

ОПТИМІЗАЦІЯ ВИТРАТ НА МАРКЕТИНГОВУ ДІЯЛЬНІСТЬ БАНКІВСЬКИХ УСТАНОВ ЗА ДОПОМОГОЮ МОДЕЛЕЙ НА ОСНОВІ АНСАМБЛІВ

Маркетингова діяльність банківських установ у сучасному світі є ключовим складником стратегії їх розвитку та конкурентоспроможності. Залучення нових клієнтів, утримання існуючих і впровадження інноваційних фінансових продуктів стає неможливим без ефективної маркетингової стратегії. В умовах зростання конкуренції на ринку банківських послуг оптимізація маркетингових витрат стає стратегічним завданням банків.

Метою дослідження є використання теоретичних знань для розроблення ефективних моделей на основі ансамблів із використанням налаштування параметрів оптимізації витрат на маркетингову діяльність банківських установ, що дасть змогу зменшити маркетингові витрати та підвищити результативність маркетингової діяльності.

У статті досліджено підхід до оптимізації витрат на маркетингову діяльність банківських установ із використанням кількох типів ансамблевих моделей, зокрема випадковий ліс, посилення градієнта, екстремальне посилення градієнта. Автори досліджують маркетингову діяльність банків і як ефективно використовувати моделі на основі ансамблю у середовищі RStudio для досягнення максимальних результатів, використовуючи набір даних із даними маркетингової кампанії португальського банку. Детальний аналіз включає вивчення чинників, зібраних банком про клієнтів, і оцінку їхнього впливу на остаточний аналіз рішення клієнта. Також у процесі дослідження для покращення результатів моделювання використовуються методи undersampling, масштабування та пошуку оптимальної кількості ознак.

У результаті дослідження встановлено, що ефективність маркетингової кампанії під час використання моделей фактично в два рази вище, ніж без них.

Результати дослідження є цінним посібником для банківських установ щодо оптимізації маркетингової діяльності. Використання модельних ансамблів може сприяти підвищенню ефективності маркетингових кампаній, зниженню витрат і підвищенню конкурентоспроможності на ринку фінансових послуг.

Ключові слова: маркетингова діяльність, банківські установи, ансамблі моделей, оптимізація витрат.

YU.V. KLEBAN, YU.O. DENYSIUK
National University "Ostroh Academy"

OPTIMIZATION OF COSTS FOR MARKETING ACTIVITIES OF BANKING INSTITUTIONS WITH THE HELP OF MODELS BASED ON ENSEMBLES

Marketing activity of banking institutions in the modern world is a key component strategy of their development and competitiveness. Attracting new customers, maintaining existing ones and introducing innovative financial products becomes impossible without an effective marketing strategy. In the conditions of growing competition in the market of banking services, optimization of marketing costs becomes a strategic task for banks

The purpose of the study is to use theoretical knowledge to develop effective models based on ensembles using parameter tuning to optimize the costs of marketing activities of banking institutions, which will reduce marketing costs and increase the effectiveness of marketing activities.

The article examines an approach to optimizing the costs of marketing activities of banking institutions using several types of ensemble models. In particular, random forest, gradient boosting, extreme gradient boosting. The authors investigate the marketing activities of banks and how to effectively use ensemble-based models in RStudio environment to achieve maximum results, using a dataset with data from a Portuguese bank's marketing campaign. Detailed analysis includes examining factors collected by the bank about customers and evaluating their impact on the final analysis of the customer's decision. Also, in the research process, the methods of under-sampling, scaling and finding the optimal number of features are used to improve the modeling results.

As a result of the study, it was found that the effectiveness of the marketing campaign when using models is actually 2 times higher than without them.

The results of the study are a valuable guide for banking institutions to optimize their marketing activities. The use of model ensembles can contribute to increasing the effectiveness of marketing senses, reducing costs and increasing competitiveness in the financial services market.

Key words: marketing activity, banking institutions, model ensembles, cost optimization.

Постановка проблеми

Банки як ключові установи фінансової системи відіграють важливу роль у стимулюванні економічного зростання. Вони забезпечують кредитування для підприємств та населення, сприяючи розвитку бізнесу та стимулюючи виробництво й споживання.

Ефективна маркетингова діяльність банків допомагає залучати нових клієнтів, розширювати клієнтську базу та забезпечувати доступ до фінансових послуг для більш широкого кола суспільства. Оптимізація витрат на маркетинг у банківській сфері не лише сприяє підвищенню їхньої конкурентоспроможності, а й дає змогу ефективніше використовувати ресурси для фінансування економічного розвитку.

