

IDEA VALIDATION

A/B TESTING METHODOLOGY



PENGERTIAN A/B TESTING



A/B testing adalah metode uji coba **eksperimen** di mana **dua versi atau variasi** dari suatu elemen (misalnya halaman web, iklan, atau produk) disajikan kepada dua kelompok pengguna yang dipilih secara acak, yaitu kelompok kontrol (A) dan kelompok variasi (B).

Tujuannya adalah untuk **mempbandingkan** kinerja kedua variasi tersebut dan menentukan mana yang lebih efektif dalam mencapai tujuan yang diinginkan.

A/B testing biasanya dilakukan dengan mengumpulkan dan membandingkan data hasil dari kedua kelompok untuk menentukan variasi mana yang memiliki **kinerja yang lebih baik dan signifikan secara statistik**.

Istilah lain dari A/B Testing:

A/B Experiment, Split Testing, Variant Testing, Controlled Experimentation, Bucket Testing, Multivariate Testing

PENGERTIAN A/B TESTING

Konsep A/B Testing sudah dipergunakan sejak lama, bahkan bukan dalam konteks online.



Agriculture

“Apa yang akan terjadi kalau saya beri pupuk dengan jumlah lebih banyak?”



Medicine

“Apakah bisa lebih efektif jika kandungan X pada obat ini dibuat 10% lebih banyak?”



Advertising

“Bentuk iklan seperti apa yang akan membuat penjualan mobil lebih besar?”

PENGETERIAN A/B TESTING

Sekarang ini, karena produk digital / online sudah marak dan mendominasi, jika kita mencari informasi tentang "A/B Testing", kebanyakan kita akan menemukan konsep A/B Testing dalam dunia produk digital / online.



