

ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ВИЗНАЧЕННЯ ЗАЛЕЖНОСТІ ДІАГНОЗУ НА ДІАБЕТ ВІД БІОЛОГІЧНИХ ПОКАЗНИКІВ ЛЮДИНИ

*Буткевич Микола Віталійович *, студент групи 355*

Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ»

Робота присвячена актуальній проблемі – визначання діагнозу пацієнта за його біологічними даними. В даній роботі розглядається залежність біологічних даних і захворювання на діабет.

Ця тема актуальна, тому що, проблема встановлення попереднього діагнозу пацієнта, ще до безпосередніх аналізів, є дуже важливою у сфері медицини, але встановити залежність між біологічними показниками які не являються фактичною діагностикою і діагнозом практично не можливо. Саме тут на допомогу лікарям і пацієнтам повинна прийти автоматизована система прийняття рішень, яка на основі великого об'єму даних, а саме попередніх діагнозів інших пацієнтів, дає відповідь на питання про захворювання на діабет даного пацієнта за його біологічними показниками, з визначеною точністю.

Враховуючи медичні дані, які ми можемо зібрати про людей, ми повинні бути в змозі краще прогнозувати, наскільки ймовірно, людина страждає на виникнення діабету, а тому діяти належним чином, щоб допомогти. Ми можемо почати аналізувати дані та експериментувати з алгоритмами, які допоможуть нам вивчити виникнення діабету у індіанців Піми.

Кожен рядок містить в собі данні зібрані відносно конкретного пацієнта. Відповідно кожна колонка відповідає певній характеристиці:

- Pregnancies - Кількість вагітностей
- Glucose - Концентрація глюкози в плазмі а 2 години при пероральному тесті на толерантність до глюкози
- BloodPressure - Діастолічний артеріальний тиск у мм рт.ст.
- SkinThickness - Товщина складової шкіри трицепса в мм
- Insulin - 2-годинний сироватковий інсулін в мю / мл
- BMI - Індекс маси тіла, виміряний як вага в кг / (висота в м) ²
- DiabetesPedigreeFunction - Функція роду діабету
- Age - Вік у роках
- Outcome - Діагноз, де 1 означає, що пацієнт хворий, 0-здоровий

Для аналізу вхідних даних обрахуємо кореляційну матрицю



Рисунок 1. Кореляційна матриця побудована по критерію Пірсона

Проводячи додаткові дослідження було зроблено висновки та оптимізовано данні для побудови моделі машинного навчання.

XGBoost - це оптимізована розподілена градієнтно-посилювальна бібліотека, розроблена, щоб бути високоефективною, гнучкою та портативною. Він реалізує алгоритми машинного навчання в рамках Gradient Boosting. XGBoost забезпечує паралельне підсилення дерев (також відоме як GBDT, GBM), яке швидко та точно вирішує багато проблем науки про дані. Той самий код працює на основних розподілених середовищах і може вирішити проблеми, що перевищують мільярди прикладів.

Результати роботи програмного забезпечення мають достатньо високу точність в середньому в діапазоні 70-80%. Це дозволяє передбачити розвиток цукрового діабету або діагностувати захворювання зібравши лише основні показники пацієнта з достатньо високою ймовірністю. В разі реалізації аналогічного програмного забезпечення для промислових компаній з достатньо великим набором статистичних даних можна використовувати результати, що ми отримали в процесі виконання цієї роботи. Також в такій ситуації можна проводити аналіз схильності пацієнта до інших хвороб після аналогічних досліджень на залежність від деяких характеристик або амбулаторних досліджень пацієнта.