УДК 519.254:004.8:681.3

Труханська В. О., здобувачка 1 курсу
ОС «Магістр» ФЗ Комп'ютерні науки,
Ніколюк П. К., д-р фіз.-мат. наук, професор
кафедри інформаційних технологій

ДОСЛІДЖЕННЯ ВПЛИВУ АНСАМБЛЕВИХ АЛГОРИТМІВ НА ПРОДУКТИВНІСТЬ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ

Донецький національний університет імені Василя Стуса, м. Вінниця

Методи статистичного навчання відіграють фундаментальну роль у створенні моделей для класифікації, прогнозування та виявлення закономірностей у даних. Однак окремі моделі, зокрема дерева рішень чи лінійні класифікатори, часто мають обмежену узагальнювальну здатність, особливо у випадку складних або шумних вибірок. Для подолання цих обмежень застосовують ансамблеві алгоритми, що поєднують декілька моделей у єдину систему з підвищеною надійністю та точністю [1].

Ансамблеві методи формують клас моделей, продуктивність якого перевищує можливості базових алгоритмів завдяки зменшенню розбіжностей у помилках та оптимізації процесу навчання. Підходи Bagging, Boosting та Random Forest є найпоширенішими й демонструють ефективність у задачах класифікації в різних галузях – від медицини до фінансового аналізу [1]. Їх спільна ідея полягає в комбінуванні декількох слабких або сильних класифікаторів для підвищення точності та стабільності результатів [1].

Bagging (Bootstrap Aggregating) передбачає побудову незалежних моделей на різних підвибірках даних, сформованих методом випадкової вибірки з поверненням. Такий підхід знижує дисперсію класифікатора завдяки усередненню результатів. Найчастіше в цьому методі застосовують дерева рішень, оскільки вони є високовариативними щодо змін у даних [2].

Random Forest поєднує ідею Bagging із додатковою випадковістю під час вибору ознак, що зменшує кореляцію між деревами та підвищує стійкість моделі до перенавчання. Завдяки такій декореляції Random Forest зберігає баланс між точністю, інтерпретованістю та стійкістю до шуму [2].

Boosting працює за іншим принципом: моделі будуються послідовно, і кожна наступна фокусується на помилках попередніх. Це зменшує зміщення та дає змогу виявляти складні закономірності. Популярні реалізації – AdaBoost, Gradient Boosting, XGBoost – забезпечують високу ефективність, але є більш чутливими до шуму та вимагають ретельного налаштування параметрів [2].

Для об'єктивного оцінювання продуктивності ансамблевих алгоритмів аналізують точність класифікації, здатність до узагальнення та стійкість до варіативності даних [3]. Узагальнені аналітичні результати свідчать, що Bagging підвищує стабільність і знижує дисперсію, Random Forest демонструє покращену узагальнювальну здатність завдяки випадковому відбору ознак, тоді як Boosting забезпечує найвищі показники точності, але за умови коректної підготовки даних [3].



Рисунок 1 – Порівняння продуктивності ансамблевих методів

Порівняння продуктивності трьох методів наведено на рис. 1. Воно демонструє поступове зростання ефективності від Bagging до Random Forest і далі до Boosting, що узгоджується з теоретичними особливостями цих моделей. Найвищі показники Boosting пояснюються його здатністю адаптивно зменшувати зміщення, тоді як Random Forest оптимально поєднує точність та стійкість до шуму.

Висновки. Ансамблеві алгоритми є одним із найефективніших інструментів сучасного статистичного навчання. Bagging знижує дисперсію моделей, Boosting покращує здатність до навчання на складних залежностях, а Random Forest забезпечує збалансований підхід до точності й стабільності. Аналітичні та теоретичні результати підтверджують їх ефективність у задачах класифікації, особливо за умов шумності чи високої складності даних. Подальші дослідження можуть бути спрямовані на оптимізацію гіперпараметрів і розробку гібридних ансамблевих підходів.

Список використаних джерел

1. Талах М. В. Ансамблеві архітектури та технології аналізу даних. Чернівці: Технодрук, 2023. 246 с.

2. Kotsiantis S. Combining bagging, boosting, rotation forest and random subspace methods. *Artificial Intelligence Review.* 2011. Vol. 35. P. 223–240. URL: https://doi.org/10.1007/s10462-010-9192-8 (дата звернення: 23.11.2025).

3. Austin P. Using ensemble-based methods for directly estimating causal effects: an investigation of tree-based g-computation. *Multivariate Behavioral Research.* 2012. Vol. 47, № 1. P. 115–135.