

Система підтримки прийняття рішень задоволення рекламного попиту на основі дерева рішень

Зоренко Олег, НТУУ “КПІ ім. І. Сікорського” ІПСА
Наук. кер.: д.т.н., проф. Данилов В. Я.

Предметна область

- **Об'єкт дослідження:** приховані закономірності в даних про користувачів. Дані про користувачів.
- **Предмет дослідження:** методи, алгоритми обробки основані на деревах рішень.
- **Мета дослідження:** аналіз інформації про користувачів які переглядають рекламу, з використанням дерев рішень задля їх класифікації

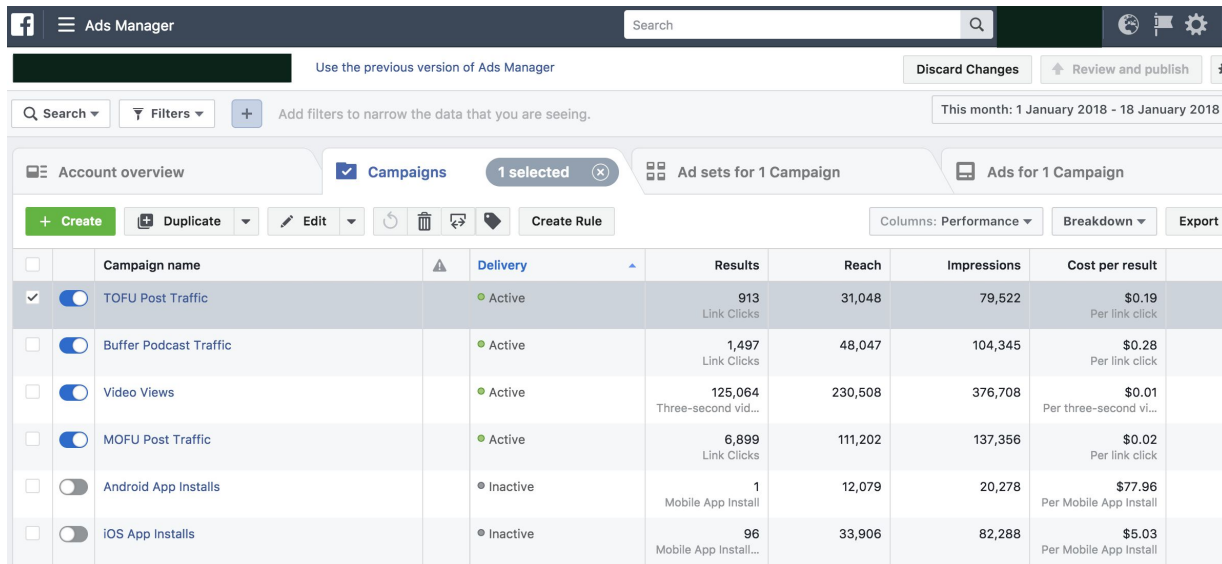
Постановка задачі

- Дослідити існуючі системи для створення реклами.
- Дослідити теоретичні джерела, що описують технології створення реклами.
- Розробити власний ПП.
- Сформулювати висновки щодо отриманого ПП та можливі напрямки його вдосконалення.

