Перевірка статистичних гіпотез як основа A/B тестування

Захист кваліфікаційної роботи

Виконала:
Варещук Марія Олегівна, ІПЗ

Наукова керівниця:
Крюкова Галина Віталіївна

- **Постановка задачі**
- Зміст роботи
- Результати

Постановка задачі

Актуальність дослідження:
кожен data-driven бізнес приймає рішення на основі результатів A/B тестування. Відповідно швидкий і коректний розрахунок метрик, необхідних для проведення тесту та аналіз результатів є критично важливою задачею.

Мета дослідження:
розробка веб-застосунку для швидкого та коректного аналізу різноманітних методів A/B тестування. Для досягнення цієї мети потрібно проаналізувати різні методи перевірки статистичних гіпотез, зробити їх порівняльний аналіз для визначення найбільш ефективних в тих чи інших випадках в бізнесі та розробити додаток, який зможе автоматично розраховувати та аналізувати результати тестів.

- Постановка задачі
- **Зміст роботи**
- Результати

Зміст роботи

Було досліджено такі методи перевірки гіпотез:

- Fixed-based horizon
- Frequentist interference
- Sequential testing
- Bayesian testing

Зміст роботи: Fixed-based horizon

Метод, що базується на попередньому визначенні розміру вибірки.

Алгоритм:
1. Обрахувати розмір вибірки
2. Запустити тест
3. Очікувати розмір вибірки
4. Оцінити тест
5. Експеримент закінчено

```
if Nreal>Sample size then
   if pvalue < α then
      if Lift > 0 then
         verdict = «B»
      else
         verdict = «A»
      end if
   else
      verdict = «no difference»
   end if
else
   verdict = «None»
end if
```

Зміст роботи: Frequentist interference

Метод, що базується на динамічному контролі значущості та потужності.

Алгоритм:
1. Запустити тест
2. Оцінити тест
3. Якщо критерій зупинки досягнуто - експеримент закінчено. В іншому випадку, потрібно продовжувати експеримент.

```
if Pvalue < a then
  if power > 1 - B then
    if Lift > 0 then
      verdict = «B»
    else
      verdict = «A»
    end if
  else
    verdict = «None»
  end if
else
  if power > 1 - B then
    verdict = «no difference»
  end if
end if
```

Зміст роботи: Sequential testing

Метод, що може застосовуватись лише для пропорційних метрик.
Так звана «гонка конверсій» – якщо різниця конверсій критична, ми зупиняємо експеримент.

Алгоритм:
1) На початку експерименту зафіксувати розмір вибірки N та тип тесту 50/50;
2) Позначимо за T кількість успіхів у тестовій групі;
3) Позначимо за C кількість успіхів у базовій групі;
4) якщо $T - C >= 2\sqrt{N}$, зупинити тест. Тестова група – переможець.
5) якщо $T + C >= N$, зупинити тест. Переможця немає.

```
if св - са > 2*sqrt(N) then
  verict = «B»
end if
if св + са > N then
  verdict = «no difference»
else
  verdict = «None»
end if
```

Зміст роботи: Bayesian testing

Метод, що базується на прийнятній для нас втраті в конверсії за допомогою дослідження апріорного розподілу.

Алгоритм:
1) Обрахувати апріорний розподіл та критичне значення
2) Запустити тест
3) Якщо кількість конверсій дозволяє прийняти рішення - завершити тест та виявити переможця. В іншому випадку, продовжувати експеримент

$$E[\mathcal{L}_{(A)}] = loss(c_A, n_A, c_B, n_B, \mathrm{A})$$
$$E[\mathcal{L}_{(B)}] = loss(c_A, n_A, c_B, n_B, \mathrm{B})$$
if $[\mathcal{L}_{(A)}] < \varepsilon$ and $[\mathcal{L}_{(B)}] < \varepsilon$ **then**
  verdict = «no difference»
else
  if $[\mathcal{L}_{(A)}] < \varepsilon$ **then**
    verdict = «A»
  end if
  if $[\mathcal{L}_{(B)}] < \varepsilon$ **then**
    verdict = «B»
  else
    verdict = «continue experiment»
  end if
end if