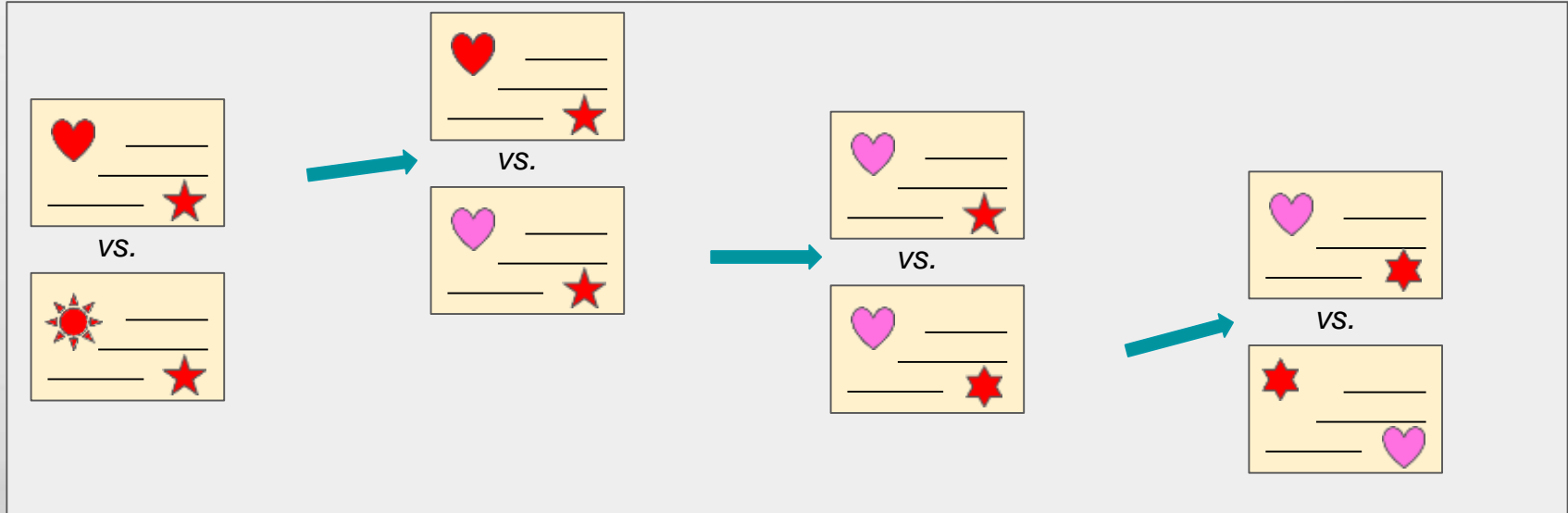
PENGETERIAN A/B TESTING

5 Elemen utama dari A/B Testing:

1. **Variasi:** Versi asli (A) dan versi yang diubah (B) dari elemen yang diuji. Versi asli biasa disebut “Control” dan versi yang diubah biasa disebut “Variant”
1. **Tujuan:** Tujuan atau metrics yang ingin dicapai, seperti meningkatkan jumlah klik, meningkatkan waktu tinggal di halaman, atau meningkatkan penjualan.
1. **Sample Size:** Jumlah pengunjung/user yang akan terlibat dalam pengujian.
1. **Durasi:** Waktu atau periode pengujian yang ditetapkan.
1. **Segmen:** Kelompok pengunjung yang akan terlibat dalam pengujian, seperti pengunjung baru (new user) atau pengunjung yang kembali (returning user).
1. **Alat pengukur / Tools:** Alat yang digunakan untuk mengukur dan menganalisis hasil pengujian. Apakah internal atau menggunakan 3rd party tools.

PENGERTIAN A/B TESTING

Dalam A/B Testing, menguji **satu** perubahan pada satu waktu membantu kita mengidentifikasi perubahan mana yang memiliki efek pada perilaku pengunjung, dan perubahan mana yang tidak berpengaruh.



Seiring waktu, kita dapat **menggabungkan efek dari beberapa perubahan** yang berhasil dari eksperimen untuk menunjukkan peningkatan yang terukur dari pengalaman baru dibandingkan yang lama.

PENTINGNYA A/B TESTING

Kinerja dari web/aplikasi menjadi lebih optimal, karena lebih tahu versi mana yang memberikan hasil terbaik.

**Meningkatkan
kinerja**

**Penghematan
biaya**

Meminimalisir pengeluaran yang tidak diperlukan.

Menghindari penggunaan asumsi atau dugaan (lebih *scientific / data-driven*).

**Pengambilan
keputusan
berdasarkan data**

**Meningkatkan *User
Experience***

Experience dari user di web/aplikasi akan lebih baik dalam jangka panjang.

PENTINGNYA A/B TESTING

Beberapa contoh penerapan A/B Testing untuk tujuan yang berbeda:



1. Untuk berhati hati dalam melakukan peningkatan UI/UX

Contoh: Evaluasi fitur lama yang akan dilakukan perubahan tampilan, dengan tujuan tidak merusak performa saat ini ataupun merusak experience dari user.

Dalam hal ini A/B Testing membantu tim product untuk dapat melakukan rilis ke sebagian kecil user terlebih dahulu. Jika ternyata varian tidak memberikan dampak negatif atau bahkan memberikan dampak positif, maka tim product dapat dengan yakin untuk merilisnya ke 100% user.



2. Untuk mengetahui dampak dari fitur baru atau perubahan di fitur

Contoh: Fitur baru yang kita yakini akan memberikan performa lebih baik pada suatu metrics dapat dibuktikan dengan melakukan A/B Testing. Dengan cara hanya merilis fitur baru ini kepada sebagian user terlebih dahulu.

Biasanya, dalam hal ini tim product sudah mengetahui hasilnya akan lebih baik dan pada akhirnya akan dirilis ke 100% juga, hanya saja untuk mengetahui detail impact yang diberikan, mereka tetap melakukan A/B Testing.

PENTINGNYA A/B TESTING

Beberapa contoh penerapan A/B Testing untuk tujuan yang berbeda:



3. Untuk meningkatkan metrics dengan eksperimen

Contoh: Tidak ada masalah UI/UX ataupun kebutuhan untuk rilis produk baru. Dengan tujuan untuk meningkatkan metrics, tim product dapat melakukan eksperimen dengan harapan, hipotesis mereka dapat terbukti sehingga metrics pun dapat meningkat.

CONTOH A/B TESTING DI INDUSTRI DIGITAL



E-commerce: Sebuah perusahaan e-commerce dapat menggunakan A/B testing untuk membandingkan dua versi dari halaman produk, misalnya, dengan mengubah ukuran gambar produk atau mengubah judul produk, untuk menentukan versi mana yang lebih efektif dalam **meningkatkan konversi** atau penjualan.



