

National University “Ostroh academy”

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

Presented by Yuliia Denysiuk

Вступ

- Банки - важливі гравці світової економіки.
- Маркетингова діяльність банків спрямована на утримання наявних клієнтів та залучення нових.
- Для оптимізації маркетингових витрат банки можуть використовувати аналітику даних та інструменти машинного навчання для прогнозування поведінки клієнтів.
- Моделі ансамблю - потужний інструмент у машинному навчанні, який дозволяє поєднувати прогнози кількох моделей для отримання більш точних і стабільних прогнозів

Моделі ансамблю

Random Forest

Заснований на ансамблі дерев рішень, де кожне дерево будується на випадковій підмножині навчальних даних та випадкових підмножинах характеристик

Gradient Boosting

Використовує дерева рішень як базові моделі. Цей алгоритм будує послідовність дерев, кожне з яких виправляє помилки попереднього дерева. Головна ідея полягає в тому, щоб на кожному кроці навчання алгоритм використовував градієнт втрат для навчання нової моделі.

Extreme Gradient Boosting

Основна ідея XGBoost полягає в 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.

Boosting

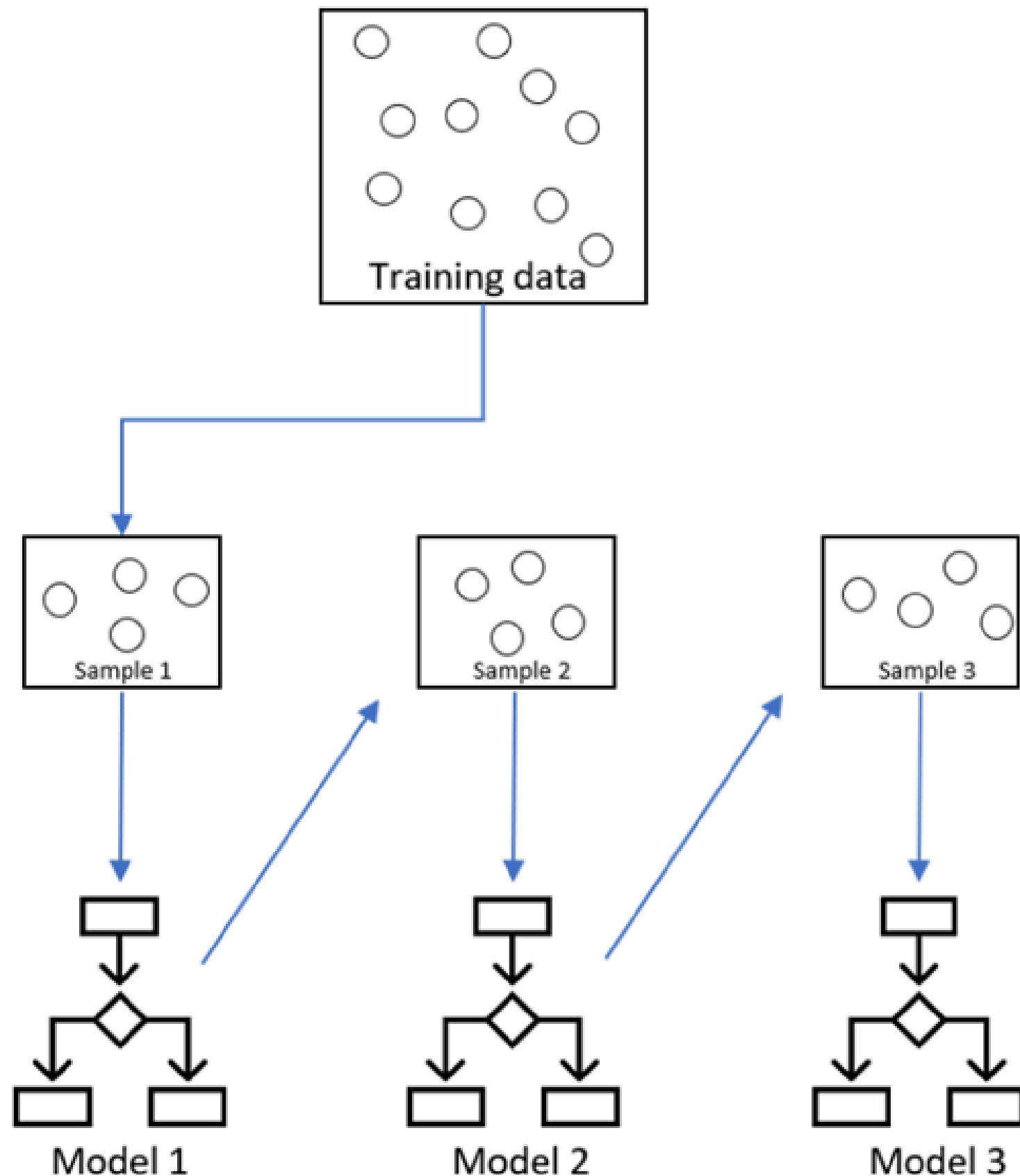


Схема роботи градієнтного бустингу

Основні етапи градієнтного бустингу:

1. Побудова першої моделі

2. Розрахунок помилок першої моделі

3. Побудова другої моделі, яка виправляє помилки першої моделі

4. Розрахунок помилок другої моделі

5. Побудова третьої моделі, яка виправляє помилки другої моделі

6. Такий процес повторюється доти, поки не буде досягнуто потрібної точності

Дані

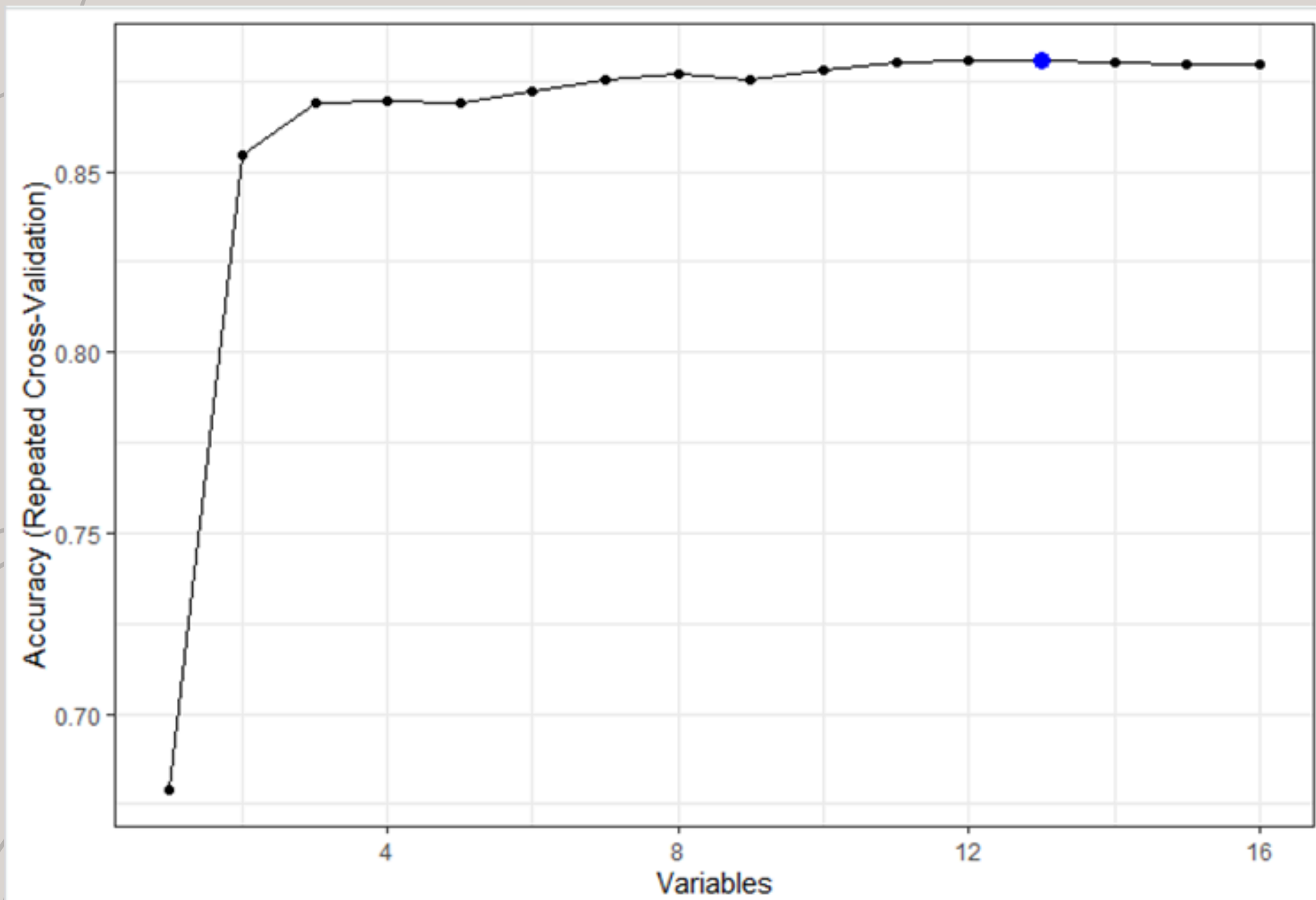
Дані стосуються прямої маркетингової кампанії (телефонних дзвінків) португальського банку. Центр обслуговування клієнтів банку телефонує клієнтам для реклами свого строкового депозитного продукту. Вони збирають основну інформацію про клієнтів та їхні рішення: чи підписалися на строковий депозит (так чи ні). Мета класифікації - передбачити їхні рішення після дзвінків.

Набір даних містить 41188 рядків і 21 стовпець, з яких 10 числових і 11 категоріальних.

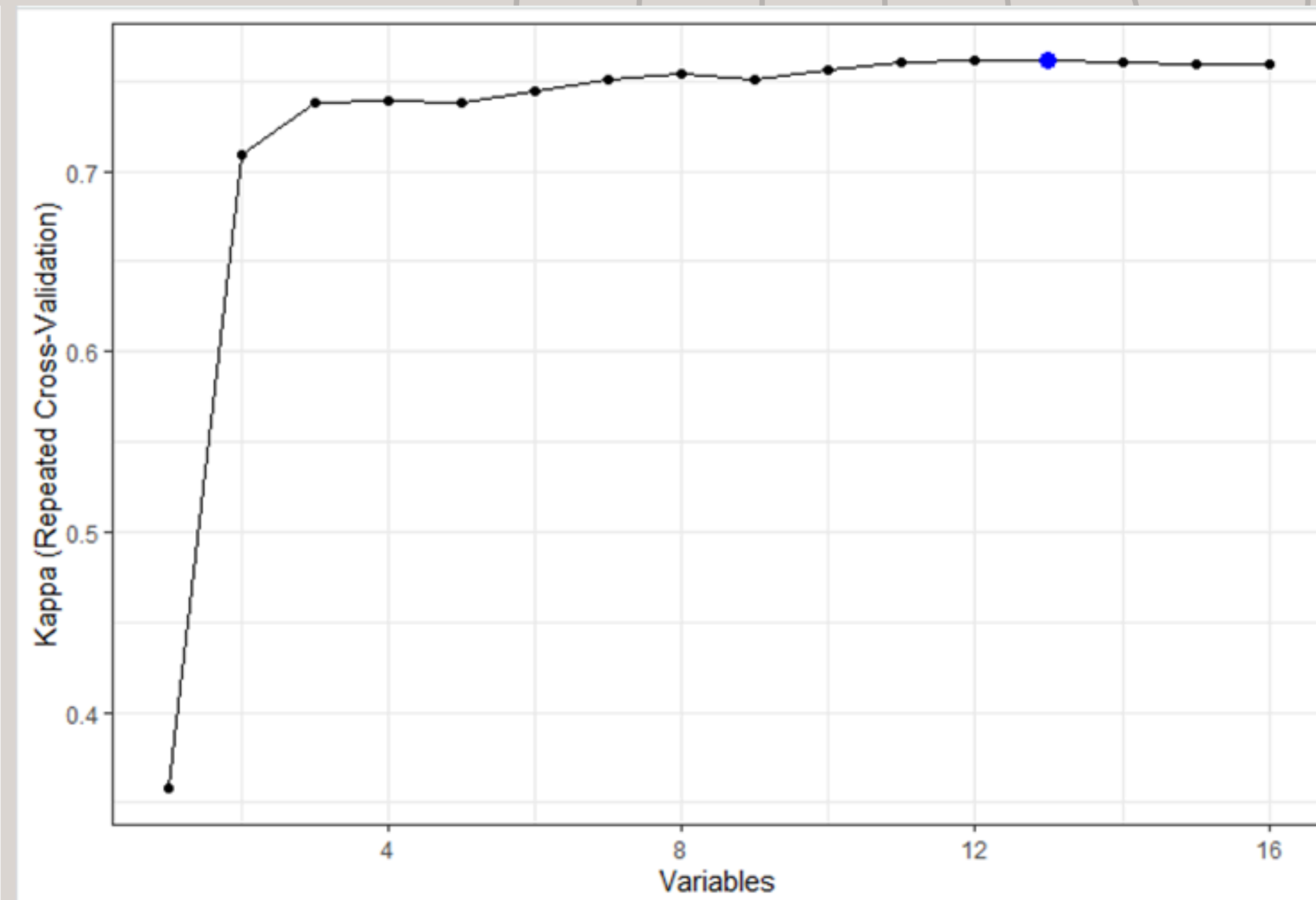
Підготовка даних до моделювання:

- Очистка невідомих і порожніх значень
- Балансування вибірки методом `under-sampling`
- Рекурсивне усунення ознак для визначення оптимальної кількості факторів для побудови моделі

Результат використання методу підбору ознак

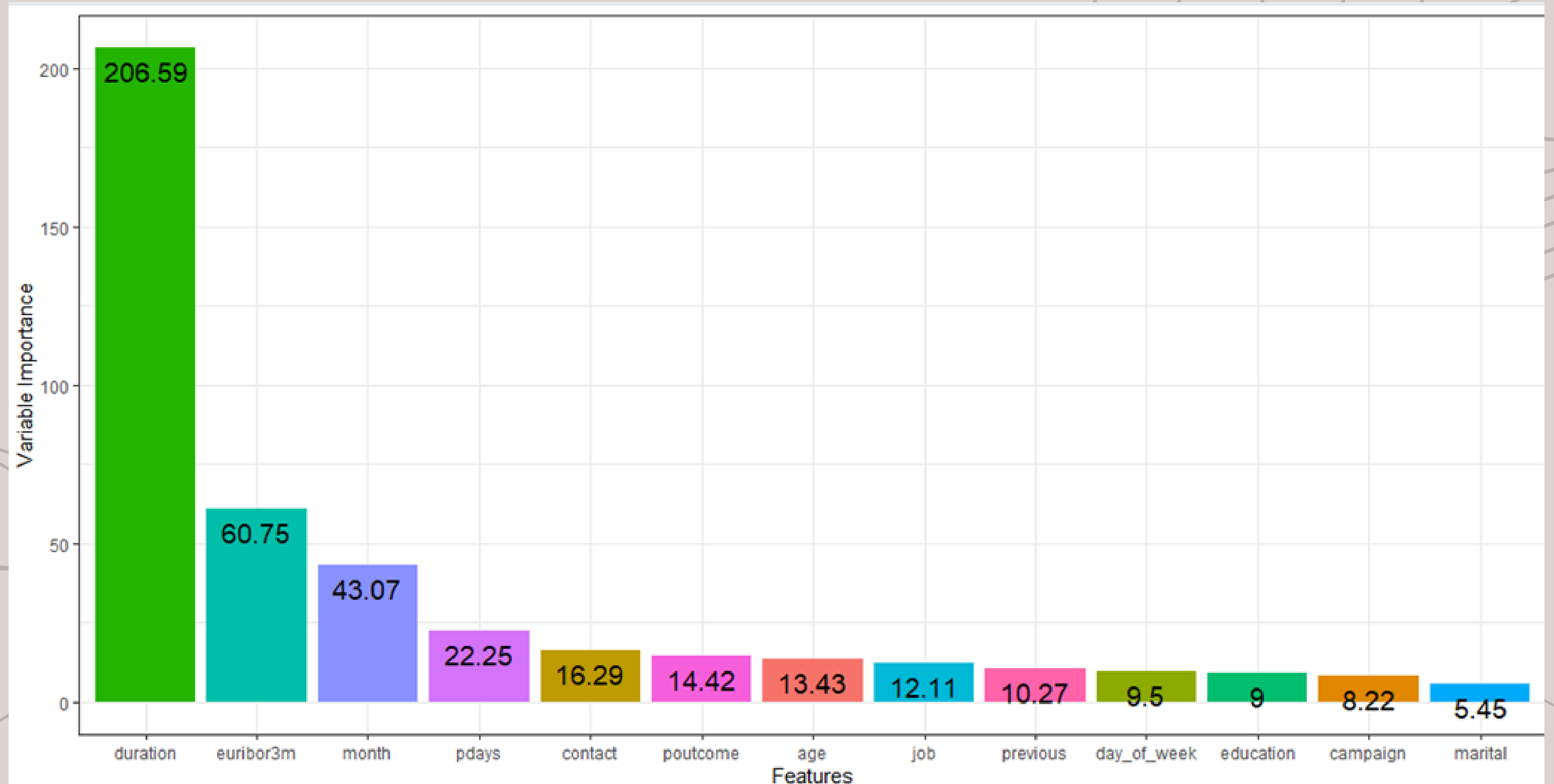


Коефіцієнт Ассурасу



Коефіцієнт Карра

Результат використання методу підбору ознак



Важливість ознак

Результати

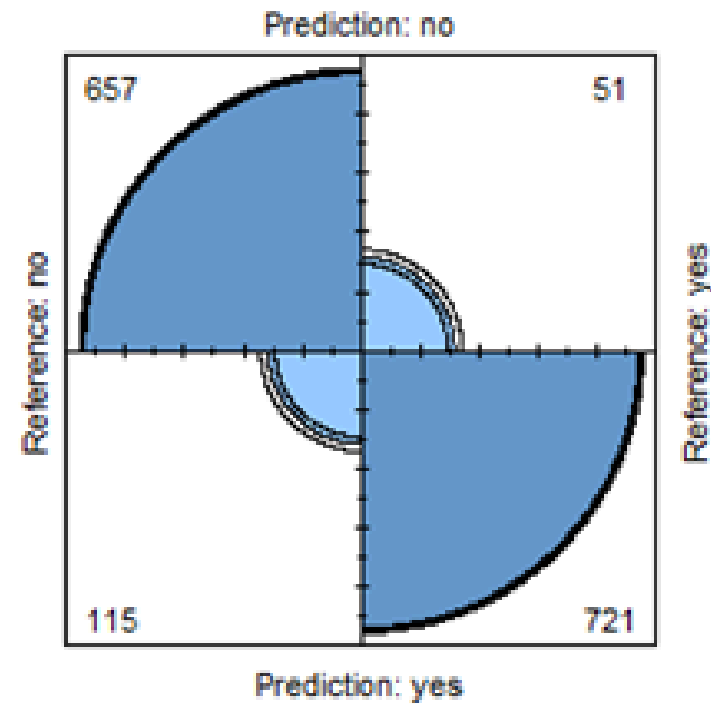
У Таблиці видно, що за показником збалансованої точності найкращою виявилась модель Random Forest з тюнінгом параметрів. Всі інші методи досить близькі до неї та все ж не кращі. Зважаючи на це увагу, далі проводиться обрахунок економічного ефекту на витрати банків на маркетингову діяльність від використання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

Співставлення результатів моделювання

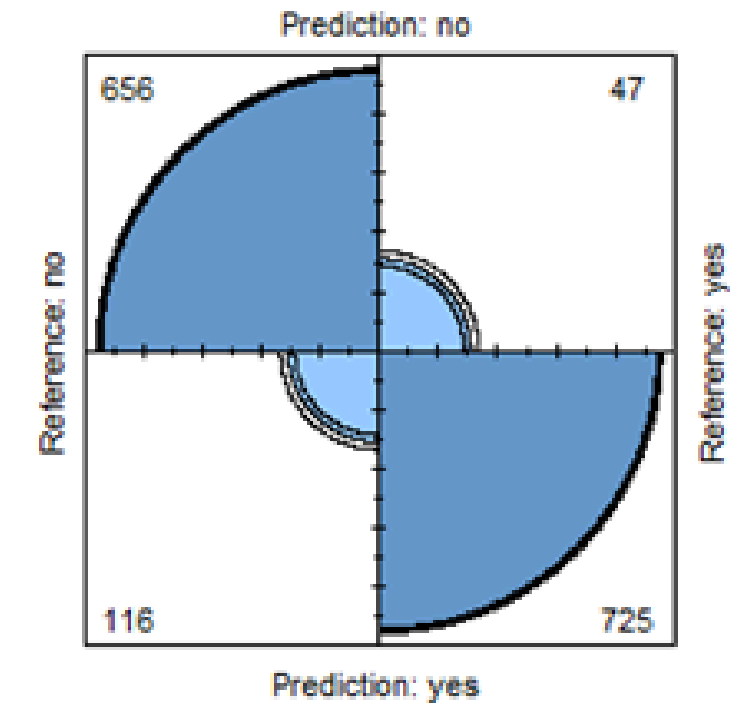
	Simple RF	Random.Forest	Basic.GBM	XGBoost
Test Accuracy	0.8924870	0.8944301	0.8808290	0.8873057
Balanced Accuracy	0.8924870	0.8944301	0.8809927	0.8885679

ЕКОНОМІЧНИЙ ефект

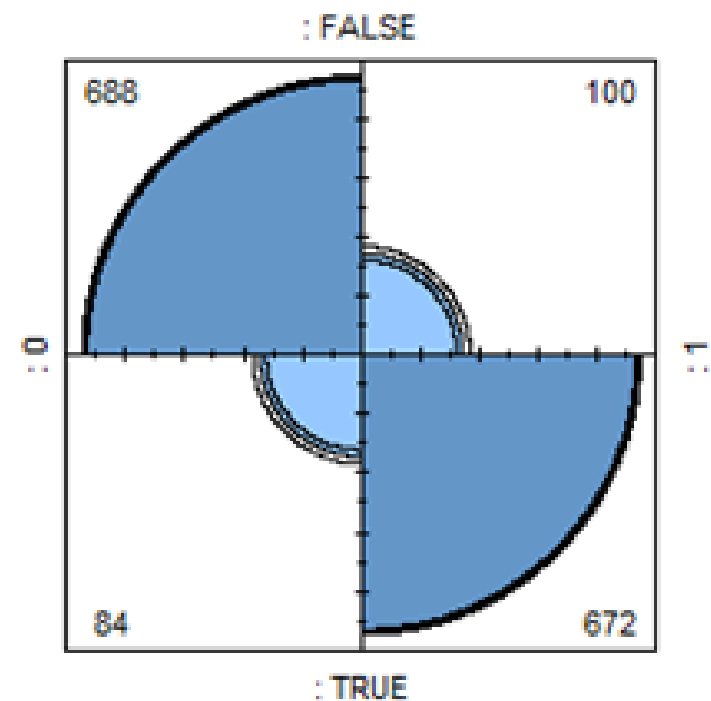
Confusion Matrix for simple Random Forest Model



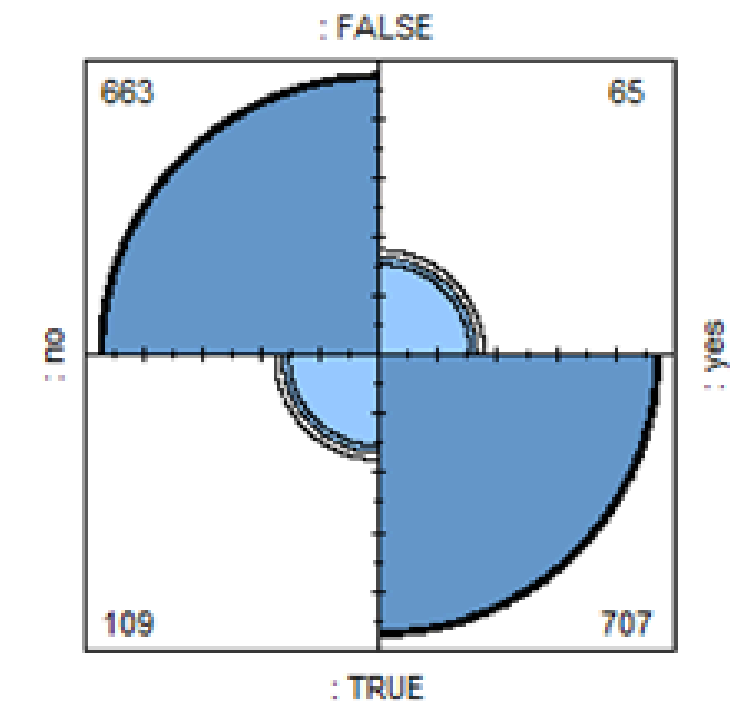
Confusion Matrix for Random Forest Model



Confusion Matrix for Gradient Bosting

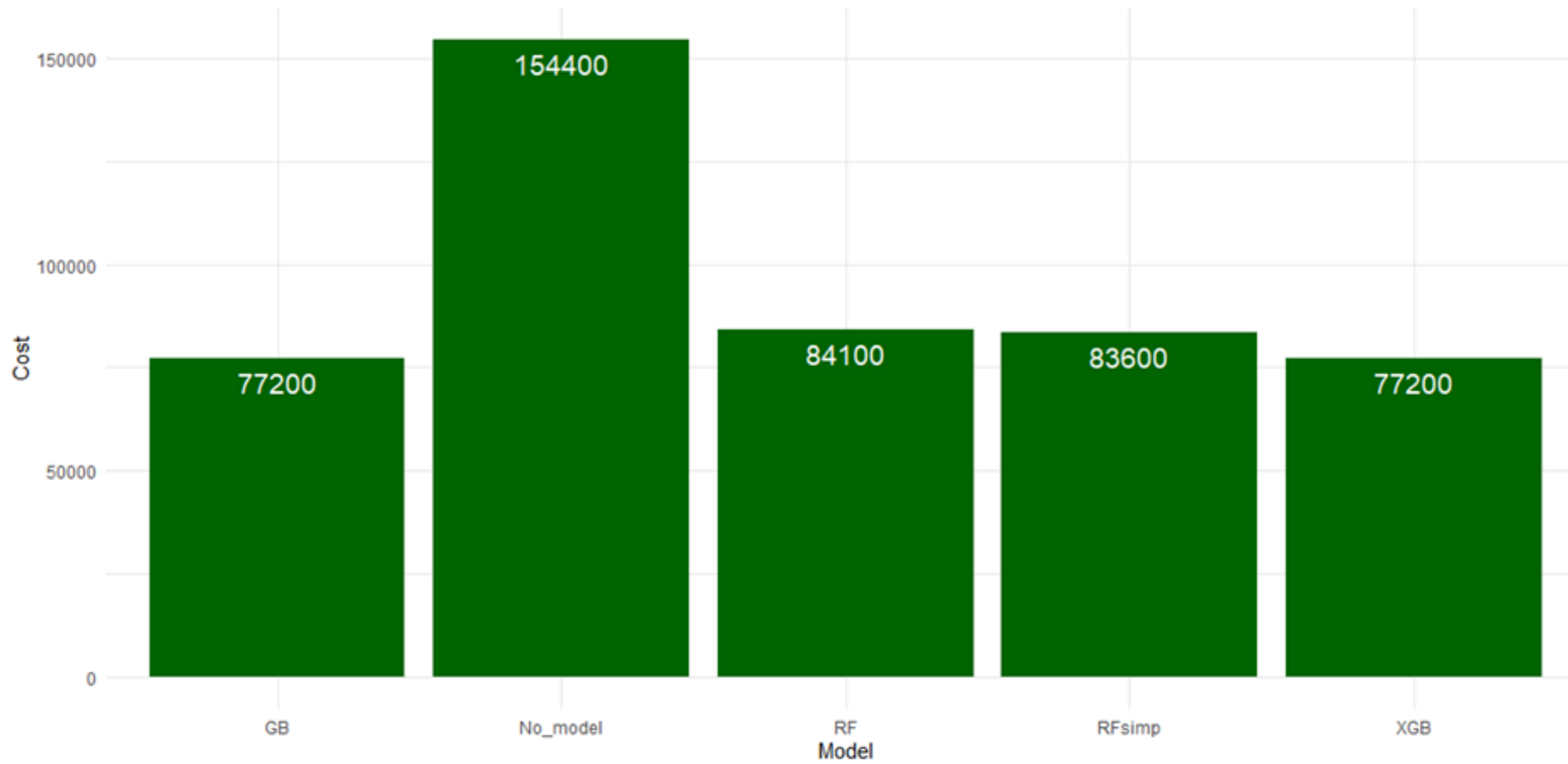


Confusion Matrix for XGBoost



Витрати

COSTS FOR MARKETING ACTIVITIES OF BANKING INSTITUTIONS WITHOUT AND WITH THE HELP OF MODELS



На графіку добре видно, що витрати на маркетингову діяльність при відсутності використання допоміжних моделей значною мірою переважають ті, що включали в себе використання моделей

Коефіцієнти економічного ефекту

No model
2.5

RF simpl
4.3122

RF
4.3104

GB
4.3523

XGB
4.5790

Висновки

Найбільшу ефективність для виконання маркетингової кампанії та оптимізації витрат надає використання моделі Extreme Gradient Boosting з тюнінгом параметрів – 4,5790. За використання цієї моделі також спостерігається найбільший чистий дохід – 276300 умовних одиниць. Але у будь-якому випадку, варто зазначити, що ефективність маркетингової кампанії при використанні моделей фактично в 2 рази більша ніж за їх відсутності.



**Thank
You**