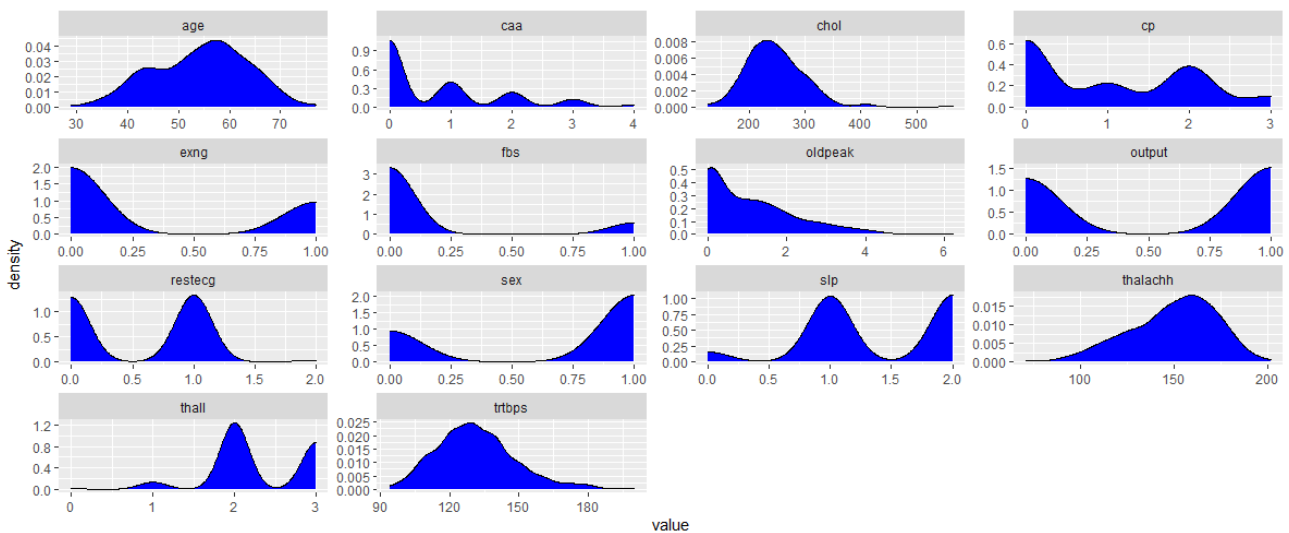


Рисунок 1. Графіки розподілу даних

Дослідження кореляції даних проводиться за допомогою функції `cor()`, а відображення на графіку – `corrplot()`. З графіку кореляції можна зробити висновок, що найбільший вплив на виникнення серцево-судинних захворювань мають значення пульсу, спуску та наслідком є сильний біль у грудях. Найменше впливають фізичні навантаження, кількість великих судин та притік крові через корональні судини.

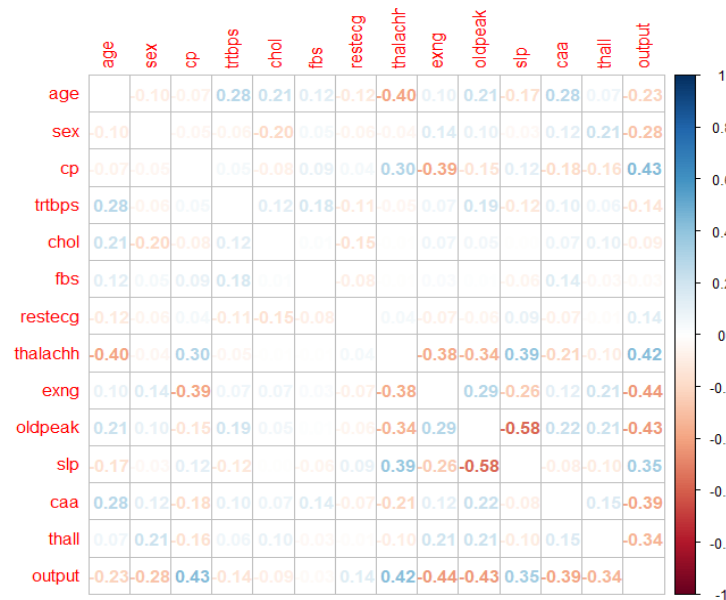


Рисунок 1. Графік кореляції даних

Також досліджувалась залежність між рівнем холестеролу та ймовірністю виникнення захворювання. З гістограми можна побачити, що рівень даного показника також досить сильно впливає на протікання хвороби та ймовірність її виникнення.

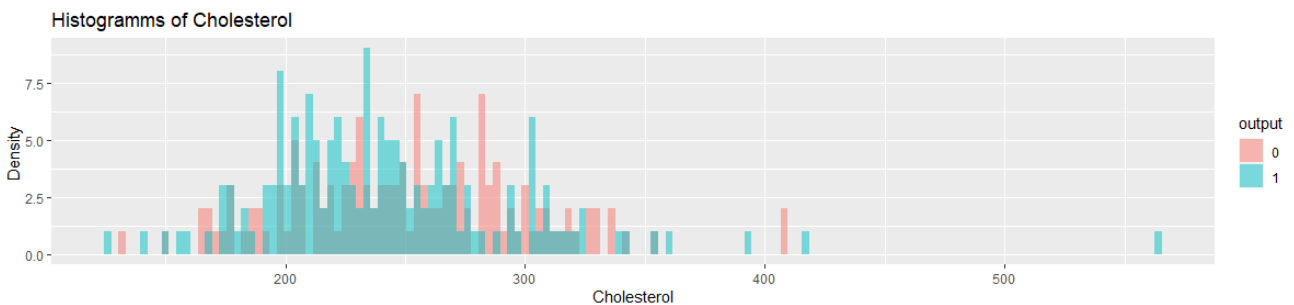


Рисунок 1. Графік залежності холестеролу та ймовірності захворювання

На даний момент існує багато методів для попередження та запобіганню серцево-судинних захворювань, визначені чинники, що призводять до їх виникнення, а також є можливість завчасного встановлення рівня схильності до них. Попри це, даний тип хвороб залишається досить небезпечним та потребує додаткового вивчення.

Список літератури.

1. Джеймс Г., Уиттон Д., Тибширани Р. Введение в статистическое обучение с примерами
2. Kaggle [Електронний ресурс]. Режим доступу – <https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>