Актуальність теми



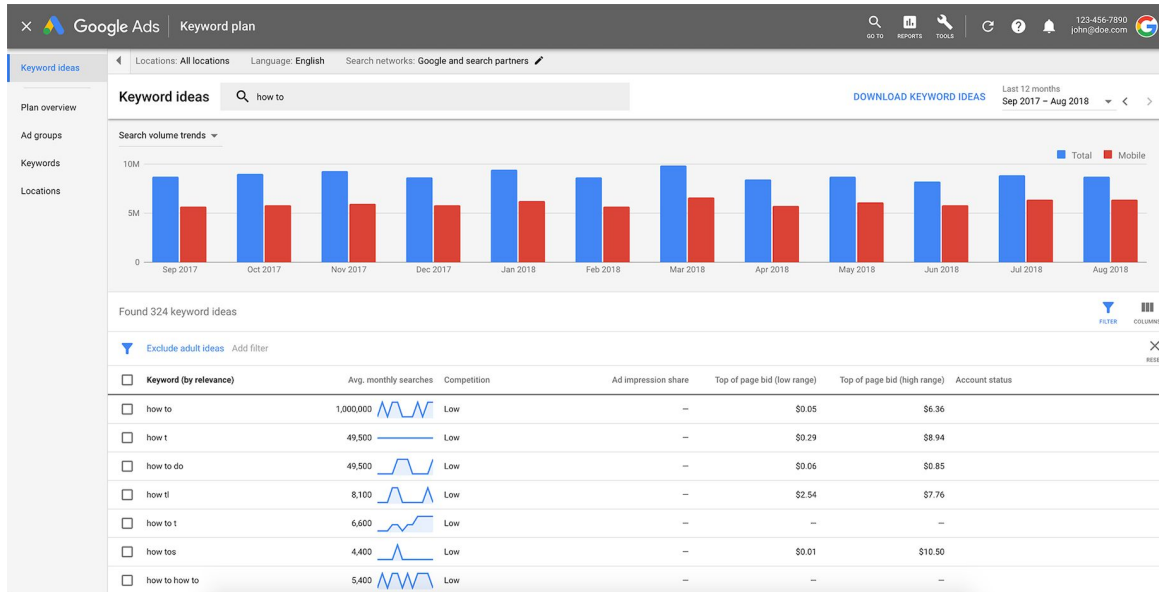
Існуючі системи для створення реклами - Facebook Ads



The screenshot displays the Facebook Ads Manager interface. At the top, there is a navigation bar with the Facebook logo, a search bar, and a date range selector set to "This month: 1 January 2018 - 18 January 2018". Below the navigation bar, there are tabs for "Account overview", "Campaigns" (1 selected), "Ad sets for 1 Campaign", and "Ads for 1 Campaign". The "Campaigns" tab is active, showing a table of campaigns. The table has columns for Campaign name, Delivery, Results, Reach, Impressions, and Cost per result. The first campaign, "TOFU Post Traffic", is selected and has a status of "Active".

	Campaign name	Delivery	Results	Reach	Impressions	Cost per result
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> TOFU Post Traffic	Active	913 Link Clicks	31,048	79,522	\$0.19 Per link click
<input type="checkbox"/>	<input checked="" type="checkbox"/> Buffer Podcast Traffic	Active	1,497 Link Clicks	48,047	104,345	\$0.28 Per link click
<input type="checkbox"/>	<input checked="" type="checkbox"/> Video Views	Active	125,064 Three-second vid...	230,508	376,708	\$0.01 Per three-second vi...
<input type="checkbox"/>	<input checked="" type="checkbox"/> MOFU Post Traffic	Active	6,899 Link Clicks	111,202	137,356	\$0.02 Per link click
<input type="checkbox"/>	<input type="checkbox"/> Android App Installs	Inactive	1 Mobile App Install	12,079	20,278	\$77.96 Per Mobile App Install
<input type="checkbox"/>	<input type="checkbox"/> iOS App Installs	Inactive	96 Mobile App Install...	33,906	82,288	\$5.03 Per Mobile App Install

Існуючі системи для створення реклами - Google Ads



Random forest

Це метод машинного навчання, що складається з безлічі дерев рішень. Замість того, щоб просто усереднювати прогнози різних дерев (така концепція називається просто «ліс»), ця модель використовує дві ключові концепції, які і роблять цей ліс випадковим.

1. Випадкова вибірка зразків з набору даних при побудові дерев.
2. При поділі вузлів вибираються випадкові набори параметрів.

