

УДК 004.942

Логін Вадим Вікторович

студент

Національний технічний університет України

«Київський політехнічний інститут»

Логин Вадим Викторович

студент

Национальный технический университет Украины

«Киевский политехнический институт»

Lohin V.

student

National Technical University of Ukraine “Kyiv Polytechnic Institute”

**ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ МАСОВОЇ РЕКЛАМНОЇ
РОЗСИЛКИ
ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ МАССОВОЙ РЕКЛАМНОЙ
РАССЫЛКИ
IMPROVING THE EFFICIENCY OF THE MASS ADVERTISING
MAILING**

Анотація: Досліджено задачу підвищення ефективності масової рекламної розсилки.

Ключові слова: масова рекламна розсилка, прямий маркетинг, підвищення ефективності реклами, методологія Data Mining.

Аннотация: Исследовано задачу повышения эффективности массовой рекламной рассылки.

Ключевые слова: массовая рекламная рассылка, прямой маркетинг, повышение эффективности рекламы, методология Data Mining.

Summary: The problem of improving the efficiency of the mass advertising mailing was investigated.

Key words: mass advertising mailing, direct marketing, increasing the effectiveness of advertising, Data Mining methodology.

Вступ. Однією з поширених сфер, у якій потрібна оцінка ефективності, підвищення ефективності та порівняння моделей, є прямий маркетинг (з англ. *direct marketing*) – пряма особиста комунікація з отримувачем повідомлення або клієнтом середовища «бізнес для бізнесу»,

з метою побудови взаємовідношень та отримання прибутку [1, 2]. Головна задача у даній сфері – мінімізація кількості розсилок клієнтам, які не відгукуються на рекламу. Якщо проводити розсилку без попереднього аналізу клієнтської бази, то затрати на таку розсилку можуть виявитись вище доходу, отриманого від продажу товарів, що рекламуються.

Бажання людини придбати той чи інший товар залежить від багатьох ознак, що описують його як потенціального клієнта. Аналітик може виділити набір таких ознак для кожного виду товарів та побудувати модель, що буде визначати, чи варто включати в список рекламної розсилки клієнта, що має певний набір значень ознак.

Тому, можна з впевненістю сказати що задача підвищення ефективності масової рекламної розсилки є дуже актуальною, оскільки рекламна розсилка є досить поширеною в наш час, причому її проведення без попереднього аналізу може спричинити збитки.

Дослідженням сутності та розробкою методик рекламної розсилки займались такі провідні зарубіжні спеціалісти у цій області як А. Розенспен, Д. Берд, Е. Неш, Б. Стоун, Р. Джейкобс [3 – 6]. Вагомий внесок у вирішення питань пошуку клієнтів та їхнього переконання в купівлі товару зробили такі відомі спеціалісти як Д. Кеннеді, З. Фегеле та ін. [7, 8].

Одним із пріоритетних засобів при проведенні сучасного маркетингу є моделювання процесів, що у ньому відбуваються, із використанням методології Data Mining.

Дослідження, що описується у даній статті, присвячене підвищенню ефективності масової рекламної розсилки. Зокрема наведено опис побудови моделей відгуку клієнтів на рекламу та вибору кращої моделі.

Постановка задачі. На початку дослідження була поставлена задача підвищення ефективності масової рекламної розсилки. Опишемо детальніше поставлену бізнес-задачу. Деяка торгова компанія, що

здійснює продаж товарів, має інформацію про своїх клієнтів та їхні покупки. Компанія провела рекламну розсилку своїм клієнтам та отримала певний відсоток відгуків. Необхідно побудувати моделі відгуку та проаналізувати результати, щоб запропонувати шляхи мінімізації витрат на нові поштові розсилки. Для того, щоб розв'язати поставлену задачу, потрібно здійснити такі кроки:

- виконати загальний огляд методів Data Mining та вибрати ті, які відповідають поставленій проблемі моделювання відгуків клієнтів на рекламну розсилку;
- побудувати моделі Data Mining за допомогою обраних підходів;
- порівняти отримані моделі та вибрати найбільш ефективну з точки зору отриманого прибутку;
- за вхідні дані використати статистичні дані деякої торгової компанії, а на виході отримати найбільш ефективну стратегію проведення рекламної розсилки.

