

**Тема: Вирішення задач  
підвищення продажів в  
маркетингу з  
використанням методів  
інтелектуального  
аналізу даних**

**Виконала: студентка гр.КА-41М  
Іванова Юлія Валеріївна  
Науковий керівник роботи:  
К.т.н., старший викладач  
Терентьев Олександр Миколайович**

**ОБ'ЄКТОМ ДОСЛІДЖЕННЯ** є вибірка з попередньої маркетингової кампанії підприємства, що являє собою набір даних про клієнтів, їх характеристики та взаємодію з компанією.

**МЕТА РОБОТИ** полягає в розробці математичних моделей для підвищення продажів в маркетингу баз даних.

**МЕТОЮ ДОСЛІДЖЕННЯ** є огляд існуючих методів прогнозування відгуку, методів статистичного аналізу даних, що можливі до застосування в даній предметній області та побудова моделей підвищення відгуку.

**МЕТОДИ ДОСЛІДЖЕННЯ:** алгоритми навчання з учителем, регресійний аналіз, методи вибору змінних.

Використано програмні пакети MATLAB, Excel

# Постановка задачі

1. Прогнозування рівня відгуку клієнтів в маркетингу баз даних
2. Огляд існуючих методів: традиційне моделювання та моделювання підвищення відгуку
3. Розробка спеціалізованого програмного продукту для моделювання підвищення відгуку за допомогою моделей логістичної регресії
4. Аналіз побудованих моделей

*“I know half the money I spend on advertising is wasted, but I can never find out which half.”*

- John Wanamaker

Які клієнти мають бути обрані для маркетингової кампанії для того щоб максимізувати рівень прибутковості?

# Класи клієнтів



# Моделювання відгуку

## Традиційне моделювання

- Визначає клієнтів, які ймовірніше відреагують на кампанію
- Цільовою аудиторією є всі можливі відповідачі
- Витрачає ресурси на клієнтів, які куплять товар незалежно від кампанії

## Моделювання підвищення відгуку

- Визначає додаткові покупки, які б не мали місця без кампанії
- Вимірює ефекти маркетингових кампаній
- Визначає клієнтів на рішення яких можна вплинути

# Математичні моделі для моделювання відгуку

## Модель різницевої оцінки

Окремо будуються моделі на контрольній і оброблені групах

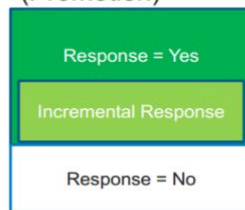
$$\widehat{Y}_T = X_T \widehat{\beta}_T$$

$$\widehat{Y}_C = X_C \widehat{\beta}_C$$

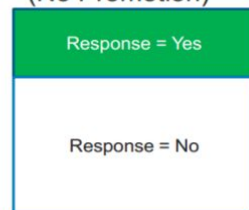
Потім обидві моделі використовуються для обрахування передбачених значень з цілого набору даних  $D = DT \cup DC$  Різниця балів може бути здобута з оцінок по формулі:

$$\widehat{DS}_i = (\widehat{Y}_T - \widehat{Y}_C)_i$$

Treatment Group  
(Promotion)



Control Group  
(No Promotion)



# Математичні моделі для моделювання відгуку

## Комбінована модель

В даній моделі використовується індикатор  $T_i$ :  $T_i = 1$  для  $DT$  і  $T_i = 0$  для  $DC$  і він дозволяє налаштовувати моделі на всіх даних ( $D = DT \cup DC$ )

$$Y = X\beta + T\gamma + XT\varphi + \epsilon$$

Різницеві оцінки отримуються з оціненого рівняння, таким чином:

$$\widehat{Y}_T = X\widehat{\beta} + \widehat{\gamma} + X\widehat{\varphi}$$

$$\widehat{Y}_C = X\widehat{\beta}$$

$$\widehat{DS}_i = (\widehat{Y}_T - \widehat{Y}_C)_i = \widehat{\gamma} + X\widehat{\varphi}$$