Media sosial: Sebuah perusahaan media sosial dapat menggunakan A/B testing untuk membandingkan dua versi iklan, misalnya, dengan mengubah gambar atau teks iklan, untuk menentukan versi mana yang lebih efektif dalam menarik perhatian pengguna dan **meningkatkan click-through rate**.

CONTOH A/B TESTING DI INDUSTRI DIGITAL



Aplikasi mobile: Sebuah perusahaan aplikasi mobile dapat menggunakan A/B testing untuk membandingkan dua versi fitur aplikasi, misalnya, dengan mengubah tata letak atau fungsi tombol, untuk menentukan versi mana yang lebih efektif dalam meningkatkan **retention rate** dan **mengurangi churn**.



Situs web: Sebuah perusahaan situs web dapat menggunakan A/B testing untuk membandingkan dua versi tata letak halaman, misalnya, dengan mengubah warna tombol atau ukuran font, untuk menentukan versi mana yang lebih efektif dalam **meningkatkan konversi** dan **mengurangi bounce rate**.

STEP A/B TESTING



STEP A/B TESTING



Pada tahap Perencanaan, perlu dilakukan:

1. Tentukan apa yang akan dites
2. Buat hipotesis
3. Tentukan variabel yang akan dites
4. Tentukan *Sample Size*
5. Tentukan durasi testing
6. Tentukan metode teknis (traditional vs 3rd party tools)

Step A/B Testing



1. Tentukan apa yang akan dites

Untuk menentukan *funnel* atau produk apa yang akan / dibutuhkan untuk dites, kita dapat menggunakan existing data analytics tools seperti Google Analytics atau internal monitoring tools (manual query).

Perhatikan angka yang ada saat ini dan pikirkan apa yang perlu dioptimalkan.

Studi Kasus Sederhana:

- Dalam flow checkout situs e-commerce X berikut adalah datanya:

○ Jumlah visitor search page	: 10.000	80%
○ Jumlah visitor product detail page	: 8.000	25%
○ Jumlah klik button “Beli”	: 2.000	90%
○ Jumlah klik klik button “Bayar”	: 1.800	

Dari sini kita bisa mengetahui funnel yang paling penting dan butuh diimprove adalah funnel visit product detail page ke klik button “Beli”

Step A/B Testing



1. Tentukan apa yang akan dites (2)

Setelah mengetahui apa yang ingin dites, selanjutnya dapat dituliskan metrics yang berkaitan.

Terdapat 3 metrics yang perlu ditentukan:

- Primary Metrics
- Secondary/Supporting Metrics
- Health Metrics

Primary Metrics

Biasanya sama dengan north-star metrics, bersifat conclusive (memberikan informasi yang cukup dan akurat untuk secara pasti menentukan keberhasilan atau kegagalan). Contoh: conversion rate.

Secondary / Supporting Metrics

Metrics yang memberi gambaran lebih lengkap, sebagai data yang mendukung/memperjelas primary metrics. Contoh: customer service ticket, order cancellation, bounce rate, retention rate.

Health Metrics

Menggambarkan “kesehatan” dari platform atau service. Penting untuk menghindari dampak negatif dari eksperimen yang sedang dijalankan pada metrics-metrics yang tidak disangka. Contohnya: Website speed, backend error, javascript error.

Step A/B Testing

1. Perencanaan

Template kalimat hipotesis:

Berdasarkan *[evidence]*, kita percaya jika kita mengganti *[your change]* untuk *[customer segment]* akan membantu mereka dalam *[impact]*.

Kita akan tahu itu benar jika kita melihat *[your expected change]* di *[primary metrics]*. Hal ini bagus untuk bisnis kita karena dengan meningkatnya/menurunnya *[primary metrics]* akan berdampak kepada peningkatan *[business KPI]*.

2. Buat hipotesis


Setelah mengetahui hal apa yang akan kita tes pada step 1 sebelumnya, selanjutnya perlu dibuat hipotesis. Pembuatan hipotesis adalah langkah yang **sangat penting** dalam keseluruhan A/B Testing, karena hipotesis memberikan **arahan yang jelas** akan tujuan yang ingin dicapai dan membuat kita **fokus** akan tujuan tersebut. Jika punya lebih dari 1 hipotesis, maka prioritaskan 1 terlebih dahulu.

