

**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., н.с.,
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.

ПРЕДМЕТ ДОСЛІДЖЕННЯ регресійні моделі, моделі підвищення відгуку

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

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

Використано програмні пакети MATLAB, Excel

Постановка задачі

1. Прогнозування рівня відгуку клієнтів та додаткових продажів в маркетингу баз даних
2. Огляд існуючих методів: традиційне моделювання та моделювання підвищення відгуку, модель додаткових продажів
3. Розробка спеціалізованого програмного продукту для моделювання підвищення продажів 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

Придбає після маркетингового стимулу		Не Турбувати 	Загублені 
		Безперечні 	Можливо переконати 
		ТАК	НІ
		Придбає без маркетингового стимулу	

Моделювання відгуку

Традиційне моделювання

- Визначає клієнтів, які ймовірно 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

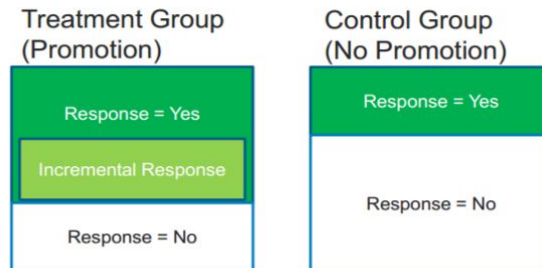
Окремо будуються моделі на контрольній і обробленій групах

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

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

Потім обидві моделі використовуються для обрахування різниці балів оцінок по формулі

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



[1]Hansotia, B. and Rukstales, B. (2002). "Incremental Value Modeling." Journal of Interactive Marketing 16:35–46.

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:

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

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

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

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

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

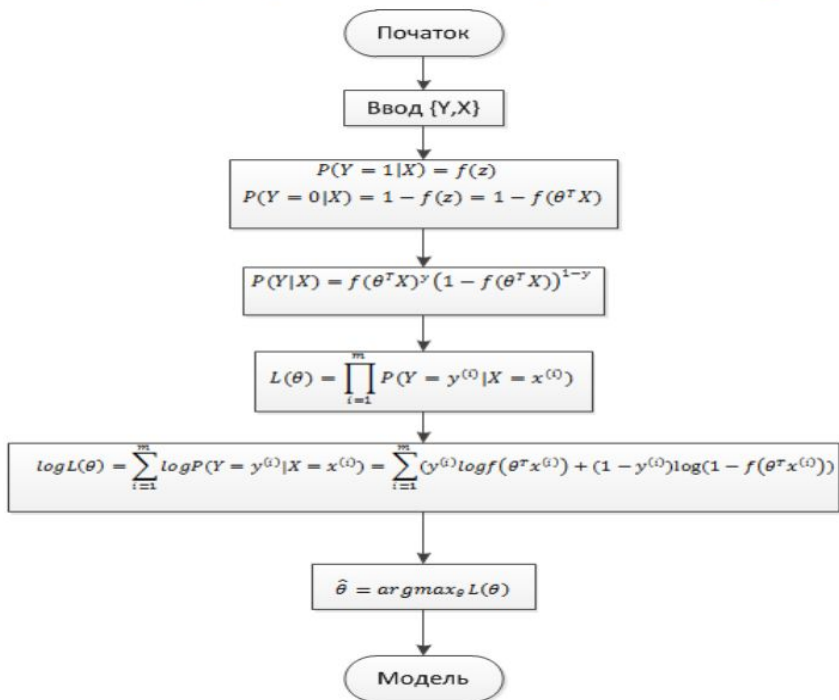
$$NIV = \sum_i (P(X = x_i | Y = 1)_T P(X = x_i | Y = 0)_C - P(X = x_i | Y = 1)_C P(X = x_i | Y = 0)_T) NWOE_i$$

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



Додаткові продажі

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:

- Z^* прихована змінна
- X та X^* не є незалежно розподіленими, вони зазвичай мають однакові або багато спільних загальних змінних