- Постановка задачі
- Зміст роботи
- **Результати**

Результати

1. Було проведено дослідження ефективності того чи іншого методу для різних бізнес-випадків та виведено оптимальну схему використання:

Fixed-based horizon:
- якщо необхідно наперед чітко знати тривалість та вартість тесту

Frequentist interference:
- Continuous метрики

Sequential sampling:
- Тест на низьку конверсію
- Швидке виявлення аномалій

Bayesian:
- Тест на низьку конверсію
- Пропорційні тести
- Негативні тести (правки від юристів, тощо)

Результати

1. Було проведено дослідження ефективності того чи іншого методу для різних бізнес-випадків та виведено оптимальну схему використання:

Приклад результатів порівняння проведення одного і того ж тесту за допомогою sequential testing та bayesian testing:

split_id	n_a	c_a	n_b	c_b	p	power	lift	z_verdict
h4trpai	47751	10576	48264	11023	0.010	0.724	3.118722	None
9i7pbtj	4541	159	4815	199	0.112	0.349	18.035097	None
yx7md19	23999	16310	23978	16128	0.102	0.374	-1.029277	None
t3m534p	34400	26097	34811	23429	0.000	1.000	-11.283355	A

split_id	n_a	c_a	n_b	c_b	P(A>B)	P(B>A)	L(A)	L(B)	b_verdict
h4trpai	47751	10576	48264	11023	0.0052	0.9948	0.0069	0.0000	B
9i7pbtj	4541	159	4815	199	0.0560	0.9440	0.0064	0.0001	B
yx7md19	23999	16310	23978	16128	0.9492	0.0508	0.0001	0.0071	A
t3m534p	34400	26097	34811	23429	1.0000	-0.0000	0.0000	0.0856	A

Результати

2. Було створено систему автоматичного розрахунку та аналізу методів A/B тестування

Інструменти, що використовувались:

Бекенд:
- Веб-фреймворк Flask на основі мови програмування Python
- База даних: MongoDB
- Бібліотека SendGrid для автоматичного відправлення електронних листів про завершення тестів
- Бібліотека APScheduler для виконання фонової задачі щогодинного перевірки результатів тестів

Фронтенд:
- Бібліотека React на основі мови програмування JavaScript
- Бібліотека Axios для взаємодії з сервером

Результати

2. Було створено систему автоматичного розрахунку та аналізу методів A/B тестування

При потраплянні на платформу, користувач може вибрати певний метод тестування, ввести необхідні поля та запустити тест