Такі інноваційні підходи, як використання ансамблевих моделей, можуть допомогти банкам покращити свою роль у сприянні стабільності та зростанню економіки.

Аналіз останніх досліджень та публікацій

Існує значна кількість досліджень, де науковці намагаються визначити найефективніший метод для прогнозування поведінки клієнтів банків із використанням їхньої персональної інформації. У статті *Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree* авторства Fereshteh Safarkhani та Sérgio Moro [1] розглядається використання дерев рішень для поліпшення точності передбачення поведінки клієнтів банку. Результати дослідження показали, що використання дерева рішень дало змогу поліпшити точність прогнозування на 10% порівняно з традиційними методами.

У статті *Bank Direct Marketing Analysis Based on Ensemble Learning* автори Ruiting Hao, Xiaoqian Xia, Siyi Shen та Xiaorong Yang [2] досліджують ефективність методів прямого маркетингу банківських послуг через аналіз даних клієнтів. У роботі використовуються ансамблеві методи машинного навчання, такі як Random Forest та Gradient Boosting, для покращення точності. Автори зазначають, що застосування ансамблевих методів машинного навчання дає змогу підвищити точність прогнозування поведінки клієнтів та зменшити ризик неправильних рішень у прямому маркетингу.

Мета дослідження

Мета дослідження – із використанням теоретичних знань розробити ефективні моделі на основі ансамблів із використанням тюнінгу параметрів для оптимізації витрат на маркетингову діяльність банківських установ, що дасть змогу знизити витрати на маркетинг та збільшити ефективність маркетингових заходів.

Виклад основного матеріалу дослідження

Банки є важливими гравцями у світовій економіці та конкурентними підприємствами на ринку фінансових послуг. Вони використовують різні стратегії для залучення клієнтів і підвищення прибутковості, такі як маркетингові кампанії, скорочення витрат та оптимізація бізнес-процесів. У сучасному світі банківські установи стають усе більш цифровими, використовуючи нові технології, такі як мобільні додатки та онлайн-банкінг, щоб забезпечити зручні та доступні послуги для клієнтів.

Маркетингова діяльність у банківській галузі – це низка заходів і стратегій, прийнятих банківською індустрією для залучення клієнтів, збільшення продажів банківського бізнесу та підвищення рівня задоволеності клієнтів.

Основними завданнями маркетингової кампанії банку є збільшення клієнтської бази, збільшення кількості та якості послуг, що надаються, підвищення лояльності клієнтів і залучення нових клієнтів.[5]

Банки витрачають значну частину своїх бюджетів на маркетингові кампанії, щоб залучити нових клієнтів та утримати існуючих. Витрати на маркетинг можуть включати різні типи

реклами, зв'язки з громадськістю, планування заходів, Інтернет-маркетинг, дослідження поведінки споживачів тощо.

Для оптимізації маркетингових витрат банки можуть використовувати аналітику даних та інструменти машинного навчання для прогнозування поведінки клієнтів і підбору персоналізованих пропозицій, що зменшить витрати на рекламні кампанії у цілому та підвищить ефективність спілкування з окремими клієнтами.

Моделі ансамблю є потужним інструментом у машинному навчанні, який дає змогу поєднувати прогнози кількох моделей для отримання більш точних і стабільних прогнозів. У R є кілька бібліотек для побудови ансамблевих моделей, таких як *caret*, *mlr*, *SuperLearner* та схожі до них.

Ансамблеві моделі можуть бути корисним інструментом для зменшення помилок передбачення та підвищення точності моделей машинного навчання. Залежно від конкретного завдання та даних можна використовувати різні типи моделей ансамблю, наприклад *Random Forest*, *Gradient Boosting*, *Extreme Gradient Boosting* та ін. Також для поліпшення результатів моделювання доцільно скористатися тюнігом гіперпараметрів для моделей [10].

Random Forest (випадковий ліс) – це алгоритм машинного навчання, що використовується для задач класифікації та регресії. Він заснований на ансамблі дерев рішень, де кожне дерево будується на випадковій підмножині навчальних даних та випадкових підмножинах характеристик [6].

