

УДК 681.004.62

Р.І. Капаціла; А. С. Сверстюк, д.т.н., проф.

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ЗАСТОСУВАННЯ ДЕРЕВ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ПРОГНОЗУВАННЯ РИЗИКУ ВИНИКНЕННЯ СЕРЦЕВИХ ЗАХВОРЮВАНЬ

R.I. Kapatsila; A.S. Sverstyuk, Dr., Prof.

APPLICATION OF DECISION TREES FOR PREDICTING THE RISK OF HEART DISEASE

Метод багатофакторного аналізу прогнозування настання серцевих захворювань з урахуванням найбільш поширених та широкодоступних медичних показників дозволяє створити математичну модель для оцінки та прогнозування можливого настання серцевих захворювань. Застосування даної моделі дозволяє використовувати доволі обмежений масив інформації для попередження потенційних ризиків пов'язаних із настанням серцево-судинних захворювань.

Для аналізу та побудови моделі на основі дерева прийняття рішень використано набір даних з 12 основних атрибутів що представлені у 746 унікальних записах.

Для побудови моделі прогнозування із використанням дерева прийняття рішень відібрано наступні чинники: вік, стать, характер болю в грудях, кров'яний тиск в стані спокою, рівень холестерину в крові, рівень цукру в крові, тип результату електрокардіограми у стані спокою, максимальний серцевий ритм зафіксований при огляді, наявність стенокардії при фізичному навантаженні, депресія ST сегменту спричинена фізичними вправами порівняно з відпочинком, тип нахилу піку навантаження ST сегменту та факт виявлення хвороби серця в результаті діагностики.

Для побудови дерева прийняття рішень використано алгоритм CART (Classification and Regression Trees).

Для оцінки якості моделі побудованої на базі вище згаданих факторів отримано значення точності класифікації моделі що становить 85%.

Для перевірки моделі використано значення F-міри: це показник, який поєднує прецизійність і чутливість в одне значення, що дозволяє легше оцінити продуктивність моделі. Середнім значенням для побудованої моделі є 0.85, що означає що у 85% випадків ми отримаємо коректний результат при аналізі вхідних даних.

Для оцінки моделі обраховано специфічність. Специфічність це спосіб виміряти, наскільки добре модель правильно ідентифікує тих, хто не має захворювання. Для обрахунку специфічності можна використати наступну формулу:

Специфічність = (кількість істинно негативних результатів) / (кількість істинно негативних результатів + кількість помилково позитивних результатів)

Для описаної моделі дане значення 0.82, що відповідає 82% специфічності.

Базуючись на проведеному дослідженні можна зробити висновок, що дерева прийняття рішень на основі CART алгоритму можуть бути ефективно застосовані у для прогнозування ризику виникнення серцевих захворювань.

Література

1. Sofogianni A, Stalikas N, Antza C, Tziomalos K. Cardiovascular Risk Prediction Models and Scores in the Era of Personalized Medicine. *J Pers Med.* 2022. doi: 10.3390/jpm12071180. PMID: 35887677; PMCID: PMC9317494.

2. Townsend N., Nichols M., Scarborough P., Rayner M. Cardiovascular disease in Europe—Epidemiological update 2015. *Eur. Heart J.* 2015;36:2696–2705. doi: 10.1093/eurheartj/ehv428.