Застосування методології Data Mining. За визначенням, Data Mining полягає у виявленні в необроблених даних раніше невідомих, нетривіальних, практично корисних і доступних для інтерпретації знань, що необхідні для прийняття рішень в різних сферах діяльності людини.

Data Mining – це не один метод, а сукупність великого числа різних методів виявлення знань. Існує кілька умовних класифікацій задач Data Mining. Варто зупинитись на чотирьох базових класах задач: *класифікація, регресія, кластеризація, асоціація*.

Оскільки, проведення рекламної кампанії без попереднього аналізу може спричинити збитки для компанії, тому аналітик повинен виділити набір певних ознак для кожного виду товарів та побудувати моделі відгуку клієнтів на рекламну розсилку.

Оскільки відгук клієнтів приймає тільки два значення – був відгук та не було відгуку, то побудова моделі відгуку зводиться до задачі

класифікації. Тому за допомогою методології Data Mining потрібно побудувати моделі, що будуть розв'язувати поставлену задачу класифікації.

Вибір методів Data Mining для моделювання відгуків клієнтів. На початку дослідження була поставлена задача підвищення ефективності масової рекламної розсилки. Проаналізуємо поставлену задачу, та визначимо шляхи її розв'язання. Оскільки ефективність розсилки напряму залежить від відгуку клієнта, то логічно, що сформульована вище задача зводиться до бінарної класифікації. Отже, потрібно розв'язати задачу класифікації методами Data Mining. Оберемо деякі методи розв'язання поставленої задачі класифікації. Так як маємо справу із бінарною класифікацією, то серед статистичних методів варто вибрати логістичну регресію, що призначена саме для таких типів задач. Поміж методів машинного навчання доцільно вибрати дерева рішень та нейронні мережі.

Вибір інструментальної платформи для здійснення моделювання.

У даному дослідженні, для здійснення моделювання відгуків клієнтів, я використовував аналітичну платформу Deductor, оскільки аналітична платформа дає змогу створювати завершені аналітичні рішення та містить в собі потужні засоби Data Mining. Я зупинився саме на платформі Deductor, тому що вона має інтуїтивно зрозумілий інтерфейс, має вбудовані сучасні методи отримання, візуалізації даних та їхнього подальшого аналізу.

Моделювання масової рекламної розсилки із використанням методології Data Mining. При побудові моделей відгуку клієнтів на рекламну розсилку можна виділити декілька етапів.

Етап 1. Вибір даних для обробки. Набір даних для аналізу є навчальним та був взятий із навчального набору даних, що постачається

разом із роботою [9]. Даний набір містить інформацію про 13504 клієнтів, включаючи відомі відгуки на рекламну розсилку.

Етап 2. Розбиття вихідної вибірки на навчальну та тестову. Після того, як вихідні дані були імпортовані до аналітичної платформи, проведемо розділення вихідної вибірки на навчальну та тестову. Спочатку розділимо у співвідношенні 60/40 %, а вже потім, після побудови моделей, повернемося до даного кроку, щоб вибрати оптимальне розбиття.

Етап 3. Проведення кореляційного аналізу. Поки що будемо працювати із навчальною вибіркою. Спочатку оцінимо вплив вхідних змінних на вихідну. Скористаємось обробником «Кореляційний аналіз» та відкриємо візуалізатор «Матриця кореляції» (рис. 1).

Матриця кореляции			
Входные поля			Корреляция с выходными полями
№	Поле	Відгук	
1	Вік	<input type="text"/>	0,000
2	Стать	<input type="text"/>	-0,007
3	Скільки років є клієнтом	<input type="text"/>	0,127
4	Кількість позицій товарів	<input type="text"/>	0,341
5	Дохід з клієнта, тис. од.	<input type="text"/>	0,363
6	Загальна кількість покупок	<input type="text"/>	0,391
7	Звернень в службу підтримки	<input type="text"/>	0,009
8	Затримки платежів	<input type="text"/>	0,011
9	Дисконтна карта	<input type="text"/>	-0,005

Рис. 1. Кореляція з полем «Відгук»

Як бачимо із останнього рисунка, поля «Скільки років є клієнтом», «Кількість позицій товарів», «Дохід з клієнта, тис. од.» та «Загальна кількість покупок» сильно впливають на відгук на розсилку та демонструють позитивний зв'язок. Тому їх потрібно включити до моделі. У той же час, усі інші поля мають низьку кореляцію із полем «Відгук», тому їх не потрібно включати до моделі.

