

ПОРІВНЯЛЬНА ХАРАКТЕРИСТИКА МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ДІАБЕТУ 2-ГО ТИПУ

Король А. С., Руденко Н. М.

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Україна, 03056, м.Київ, пр-т Перемоги 37,

e-mail: akorolq@gmail.com

Стрімкий розвиток технологій та зростання популяції призводить до нових методів так технік діагностування хвороб. Наука про дані вивчає шаблони та закономірності, які є в наявності у кожного індивіда з певною хворобою.

Дана робота орієнтована на вивчення даних щодо діабету 2-типу, який займає 90% випадків діабету. Якщо людина має генетичну схильність до діабету, то надмірна вага слугує тригером та призводить до захворювання. Характерним показником є високий рівень глюкози в крові за умови відторгнення інсуліну та його нестачі в організмі. Рівень захворювання зростає паралельно зі збільшення ваги у дорослої частини населення. Підвищений рівень цукру в крові протягом довго часу призводить до захворювань серця, інсульту, діабетичної ретинопатії, яка вражає органи зору, ниркової недостатності, що веде до необхідності діалізу, а недостатній кровообіг в кінцівках може призвести до їх ампутації [1].

Мета роботи проаналізувати різні методи машинного навчання для передбачення діабету 2-го типу.

У даній роботі було проаналізовано методи машинного навчання та виконана їх порівняльна характеристика за даними індіанського народу Піма, що населяє південь штату Аризона (США). База даних знаходиться у вільному доступі на UCI Machine Learning Repository та є у переліку літератури [2]. Класифікація здійснювалася за 8-ми показникам, а саме: число вагітності як “Pregnancies”, рівень концентрації глюкози у плазмі крові як “Glucose”, значення діастолічного артеріального тиску як “BloodPressure”, значення товщини шкіри в області трицепсу як “SkinThickness”, значення 2-ох годинного сироваткового інсулін як “Insulin”, індекс маси тіла як “BMI”, значення функції спадковості як “DiabetesPedigreeFunction” та вік як “Age”. База даних включає 768 жінок. Значення “Outcome” показує чи пацієнтка була діагностована з діабетом 1, та без – 0. Задача зводиться до контрольованої бінарної класифікації.

Робота включає етап підготовки даних для обробки (зменшення розмірності даних, усунення артефактів у даних, вибір найбільш вагомих показників та маніпуляції з даними), етап безпосередньої побудови моделі для класифікації, візуалізація отриманих результатів.

У дані роботі було реалізовано такі методи: логістична регресія, метод k -найближчих сусідів, наївний баєсів класифікатор, метод опорних векторів, метод випадкового лісу, метод регресора дерева рішень [3-4].

Табл.1 - Аналіз методів машинного навчання

Назва методу	Точність (accuracy)
Логістична регресія	75.56%
Метод k -найближчих сусідів	74.10%
Наївний баєсів класифікатор	73.95%
Метод опорних векторів	75.72%
Метод випадкового лісу	72.30%
Метод регресора дерева рішень	74.75%

Отримавши значення точності, спробували поліпшити результати. Оокремо працюючи з методом опорних векторів [5] та функціями підбору параметрівм, було досягнуто точності 77.51%. Подальші дослідження включають застосування інших технік обробки даних, застосування алгоритмів, що не були протестовані у цій роботі. Процес тренування моделі машинного навчання вимагає постійної роботи з даними.

Результати роботи дозволяють на ранніх етапах сфокусуватися на стилі життя, а саме належній дієті та фізичним вправам, зниженню серцево судинних факторів ризику та стабілізування рівню глюкози в крові.

ЛІТЕРАТУРА

1. *Leonid Poretsky*; Principles of diabetes mellitus (вид. 2nd). Springer, New York: 2009, ISBN 978-0-387-09840-1, p. 3.
2. [Електронний ресурс] – Режим доступу:Database: <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
3. [Електронний ресурс] – Режим доступу:Code source: <https://github.com/qkorol/Pima-Indians-Diabetes-Project-DS-ML->
4. *Лавренюк М.С., Новіков О.М.* Огляд Методів Машинного Навчання Для Класифікації Великих Обсягів Супутникових Даних, System Research & Information Technologies, 2018, ISSN 1681–6048, p. 52-72
5. *Ingo Steinwart, Andreas Christmann.* Support Vector Machines, Springer-Verlag, New York: 2008, p. 602.