

ЗАСТОСУВАННЯ ІНСТРУМЕНТІВ DATA MINING ДЛЯ ПРОГНОЗУВАННЯ ВИЖИВАННЯ ПАЦІЄНТІВ

Литвин І.І., Поворознюк А.І.

*Національний технічний університет
«Харківський політехнічний інститут», м. Харків*

Сучасні електронні медичні картки містять різноманітні клінічні дані, обробка яких за допомоги інструментів Data Mining дозволяє виявити приховані закономірності та кореляції. Прикладом цього є прогнозування виживання пацієнтів на основі їх клінічних даних.

В даній роботі було проаналізовано набір даних 299 пацієнтів із серцевою недостатністю [3]. Для цього було використано програмний інструмент Orange [1] та наступні класифікатори: Logistic Regression, k-Nearest Neighbors, Decision Trees, Support Vector Machine, Naive Bayes, Random Forest, Gradient Boosting. Для отримання достовірних результатів прогнозування виживання пацієнтів всі методи були перехресно валідовані з 10-кратним повторенням. Загальний процес класифікації наведено на рис. 1.

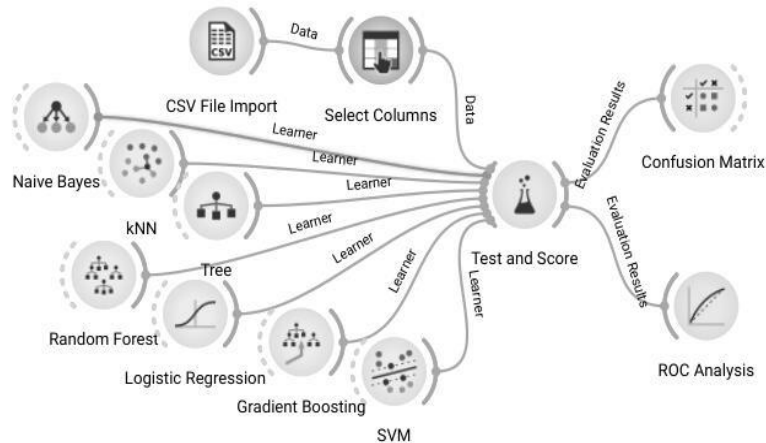


Рис.1 – Схема процесу дослідження

Результатами прогнозування показали, що найкращу точність (0,837) та AUC (0,906) отримав Naive Bayes класифікатор. Найгірше з прогнозуванням впоралась модель з k-Nearest Neighbors, отримавши точність (0.422) та AUC (0.557). Отримані результати було порівняні з результатами суміжних досліджень [2, 3] на цьому наборі даних.

Експеримент довів успішність та актуальність використання Data Mining, як допоміжного інструменту для лікарів при прогнозуванні того, чи виживе пацієнт із серцевою недостатністю.

Література:

1. Orange: Data Mining Toolbox in Python / Demsar J, Curk T, Erjavec A and etc. // Journal of Machine Learning Research – 14(71) – 2013. – pp. 2349–2353.
2. Survival analysis of heart failure patients: A case study / Ahmad T, Munir A, Bhatti SH, and etc. // PLoS ONE – 12(7) – 2017. – e0181001. <https://doi.org/10.1371/journal.pone.0181001>
3. Chicco, D., Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. // BMC Med Inform Decis Mak – 20(16) – 2020. <https://doi.org/10.1186/s12911-020-1023-5>