Contoh hipotesis yang relevan dengan studi kasus sebelumnya:

Berdasarkan *hasil riset kualitatif*, kita percaya jika kita mengganti *user interface* dari button “Beli” untuk *pembeli produk di kategori F&B* akan membantu mereka dalam *menemukan dan menentukan pilihan mereka dalam berbelanja*.

Kita akan tahu itu benar jika kita melihat *10% peningkatan di purchase button (click-through-rate) CTR*. Hal ini bagus untuk bisnis kita karena dengan meningkatnya *purchase button CTR* akan berdampak kepada peningkatan *gross merchandise value (GMV)*.

Step A/B Testing



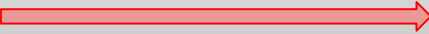
1. Perencanaan

3. Tentukan variabel yang akan dites

Variabel ini dapat termasuk apapun yg ada dalam *design* atau *layout* dari web/aplikasi. Contohnya seperti warna, copywriting, ukuran, posisi/layout dari sebuah objek, jumlah suatu objek, dll.

Jika ada beberapa variabel yang terpikirkan, **prioritaskan 1 variabel** terlebih dahulu yang kamu anggap akan memberikan **dampak terbesar** sesuai dengan tujuan yang ingin dicapai.

Studi Kasus Sederhana:

- Dalam product detail page, berikut adalah beberapa variabelnya:
 - UI tombol beli 
 - Warna/ukuran/copywriting
 - UI gambar produk
 - UI harga produk

Dari aplikasi e-commerce X, saat ini dinilai user interface dari tombol beli masih dapat di-improve, misal dari segi copywriting dapat dites antara “Beli” vs “Beli Sekarang”

Step A/B Testing

1. Perencanaan

Jika sample size telah tercapai, maka penelitian atau pengujian **dapat dihentikan** atau **dilanjutkan**.

Sample size yang telah tercapai dapat dianggap sebagai **titik awal untuk melakukan analisis** data dan mengambil kesimpulan dari penelitian atau pengujian yang dilakukan.

4. Tentukan Sample Size

Sample Size → jumlah user yang diambil dari populasi untuk dijadikan sampel dalam eksperimen. Ditentukan dari:

- **Baseline conversion rate** → Angka conversion rate dari versi saat ini (control group). Didapatkan dari data yang dimiliki saat ini.
- **Minimum detectable effect (MDE)** → “Minimum lift”. Nilai terkecil perbedaan yang dapat dideteksi oleh sebuah eksperimen atau penelitian. Semakin kecil angka MDE, maka akan semakin besar angka sample size (semakin lama punya waktu eksperimen yang dibutuhkan).
- **Statistical significance** → Menunjukkan bahwa perbedaan antara dua kelompok atau perlakuan yang diuji adalah hasil dari perbedaan yang sebenarnya, dan **bukan karena kesalahan pengamatan** atau **kebetulan** semata. Biasanya digunakan **95%** (kemungkinan perbedaan antara dua kelompok atau perlakuan yang diuji terjadi secara kebetulan hanya sebesar 5%, yang dianggap cukup rendah untuk menentukan bahwa perbedaan tersebut signifikan secara statistik).

Untuk menghitung sample size dapat menggunakan tools online seperti:
<https://www.optimizely.com/sample-size-calculator>

Step A/B Testing

1. Perencanaan

Gunakan MDE sebagai panduan bukan prediksi yang pasti.

Konsep eksperimen didasarkan pada kenyataan bahwa **Anda tidak tahu efek apa yang akan dihasilkan dari perubahan yang diberikan**. Alih-alih mencoba untuk menentukan MDE dengan tepat, gunakan perhitungan sebagai panduan untuk menetapkan batasan pada waktu yang Anda bersedia investasikan dan nilai yang Anda harapkan dapat dihasilkan.

