

## Data Science bidang Pemasaran : Analisis Prilaku Pelanggan

Mawaddah Harahap<sup>1\*</sup>, Yusniar Lubis<sup>2</sup>, Zakarias Situmorang<sup>3</sup>

<sup>1</sup>Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika

<sup>1,2,3</sup>Universitas Prima Indonesia, Fakultas Ekonomi, Doktor Manajemen

<sup>1\*</sup>mawaddah@unprimdn.ac.id

Received: xx xx xxxx

Accepted: 9 May 2019

Published: 14 May 2019



### \*Mawaddah Harahap

**Keywords:** Data Science, Pemasaran, Analisis Prilaku Pelanggan, Segmentasi

**DSI: Jurnal Data Science Indonesia** is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

**Abstrak :** Dalam kegiatan pemasaran digital, data Science (DS) memiliki peran penting dalam memahami kinerja industri pemasaran sebelum menerapkan teknik pemasaran digital pada pemasaran produk. Hal ini dikarenakan setiap pelanggan merespons secara berbeda setiap penawaran. Perilaku pelanggan juga berubah berdasarkan waktu karena mereka mungkin memiliki kebutuhan yang berbeda pada situasi yang berbeda. Pada makalah ini fokus menyajikan analisis bisnis dengan penerapan DS untuk mengeksplorasi pola perilaku dan juga memprediksi bagaimana pelanggan akan merespons penawaran yang berbeda. Penerapan analisis data eksplorasi juga diterapkan untuk menjawab beberapa pertanyaan bisnis, dari hasil pengamatan menghasilkan lima kelompok pelanggan yang disajikan dalam bentuk visualisasi dan model *Random Forest Classifier* memiliki skor akurasi prediksi terbaik sebesar 91%, kemudian *K neighbors Classifier* dan *Logistic Regression*.

## PENDAHULUAN

Di era teknologi, industri digital marketing terus berkembang. Dalam kegiatan pemasaran digital, data Science (DS) memiliki peran penting dalam memahami kinerja industri pemasaran sebelum menerapkan teknik pemasaran digital pada pemasaran produk [1]. Dengan menganalisis dan melacak teknologi pemasaran digital, perusahaan dapat lebih mudah mengetahui alat pemasaran digital mana yang terbaik untuk memasarkan produknya [2]. Pemasaran digital dan DS adalah dua jenis pemasaran yang tidak dapat dipisahkan, karena kegiatan pemasaran memerlukan data untuk mengidentifikasi kebutuhan dan preferensi pelanggan terhadap produk yang ditawarkan untuk memenuhi kebutuhan pelanggan. Digital marketing adalah kegiatan untuk membangun koneksi melalui jaringan yang terus berkembang dan tersebar luas [3], [4].

Munculnya teknologi baru dan penggunaan teknik DS untuk data dalam jumlah besar memudahkan perusahaan untuk fokus pada faktor-faktor yang meningkatkan loyalitas pelanggan terhadap produk mereka. Saat ini, perusahaan di seluruh dunia menginvestasikan waktu dan uang mereka dalam DS, analitik, dan pengujian statistik. DS dapat membantu perusahaan menavigasi lautan luas data yang tersedia untuk membuat keputusan bisnis yang terinformasi dan tepat waktu. Tidak hanya bisnis besar, tetapi hampir setiap pengecer online yang menggunakan platform e-commerce mapan seperti Shopify atau solusi e-commerce lainnya memiliki banyak data [5], tetapi biasanya tidak memiliki keterampilan dan orang untuk menganalisisnya dan menggunakannya secara efisien [6]. Perusahaan-perusahaan yang dapat melakukannya memiliki keunggulan kompetitif yang jelas terutama dalam hal pemasaran dan penjualan [7], karena sekitar 90% data yang dikumpulkan perusahaan berkaitan dengan tindakan pelanggan dalam konteks aktivitas pemasaran. Namun, pada kenyataannya untuk mencapai hal tersebut seringkali tidak berjalan sesuai rencana bahkan proses eksekusinya sangat rumit.

Dengan adanya DS perusahaan dapat mengetahui dimana pasar yang tepat untuk produknya dan dapat

dimanfaatkan dalam memahami perilaku pengguna. Analisis data adalah sumber wawasan berharga yang dapat mengungkapkan bagaimana perusahaan membuat keputusan [8]. Sementara itu, setiap perusahaan memiliki tantangannya sendiri, seperti proses mengumpulkan, memvisualisasikan, dan menganalisis data pelanggan, yang merupakan aset yang paling berharga [9]. Setiap perusahaan melakukan analisis data untuk membuka jalan baru bagi lebih banyak pelanggan, lebih banyak keuntungan, dan pengambilan keputusan yang lebih baik [10], [11]. Umumnya, perusahaan memiliki banyak data pelanggan, sehingga penerapan DS dapat memanfaatkan data dari umpan balik pelanggan dunia nyata untuk mengembangkan dan menginformasikan produk dan strategi pemasaran mereka [12].

Pada makalah ini kami fokus pada Studi kasus bisnis analisis dan ilmu data untuk mengeksplorasi pola perilaku dan juga memprediksi bagaimana pelanggan akan merespons penawaran yang berbeda. Pelanggan yang berbeda merespons secara berbeda setiap penawaran. Perilaku pelanggan juga berubah berdasarkan waktu karena mereka mungkin memiliki kebutuhan yang berbeda pada situasi yang berbeda.