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

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

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

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

$$E[Y|X, Z = 1] = X\beta + E[\epsilon | \omega > -X^*\gamma]$$

$$E[\epsilon | \omega > -X^*\gamma] = \rho\sigma \frac{\phi(-X^*\gamma)}{1 - \Phi(-X^*\gamma)} = \rho\sigma \frac{\phi(X^*\gamma)}{\Phi(X^*\gamma)}$$

$$\epsilon \sim N(0, \sigma)$$

$$\omega \sim N(0, 1)$$

$$\text{Corr}(\epsilon, \omega) = \rho$$

Додатковий прибуток

Модель додаткових продажів має дві цільові змінні:

- \hat{Y}_{Resp}
- $\hat{Y}_{Outcome}$

тому додатковий прибуток може бути розрахований наступним чином, де C це вартість маркетингового стимулу на контакт:

$$\hat{R}_T = \hat{Y}_{Resp_T} * \hat{Y}_{Outcome_T} - C$$

$$\hat{R}_C = \hat{Y}_{Resp_C} * \hat{Y}_{Outcome_C}$$

$$\hat{R}_{incre} = \hat{R}_T - \hat{R}_C$$

Структура вибірки

5397 навчальна вибірка

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

3603 для перевірки

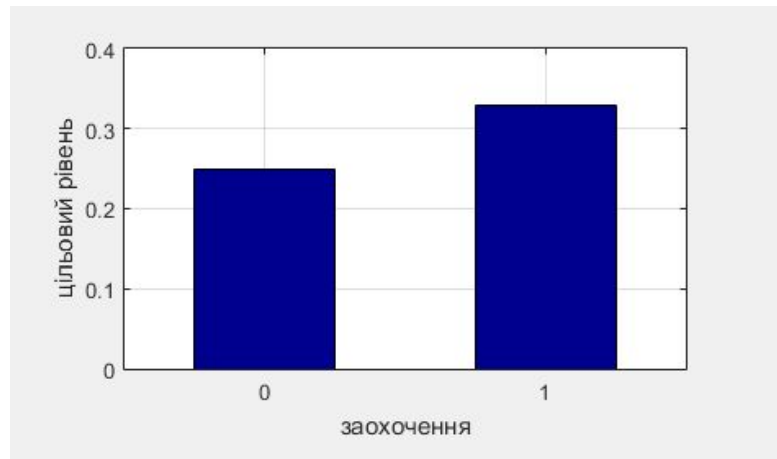
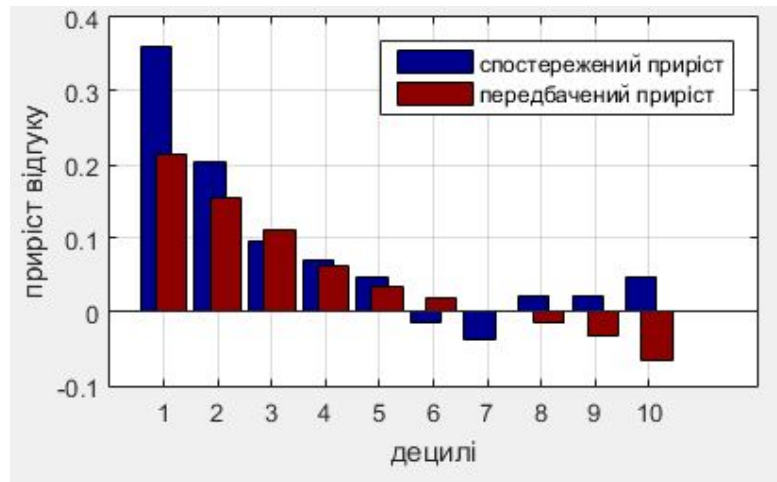
- Promotion = 0 для 1802 записів
- Promotion = 1 для 1801 записів

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 %)