Етап 4. Оцінка рівномірності вихідних класів. Після цього приступимо до ще одного важливого етапу. Відкриємо статистичні характеристики для навчальної вибірки (рис. 2).



Рис. 2. Гістограма для атрибута «Відгук» для навчальної вибірки

Як бачимо, доля клієнтів із позитивним відгуком в навчальній множині складає 14,4 %. Тобто розподіл класів у вихідній змінній є нерівномірним. Як відомо, в такому випадку бажано будувати модель не на всій доступній множині прикладів, а рекомендується попередньо врівноважити їх.

Етап 5. Проведення моделювання відгуку клієнтів. Розпочнемо моделювання із побудови логістичної регресії. Оскільки розподіл вихідних класів є нерівномірним, тому здійснимо їхнє врівноваження. В логістичній регресії для збалансування класів потрібно відповідним чином розрахувати поріг відсікання. Для цього будемо використовувати правило Байєса. Відомо, що відношення втрат обох типів дорівнює $\frac{C_{FN}}{C_{FP}} = 17/3$. Тоді, відповідно до правила Байєса, поріг відсікання потрібно встановити рівним:

$$\frac{1}{1 + (C_{FN}/C_{FP})} = \frac{1}{1 + 17/3} = \frac{3}{20} = 0,15.$$

Тепер побудуємо модель на основі логістичної регресії на даних навчальної вибірки. При цьому вхідні поля були відібрані при кореляційному аналізі, а вихідне поле – це відгук. У нашому випадку позитивним результатом буде вважатись наявність відгуку.

В задачі даного типу потрібно також провести аналіз Lift-діаграми, яка дає змогу визначити розмір вибірки, який дасть найкраще співвідношення витрат (кількості розсилок) та прибутку (кількості отриманих відгуків). На рис. 3 зображена отримана Lift-крива. Діагональна лінія відображає роботу некорисного класифікатора, тобто ситуацію, коли списки отримувачів розсилки формуються випадковим чином. Як відомо,

чим вище над діагональною лінією проходить графік кривої, тим кращою є модель.

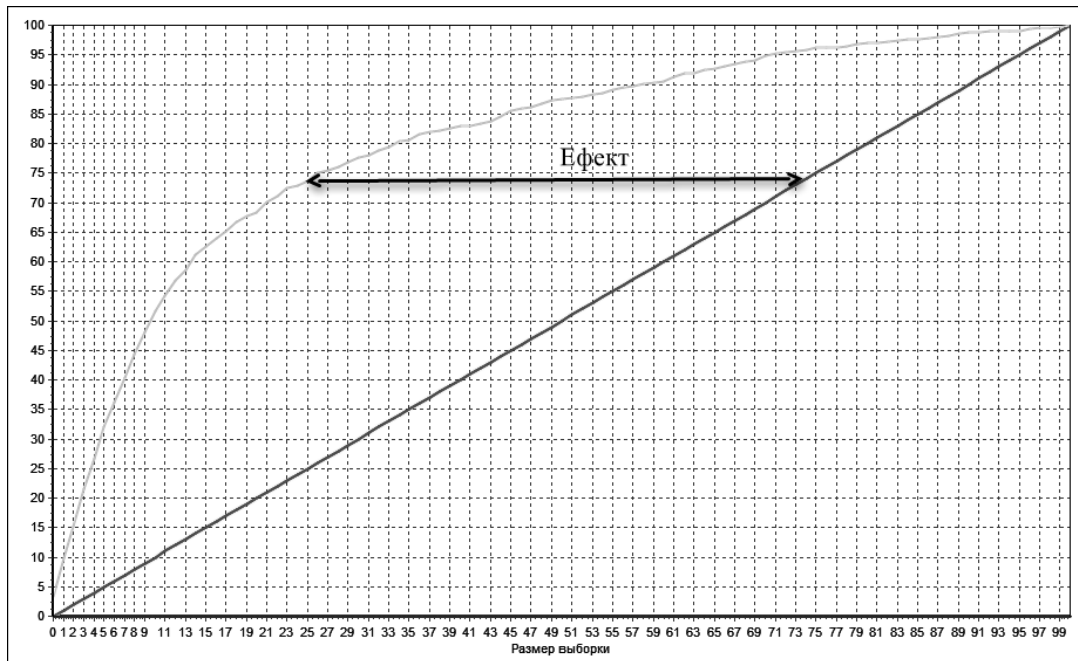


Рис. 3. Lift-крива для моделі відгуку на основі логістичної регресії

Також у даній ситуації корисно припустити, що компанія могла раніше використовувати правило, згідно з яким розсилка здійснювалась в першу чергу тим клієнтам, які принесли найбільші доходи. Тому побудуємо Lift-криву для такого випадку розсилки (рис. 4). Як виявилось, дана Lift-крива є гіршою від кривої для логістичної регресії. Але при невеликих об'ємах розсилки (до 7 %) не поступається їй за ефективністю.

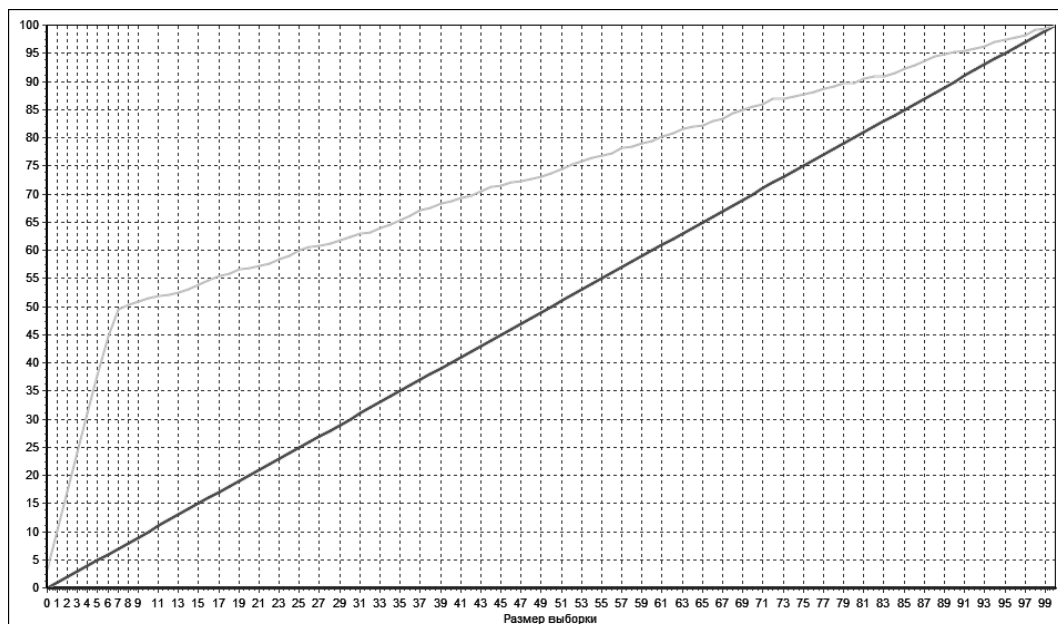


Рис. 4. Lift-крива для випадку першочергової розсилки клієнтам, від яких був найбільший дохід

Тепер побудуємо модель на основі дерева рішень. Як вже було зазначено раніше, класи є незбалансованими, тому проведемо спершу врівноваження. У даному випадку ми використовуємо стратегію семплінгу, а саме – undersampling, оскільки маємо досить велику кількість прикладів (13504). Як відомо, відношення втрат від помилкової класифікації дорівнює 17/3. Тоді, відповідно до правила, потрібно пожертвувати приблизно 14/17 прикладів з клієнтами, від яких не було відгуку.