## TINJAUAN LITERATUR

Dalam manajemen pemasaran, potensi dalam membuka wawasan untuk memenangkan, mempertahankan pelanggan, mendorong efisiensi bisnis, dan pada akhirnya meningkatkan kinerja dalam hal penjualan dan minat pelanggan. Dalam beberapa tahun ini, banyak peneliti mengusulkan analisis *recency*, *frequency* dan *monetary* (RFM) seperti Rahim, M et al [10] menerapkan pemodelan RFM dan teknik pemodelan data untuk mendeteksi pola perilaku pelanggan. Mereka mengusulkan tiga model berbeda untuk klasifikasi pelanggan yaitu *Multi-Layer Perceptron* dengan tingkat akurasi 99% lebih unggul dari *metode Suport Vector Machine* (SVM) dengan akurasi 95% dan *Decision Tree Classification* (DTC) sebesar 98%. Sedangkan Zhang et al [7] mengembangkan pemodelan RFMC untuk mengoptimalkan kekurangan dalam model RFM. Sementara itu, Qiu et al [13] membangun kerangka prediktif untuk perilaku pembelian pelanggan dalam konteks e-commerce. Mereka menggunakan pemodelan dua tahap, yaitu pertama, asosiasi di antara produk diselidiki dan dieksploitasi untuk menentukan motivasi pelanggan dan membangun koleksi kandidat produk. Kemudian, preferensi pelanggan untuk fitur produk dipelajari dan selanjutnya digunakan untuk mengidentifikasi calon produk yang paling mungkin dibeli. Hasil eksperimen mereka menyatakan bahwa preferensi pelanggan untuk fitur produk tertentu memainkan peran kunci dalam pengambilan keputusan. Selanjutnya, Bansal et al [14] mengusulkan analisis keranjang pasar untuk membantu pengguna dengan pilihan yang tepat saat membeli produk bersama dengan barang yang sudah dibeli, sehingga menawarkan pengalaman membeli yang lebih baik. Pendekatan data mining untuk menganalisis perilaku pelanggan diusulkan oleh Gull, M and Pervaiz, M [15]. Dalam makalah ini, kami melakukan analisis eksplorasi dasar, diikuti dengan mengelompokkan pelanggan ke dalam kelompok demografis menggunakan pengelompokan K-means, dan menganalisis beberapa dari kelompok tersebut, selanjutnya menyajikan hasil beberapa metrik. Akhirnya kami memprediksi bagaimana pelanggan akan merespons penawaran yang berbeda.

## BAHAN DAN METODE

Pada bagian ini, kami uraikan bahan dan metode yang digunakan untuk menganalisis perilaku berbelanja pelanggan yaitu:

### 1. Himpunan Data

Himpunan data yang digunakan merupakan data simulasi yang meniru perilaku pelanggan di aplikasi seluler hadiah dari perusahaan Starbucks. Starbucks adalah Perusahaan multinasional Amerika dengan berbagai rantai kedai kopi yang berkantor pusat di Seattle, Washington. Setiap beberapa hari sekali, Starbucks mengirimkan penawaran kepada pengguna aplikasi seluler. Penawaran dapat berupa iklan minuman atau penawaran nyata seperti diskon atau BOGO (*Buy one get one*). Beberapa pengguna mungkin tidak menerima penawaran apa pun selama minggu-minggu tertentu. Karakteristik dari data ini terdiri dari tiga bagian yaitu:

- a) Portofolio, yaitu kumpulan data yang menjelaskan karakteristik setiap jenis penawaran, termasuk jenis penawaran, kesulitan, dan durasinya. Kerangka data ini memiliki 10 baris dan 6 kolom dengan Pengeluaran minimum rata-rata yang diperlukan untuk menyelesaikan penawaran adalah \$7,70 dengan reward rata-rata \$4,20.
- b) Profil, yaitu kumpulan data yang berisi informasi mengenai demografi pelanggan termasuk usia, jenis kelamin, pendapatan, dan tanggal membuat akun untuk Starbucks Rewards. Kerangka data memiliki 17.000 baris dan 5 kolom dengan usia rata-rata 63 tahun dan pendapatan rata-rata \$65.404.
- c) Transkrip, yaitu kumpulan data yang berisi semua kejadian saat pelanggan melakukan pembelian, melihat penawaran, menerima penawaran, dan menyelesaikan penawaran.

## 2. Metode yang diusulkan

Metode yang diusulkan dalam penelitian ini terdiri dari beberapa tahapan, yaitu:

### a) Analisis bisnis

Kepuasan pelanggan mendorong kesuksesan bisnis, dan analitik data memberikan wawasan tentang apa yang dipikirkan oleh pelanggan. Pada umumnya, setiap orang memiliki sifat yang berbeda dan bertindak terhadap penawaran, sehingga sangat penting untuk memahami perilaku pelanggan dalam menawarkan sebuah produk. Pemasaran tidak terlepas dengan adanya biaya dan umumnya para pelaku bisnis suka mempelajari bagaimana setiap segmen pelanggan merespons penawaran yang berbeda dan bagaimana menargetkan/menyesuaikan penawaran kepada individu untuk meningkatkan pendapatan dengan mengubah pelanggan potensial menjadi pelanggan yang tersenyum. Kumpulan data yang digunakan pada penelitian ini adalah simulasi perilaku pelanggan pada aplikasi seluler penghargaan Starbucks. Secara berkala, Starbucks mengirimkan penawaran kepada pengguna yang dapat berupa iklan, diskon, atau beli satu gratis. Karakteristik penting mengenai kumpulan data ini adalah tidak semua pengguna menerima penawaran yang sama. Secara keseluruhan analisis bisnis ini fokus pada tiga pertanyaan penting, yaitu:

1. Berapa kerugian perusahaan Starbucks karena penawaran?
2. Pelanggan seperti apa yang sering menyelesaikan penawaran tanpa melihatnya?
3. Bagaimana distribusi pendapatan antar tipe pelanggan?