Градiєнтний бустинг використовує дерева рішень як базові моделі, тому його також називають градiєнтним бустингом дерев рішень (*Gradient Boosted Trees*). Цей алгоритм будує послідовність дерев, кожне з яких виправляє помилки попереднього дерева. Головна ідея полягає у тому, щоб на кожному кроці навчання алгоритм використовував градiєнт утрат для навчання нової моделі [6].

XGBoost (*eXtreme Gradient Boosting*) – це бібліотека машинного навчання, яка використовує градiєнтний бустинг для вирішення задач класифікації та регресії. Основна ідея *XGBoost* полягає у тому, щоб послідовно додавати до моделі нові дерева рішень, які виправляють помилки, зроблені попередніми деревами. Кожне нове дерево намагається зменшити помилки, які залишилися після попередніх дерев, і додається до моделі з вагою, яка динамічно залежить від ефективності моделі [7].

Є багато різних аспектів, які слід урахувати під час розроблення машинного навчання, наприклад розмір даних, тип даних, призначення моделі та налаштування її гіперпараметрів, деталі проблеми тощо.

Налаштування гіперпараметрів – це завдання знаходження оптимальних гіперпараметрів для алгоритму навчання для певного набору даних і, зрештою, поліпшення продуктивності моделі. Існує три основні методи налаштування/оптимізації гіперпараметрів: метод пошуку в сітці, метод випадкового пошуку, метод інформованого пошуку (найпопулярніший метод – байєсівська оптимізація).

Перед початком побудови власне моделей було проаналізовано датасет та його складники. Дані стосуються прямої маркетингової кампанії (телефонних дзвінків) португальського банку. Центр обслуговування клієнтів банку телефонує клієнтам для реклами свого строкового депозитного продукту. Вони збирають основну інформацію про клієнтів та їхні рішення: чи підписалися на строковий депозит (так чи ні). Мета класифікації – передбачити їхні рішення після дзвінків.

Набір даних містить 41 188 рядків і 21 стовпець, із яких 10 числових і 11 категоріальних [9].

Після дослідження було виявлено, що бракує деяких даних, позначених невідомими. Оскільки набір даних досить великий, можна просто видалити їх із моделювання. Після видалення, зважаючи на те, що набір даних незбалансований (тобто вибірку залежної змінної,

що відображає рішення клієнтів, становлять два варіанти і один дуже сильно переважає над іншим), до класу no належить 26 629 спостережень, а до класу yes – 3 859 спостережень, його було вирішено збалансувати методом under-sampling.

Under-sampling – це метод балансування даних, який використовується в задачах класифікації для вирішення проблеми незбалансованості класів. Незбалансованість класів виникає, коли кількість прикладів одного класу значно переважає кількість прикладів іншого класу. У методі under-sampling видаляються деякі приклади з більш численного класу, щоб зрівняти кількість прикладів у двох класах. Це зменшує дисбаланс, але також може призвести до втрати корисної інформації з даних [8].

Після поділу виконано розроблення функцій. По-перше, видаляються змінні відхилення, близькі до нуля, які є rdays і default. У цих даних є лише три спостереження, які раніше не виконувалися. По-друге, стандартизовано числову змінну, позначено кодуванням категорійних змінних, які мають упорядковані рівні і створено фіктивні змінні для тих, які не мають порядку.

У датасеті лишилося 16 незалежних змінних, тому далі був використаний метод рекурсивного усунення ознак [3], щоб визначити оптимальну кількість ознак для побудови моделей. Підбір оптимальних ознак було здійснено на базі моделі random forest та застосовано стратифіковану десятикратну перехресну перевірку.

Стратифікована 10-кратна перехресна перевірка (Stratified 10-fold cross-validation) є одним із методів оцінки продуктивності моделей машинного навчання. Вона використовується для оцінки якості моделі та перевірки її універсальності на незалежних даних.