# Вибір змінних

Для бінарної цільової змінної  $Y \in \{0,1\}$

$$WOE_i = \log \frac{P(X = x_i | Y = 1)}{P(X = x_i | Y = 0)}$$

$$IV = \sum_i (P(X = x_i | Y = 1) - P(X = x_i | Y = 0)) * WOE_i$$

IV вимірює силу кореляції між пояснювальною змінною, та відгуком в випадку бінарної змінної. Оцінки IV показують : значення менше 0,02 вказує, що змінна не прогностичною, а значення більше 0,3 показує, що змінна має сильну прогностичну силу

## Вибір змінних

Враховуючи оброблену (Т) та контрольну (С) групи поняття WOE і IV можуть бути застосовані до моделі підвищення відгуку

$$NWOE = \log \frac{P(X = x_i | Y = 1)_T / P(X = x_i | Y = 0)_T}{P(X = x_i | Y = 1)_C / P(X = x_i | Y = 0)_C}$$

$$NIV = \sum_i (P(X = x_i | Y = 1)_T P(X = x_i | Y = 0)_C - P(X = x_i | Y = 1)_C P(X = x_i | Y = 0)_T) NWOE_i$$

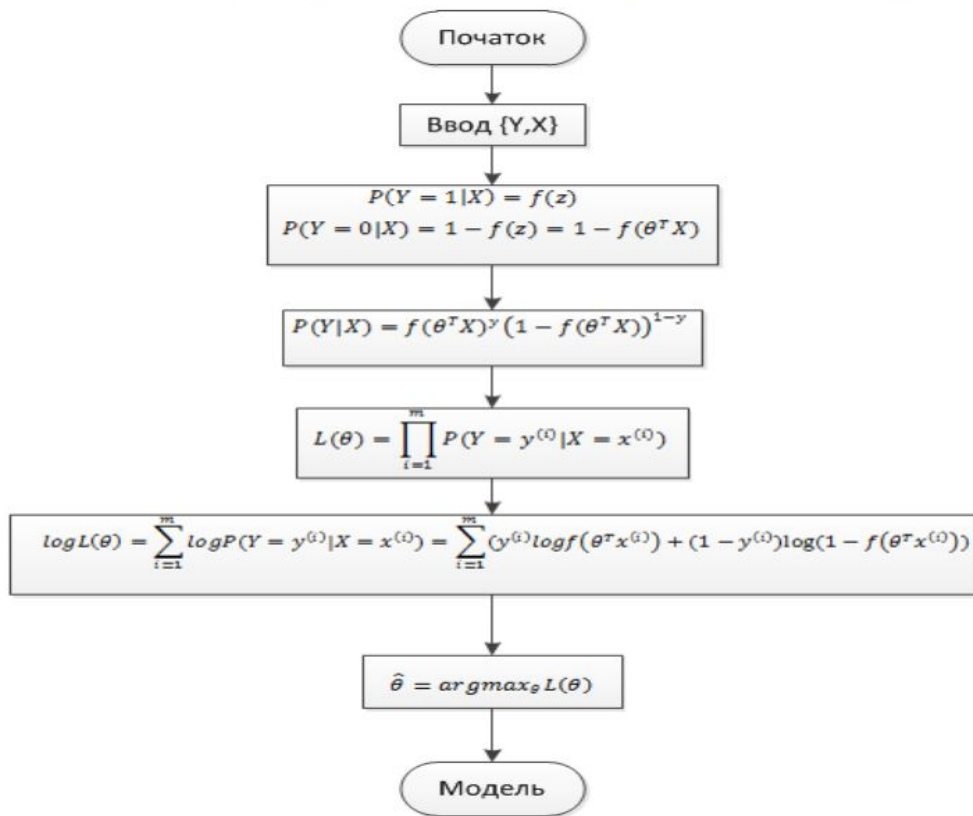
# Логістична регресія

Якщо  $Y_i$  бінарна змінна відгуку, яка представляє чи відповів  $i$  клієнт на кампанію, розглядаємо наступний набір незалежних змінних:  $X_i, T_i, X_i * T_i$ , де  $T_i = 1$  якщо  $i$  клієнт в обробленій групі і якщо в контрольній. Можемо змодельовати рівень відгуку за допомогою логістичної регресії:

$$P_i = E(Y_i | X_i) = \frac{\exp(\alpha + \beta' X_i + \gamma T_i + \delta' X_i T_i)}{1 + \exp(\alpha + \beta' X_i + \gamma T_i + \delta' X_i T_i)}$$

де  $\alpha, \beta, \gamma, \delta$  параметри які треба оцінити

# Логістична регресія



# Додаткові продажі

Модель підвищення відгуку, яка підтримує обидві цілі також називається моделлю додаткового продажу, або моделю підвищення продажів. Інтервальну змінну можна оцінити тільки коли отриманий відгук, що вводить відхилення відбору у модель. Щоб подолати цю проблему використовується модель вибору Хекмана. Припустимо що є дві цільові змінні:  $Y$  - обсяг продажів на одного клієнта, і  $Z$  - змінна, яка відображає реакцію клієнта на заохочення.  $Y$  спостерігається тільки тоді, коли клієнт реагує на заохочення. Тому ця модель може бути реалізована наступним чином:

$$Y = X\beta + \epsilon \text{ if } Z = 1$$

$$Y = \text{not observed if } Z = 0$$

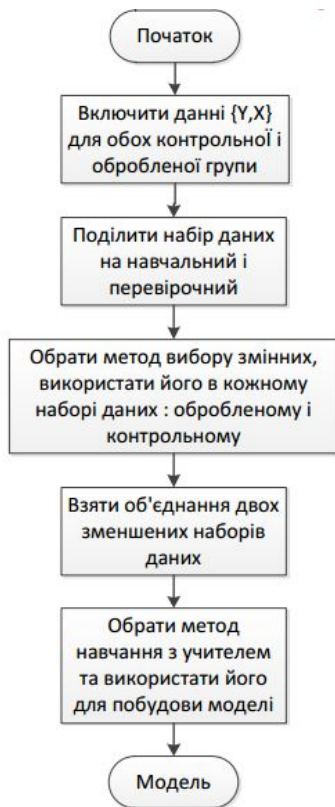
$$Z^* = X^*\gamma + \omega$$

$$Z = \begin{cases} 1, & \text{if } Z^* > 0 \\ 0, & \text{if } Z^* \leq 0 \end{cases}$$

де  $Z^*$  прихована змінна;

$X$  та  $X^*$  не є незалежно розподіленими, вони зазвичай мають однакові або багато спільних загальних змінних.

# Загальний алгоритм побудови моделі



# Загальний алгоритм побудови коефіцієнту підвищення



# Структура вибірки

5397 навчальна вибірка

- Promotion = 0 для 2698 записів
- Promotion = 1 для 2699 записів

3603 для перевірки

- Promotion = 0 для 1802 записів
- Promotion = 1 для 1801 записів



# Модифікована вибірка

Навчальна вибірка:

Promotion	Response	Count
0	0	823
0	1	674
1	0	742
1	1	890

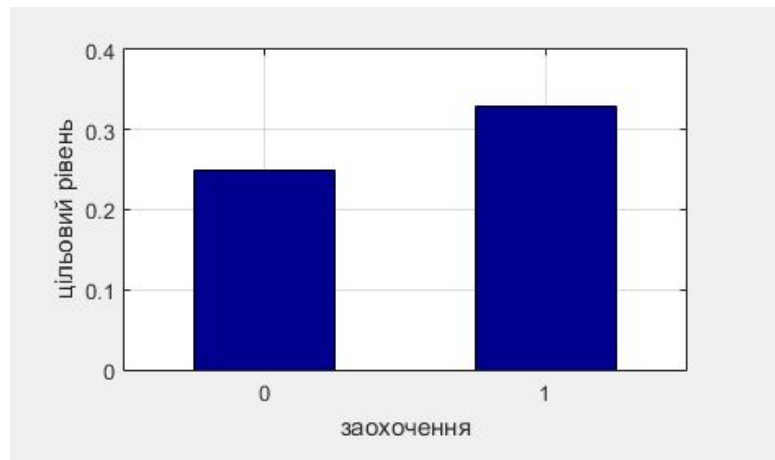
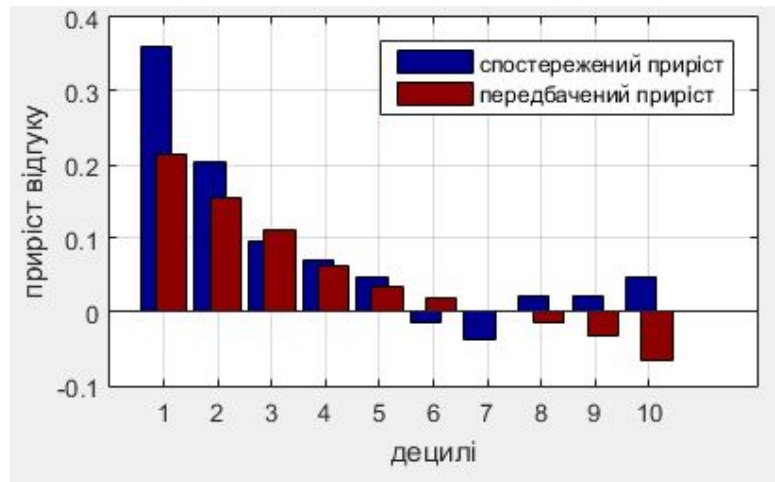
Перевірочна вибірка:

Promotion	Response	Count
0	0	549
0	1	451
1	0	496
1	1	595

# Результати: покрокова регресія

- Обробка : 33,4%
- Контроль : 25,1%
- Приріст : 8,3%

Приріст : Верхній дециль має найвищу частоту додаткових відгуків, з передбаченим прирощенням 21,2 % і спостережним приростом 35,2 %.  
Передбачене прирощення приблизно в три рази вище, ніж середній приріст відгуків (8,3 %)

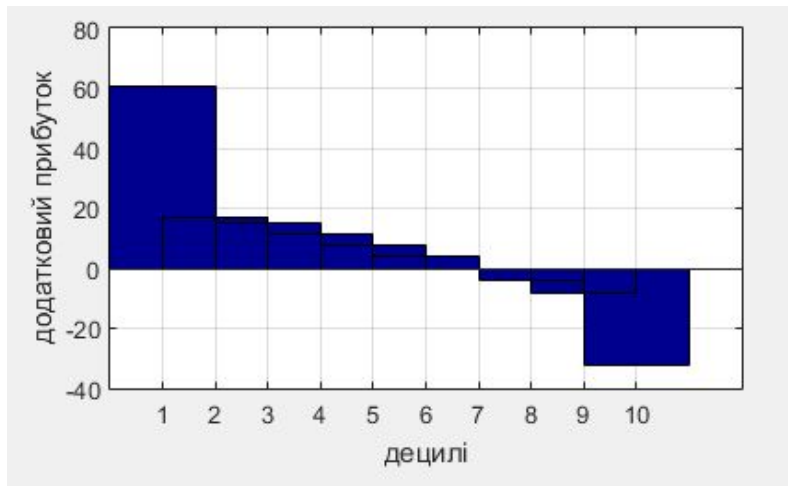
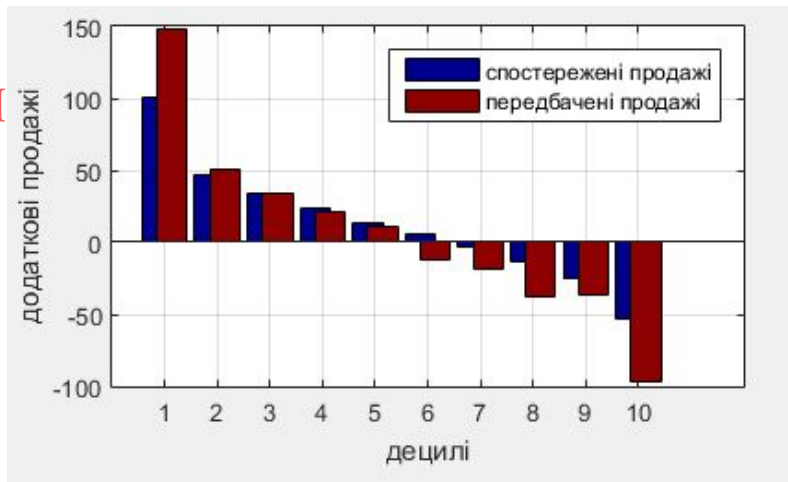


# Результати: покрокова регресія

Графіки ілюструють продуктивність моделі з боку продажів. Якщо клієнт купляє будь-який товар під час маркетингової кампанії, то вартість покупки також стає цільовою змінною і будується двуступінчаста модель.

Перші 10% клієнтів демонструють додаткові продажі у 100,98\$.

Клієнти вважаються вигідними, якщо додатковий прибуток перевищує вартість маркетингового контакту, як бачимо 70% клієнтів гарантує позитивний прибуток, перші 10% генерують 60\$ додаткового прибутку.



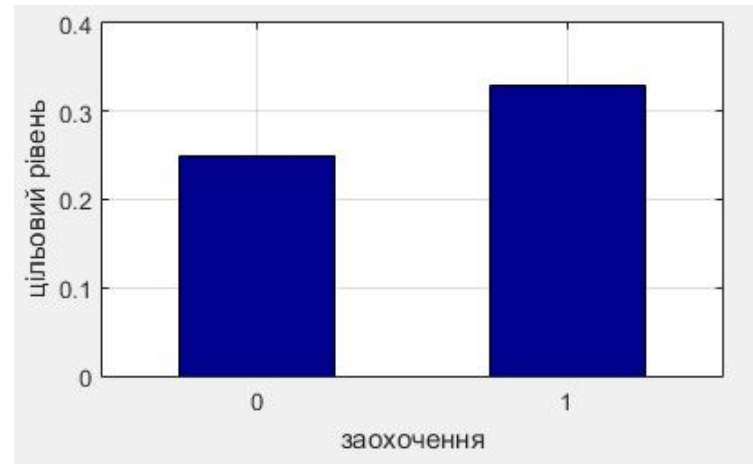
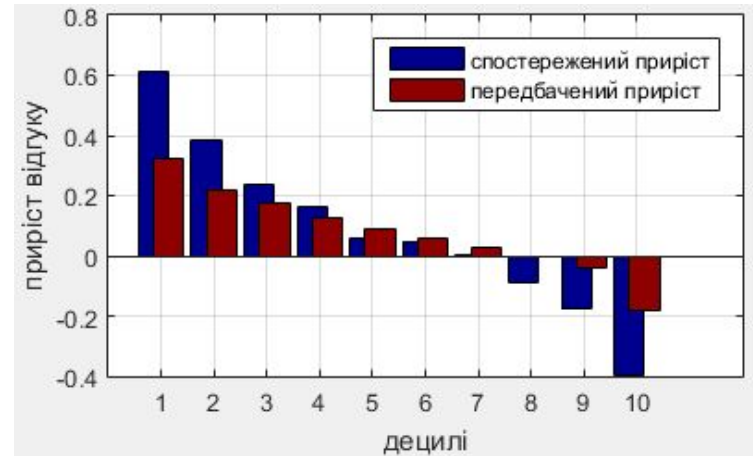
## Результати: попередній вибір змінних

Назва змінної	Чиста інформативність
CUST_TENURE	2056.967
MEMBERSHIP	477.9946
SPEND_CAT5	450.1365
SPEND_CAT2	357.7655
ORDER_ONLINE	336.199
ORDER_PHONE	329.5455
SPEND_CAT4	297.8781
ITEM_TOTAL	297.8762
LAST_YEAR_SPEND	297.6157
SPEND_CAT3	291.5341
SPEND_CAT1	279.4105
ITEM_ONLINE	278.6421
FREQUENT_BUYER	276.9668
ORDER_TOTAL	230.2379
ORDER_APR	207.8365
RECENCY	136.1352

# Результати: з вибором змін

- Обробка : 32,91%
- Контроль : 25,08%
- Приріст : 7,83%

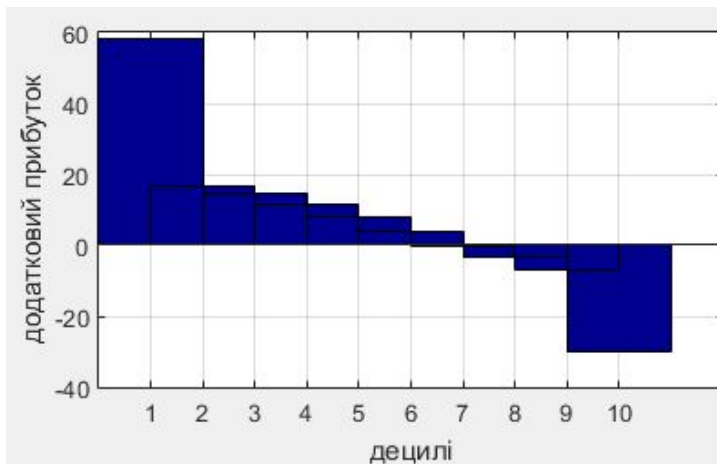
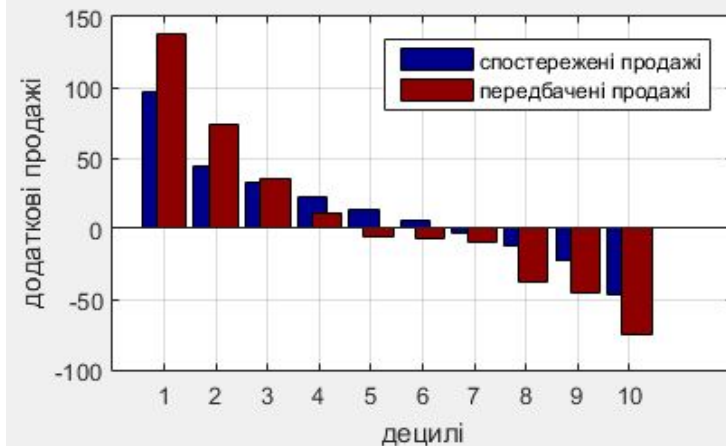
Приріст : Верхній дециль має найвищу частоту додаткових відгуків, з передбаченим прирощенням 32,29 % і спостережним приростом 59,77 %. Передбачене прирощення приблизно в чотири рази вище, ніж середній приріст відгуків (7,83 %)



# Результати: з вибором змін

Перші 10% клієнтів демонструють додаткові продажі у 96,76\$.

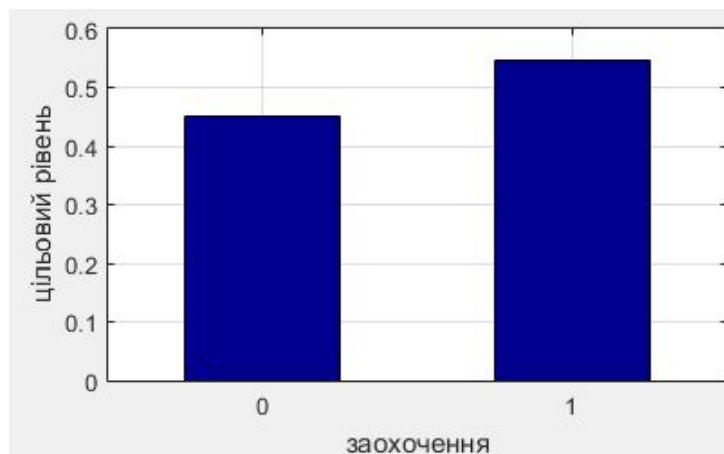
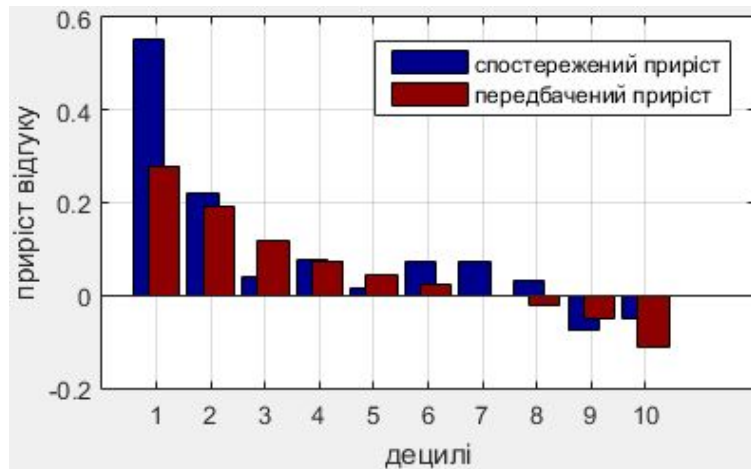
Клієнти вважаються вигідними, якщо додатковий прибуток перевищує вартість маркетингового контакту, як бачимо 70% клієнтів гарантує позитивний прибуток, перші 10% генерують 58,28\$ додаткового прибутку.



