

АНАЛІЗ ЕФЕКТИВНОСТІ АНСАМБЛЕВИХ МЕТОДІВ ТА СТАТИСТИЧНИХ МОДЕЛЕЙ У СИСТЕМАХ ФІНАНСОВОГО ПЛАНУВАННЯ

Вінницький національний технічний університет

Анотація

Запропоновано підхід до побудови інтелектуальної системи персонального фінансового планування, що дозволив вирішити проблему обробки гетерогенних даних та виявлення нелінійних залежностей у фінансовій поведінці користувачів. Розглянуто застосування методів градієнтного бустингу для задач регресійного аналізу витрат та бінарної класифікації досяжності цілей. Обґрунтовано переваги використання ансамблевих моделей над класичними статистичними алгоритмами та методами бегінгу в умовах незбалансованих вибірок.

Ключові слова: інформаційна технологія, градієнтний бустинг, регресійний аналіз, класифікація, гетерогенні дані, фінансове планування.

Abstract

An approach to building an intelligent personal financial planning system is proposed, which solves the problem of processing heterogeneous data and identifying nonlinear dependencies in users' financial behavior. The application of gradient boosting methods for regression analysis of costs and binary classification of goal achievability is considered. The advantages of using ensemble models over classical statistical algorithms and bagging methods in conditions of unbalanced samples are substantiated.

Keywords: information technology, gradient boosting, regression analysis, classification, heterogeneous data, financial planning.

Вступ

Існуючі засоби автоматизації фінансового планування переважно базуються на детермінованих алгоритмах та евристичних правилах. Такий підхід має суттєві обмеження при роботі зі складними патернами фінансової поведінки, що характеризуються високою варіативністю та наявністю прихованих залежностей. Класичні лінійні статистичні моделі є недостатньо гнучкими при обробці гетерогенних даних, які містять як кількісні, так і категоріальні ознаки [1]. Актуальним науковим завданням є дослідження ефективності застосування ансамблевих методів машинного навчання, зокрема градієнтного бустингу, для підвищення адаптивності та здатності прогнозування в системах персонального фінансового планування.

Результати дослідження

У роботі досліджено дві взаємопов'язані задачі персонального фінансового планування: оцінювання доцільності витрат та прогнозування ймовірності досягнення фінансової мети.

Для вирішення задачі оцінки доцільності витрат запропоновано перехід від дискретної класифікації до регресійного аналізу з подальшою інтерпретацією результатів. В якості базового алгоритму обґрунтовано вибір методу LightGBM. Гіпотеза дослідження полягала у тому, що архітектурні особливості даного методу, зокрема стратегія росту дерев по листках, забезпечують більш ефективну мінімізацію функції втрат порівняно з алгоритмами, що використовують ріст по рівнях, або класичними лінійними моделями регресії. Додатковою перевагою обраного підходу виявлено алгоритм автоматично опрацьовувати категоріальні змінні без необхідності їх попереднього перетворення у розріджені матриці великої

розмірності, що оптимізує обчислювальний процес [2].

Для задачі прогнозування досяжності фінансової мети проведено порівняльний аналіз ефективності паралельних та послідовних ансамблевих методів. Специфіка предметної області передбачає роботу з незбалансованими класами даних. Теоретично обґрунтовано, що алгоритми бустингу, які базуються на ітеративному виправленні помилок попередніх моделей ансамблю та послідовному зменшенні зміщення, є більш ефективними для фінансових задач порівняно з методами бегінгу, спрямованими переважно на зменшення дисперсії [3].

В ході дослідження проведена апробація підтвердила, що моделі на базі градієнтного бустингу забезпечують кращу апроксимацію нелінійних залежностей у фінансових даних порівняно з лінійними моделями та поодинокими деревами рішень.

Висновки

Підтверджено доцільність використання алгоритмів градієнтного бустингу при розробці систем персонального фінансового планування. Встановлено, що перехід до регресійної моделі оцінки витрат дозволяє отримати неперервну метрику доцільності, яка є більш інформативною для користувача, ніж бінарна оцінка.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Рассел, С., & Норвіг, П. (2020). Штучний інтелект: сучасний підхід (4-те вид.). Видавництво Педагог.
2. Ke G., Meng Q., Finley T. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems. 2017. Vol. 30.
3. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016.

Щетинін Данило Сергійович – студент групи ІКН-24м, факультет інтелектуальних інформаційних технологій та автоматизації Вінницький національний технічний університет, Вінниця, e-mail: mail.redoge@gmail.com

Прудивус Олександр Сергійович – студент групи ІКН-24м, факультет інтелектуальних інформаційних технологій та автоматизації Вінницький національний технічний університет, Вінниця, e-mail: jir4fchick@gmail.com

Колесницький Олег Костянтинович – професор, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Shchetinin Danylo - student of group IKN-24m, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: mail.redoge@gmail.com

Prudyvus Oleksandr - student of group IKN-24m, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: jir4fchick@gmail.com

Kolesnytskyi Oleh - Professor, Professor of the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia.