

Stage 0

Brainstorm

(Kamis, 5 Januari 2023)

Homework Week 10

1. Problem statement : Tingkat penerimaan campaign rendah
 2. Role : data scientist / konsultan data scientist
 3. Goal : Peningkatan penerimaan campaign
 4. Objective : Memberikan insight agar dapat menentukan strategi marketing yang tepat sasaran dan untuk mengoptimalkan penerimaan campaign hingga 2x (or 100%) lipat pada campaign selanjutnya.
 5. Business metrics : Campaign acceptance rate
- Accept campaign = beli barang or visit web/melihat barang. *Define accept campaign*
 - **Current accept campaign definition:** customer at least see the campaign via web, then can proceed to buy products or not
 - **Diasumsikan** timeline 2 tahun dan telah dilakukan 6x campaign, sehingga setahun ada 3 campaign dengan durasi 1 campaign per 4 bulan
 - **Fact 1 for problem statement:** dilihat dari 6 campaign sebelumnya, customer yang menerima campaign rata-rata kurang lebih hanya 150 orang
 - **Facts for business understanding:** barang yang dijual apa saja, perbandingan penjualan online dan offline
 - Possible suggestion: kasih campaign ke yg lebih tua bcs diduga yg lebih tua means lebih banyak uang dan lebih possible untuk belanja
 - Marital status dibagi jadi hidup sendiri dan berkeluarga as one of the ways to see possible suggestion

Mentoring

(Jumat, 6 Januari 2023)

Homework Week 10 FIX

1. Problem statement : Tingkat penerimaan campaign rendah
 2. Role : Konsultan data scientist
 3. Goal : Peningkatan penerimaan campaign
 4. Objective : Memprediksi Customer yang berpotensi untuk menerima penawaran campaign untuk memberikan target marketing yang akurat.
 5. Business metrics : Campaign acceptance rate
- Objective sebelumnya (*Memberikan insight agar dapat menentukan strategi marketing yang tepat sasaran dan untuk mengoptimalkan penerimaan campaign hingga 2x (or 100%) lipat pada campaign selanjutnya*) bisa dijadikan business recommendation
 - Problem dan goal cukup 1, tapi harus selaras. Overall sudah oke
 - **Accept campaign definition** = customer menerima tawaran dari campaign dan membeli barang

Stage 1

Brainstorm

(Kamis, 12 Januari 2023)

Homework Week 11

1. A. Kolom 'DT Customer' bertipe string, seharusnya bertipe date
B. Kolom 'income' memiliki 24 baris data null
C. Kolom 'income' memiliki nilai max yang berbeda jauh dari nilai rata-ratanya
2. - Kolom 'year_birth' terdistribusi **negative skew**
- Kolom 'income' memiliki outliers yang sangat jauh dan grafik cenderung ke kiri
- Kolom 'kidhome' dan 'teenhome' terdistribusi **bimodal recency**
- Seluruh kolom **amount products** terdistribusi **positive skew**
- Kolom 'numwebvistmonth' terdistribusi **negative skew**
3. A. Terdapat 14 feature yang relevan, yaitu **acceptedcampaign 1-5, income, mntproducts (6), numweb, numcatalog.**
B.

- acp4, acp2 x mnt wines, yg menerima cmp4, orang yg suka beli wine
- acp5, acp1 x mnt produk, terkecil 0.2, orang yg menerima cmp5 orang yg suka belanja produk2 dengan harga tinggi
- income x kidhome, -0.43, yang kaya jarang punya anak kecil
- income x mnt produk, semuanya hampir diatas 0.44 kecuali mntgold semakin tinggi income, semakin tinggi spent produk
- income x numwebpurchase, numcatalogpurchase, numstorepurchase, semuanya diatas 0.53 kecuali numwebpurchase semakin tinggi income, semakin tinggi sering beli tetapi income tidak ada hubungan dgn numdealspurchase (deal dgn diskon)
- income x numweb, -0.55, yang kaya jarang belanja online
- kidhome x numwebvisit, 0.45, semakin banyak anak semakin sering mengunjungi web
- kidhome x numdealspurchase, 0.22, semakin banyak anak semakin suka makai diskon
- kidhome x mntproduk, numpurchase, terkecil -0.35, yang punya anak jarang belanja
- teenhome x numdealspurchase, 0.22, semakin banyak anak semakin suka makai diskon
- mnt produk saling berkorelasi dg nilai min 0.33
- mnt produk saling berkorelasi dengan numwebpurchase, numcatalogpurchase, numstorepurchase, minimal 0.29,
- mnt produk x numwebvisit, terkecil -0.25, webvisit tidak menghabiskan banyak uang diproduksi
- numdealspurchase x numwebvisit, 0.35, banyak yang menggunakan diskon adalah orang belanja dgn web

4. Business insight

1. Semakin tinggi income, semakin banyak jumlah belanja produk
2. Orang dengan income tinggi lebih suka belanja dari store dan katalog
3. Orang yang nerima campaign terakhir (accepted campaign 5) banyak dari golongan income tinggi
4. Orang dengan income tinggi paling banyak melakukan pembelian melalui katalog dengan produk favorit wine dan meat
5. Orang yang menerima campaign terakhir (accepted campaign 5) lebih berpotensi untuk menerima campaign selanjutnya (response)

Business rec

1. Next campaign bisa ditargetkan ke orang dengan income yang tinggi dan difokuskan pada pembelian melalui katalog berkategori wine dan meat
- 5.
- No. 2 masih ragu
 - No. 3 (multivariate). Ada 14 kolom yg dirasa relevan, kebanyakan atau tidak? How to choose features?
 - Korelasi no.3 terlalu banyak?
 - **Additional insights for no. 4:**
 - income x kidhome, -0.43, yang kaya jarang punya anak kecil
 - income x mnt produk, semuanya hampir diatas 0.44 kecuali mntgold semakin tinggi income, semakin tinggi spent produk
 - income x numwebpurchase, numcatalogpurchase, numstorepurchase, semuanya diatas 0.53 kecuali numwebpurchase semakin tinggi income, semakin tinggi sering beli tetapi income tidak ada hubungan dgn numdealspurchase (deal dgn diskon)
 - income x numweb, -0.55, yang kaya jarang belanja online
 - income x acptcmp5, 0.34, ada korelasi antara income tinggi dgn acptcmpnya
 - basic memiliki income relative kecil, sedangkan yang lain hampir sama di 100000
 - basic jarang beli (mntproduk, numpurchase, numvisit : kecil)
 - customer yg acceptcmp1,2,5, itu minimal sudah datang ke store 3 sampai 4 kali

Mentoring

(Sabtu, 14 Januari 2023, 14.00)

Homework Week 11 FIX

1. A. Kolom '**DT Customer**' bertipe string, seharusnya bertipe date
B. Kolom '**income**' memiliki **24** baris data null
C. - Data yang mean dan median berbeda cukup jauh: '**Kidhome**', '**Teenhome**', '**MntWines**', '**MntFruits**', '**MntMeatProducts**', '**MntFishProducts**', '**MntSweetProducts**'
 - Nilai max yang tidak normal: '**income**'
2. - Kolom '**year_birth**' terdistribusi negative skew karena mean lebih kecil dibandingkan nilai mediannya (ada user yang tahun kelahirannya sangat jauh dibanding user lainnya). Untuk kolom tersebut bisa diproses agar terdistribusi normal
 - Kolom '**income**' memiliki outlier karena nilai maxnya sangat jauh dibandingkan dari nilai mean & median. Nilai outlier kolom ini perlu diproses pada tahap selanjutnya
 - Kolom '**NumWebPurchases**' terdapat outliers, mean hampir sama dengan median dan terdistribusi normal
 - Kolom '**NumCatalogPurchases**', terdapat outliers, mean > median, terdistribusi positif skew
 - Kolom '**NumWebVisitsMonth**' terdapat outliers, mean < median terdistribusi negatif skew
 - Kolom '**NumStorePurchases**' tidak ada outliers, mean > median, terdistribusi positif skew

- Kolom '**kidhome**' dan '**teenhome**' terdistribusi bimodal karena ada 2 puncak (nilai modus tertinggi) pada user yang memiliki 1 anak dan tidak punya anak
 - Kolom '**recency**' terdistribusi normal
 - Seluruh kolom 'amount products' terdistribusi positive skew karena nilai mean lebih besar dibanding nilai mediannya, namun terdapat outlier yang perlu diproses pada tahap selanjutnya
 - Kolom lainnya terdistribusi positive skew karena mean > median
3. A. 11 feature yang dipertahankan: '**Income**', '**MntWines**', '**MntFruits**', '**MntMeatProducts**', '**MntFishProducts**', '**MntSweetProducts**', '**MntGoldProds**', '**NumWebPurchases**', '**NumCatalogPurchases**', '**NumStorePurchases**', '**Marital_Status**'
- B. Korelasi antar feature:
- Korelasi antara kolom '**Income**' dan seluruh kolom 'amount products' berada pada nilai korelasi di atas 0.44, kecuali kolom '**MntGoldProds**', sehingga terlihat trend bahwa semakin tinggi income maka semakin tinggi pembelian produk
 - Korelasi antara kolom '**Income**' dan keseluruhan purchase ('**NumWebPurchases**', '**NumCatalogPurchases**', '**NumStorePurchases**'), semuanya berada di atas 0.53 kecuali pada kolom '**NumWebPurchases**' yang berarti pelanggan dengan income tinggi lebih sering melakukan pembelian melalui toko dan katalog
 - Kolom '**Income**' tidak berkorelasi dengan kolom '**NumDealsPurchases**' dan '**NumWebVisitsMonth**'

4. Business Insight

- Semakin tinggi pendapatan pelanggan maka semakin banyak jumlah produk yang dibeli
- Pelanggan dengan pendapatan tinggi lebih suka berbelanja melalui store dan katalog
- Orang yang menerima campaign terakhir (accepted campaign 5) banyak berasal dari golongan income tinggi
- Orang dengan income tinggi paling banyak melakukan pembelian melalui katalog dengan produk favorit wine dan meat
- Orang yang menerima campaign terakhir (accepted campaign 5) lebih berpotensi untuk menerima campaign selanjutnya (response)

Business Recommendation

Next campaign bisa ditargetkan ke orang dengan income yang tinggi dan difokuskan pada pembelian melalui katalog berkategori wine dan meat

5.

- (No. 2) Kalau distribusi pakai boxplot
- (No. 2) Perbandingan jumlah produk (**kolom mntproduct**) pakai histogram dipisah atas bawah, atau multiline plot dalam 1 frame
- Pembuatan plot korelasi hanya menggunakan data numerical. **Kidhome etc, acceptedcampaign, response tidak termasuk**, tidak perlu dimasukkan dalam korelasi
- *Tipe machine learning klasifikasi, output true or false*
- Pemilihan feature menggunakan evidence dari hasil korelasi response dan tiap kolom, tapi evidence lemah krn korelasi rata-rata hanya 0.2.

- **Business sense** bisa jadi alasan pemilihan feature selain hard evidence.
- Bisa gunakan artikel untuk mendukung business sense