### b) Pemahaman Data

Salah satu segmen penting dari bisnis apa pun adalah mencari area yang dapat ditingkatkan untuk menghasilkan lebih banyak pendapatan. Dalam penelitian ini, kami menganalisis kumpulan dataset untuk melihat seberapa baik promosi tertentu digunakan oleh semua pelanggan dan untuk mencoba dan melihat metode apa yang lebih efektif untuk menawarkan promosi. Memahami bagaimana pelanggan biasanya bereaksi, akan menemukan cara yang lebih baik untuk memastikan bahwa orang yang tepat menerima promosi yang tepat.

### c) Analisis Data Eksplorasi

Analisis data eksplorasi bertujuan untuk melihat tipe pelanggan seperti apa yang ada di dunia Starbucks, kebiasaan transaksi/keterlibatan promosi mereka, dan membuat kerangka data matriks pelanggan yang memberikan gambaran keseluruhan tentang tipe pelanggan seperti apa mereka. Pendekatan pembelajaran mesin terbaik kemudian menggunakan metode pengelompokan yang dapat memecah pelanggan menjadi segmen yang berbeda yang akan merespon secara berbeda terhadap penawaran promosi yang berbeda

### d) Segmentasi Pelanggan

Pada tahapan ini dilakukan untuk mengelompokkan pelanggan ke dalam demografi yang berbeda, kemudian memahami penawaran mana yang mereka tanggap lebih banyak (penyelesaian penawaran).

### e) Modelling

Pada tahapan ini diusulkan tiga model berbeda yang dievaluasi berdasarkan nilai akurasi, precision, recall dan f1-score menjadi metrik utama untuk memprediksi apakah pelanggan akan menanggapi tawaran. Sebelum membuat model, kumpulan data dibagi menjadi set data latih dan uji. Set data latih

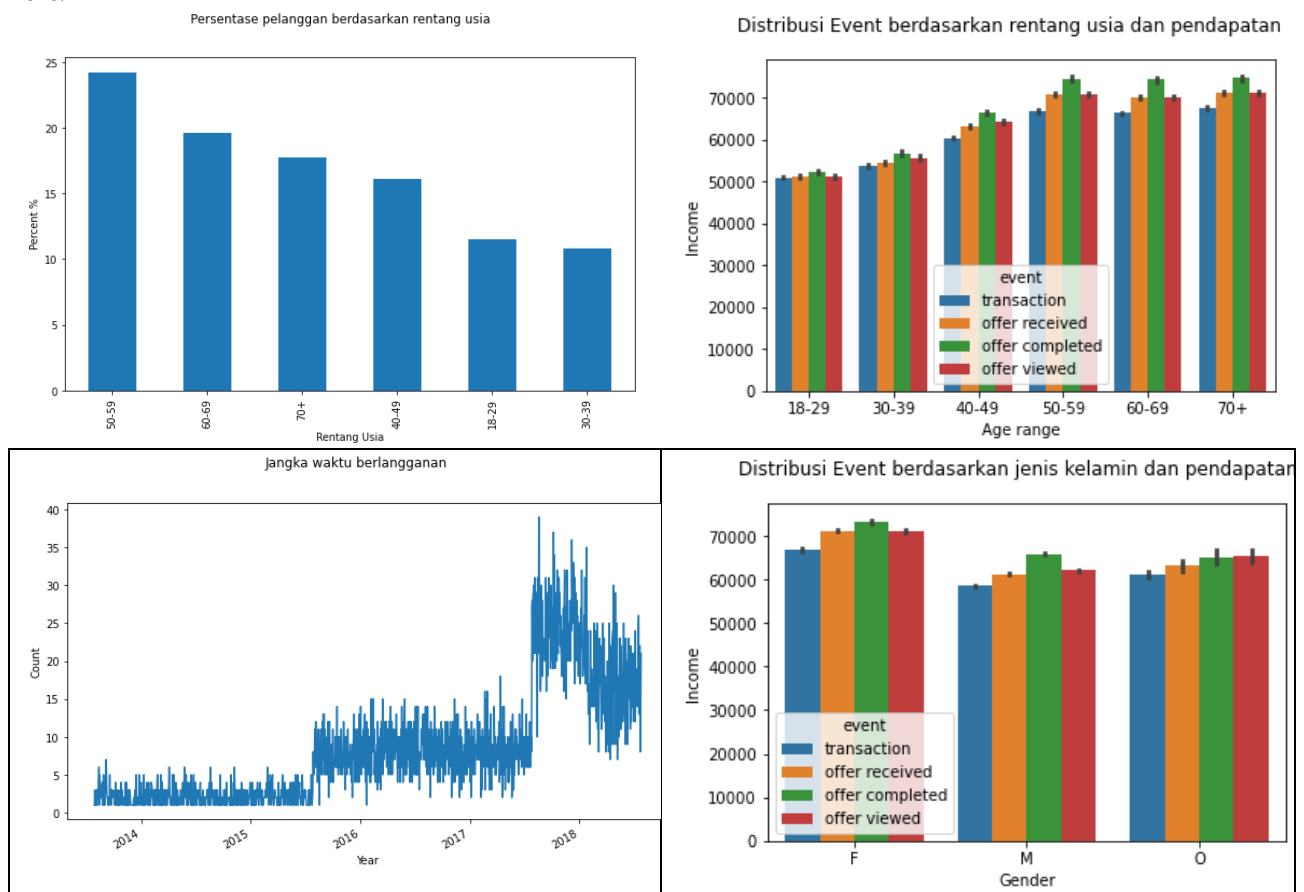
merupakan tahapan untuk mengevaluasi model dan set data uji digunakan untuk mengevaluasi seberapa baik kinerja model.

## HASIL PENELITIAN

Pengujian dilakukan dengan menggunakan spesifikasi perangkat Intel Core i7, 16.0 GB RAM, Python 3.0 dan beberapa pustaka seperti numpy, pandas, matplotlib dan lainnya. Hasil pengujian dibagi menjadi 3 bagian, pertama menjawab tiga pertanyaan yang diuraikan pada bagian analisis sistem, kemudian analisis pengelompokan pelanggan dan terakhir adalah memprediksi repon pelanggan terhadap penawaran. Untuk mendapatkan hasil dari ketiga pertanyaan, maka dilakukan beberapa tahapan seperti menjelajahi data untuk menemukan pola dan menyelami detail set data yang disediakan, pembersihan dan transformasi data untuk membantu mendapatkan dataset gabungan untuk memudahkan tahapan analisis.

### 1. Analisis Data

Berdasarkan analisis data eksplorasi, kami memperhatikan bahwa 24% pelanggan adalah pelanggan berusia antara 50 hingga 59 tahun yang lebih tinggi daripada semua pelanggan berusia antara 18 hingga 39 tahun (22%). Kelompok terbesar kedua adalah pelanggan yang berusia antara 60 hingga 69 tahun yang menyumbang 20%.



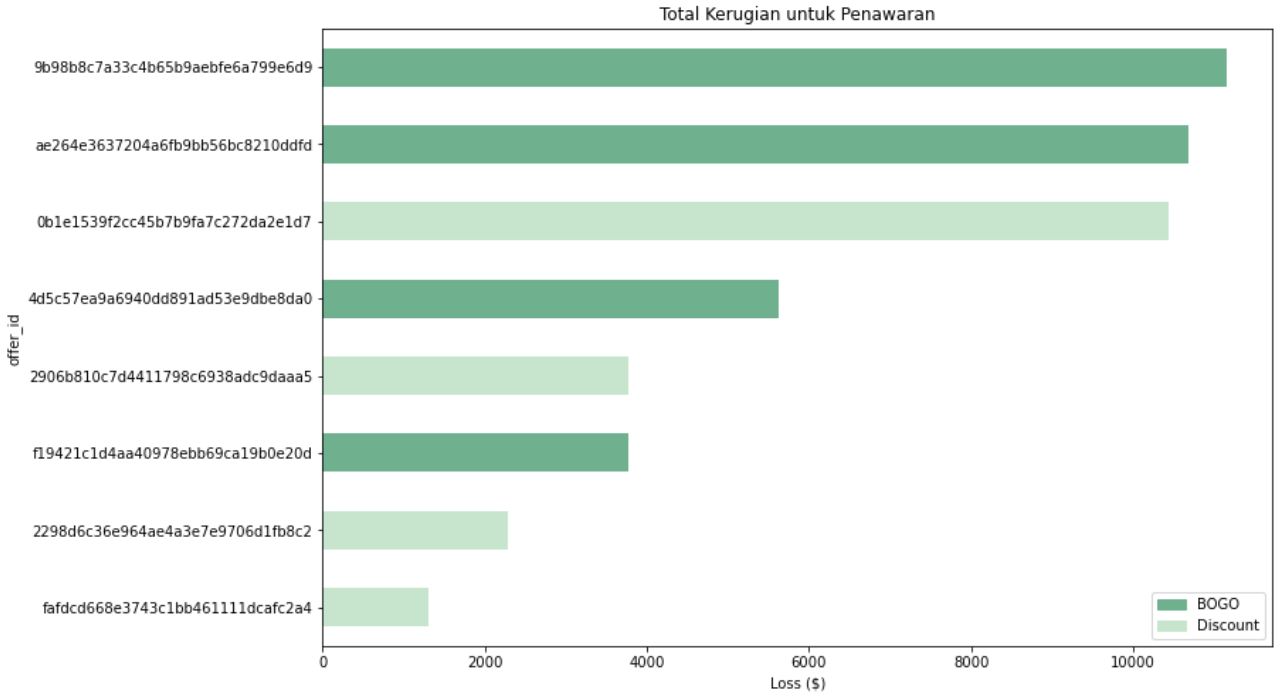
Gambar 1 Analisis Data Eksplorasi

Pada gambar 1 dapat dilihat hasil berdasarkan pendapatan ada peningkatan umum dalam transaksi seiring dengan bertambahnya usia. Hasil analisis ini dapat diasumsikan bahwa lebih banyak pelanggan yang melakukan transaksi adalah pelanggan yang bekerja dengan tingkat pendapatan menengah ke atas dan pelanggan ini sudah berlangganan dalam waktu 5 tahun. Menarik untuk dicatat bahwa meskipun banyak pelanggan pria, dalam hal transaksi aktual, melihat penawaran dan menyelesaikan penawaran kampanye, mayoritas adalah wanita. Dari hasil pengamatan analisis data ini, memahami tingkat keberhasilan penawaran

yang dikirim dan analisis penawaran dengan menjawab pertanyaan-pertanyaan berikut:

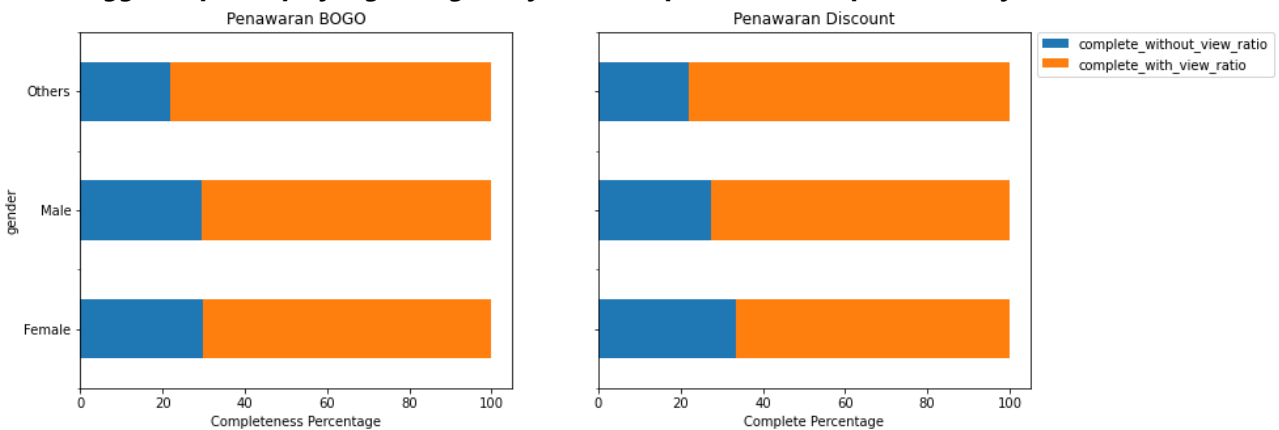
**a) Berapa kerugian perusahaan Starbucks karena penawaran?**

Perusahaan kehilangan pendapatan sebesar \$49.032. Jika dilihat kolom waktu di kerangka data, penawaran ini mendapatkan nilai maksimum 714. Dengan demikian, rata-rata perusahaan memiliki potensi kerugian sebesar  $\$49.032 \times 12 = 588.384$  pendapatan dalam setahun dengan mempertimbangkan waktu eksperimen maksimum adalah 714 jam = 30 hari, sehingga kita dapat mengasumsikan bahwa eksperimen berjalan selama 30 hari. Terdapat delapan penawaran dimana penawaran diskon sebesar \$5.391 menghasilkan kerugian sebesar \$17.802, sedangkan penawaran BOGO (*Buy one get one*) sebesar \$4.616 mengalami kerugian sebesar \$31.230. Dari analisis ini kita menyimpulkan bahwa sebagian besar kerugian berasal dari penawaran BOGO.



Gambar 2 Analisis Kerugian Penawaran

**b) Pelanggan seperti apa yang sering menyelesaikan penawaran tanpa melihatnya?**

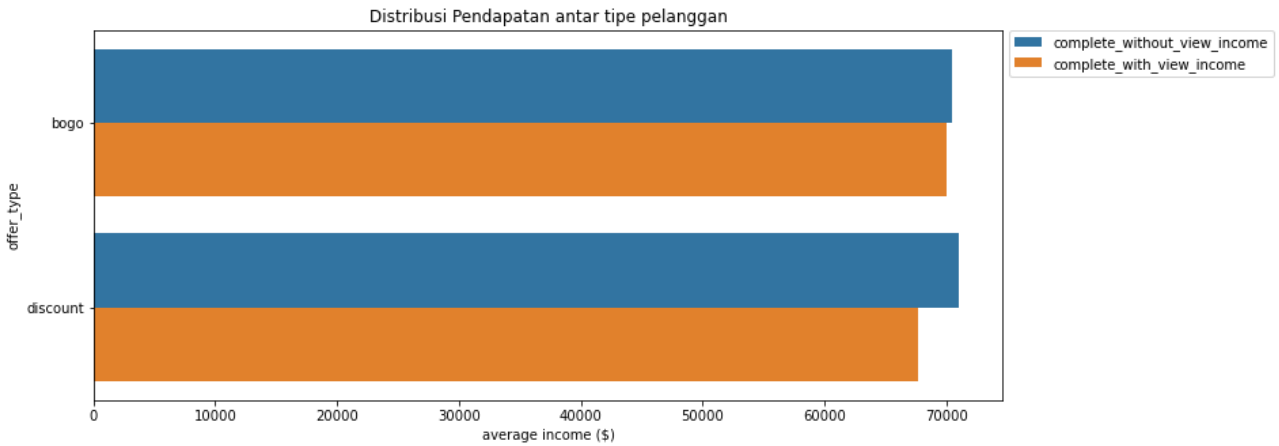


Gambar 3 Analisis penawaran tanpa melihatnya

Visualisasi pada gambar 3 menunjukkan persentase pelanggan yang menyelesaikan penawaran tanpa melihat penawaran. Seperti hasil pada gambar 3, dalam Penawaran BOGO tidak ada banyak perbedaan antara pelanggan pria dan wanita, tetapi dalam penawaran diskon dapat melihat bahwa wanita memiliki rasio sedikit lebih tinggi daripada pria dan lainnya, dengan 33,4% dibandingkan dengan 27,5% dan 21,9% dari masing-

masing. Hal ini menunjukkan bahwa wanita cenderung kurang "didorong terhadap diskon" daripada pria dan lainnya.

**c) Bagaimana distribusi pendapatan antar tipe pelanggan?**



Gambar 4 Analisis Distribusi Pendapatan antar tipe pelanggan

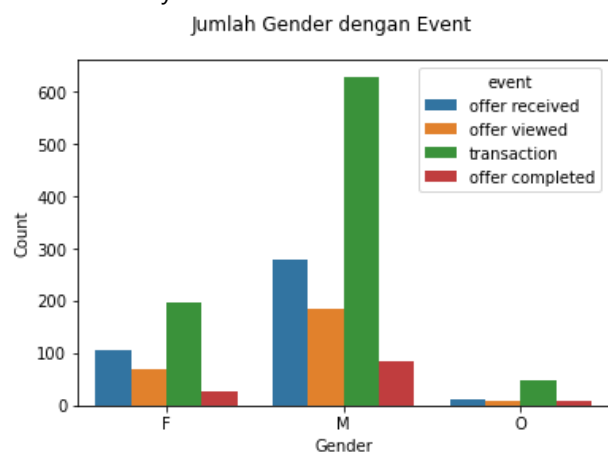
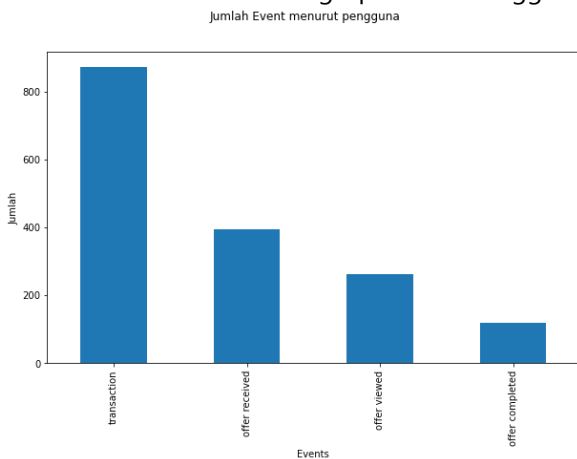
Jika dilihat dari visualisasinya, tidak banyak perbedaan antara pelanggan yang menyelesaikan penawaran tanpa melihatnya atau tidak. Namun secara keseluruhan, pelanggan yang menyelesaikan penawaran tanpa melihat penawaran terlebih dahulu memiliki pendapatan rata-rata yang lebih tinggi. Hal Ini cukup intuitif karena dapat diasumsikan bahwa orang-orang yang memiliki pendapatan lebih tinggi tidak terlalu peduli dengan penawaran.

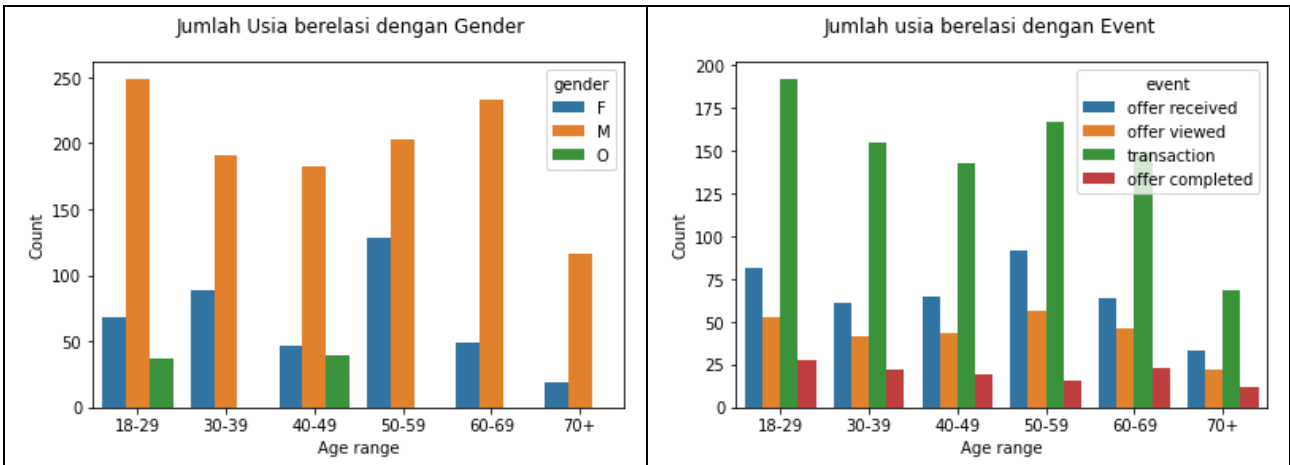
**2. Pengelompokan Pelanggan**

Pengelompokan pelanggan merupakan salah satu tahapan penting dalam hasil penelitian ini, kami mempertimbangkan penerapan fitur demografi pendapatan untuk menyegmentasikan pelanggan karena saya akan berasumsi bahwa melihat daya beli pelanggan adalah cara yang lebih baik. Pengelompokan pelanggan ini dibagi menjadi 5 (lima) kelompok berdasarkan demografi, yaitu:

**a) Pendapatan Rendah lebih kecil atau sama dengan nilai \$30.000**

Kelompok ini memiliki 88 pelanggan dengan rentang usia 19 sampai 92 dan 1.652 aktivitas (interaksi) yang terdiri dari 874 transaksi. 396 penawaran diterima dengan 262 dari penawaran tersebut dilihat sekitar 66%, 120 penawaran diselesaikan hampir setengah (46%) dari penawaran yang dilihat. Penawaran diskon diterima dengan baik oleh grup ini sebesar 53% (63) dari penawaran yang telah selesai dan BOGO sebesar 48% (57). Penawaran informasi dalam grup ini lebih tinggi dari penawaran lainnya.