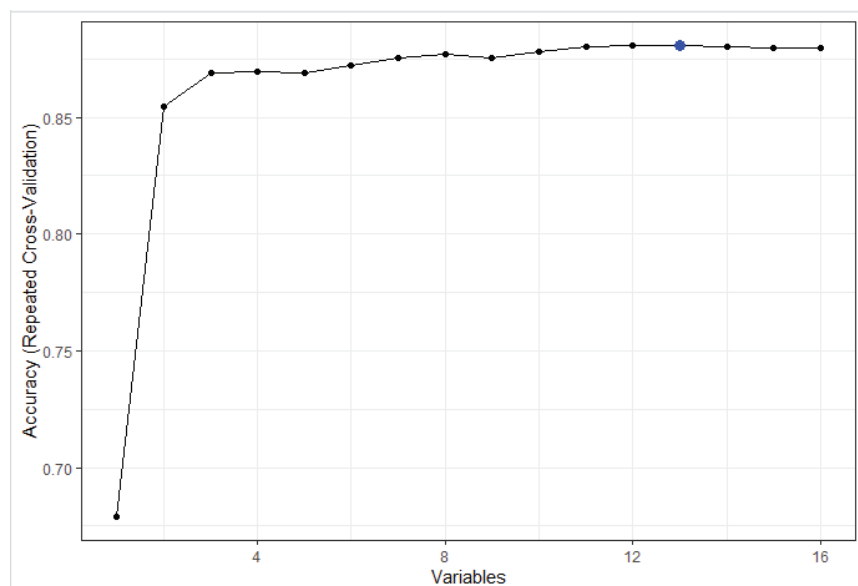


Рис. 1. Результат використання методу рекурсивного усунення ознак із відображенням коефіцієнта Accuracy

*Джерело: розроблено авторами

Отже, використаний метод підбору ознак порекомендував будувати моделі, користуючись тринадцятьма ознаками, оскільки значення Accuracy (рис. 1) та Карра (рис. 2) у цій точці максимальні. Зокрема, рекомендовано використати такі ознаки, як duration, euribor3m, month, rdays, contact, poutcome, age, job, previous, day_of_week, education, campaign, marital. Варто переглянути важливість цих ознак для моделювання та відкинути кілька, що мають найменший вплив.

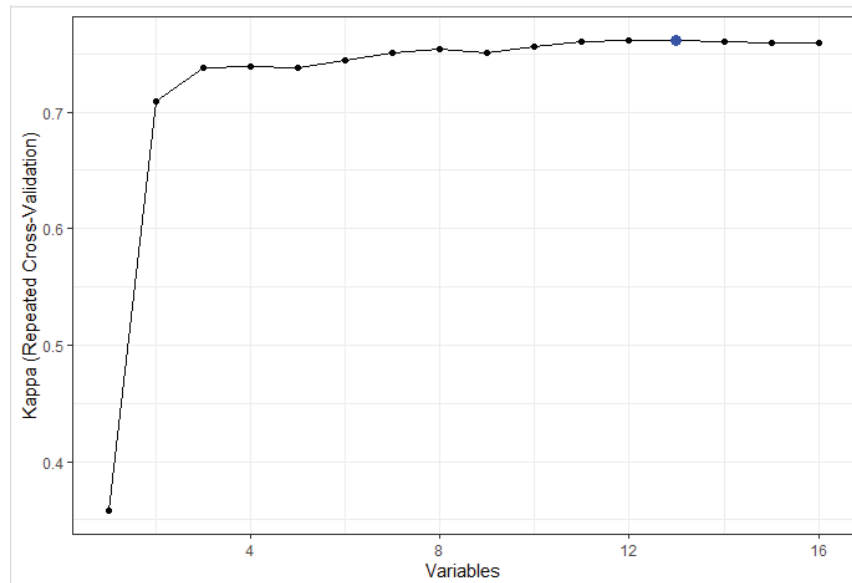


Рис. 2. Результат використання методу рекурсивного усунення ознак із відображенням коефіцієнта Карра