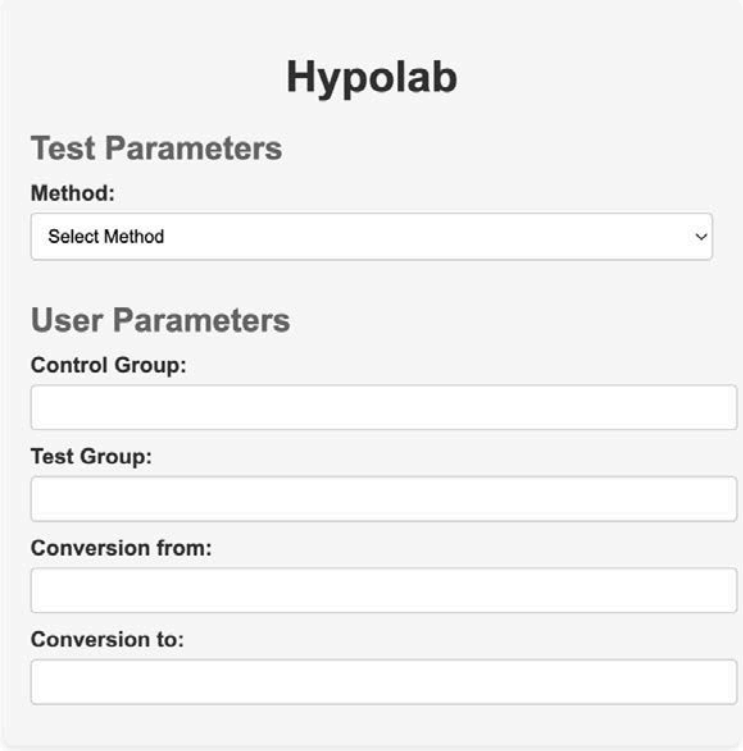

Рисунок 3 - Початкова форму після потрапляння на платформу

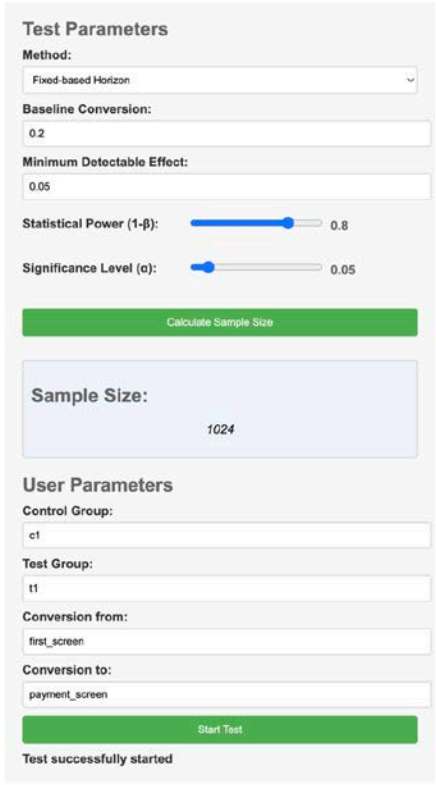

Рисунок 4 - Форма вводу при виборі Fixed-based horizon

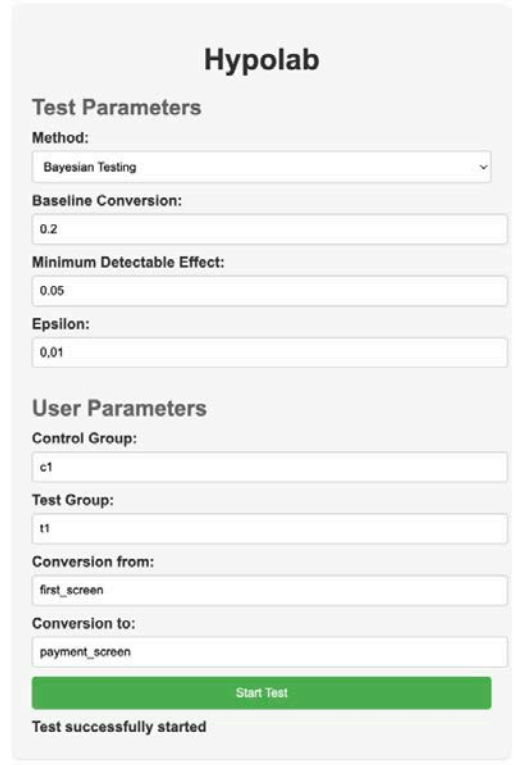

Рисунок 5 - Форма вводу при виборі Bayesian testing

Результати

2. Було створено систему автоматичного розрахунку та аналізу методів A/B тестування

При старті тесту створюється новий запис в базі даних та фонова задача щогодини розраховує результати тесту. Як тільки тест завершено – сповіщаємо користувача електронним листом.

Рисунок 6 - Приклад документу в MongoDB

Рисунок 7 - Приклад електронного листа про завершення тесту

Дякую за увагу!