4. Tentukan Sample Size (contoh)

Dihitung dengan menggunakan <https://www.optimizely.com/sample-size->

Baseline Conversion Rate	Your control group's expected conversion rate. [?]
<input type="text" value="25"/> %	
Minimum Detectable Effect	The minimum relative change in conversion rate you would like to be able to detect. [?]
<input type="text" value="10"/> %	
Statistical Significance	95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [?]
<input type="text" value="95%"/>	
Sample size per variation	
3,800	

25% conversion rate dengan 10% MDE berarti perubahan +/- 2.5% sudah mencukupi yaitu conversion rate menjadi 22.5% atau 27.5%.

Baseline Conversion Rate	Your control group's expected conversion rate. [?]
<input type="text" value="25"/> %	
Minimum Detectable Effect	The minimum relative change in conversion rate you would like to be able to detect. [?]
<input type="text" value="5"/> %	
Statistical Significance	95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [?]
<input type="text" value="95%"/>	
Sample size per variation	
18,000	

25% conversion rate dengan 10% MDE berarti perubahan +/- 1.25% sudah mencukupi yaitu conversion rate menjadi 23.75% atau 26.25%.

Step A/B Testing

1. Perencanaan

5. Tentukan Durasi Testing

Setelah mendapatkan sample size, untuk mengetahui lamanya (estimasi) waktu yang dibutuhkan dalam melaksanakan A/B testing, dapat digunakan rumus berikut:

#1

$$\text{Sample Size} \times \frac{\text{Total Variasi}}{\text{Total visitor yang dibutuhkan}} = \text{Total visitor yang dibutuhkan}$$

#2

$$\frac{\text{Total visitor yang dibutuhkan}}{\div \text{menjalankan A/B Testing}} \div \text{Rata-rata visitor per hari} = \text{Lamanya hari untuk menjalankan A/B Testing}$$

Pada studi kasus sebelumnya perhitungannya akan menjadi seperti berikut:

Diketahui:

- Sample Size = 3.800
- Total variasi = 2
- Rata-rata visitor per hari = 500

#1

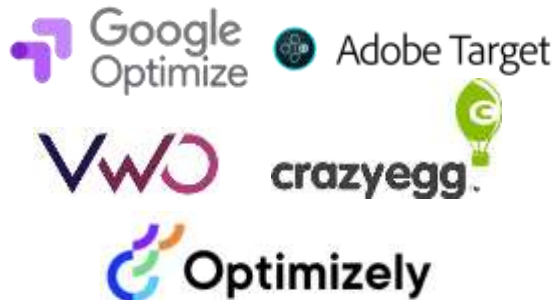
$$3800 \times 2 = 7600$$

#2

$$7600 \div 500 = 15.2 \text{ hari}$$

Step A/B Testing

1. Perencanaan



6. Tentukan Metode Teknis

Proses pelaksanaan A/B Testing dapat dilakukan secara tradisional maupun menggunakan 3rd party tools.

- **Tradisional** berarti tim product dan engineering menggunakan semua internal resource yang ada, mulai dari proses penentuan acak user yang mendapatkan control dan variant, track & store aktivitas user, sampai analisis datanya.
- Jika menggunakan **3rd party tools** untuk eksperimen, tim product dan engineering perlu melakukan beberapa implementasi yang dibutuhkan untuk menghubungkan web/aplikasi dengan 3rd party tools tersebut, sedangkan proses penentuan acak user yang mendapatkan control dan variant, track & store aktivitas user, serta analisis data akan **“dibantu”** oleh 3rd party tools tersebut.
 - Beberapa contoh 3rd party tools untuk eksperimen A/B Testing:
 - Google Optimize
 - Optimizely
 - VWO
 - Crazy Egg
 - Adobe Target

- dari Dalam tahap ini juga perlu ditentukan pembagian persentase (%) masing-masing varian. Normalnya 50:50, atau 33:33:33 jika terdapat 3 varian.