Random forest

Алгоритм:

1. Генеруємо випадкову підвибірку з повторенням розміром n з навчальної вибірки. (Таким чином, деякі приклади потраплять в неї кілька разів, а приблизно $N/3$ прикладів не ввійдуть у неї взагалі).
2. Побудуємо дерево рішень, яке класифікує приклади даної підвибірки, причому в ході створення чергового вузла дерева будемо вибирати ознаку, на основі якої проводиться розбиття, не з усіх M ознак, а лише з m випадково вибраних. Для вибору найкращого з цих m ознак використовується критерій Джині.
3. Дерево будується до повного вичерпання підвибірки

Критерій Джині:

$$Gini(T) = 1 - \sum_{i=1}^n p_i^2,$$

де p_i – ймовірність (відносна частота) класу i в T .

Якщо набір T розбивається на дві частини T_1 та T_2 з числом прикладів в кожному N_1 та N_2 відповідно, тоді показник якості розбиття буде дорівнювати:

$$Gini_{split}(T) = \frac{N_1}{N} \cdot Gini(T_1) + \frac{N_2}{N} \cdot Gini(T_2)$$

Найкращим вважається те розбиття, для якого $Ginisplit(T)$ мінімальне.

Дерева рішень

Загальна схема побудови дерева ухвалення рішень за тестовими прикладам виглядає таким чином:

1. Вибираємо черговий атрибут Q, поміщаємо його в корінь.
2. Для всіх його значень i:
 - 2.1 Залишаємо з тестових прикладів тільки ті, у яких значення атрибута Q дорівнює i
 - 2.2 Рекурсивно будуємо дерево в цьому нащадку



Дерева рішень

В інтелектуальному аналізі даних, дерева рішень можуть бути використані як математичні та обчислювальні методи, щоб допомогти описати, класифікувати і узагальнити набір даних, які можуть бути записані таким чином:

$$(x, Y) = (x_1, x_2, x_3 \dots x_k, Y)$$

Залежна змінна Y є цільовою змінною, яку необхідно проаналізувати, класифікувати й узагальнити. Вектор x складається з вхідних змінних x_1, x_2, x_3 тощо, які використовуються для виконання цього завдання.

Набір характеристик користувача: вікова група, стать, місце знаходження, тип реклами, і т.д. Нам потрібно навчити класифікатор відносити ці дані до одного з 2 класів:

1. Кліку не буде (0)
2. Клік буде (1)

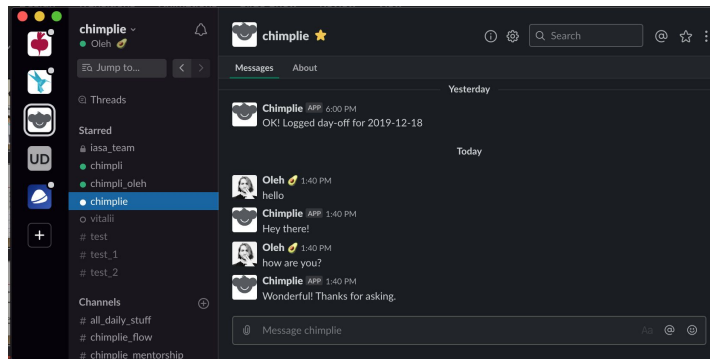
Технології реалізації програмного продукту

Запит:

```
{  
  "age_group": 0,  
  "sex": 1,  
  "location": 100,  
  "type": 2,  
  "theme": 6  
}
```

Відповідь:

```
{  
  "click": 0  
}
```



Python.
Scikit-Learn.
NumPy.
Pandas.
Py2neo.

Висновки

В рамках магістерської роботи було:

- зроблено огляд існуючих систем створення реклами;
- розробка структури СППР та математичної моделі;
- розробка програмного забезпечення;
- розробка стартап-проекту виведення на ринок результатів дослідження;

Покращення

1. Створення веб-порталу.
2. Використання більшої кількості параметрів при навчанні моделі
3. Використання декількох значень для параметра (наприклад, декілька вікових груп)

Дякую за увагу!