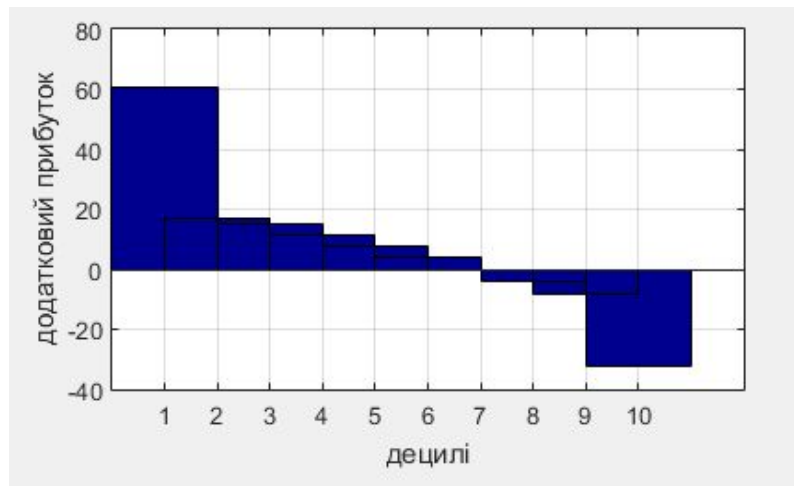
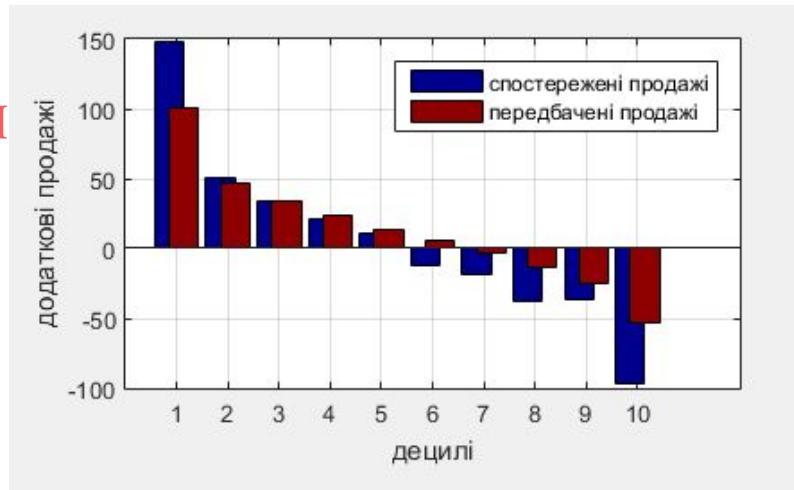


Результати: покрокова регресія

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.

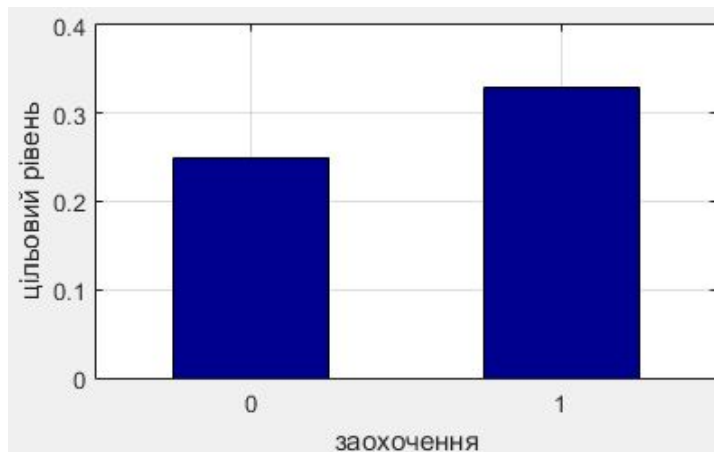
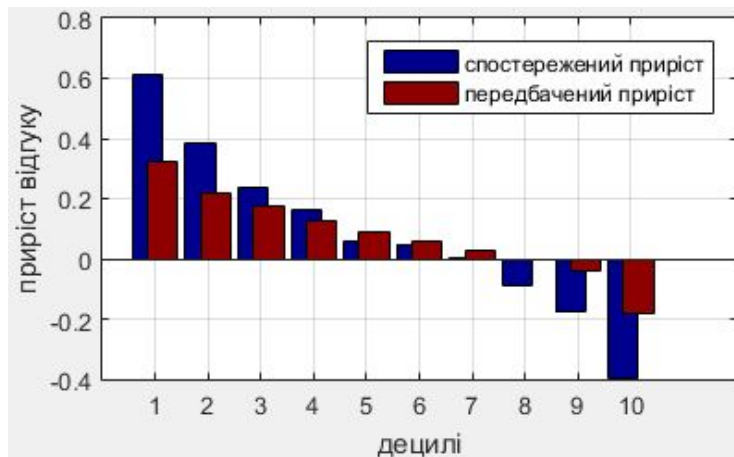


Результати: з вибором змін

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

Приріст : Верхній дециль має найвищу частоту додаткових відгуків, з передбаченим приростом 32,29 % і спостережним приростом 59,77 %.