При налаштуванні обробника «Дерево рішень» вхідні та вихідні поля встановлюємо аналогічно із моделлю логістичної регресії. Побудова дерева буде здійснюватися за алгоритмом C4.5. В результаті роботи алгоритму було виявлено 7 правил; точність класифікації на навчальній вибірці становить 81,83 %. Візуалізатор «Дерево рішень» дозволяє побачити отриманий набір правил в схематичному вигляді, а також виводить показники достовірності та підтримки для кожного вузла (рис. 5). Це і є модель відгуку на основі дерева рішень.

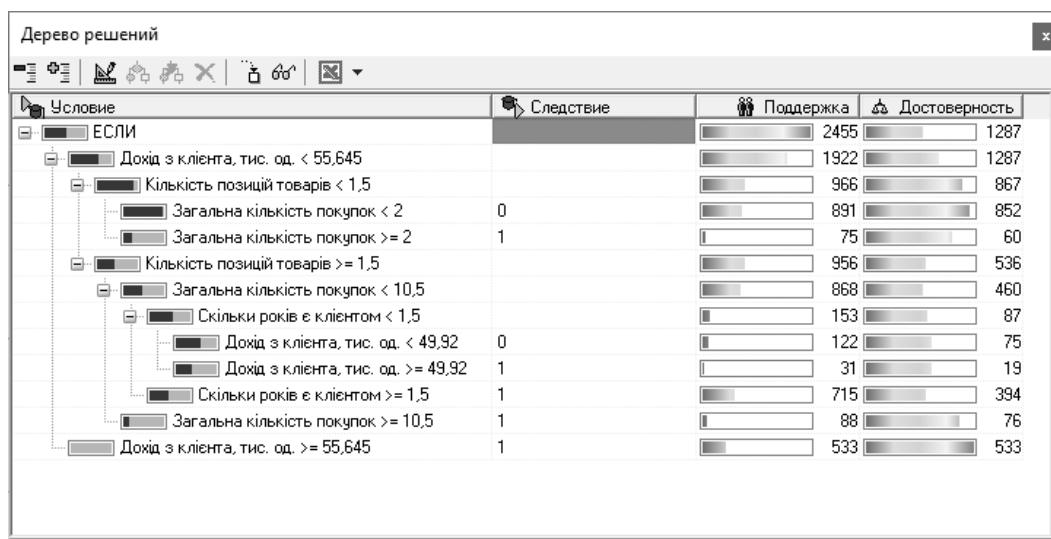


Рис. 5. Дерево рішень для моделі відгуку

Нарешті, приступимо до побудови моделі на основі нейронної мережі. Спершу виконаємо балансування вихідних класів. Як і у випадку з

деревом рішень, використаємо процедуру undersampling. Відповідно до неї потрібно відкинути приблизно 14/17 прикладів з клієнтами, від яких не було відгуку.

При налаштуванні обробника «Нейромережа» вхідні та вихідні поля встановлюємо аналогічно із моделлю логістичної регресії. Як відомо, із усіх архітектур нейронних мереж, багат шаровий перцептрон разом із сигмоїдною функцією активації є базовим для розв'язання задач класифікації та регресії. У задачах Data Mining найбільш популярним є тришаровий перцептрон. Кількість нейронів у внутрішньому рівні, відповідно до деяких обґрунтованих рекомендацій, має дорівнювати половині суми чисел вихідних та вхідних нейронів [10].

Тобто можна зробити висновок, що для нашої задачі достатньо одного прихованого шару із трьома нейронами. Для навчання виберемо алгоритм Back Propagation (BackProp), після навчання – візуалізатор «Граф нейромережі» (рис. 6).