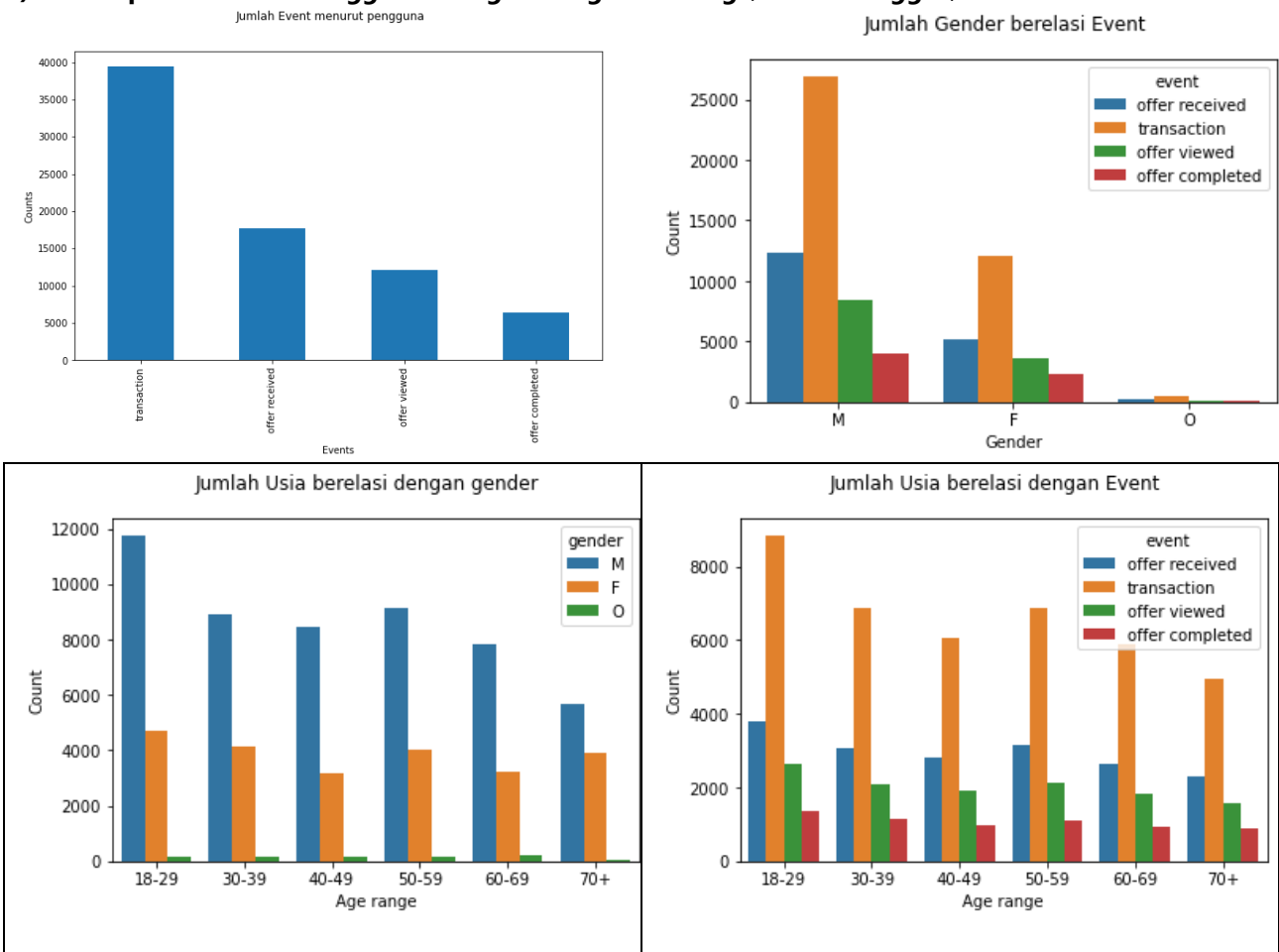




Gambar 5 Analisis Cluster 0

Total transaksi pada kelompok ini sekitar \$4.335,32, sebagian besar transaksi ini dilakukan oleh pelanggan pria. Dari hasil pengamatan pada kelompok ini, lebih banyak penawaran dikirimkan kepada pelanggan pria. Dalam kelompok ini, terdapat banyak pelanggan berusia antara 18 hingga 29 yang didominasi oleh pria, sedangkan pelanggan wanita berusia 50 hingga 59 dan Pelanggan yang berusia 70 tahun ke atas paling sedikit. Pelanggan yang berusia rentang 18 hingga 29 tahun lebih banyak transaksi diikuti oleh pelanggan berusia 50 hingga 59 tahun.