Step A/B Testing



Pada tahap Pelaksanaan, perlu dilakukan:

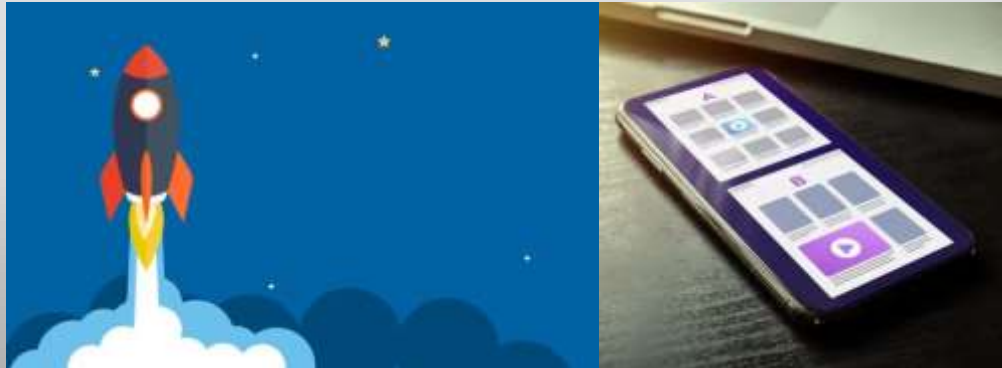
1. Luncurkan A/B Testing yang sudah direncanakan
2. Monitor tes
3. Kontrol faktor eksternal

Step A/B Testing

2. Pelaksanaan

1. Luncurkan A/B Testing yang sudah direncanakan

Berdasarkan perencanaan sebelumnya, eksperimen A/B Testing perlu diluncurkan. Peluncuran ini bisa dilakukan langsung dari dashboard 3rd party tools jika memang menggunakannya, ataupun dengan melakukan deployment jika melakukannya tanpa 3rd party tools. Pastikan konfigurasi segmen user dan varian sudah benar.



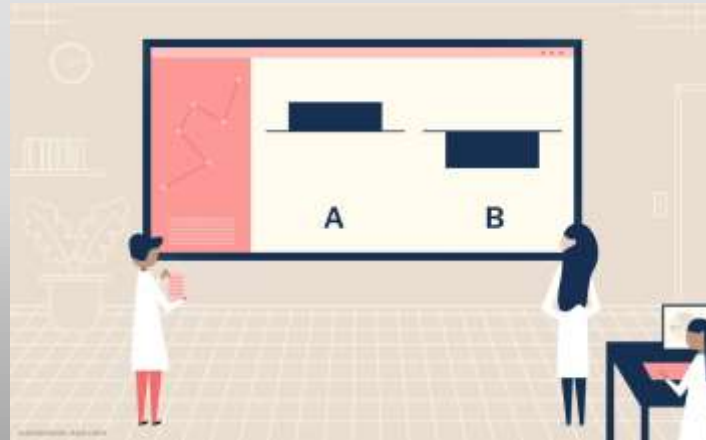
Step A/B Testing

2. Pelaksanaan

2. Monitor tes

Data akan mulai masuk dan tercatat dari waktu peluncuran eksperimen A/B Testing ini. Lakukan pengecekan untuk memastikan data sudah benar-benar masuk dan user sudah terbagi sesuai dengan persentase yang sudah ditentukan di awal.

Jika data sudah mencukupi (sample size terpenuhi), maka kita dapat memutuskan untuk menghentikan atau melanjutkan proses eksperimen.



Step A/B Testing

2. Pelaksanaan

3. Kontrol faktor eksternal

Coba kendalikan faktor eksternal yang dapat mempengaruhi hasil tes. Misalnya, hindari menjalankan eksperimen lain yang bersinggungan langsung dengan eksperimen yang sedang berjalan atau membuat perubahan signifikan pada website atau aplikasi selama periode eksperimen.



Step A/B Testing



Pada tahap Analisis Hasil, perlu dilakukan:

1. Interpretasikan hasil akhir
2. Tentukan pemenang dan implementasikan

Step A/B Testing

3. Analisis Hasil



1. Interpretasikan hasil akhir

Setelah jumlah sample size telah tercapai, kita dapat menghentikan jalannya eksperimen dan mulai menginterpretasikan data yang sudah kita dapat.

Jika menggunakan **3rd party tools** → biasanya hasil sudah dapat langsung dibaca dengan mudah, bahkan sudah langsung ditunjukkan hasilnya varian mana yang “Win” dan mana yang “Lose”.

Jika menggunakan **internal resource** (tanpa 3rd party tools), kita dapat bekerja sama dengan data scientist / data analyst untuk dapat menarik data hasil A/B Testing dan melakukan analisis.