*Джерело: розроблено авторами

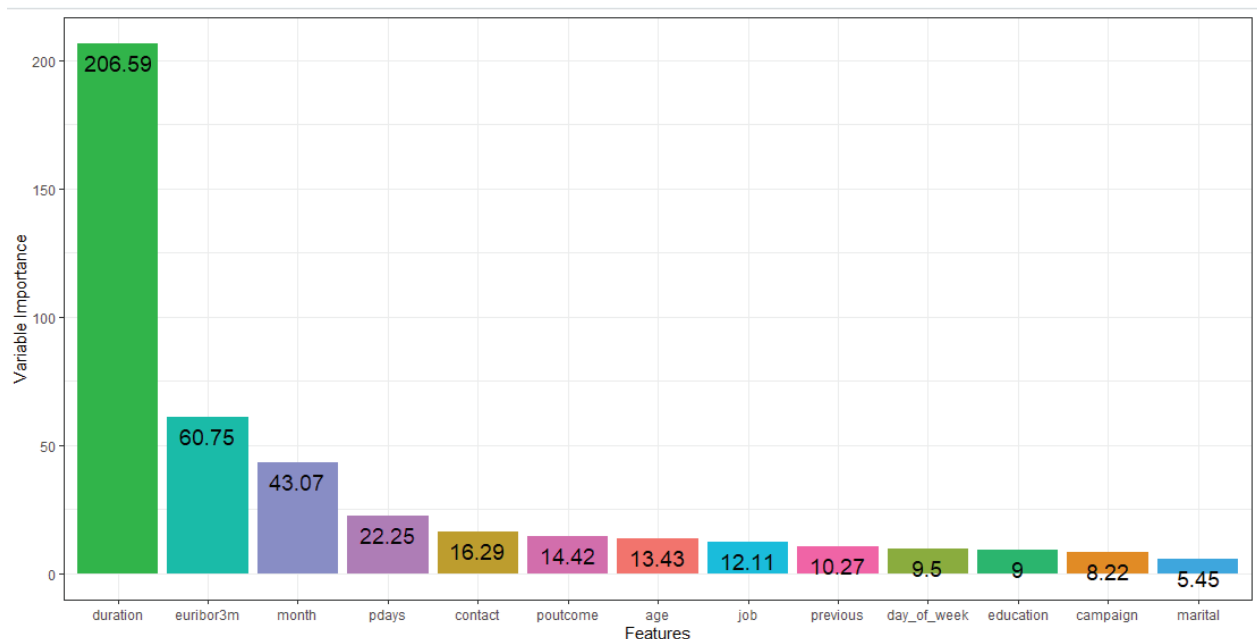


Рис. 3. Важливість ознак

*Джерело: розроблено авторами

Зважаючи на рис. 1–3, було вирішено відібрати для моделювання сім факторних ознак, що мають найбільший вплив на результуюче рішення клієнтів. Оскільки значення цих ознак знаходяться в занадто різних числових діапазонах, їх було нормалізовано методами центрування та скейлінгу.

Скейлінг у R – це процедура, яка використовується для нормалізації або стандартизації даних. Вона дає змогу привести значення змінних до одного масштабу, щоб полегшити їх порівняння та обробку.

Тепер, коли дані повністю готові, можна будувати моделі ансамблів. Збалансована точність прогнозу простої моделі random forest приблизно дорівнює 89,2%, для такої ж моделі з тюнінгом гіперпараметрів – 89,4%, для моделі градієнтного бустингу з тюнінгом параметрів це значення дорівнює 88,1%, а для моделі XGBoost з тюнінгом – 88,9%. Для фінального порівняння моделей та вибору найкращої було створено таблицю результатів моделювання.

Таблиця 1

Співставлення результатів моделювання

	Simple RF	Random.Forest	Basic.GBM	XGBoost
Test Accuracy	0.8924870	0.8944301	0.8808290	0.8873057
Balanced Accuracy	0.8924870	0.8944301	0.8809927	0.8885679

У табл. 1 видно, що за показником збалансованої точності найкращою виявилася модель Random Forest із тюнінгом параметрів. Усі інші методи досить близькі до неї та все ж не кращі. Зважаючи на це, далі проводиться обрахунок економічного ефекту на витрати банків на маркетингову діяльність від використання моделей.

Підрахування економічного ефекту було здійснено на основі матриць помилок моделей (Confusion Matrix) для моделі Random Forest із тюнінгом параметрів, оскільки її визнано найкращою (рис. 4).