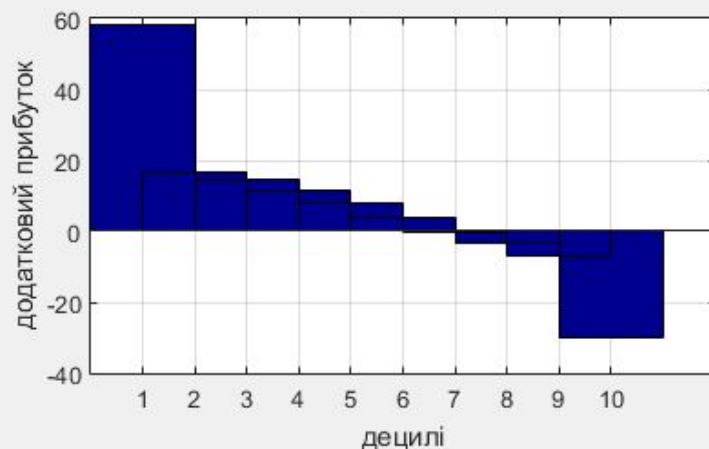
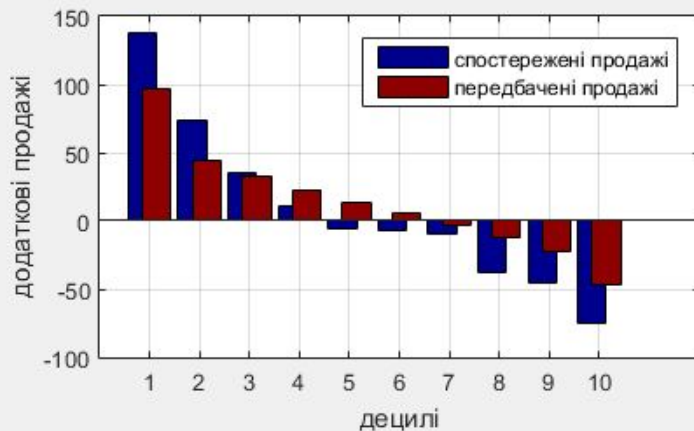
Передбачене приростання приблизно в чотири рази вище, ніж середній приріст відгуків (7,83 %)



Результати: з вибором змін

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

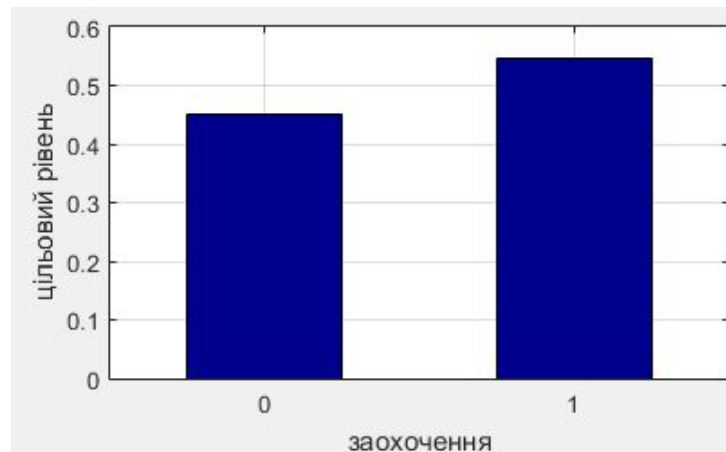
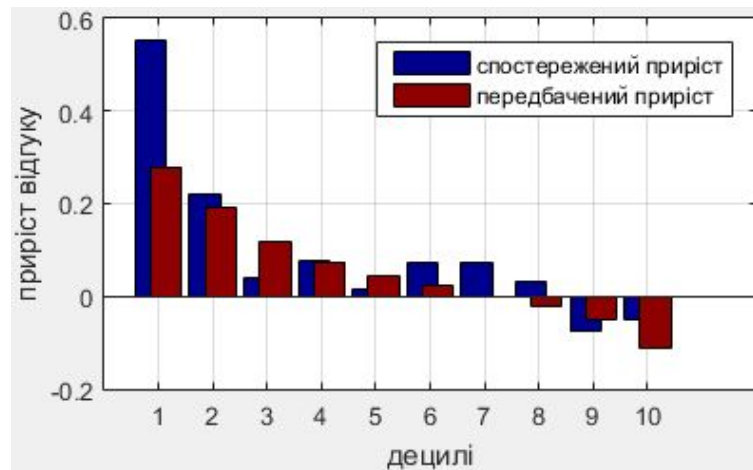
Клієнти вважаються вигідними, якщо додатковий прибуток перевищує вартість маркетингового контакту, як бачимо 70% клієнтів гарантує позитивний прибуток, перші 10% генерують 58,28\$ додаткового прибутку.