# Результати: комбінована модель

- Обробка : 51,68%
- Контроль : 42,18%
- Приріст : 9,5%

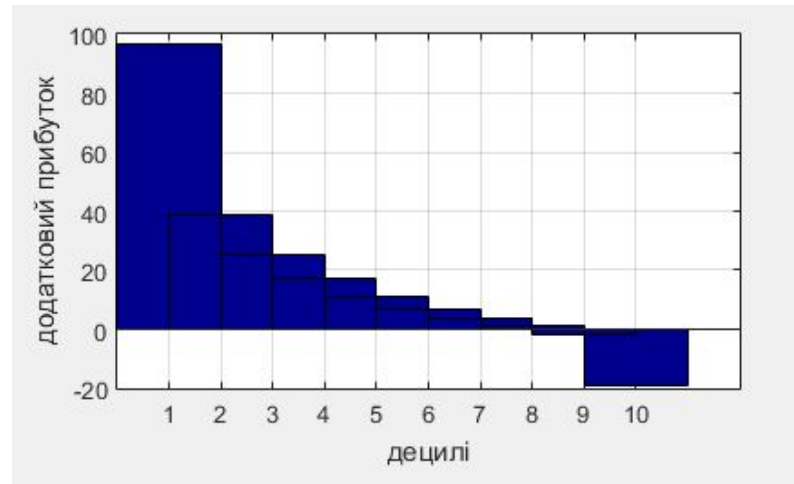
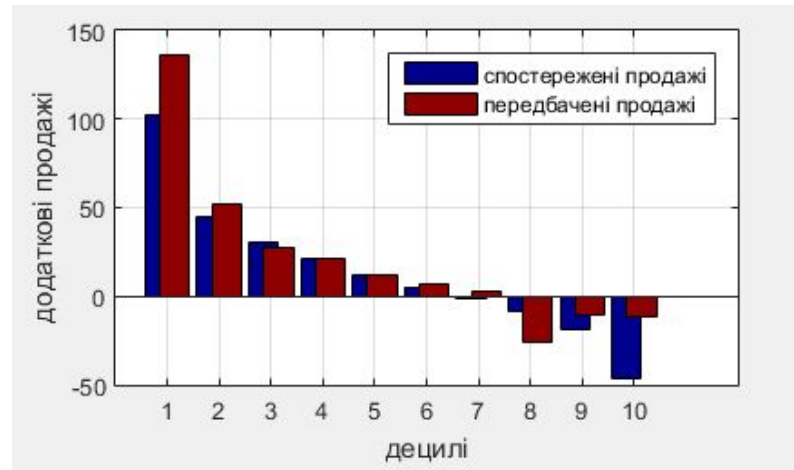
Приріст : Верхній дециль має найвищу частоту додаткових відгуків, з передбаченим прирощенням 24,5 % і спостережним приростом 57,2 %. Передбачене прирощення приблизно в три рази вище, ніж середній приріст відгуків (9,5 %)



# Результати: комбінована модель

Перші 10% клієнтів демонструють додаткові продажі у 102,50\$.

Клієнти вважаються вигідними, якщо додатковий прибуток перевищує вартість маркетингового контакту, як бачимо 80% клієнтів гарантує позитивний прибуток, перші 10% генерують 96,33\$ додаткового прибутку.





# Аналіз результатів

Назва моделі	SSE	Misc	Roc	Gini
Модель1	0.1878	0.2731	0.7106	0.4212
Модель2	0.1962	0.2837	0.6850	0.3701
Модель3	0.2244	0.3864	0.6217	0.2435

# Висновки

- Розглянуті методи прогнозування відгуку в маркетингу баз даних : традиційні методи моделювання та моделювання підвищення відгуку
- Розроблено програму для моделювання підвищення відгуку та знаходження оптимальних цільових клієнтів, використовуючи різні структури вибірок та підходи до моделювання за допомогою логістичної регресії
- Виконано порівняльний аналіз якості моделі на основі статистичних критеріїв (SSE,Misc,Roc,Gini)
- Подальші кроки дослідження : моделювання на базі інших алгоритмів таких як: нейроні мережі, наївний байєсовський класифікатор, дерева рішень, метод головних компонент

**Дякую за увагу!**