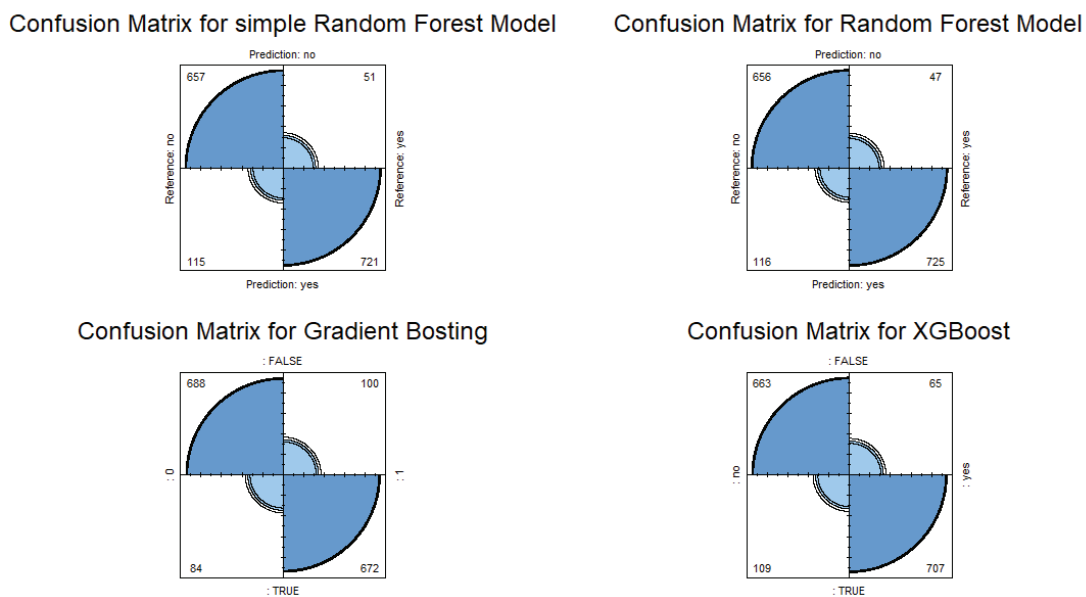


Рис. 4. Матриці помилок (Confusion Matrix) моделей

*Джерело: розроблено авторами

У матриці помилок дані можуть мати чотири варіації:

- True Positive – правильно передбачені позитивні відповіді.
- False Positive – ситуація, коли прогноз дав позитивну відповідь, а насправді вона була негативною.
- True Negative – правильно передбачені негативні відповіді.
- False Negative – ситуація, коли прогноз дав негативну відповідь, а насправді вона була позитивною [11].

Отже, передусім було підраховано умовні витрати на маркетингову діяльність банків без використання моделей. Було зроблено припущення, що під час проведення

опитування на кожного клієнта витрачається 100 умовних одиниць. За такої умови за невикористання моделей витрачається 154 400 умовних одиниць. Також нехай із кожної згоди клієнта на строковий депозит банк отримує 500 умовних одиниць. У такому разі дохід становитиме 386 000 одиниць. Звідси, варто обчислити ефективність такого підходу, поділивши доходи на витрати. Отже, коефіцієнт ефективності такого підходу за невикористання моделей становить 2,5. Чистий дохід за такого розвитку подій становить 231 600 умовних одиниць. Такі ж умови було застосовано для обрахунку ефективності використання моделей.

Для звичайної моделі Random Forest витрати становлять 83 600 умовних одиниць, а доходи – 360 500 умовних одиниць. Ефективність використання такого методу становить 4,3122, що майже в два рази краще, ніж за відсутності використання моделей. Звідси, чистий дохід від маркетингової кампанії становить 276 900 умовних одиниць.

Під час використання моделі Random Forest із тюнінгом параметрів витрати становили 84 100 умовних одиниць, а доходи – 362 500. Звідси, ефективність від використання моделі рівна 4,3104. Чистий дохід від кампанії з використанням цієї моделі становитиме 278 400 умовних одиниць.

Під час використання градієнтного бустингу з тюнінгом параметрів витрати на рекламну кампанію становили 77 200 умовних одиниць при доходах у 336 000 одиниць. За таких результатів коефіцієнт ефективності буде рівним 4,3523. Отже, чистий дохід від такої кампанії становитиме 258 800 умовних одиниць.

За використання Extreme Gradient Boosting для прогнозування відповідей клієнтів на пропозицію строкового депозиту витрати становили 77 200 умовних одиниць, а доходи – 353 500. За такого результату коефіцієнт ефективності використання даної моделі становить 4,5790, а чистий дохід – 276 300 умовних одиниць.