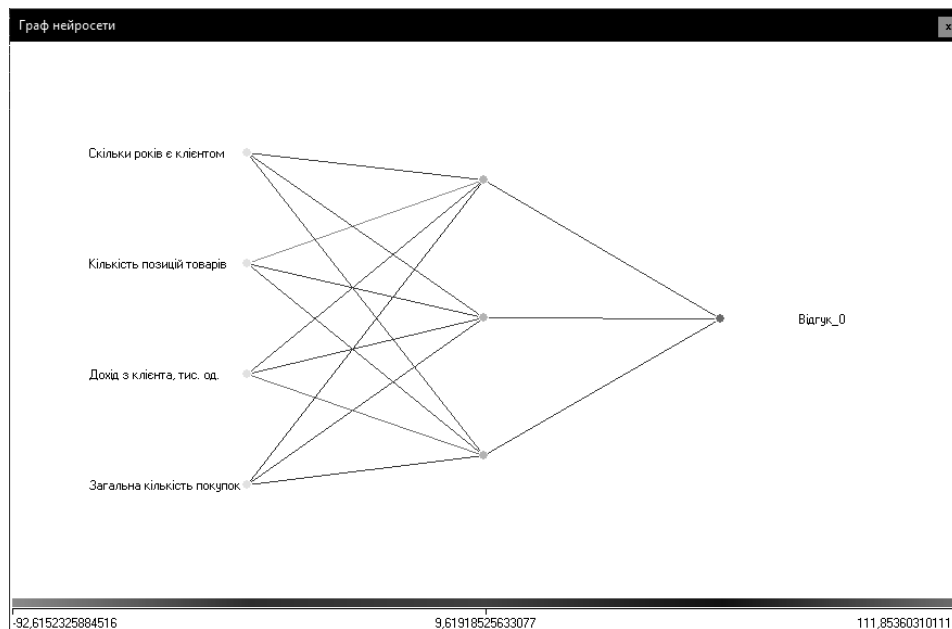


Рис. 6. Граф нейромережі

Етап 6. Вибір оптимального розбиття вихідної вибірки. Для порівняння варіантів розбиття, здійснимо наступні кроки. Як вже зазначалось раніше, будемо розбивати вихідну множину в таких

співвідношеннях: 60/40 %, 70/30 % та 80/20 %. Після цього порівнюємо отримані варіанти розбиття за точністю класифікації та виберемо найкращий. Саме з цим варіантом будуть проводитись подальші дослідження.

У таблицях 1, 2, 3 наведені показники якості моделей для варіантів розбиття 60/40 %, 70/30 %, 80/20 % відповідно.

Таблиця 1

Показники якості моделей для розбиття 60/40 %

Тип вибірки	Модель	AUC	Gini, %	KS, %	Точність класифікації, %
Навчальна	Логістична регресія	0,84	68,75	55,57	81,03
	Дерево рішень	–	–	–	82,28
	Нейронна мережа	–	–	–	80,52
Тестова	Логістична регресія	0,85	70,5	57,47	82,01
	Дерево рішень	–	–	–	76,47
	Нейронна мережа	–	–	–	87,86

Таблиця 2

Показники якості моделей для розбиття 70/30 %

Тип вибірки	Модель	AUC	Gini, %	KS, %	Точність класифікації, %
Навчальна	Логістична регресія	0,85	70,1	57,16	81,6
	Дерево рішень	–	–	–	82,67
	Нейронна мережа	–	–	–	80,14
Тестова	Логістична регресія	0,84	68,3	54,51	80,99
	Дерево рішень	–	–	–	74,94
	Нейронна мережа	–	–	–	87,09

Таблиця 3

Показники якості моделей для розбиття 80/20 %

Тип вибірки	Модель	AUC	Gini, %	KS, %	Точність класифікації, %
Навчальна	Логістична регресія	0,85	69,3	55,75	81,1
	Дерево рішень	–	–	–	81,56
	Нейронна мережа	–	–	–	80,16
Тестова	Логістична регресія	0,85	70,4	57,8	82,71
	Дерево рішень	–	–	–	75,53
	Нейронна мережа	–	–	–	87,97

Отже, після порівняння даних в наведених вище таблицях, можна зробити висновок, що зміна тестової вибірки в межах 20-40 % незначною мірою впливає на якість класифікації. Проте, серед наведених варіантів розбиття кращим виявився варіант 60/40 %.

