

УДК 004.85:61

ДОСЛІДЖЕННЯ МЕТОДІВ ПРОГНОЗУВАННЯ ЗАХВОРЮВАНЬ З ВИКОРИСТАННЯМ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

Василенко О.О.

Науковий керівник – канд. техн. наук, доц. Ревенчук І.А.

Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна

тел. +38(096) 283-65-34

This work is devoted to the assessment of the effectiveness of disease forecasting methods using machine learning algorithms. The main algorithms for forecasting diseases are considered. Comparative calculations of efficiency in various diseases were carried out. It has been established that the KNN algorithm is the most efficient algorithm for working with a large amount of data and is most suitable for working with medical data for further analysis.

У машинному навчанні існує кілька методів, які в різних галузях можуть виконувати прогнозу аналітику великих обсягів даних. Прогностична аналітика в охороні здоров'я є складним заходом, але зрештою вона може допомогти практикуючим лікарям приймати своєчасні рішення щодо здоров'я та лікування пацієнтів на основі масивних даних. Такі захворювання, як рак молочної залози, діабет і серцево-судинні захворювання, є причиною багатьох смертей у всьому світі, але більшість із цих смертей відбувається через відсутність своєчасних оглядів. Зазначена проблема виникає через відсутність медичної інфраструктури та низьку кількість лікарів у чисельності населення. Статистика чітко показує те саме, рекомендоване ВООЗ, співвідношення лікарів до пацієнтів становить 1:1000, тоді як співвідношення лікарів до населення в Україні становить 1:1456, це свідчить про нестачу лікарів.

Ми будемо розглядати три алгоритма:

– алгоритм логістичної регресії використовує логістичну функцію, тому цей алгоритм називається логістичної регресії. Логістична функція — це S-подібна крива, розроблена для статистичних функціональних можливостей, і крива зображена між 0 і 1. Для цілей представлення логістична регресія використовує рівняння, такі як лінійна регресія [2]:

$$y = \exp(u_0 + u_1x) / (1 + \exp(u_0 + u_1x)),$$

– KNN — це техніка машинного навчання, яка використовується як для регресії, так і для класифікації. Алгоритм вважається дорогим з обчислювальної точки зору, оскільки він передбачає кілька ітерацій для отримання найкращої можливої точності. Ця техніка є технікою керованого машинного навчання, яка означає, що дані позначаються, а алгоритм вчиться передбачати результат на основі вхідних даних. Алгоритм також добре працює, навіть якщо навчальні дані великі та містять зашумлені значення;

– SVM — це методика під наглядом. Цей алгоритм використовується як для класифікаційних, так і для регресійних досліджень. У цьому алгоритмі дані відображаються в n-вимірному просторі (використовуються координати). SVM можна класифікувати на лінійні та нелінійні типи.

Для прогнозування виникнення захворювань ми використовували «Набір даних про захворювання» UCI. Цей набір даних складається з 13 медичних прогностичних ознак і однієї цільової функції. Набір даних містить 1303 екземпляри та 75 атрибутів.[3]

У цій роботі використовуються алгоритми логістичної регресії, KNN, SVM, а точність обчислюється за допомогою перехресної перевірки з коефіцієнтом cv 10.

Таблиця 1 – Графік точності алгоритмів при різних захворюваннях

Захворювання	Логістична регресія	SVM	KNN	Найкраща точність
Хвороба серця	82.50	78.87	83.84	83.84
Діабет	77.60	75.64	75.52	77.60
Рак	90.55	91.38	92.55	94.55

На закінчення ми дослідили використання різних алгоритмів машинного навчання, включаючи логістичну регресію, машину опорних векторів (SVM) і K-Nearest-Neighbors (KNN), для прогнозування захворювань. Після оцінки продуктивності кожного алгоритму ми виявили, що KNN є найбільш підходящим алгоритмом. KNN продемонстрував вищу точність і нижчий рівень помилок порівняно з іншими алгоритмами в нашому експерименті

Список використаних джерел:

1. Sneha Grampurohit (2023 March 25). *Disease Prediction using Machine Learning Algorithms*. <https://ieeexplore.ieee.org/document/9154130/metrics#metrics>
2. Rameez Bhajjee (2023, 25 березня). *Medical Disease Prediction using Machine Learning Algorithms*. <https://www.ijraset.com/research-paper/medical-disease-prediction-using-ml-algorithms>
3. Jonathan G. Richens (2023, 25 березня). *Improving the accuracy of medical diagnosis with causal machine learning*. <https://www.nature.com/articles/s41467-020-17419-7>