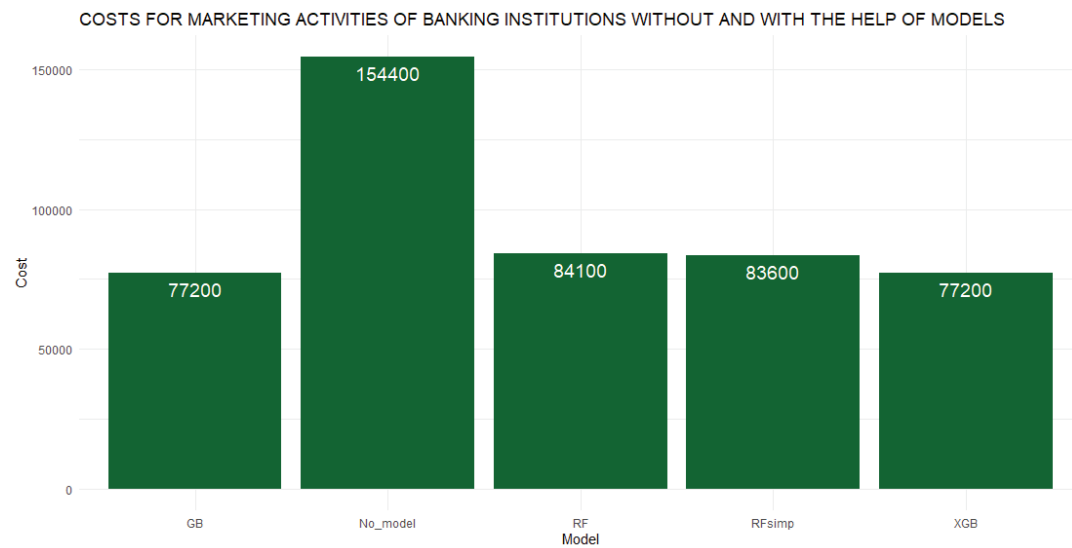


Рис. 5. Графік витрат банків на маркетингову діяльність із використанням розроблених моделей та без них

*Джерело: розроблено авторами

На графіку добре видно, що витрати на маркетингову діяльність за відсутності використання допоміжних моделей значною мірою переважають ті, що включали в себе використання моделей. Зокрема, витрати на маркетинг без використання моделей у 1,85 рази переважають витрати з використанням звичайної моделі Random Forest, у 1,84 рази переважають витрати з використанням тюнінгової моделі Random Forest, у 2,00 рази переважають витрати

з використанням моделі градієнтного бустингу з тюнінгом параметрів та в 2,00 рази переважають витрати за використанням XGBoost моделі з тюнінгом параметрів.

Отже, найбільшу ефективність для виконання маркетингової кампанії та оптимізації витрат надає використання моделі Extreme Gradient Boosting із тюнінгом параметрів 4,5790. За використання цієї моделі також спостерігається найбільший чистий дохід – 276 300 умовних одиниць.

Висновки

У процесі дослідження було розроблено ефективні моделі на основі ансамблів для оптимізації витрат на маркетингову діяльність банківських установ, що дали змогу знизити витрати на маркетинг та збільшити ефективність маркетингових заходів.

Зокрема, було побудовано моделі на основі звичайного Random Forest, Random Forest із тюнінгом параметрів, градієнтного бустингу з тюнінгом параметрів та XGBoost із тюнінгом параметрів, серед яких найкращою за показником збалансованої точності виявилася модель на основі Random Forest із тюнінгом параметрів 89,44%.

Під час обрахунку економічного ефекту виявилось, що витрати на маркетинг без використання моделей у 1,85 рази переважають витрати з використанням звичайної моделі Random Forest, у 1,84 рази переважають витрати з використанням тюнінгової моделі Random Forest, у 2,00 рази переважають витрати з використанням моделі градієнтного бустингу з тюнінгом параметрів та в 2,00 рази переважають витрати за використанням XGBoost моделі з тюнінгом параметрів.

Найбільшу ефективність для виконання маркетингової кампанії та оптимізації витрат надає використання моделі Extreme Gradient Boosting із тюнінгом параметрів 4,5790. За використання цієї моделі також спостерігається найбільший чистий дохід – 276 300 умовних одиниць. Але у будь-якому разі варто зазначити, що ефективність маркетингової кампанії за використання моделей фактично в два рази більша, ніж за їх відсутності.

Список використаної літератури

1. Safarkhani F., Moro S. Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree. *Applied Sciences*. 2021. 11(19). 9016.
2. Hao Ruiting, Xia Xiaoqian, Shen Siyi, Yang Xiaorong. Bank Direct Marketing Analysis Based on Ensemble Learning. *Journal of Physics: Conference Series*, 2020.
3. Chen X.-W., Jeong J.C. Enhanced recursive feature elimination, *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Cincinnati, OH, USA, 2007, pp. 429–435, doi: 10.1109/ICMLA.2007.35.
4. Jović A., Brkić K., Bogunović N. A review of feature selection methods with applications, 2015 *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 2015, pp. 1200–1205.
5. Meidan A. Bank Marketing Strategies, *International Journal of Bank Marketing*, 1983. Vol. 1. No. 2, pp. 3–17.
6. González-Recio O., Jiménez-Montero J.A., Alenda R. The gradient boosting algorithm and random boosting for genome-assisted evaluation in large data sets, *Journal of Dairy Science*, 2013. Volume 96, Issue 1, p. 614–624.
7. Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, p. 785–794.
8. Bunkhumpornpat C., Sinapiromsaran K., Lursinsap C. MUTE: Majority under-sampling technique, *2011 8th International Conference on Information, Communications & Signal Processing*, Singapore, 2011, pp. 1–4.