Етап 7. Вибір кращої моделі відгуку клієнтів на рекламну розсилку. На даному етапі зведемо разом результати класифікації та прибутки для побудованих моделей на тестовій множині та для моделі «Розіслати всім» (табл. 4).

Таблиця 4

Результати класифікації

Модель	TN	TP	FN	FP	Загальна помилка	Прибуток, од.
Логістична регресія (поріг 0,15)	3859	571	213	759	17,99 %	7430
Дерево рішень (undersampling, 17/3)	3430	701	83	1188	23,53 %	8353
Нейронна мережа (undersampling, 17/3)	4206	540	244	412	12,14 %	7944
«Розіслати всім»	0	784	0	4618	—	-526

Як бачимо із табл. 4, зусилля, що буди прикладені на врівноваження вихідних класів, призвели до того, що на тестовій множині класифікатори частіше помилялись в бік псевдопозитивних випадків, що добре у нашому випадку, оскільки втрати від пропусків потенціального клієнта є більшими, ніж втрати від розсилки тим, хто не реагує на неї.

Також із таблиці 4 видно, що з точки зору отриманого прибутку найкращою моделлю для передбачення відгуку клієнта є дерево рішень, що дає прибуток 8353 од. Хоча варто зазначити, що з точки зору мінімальної похибки кращою є модель «Нейронна мережа».

При використанні моделі, що заснована на дереві рішень, ми можемо спостерігати підвищення ефективності розсилки на $8353 - (-526) = 8879$ од.

Висновки. Отже, в результаті проведення дослідження, для розв'язання поставленої задачі було побудовано 3 моделі з використанням методології Data Mining: логістична регресія; дерево рішень; нейронна мережа.

Після порівняння побудованих моделей за передбачуваним прибутком, виявилось, що найбільш ефективною є модель на основі дерева рішень. При використанні даної моделі спостерігається підвищення ефективності рекламної розсилки, у порівнянні з моделлю «Розіслати всім», на 8879 од.

Таким чином, ми отримали нову стратегію рекламної розсилки, що є найбільш ефективною та повинна приносити прибуток, на відміну від стратегії «Розіслати всім». При наступних рекламних розсилках рекомендовано використовувати стратегію, що заснована на моделі дерева рішень.

Новизна отриманих результатів полягає у тому, що при проведенні рекламної розсилки пропонується використовувати не стратегію «Розіслати всім» чи першочергової розсилки клієнтам, від яких був найбільший дохід, а стратегію, що заснована на моделі відгуку клієнтів на основі дерева рішень.

Література:

1. Директ-маркетинг [Електронний ресурс] – Режим доступу: http://www.marketch.ru/marketing_dictionary/marketing_terms_d/direct_marketing/.
2. Прямий маркетинг [Електронний ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/Прямий_маркетинг.
3. Розенспен А. Исповедь одержимого эффективностью / Алан Розенспен. – М.: Ин-Октаво, 2005. – 368 с.

4. Бёрд Д. Прямой маркетинг. Бизнес здравомыслящих / Драйтон Бёрд. – М.: Олимп-Бизнес, 2004. – 400 с.
5. Nash E. Direct Marketing: Strategy, Planning, Execution / Edward Nash. – Columbus: McGraw-Hill Education, 2000. – 600 p.
6. Стоун Б. Директ-маркетинг: эффективные приемы / Б. Стоун, Р. Джейкобс. – М.: Гребенников, 2005. – 616 с.
7. Кеннеди Д. Жесткий директ-маркетинг: Заставьте покупателя достать бумажник / Дэн Кеннеди. – М.: Альпина Паблишер, 2014. – 175 с.
8. Фёгеле З. Директ-маркетинг. 99 практических советов, как найти потребителя / Зигфрид Фёгеле. – М.: Интерэксперт, 2001. – 256 с.
9. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям (+CD): учебное пособие. 2-е изд., испр. / Н. Б. Паклин, В. И. Орешков. – СПб.: Питер, 2013. – 704 с.
10. Ситник В. Ф. Интелектуальний аналіз даних (дейтамаїнінг): навч. посібник / В. Ф. Ситник, М. Т. Краснюк. – К.: КНЕУ, 2007. – 376 с.