A screenshot of an A/B testing results dashboard. The dashboard shows a table with columns for 'Test Conversion Rate', 'Conversion Rate', 'Significance', and 'Confidence Interval'. There are three rows of data representing different variants. A red box highlights the 'Statistical Significance' column, which shows values of +7% for all three variants. The dashboard also includes a 'Results Summary' button and a 'Full' link.

Variant	Test Conversion Rate	Conversion Rate	Significance	Confidence Interval	Statistical Significance
check_out_flow_1	20.691 (95.91)	0.673	---	---	+7%
check_out_flow_2	20.318 (95.91)	0.688	-0.68%	---	+7%
check_out_flow_3	20.605 (95.91)	0.670	-0.42%	---	+7%

Step A/B Testing

3. Analisis Hasil

2. Tentukan pemenang dan implementasikan

Jika kita sudah memiliki varian mana yang memiliki hasil lebih baik sesuai yang sejalan dengan tujuan eksperimen di awal. Kita dapat menentukan (memenangkan) salah satu dari varian tersebut.

Selanjutnya kita dapat menerapkan secara penuh (100%) varian tersebut untuk dapat digunakan oleh user.

Terdapat kemungkinan versi control yang ternyata memenangkan A/B Testing, jika hal ini terjadi kita dapat mengembalikan aplikasi ke versi control sepenuhnya. Bisa dikatakan jika hipotesis yang sudah dibuat di awal tidak terbukti.



Kesalahan yang sering terjadi pada A/B Testing

1. Melakukan A/B Testing pada periode yang low traffic

Perhatikan faktor eksternal. Sebisa mungkin hindari melakukan A/B Testing pada kondisi traffic yang sedang rendah, seperti misalnya pada e-commerce, biasanya periode libur hari raya memiliki traffic lebih rendah dari biasanya.

Maret						
M	S	S	R	K	J	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

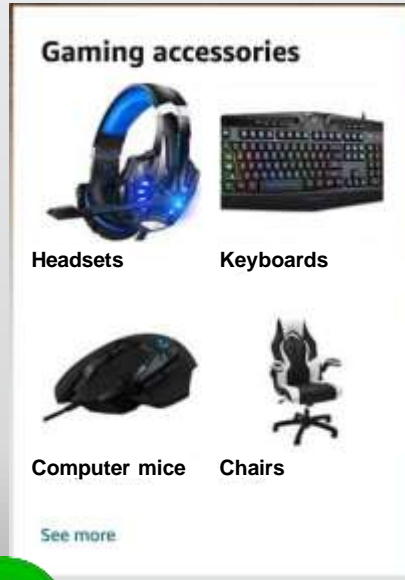
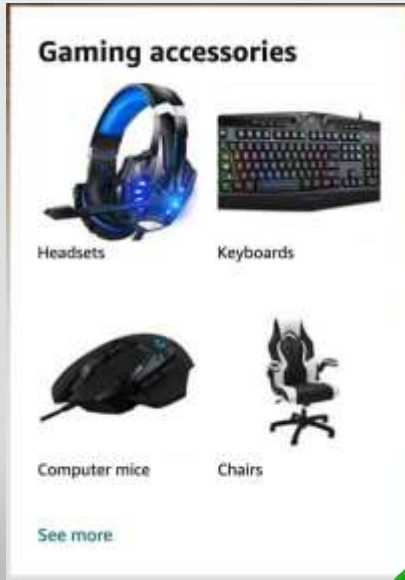
- 20 Maret** ● : Rekomendasi ambil cuti
- 21 Maret** ● : Rekomendasi ambil cuti
- 22 Maret** ● : Hari Raya Nyepi
- 23 Maret** ● : Cuti Bersama Nyepi
- 24 Maret** ● : Rekomendasi ambil cuti