9. Moro S., Rita P., Cortez P. Bank Marketing. UCI Machine Learning Repository. 2012.
10. Alibrahim Hussainand, Ludwig Simone A. Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. In 2021 *IEEE Congress on Evolutionary Computation (CEC)*. IEEE Press, 2021, p. 1551–1559.
11. Amalia Luque, Alejandro Carrasco, Alejandro Martín, Ana de las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, *Pattern Recognition*, Volume 91, 2019, Pages 216–231.

References

1. Safarkhani, F., & Moro, S. (2021). Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree. *Applied Sciences*. 11(19). 9016. doi: 10.3390/app1119901 [in English].
2. Ruiting, Hao, Xiaoqian, Xia, Siyi, Shen, & Xiaorong Yang. (2020). Bank Direct Marketing Analysis Based on Ensemble Learning. *Journal of Physics: Conference Series*. 1627. 012026. doi: 10.1088/1742-6596/1627/1/012026 [in English].
3. Chen, X.-W., & Jeong, J.C. (2007). Enhanced recursive feature elimination. *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Cincinnati, OH, USA. 429–435. doi: 10.1109/ICMLA.2007.35 [in English].
4. Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia. 1200–1205, doi: 10.1109/MIPRO.2015.7160458 [in English].
5. Meidan, A. (1983). Bank Marketing Strategies. *International Journal of Bank Marketing*. Vol. 1 No. 2. 3–17. doi: 10.1108/eb010721 [in English].
6. González-Recio, O., Jiménez-Montero, J.A., & Alenda, R. (2013). The gradient boosting algorithm and random boosting for genome-assisted evaluation in large data sets. *Journal of Dairy Science*. Vol. 96, Issue 1, 614–624, ISSN 0022-0302, doi: 10.3168/jds.2012-5630 [in English].
7. Tianqi, C., & Carlos, G. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA. 785–794. doi: 10.1145/2939672.2939785 [in English].
8. Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2011). MUTE: Majority under-sampling technique. *8th International Conference on Information, Communications & Signal Processing, Singapore*. 1–4. doi: 10.1109/ICICS.2011.6173603 [in English].
9. Moro, S., Rita, P., & Cortez, P. (2012). Bank Marketing. *UCI Machine Learning Repository*. doi: 10.24432/C5K306 [in English].
10. Alibrahim, H., & Ludwig, S.A. (2021). Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. In 2021 *IEEE Congress on Evolutionary Computation (CEC)*. IEEE Press, 1551–1559 [in English].
11. Luque, A., Carrasco, A., Martín, A., & Ana, de las Heras. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*. Volume 91. 216–231. doi: 10.1016/j.patcog.2019.02.023 [in English].

Клебан Юрій Вікторович – старший викладач кафедри економіко-математичного моделювання та інформаційних технологій Національного університету «Острозька академія». E-mail: yuriy.kleban@oa.edu.ua, ORCID: 0000-0002-7070-5175, Researcher ID: AAN-3865-2021.

Денисюк Юлія Олександрівна – здобувачка першого рівня (бакалаврського) вищої освіти кафедри економіко-математичного моделювання та інформаційних технологій

Національного університету «Острозька академія». E-mail: yuliia.denysiuk@oa.edu.ua, ORCID: 0009-0007-0854-3084, Researcher ID: KCK-7199-2024.

Kleban Yuriy Viktorovych – Senior Lecturer, Department of Economic and Mathematical Modeling and Information Technologies of National University "Ostroh Academy". E-mail: yuriy.kleban@oa.edu.ua, ORCID: 0000-0002-7070-5175, Researcher ID: AAN-3865-2021.

Denysiuk Yuliia Oleksandrivna – student of the first level (bachelor) of higher education, Department of Economic and Mathematical Modeling and Information Technologies of National University "Ostroh Academy". E-mail: yuliia.denysiuk@oa.edu.ua, ORCID: 0009-0007-0854-3084, Researcher ID: KCK-7199-2024.