

УДК 004

Індик А.А.¹, Леошенко С.Д.²

¹ студ. гр. КНТ-214м НУ «Запорізька політехніка»

² доц. НУ «Запорізька політехніка»

РОЗРОБКА ТА ДОСЛІДЖЕННЯ СИСТЕМИ ПРОГНОЗУВАННЯ РИЗИКІВ ЗАХВОРЮВАНЬ НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Об'єкт дослідження – процес прогнозування захворювань методами машинного навчання.

Предмет дослідження – методи та засоби для прогнозування ризиків захворювань.

Мета роботи – підвищення точності прогнозування захворювань методами машинного навчання.

Матеріали, методи та технічні засоби: структурне та об'єктно-орієнтоване програмування, клієнт-серверна архітектура, мови розмітки HTML і CSS, мови програмування Python, JavaScript, вибірка даних із медичними показниками для навчання моделей машинного навчання, персональний комп'ютер з процесором Intel Core I7 під управлінням операційної системи Microsoft Windows 11.

У процесі виконання роботи було проведено комплексне дослідження проблеми прогнозування ризиків захворювань на основі медичних показників та проаналізовано можливості використання методів машинного

навчання у медичних інформаційних системах. Розглянуто предметну область медичної діагностики, окреслено основні труднощі, пов'язані з обробленням значних обсягів даних і необхідністю підтримки прийняття рішень лікарями.

Огляд існуючих наукових досліджень показав, що для задач прогнозування ризиків захворювань найчастіше застосовують логістичну регресію, дерева рішень, випадкові ліси, SVM, наївний баєсівський класифікатор й інші ансамблеві підходи. У багатьох роботах продемонстровано, що комбіновані або ансамблеві моделі забезпечують помітне підвищення точності порівняно з окремими базовими алгоритмами.

У роботі розроблено архітектуру вебзастосунку, що включає вебінтерфейс користувача, серверну частину, модулі машинного навчання та базу даних. Побудовано структурну схему системи, UML-діаграму компонентів та ER-діаграму таблиць, що забезпечило чітке уявлення про внутрішню організацію застосунку та взаємодію його складових.

У ході роботи створено повноцінний вебзастосунок для прогнозування ризиків захворювань, що поєднує серверну частину на Flask, клієнтський інтерфейс на основі HTML-шаблонів та інтегровані моделі машинного навчання. Реалізовано механізми автентифікації користувачів, обробки введених медичних показників, формування прогнозів та збереження історії результатів у базі даних SQLite. Окремий модуль машинного навчання забезпечує підготовку даних, навчання та оновлення моделей, а також автоматичну реєстрацію отриманих параметрів у системі. Для розв'язання задачі прогнозування впроваджено та налаштовано декілька алгоритмів, а саме Support Vector Machine, Naïve Bayes, Random Forest, XGBoost, Logistic Regression та зважену ансамблеву модель, що дозволило створити функціональну, інтегровану програмну систему, придатну для подальшого використання й розширення.

Створено окремий модуль тестування, який оцінює отримані моделі, на спільному тестовому наборі та порівнює за низкою метрик: точністю, повнотою, точністю позитивного прогнозу, F1-мірою, AUC ROC та AUC PRC. За результатами експериментів встановлено, що найвищу якість показали моделі XGBoost та Random Forest, які не розглядалися як ключові на етапі відбору методів, але продемонстрували найкращі результати серед усіх. Власна реалізація зваженого ансамблю забезпечила стабільно високі результати, близькі до найсильніших окремих моделей, і може бути корисною у ситуаціях, де окремі алгоритми дають нестійкі прогнози.

У результаті виконаної роботи створено повноцінний вебзастосунок, який забезпечує введення медичних показників користувачем, їх оброблення, аналіз та прогнозування ризику захворювання із застосуванням моделей

машинного навчання, що дає змогу визначати його стан і підтримувати процес моніторингу здоров'я.

Система може застосовуватися у медичних установах як інструмент підтримки прийняття рішень або використовуватися пацієнтами для самостійного моніторингу свого стану.

Галузь використання – медична діагностика та моніторинг стану здоров'я користувачів у медичних закладах лікарями або в домашніх умовах пацієнтами.