Результати: комбінована модель

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

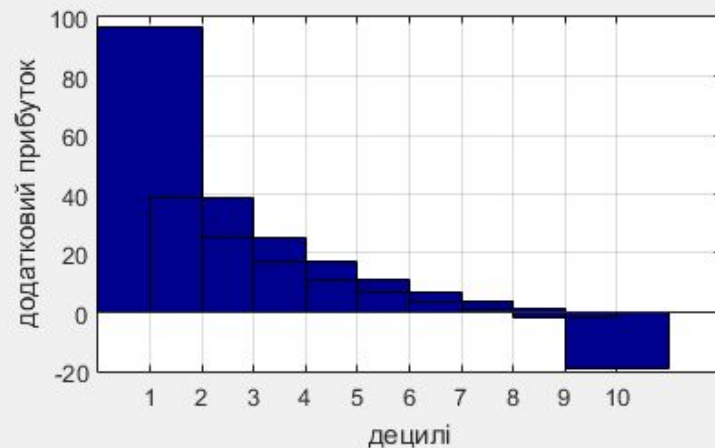
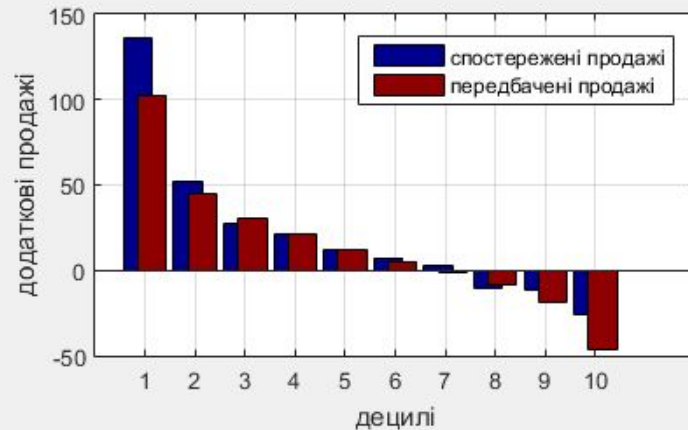
Приріст : Верхній дециль має найвищу частоту додаткових відгуків, з передбаченим прирощенням 24,5 % і спостережним приростом 57,2 %. Передбачене прирощення приблизно в три рази вище, ніж середній приріст відгуків (9,5 %)



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	0.1962	0.2837	0.6850	0.3701
Модель2	0.1878	0.2731	0.7106	0.4212
Модель3	0.2244	0.3864	0.6217	0.2435

Висновки

- Розглянуті методи прогнозування відгуку в маркетингу баз даних : традиційні методи моделювання, моделювання підвищення відгуку, моделювання додаткових продажів
- Розроблено програму для знаходження оптимальних цільових клієнтів, використовуючи різні 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

Новизна

- Запропонована модифікація методу прогнозування підвищення відгуку на основі використання цілого набору даних
- Розроблена власна програма на мові програмування Matlab для пошуку оптимальних клієнтів із використанням запропонованої модифікації
- Виконана апробація запропонованої методики на даних продажів попередньої маркетингової кампанії

Публікації

- Using modified logistic regression to increase customer response rate to marketing campaigns/ Y.Ivanova, O.Terentiev, L.Korshevnyuk , T.Prosyankina-Zharova // САИТ - 2016. - 304 с.
- Моделювання відгуку клієнтів - покупців з використанням методів регресійного аналізу/ Ю.В.Іванова, О.М. Терентьев,Т.І. Просянкінa-Жарова // Наукові вісті НТУУ “КПІ” №5 - 2016. [подано]

Дякую за увагу!