April						
M	S	S	R	K	J	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

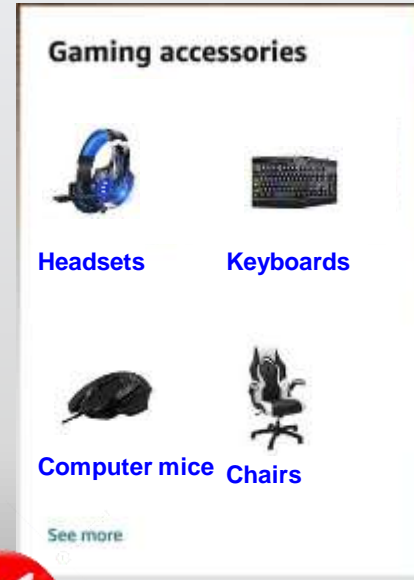
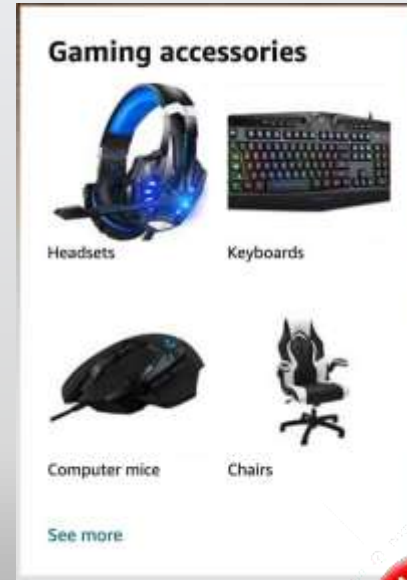
- 21-26 April** ● : Cuti Bersama Idul Fitri
- 27 April** ● : Rekomendasi ambil cuti
- 28 April** ● : Rekomendasi ambil cuti

Kesalahan yang sering terjadi pada A/B Testing

2. Menguji terlalu banyak variabel sekaligus



-Ubah judul kategori menjadi bold



-Ubah judul kategori menjadi bold
-Perbesar ukuran font
-Ubah warna font
-Perkecil gambar ilustrasi

Kesalahan yang sering terjadi pada A/B Testing

3. Tidak Memisahkan Data

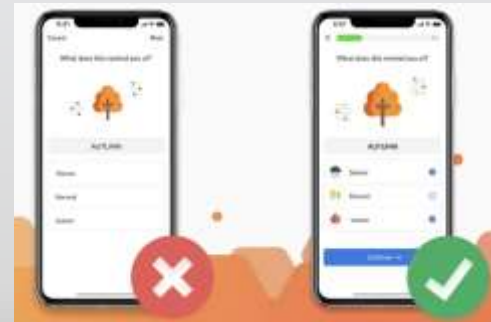
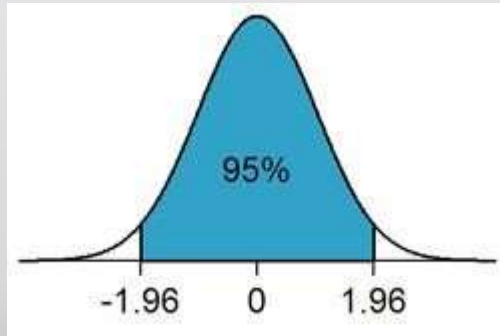
Tidak memisahkan data berdasarkan demografi pengguna, lokasi, atau faktor lainnya bisa menyebabkan hasil yang tidak akurat. Memisahkan data memungkinkan pemahaman yang lebih akurat tentang bagaimana kelompok pengguna yang berbeda merespon perubahan yang diuji.



Kesalahan yang sering terjadi pada A/B Testing

4. Terlalu Bergantung pada Signifikansi Statistik

Signifikansi statistik adalah faktor penting dalam A/B testing, tetapi bukan satu-satunya faktor yang perlu dipertimbangkan. Terlalu fokus pada signifikansi statistik bisa menyebabkan mengabaikan faktor penting lainnya seperti signifikansi praktis dan dampak keseluruhan pada pengalaman pengguna.



Cara menemukan A/B Testing yang sedang berjalan

Studi Kasus

1. Buka website Amazon.com di browser mode normal (non-incognito)
2. Buka website Amazon.com di browser mode incognito
3. Temukan perbedaan yang ada pada kedua website yang sudah dibuka tersebut (Jika belum ada, kamu dapat close seluruh window mode incognito lalu buka kembali websitenya hingga menemukan adanya perbedaan, lakukan berkali-kali jika masih belum menemukannya)

Cara menemukan A/B Testing yang sedang berjalan