**b) Pendapatan Rendah hingga Menengah dengan rentang \$31.000 hingga \$50.000**



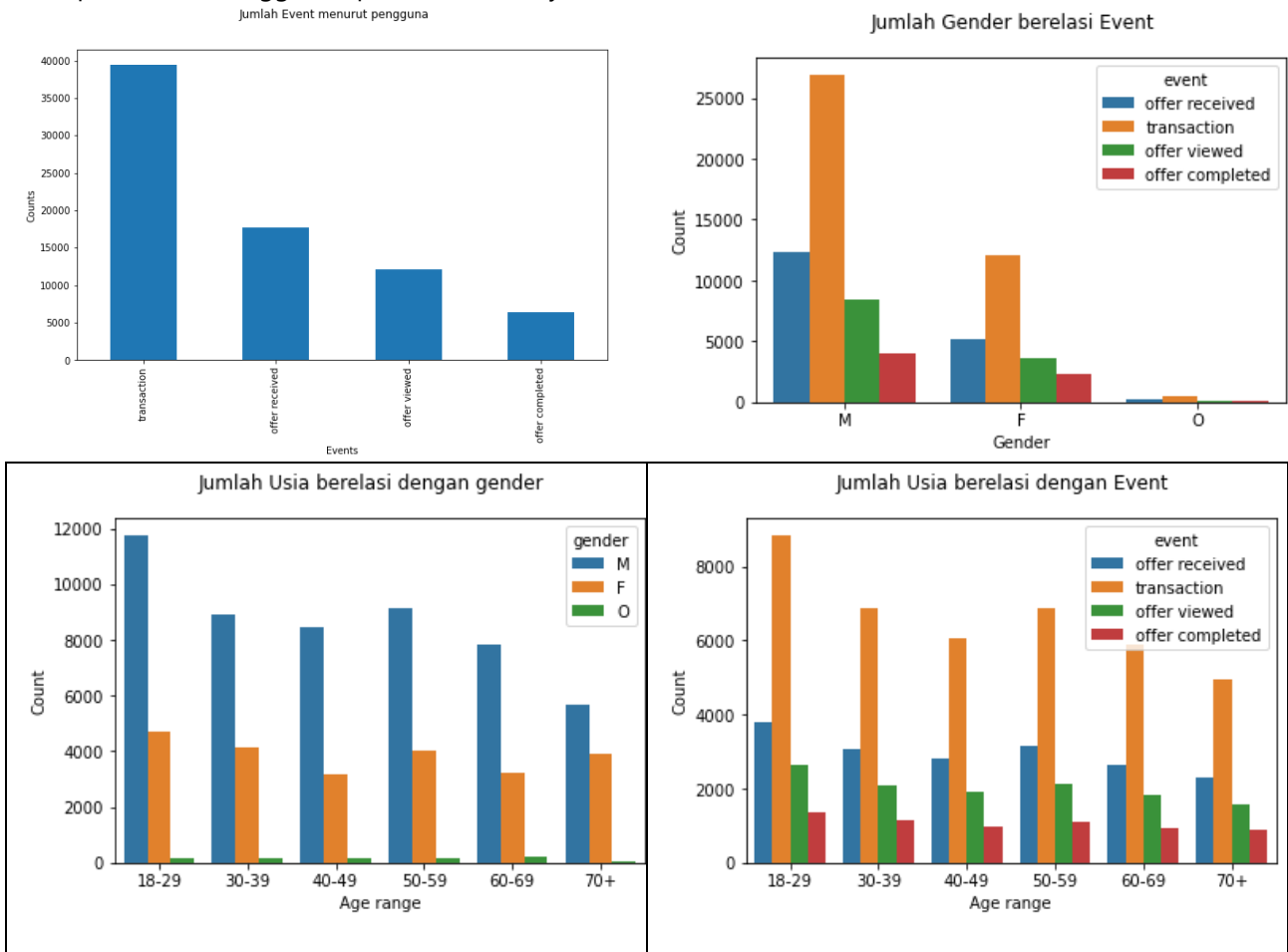
Gambar 5 Analisis Cluster 1

Kelompok ini memiliki 3.946 pelanggan dengan rentang usia 18 hingga 101 dan usia rata-rata 48 tahun. Kelompok ini menyumbang 75.791 aktivitas yang terdiri dari 39.461 transaksi. 17.715 penawaran diterima dengan 12.183 dari penawaran tersebut dilihat atau sekitar 69% dan dari 12.183 penawaran yang dilihat, 6.432 penawaran diselesaikan (53%) dari penawaran dilihat. Penawaran diskon mendapat tanggapan lebih dari kelompok ini sebesar 56% (3.606) dari penawaran yang diselesaikan dan BOGO sebesar 44% (2.826). Penawaran informasi yang dikirim pada kelompok ini lebih tinggi dari penawaran lainnya sekitar 5.762.

Total yang dihabiskan untuk transaksi sebesar \$525.708,12 dimana pria melakukan sebagian besar transaksi ini sekitar \$283.007,82 dan wanita sekitar \$231.965,85. Sama seperti kelompok sebelumnya, penawaran lebih banyak dikirim pada pelanggan pria. Pelanggan berusia 50 hingga 59 lebih banyak melakukan transaksi, kemudian berusia 40 hingga 49, namun kelompok ini pelanggan berusia 18-29 tahun mengalami penurunan sedangkan pelanggan wanita meningkat di semua rentang usia.

**c) Demografi Pendapatan Menengah rentang \$51.000 hingga \$70.000**

Kelompok ini memiliki 3.591 pelanggan dengan rentang usia 18 hingga 101 dan usia rata-rata 59 tahun. Kelompok ini menghasilkan 63.044 peristiwa yang terdiri dari 25.183 transaksi. 16.081 penawaran diterima dengan 12.508 dari penawaran yang dilihat (77%) dan dari 12.508 penawaran yang dilihat, 9.272 penawaran diselesaikan (74%) dari penawaran. Penawaran diskon mendapat tanggapan lebih dari kelompok ini sebesar 52% (4.832) dari penawaran yang diselesaikan dan BOGO sebesar 48% (4.440). Penawaran yang dikirim dalam kelompok ini lebih tinggi dari penawaran lainnya sekitar 5.442.



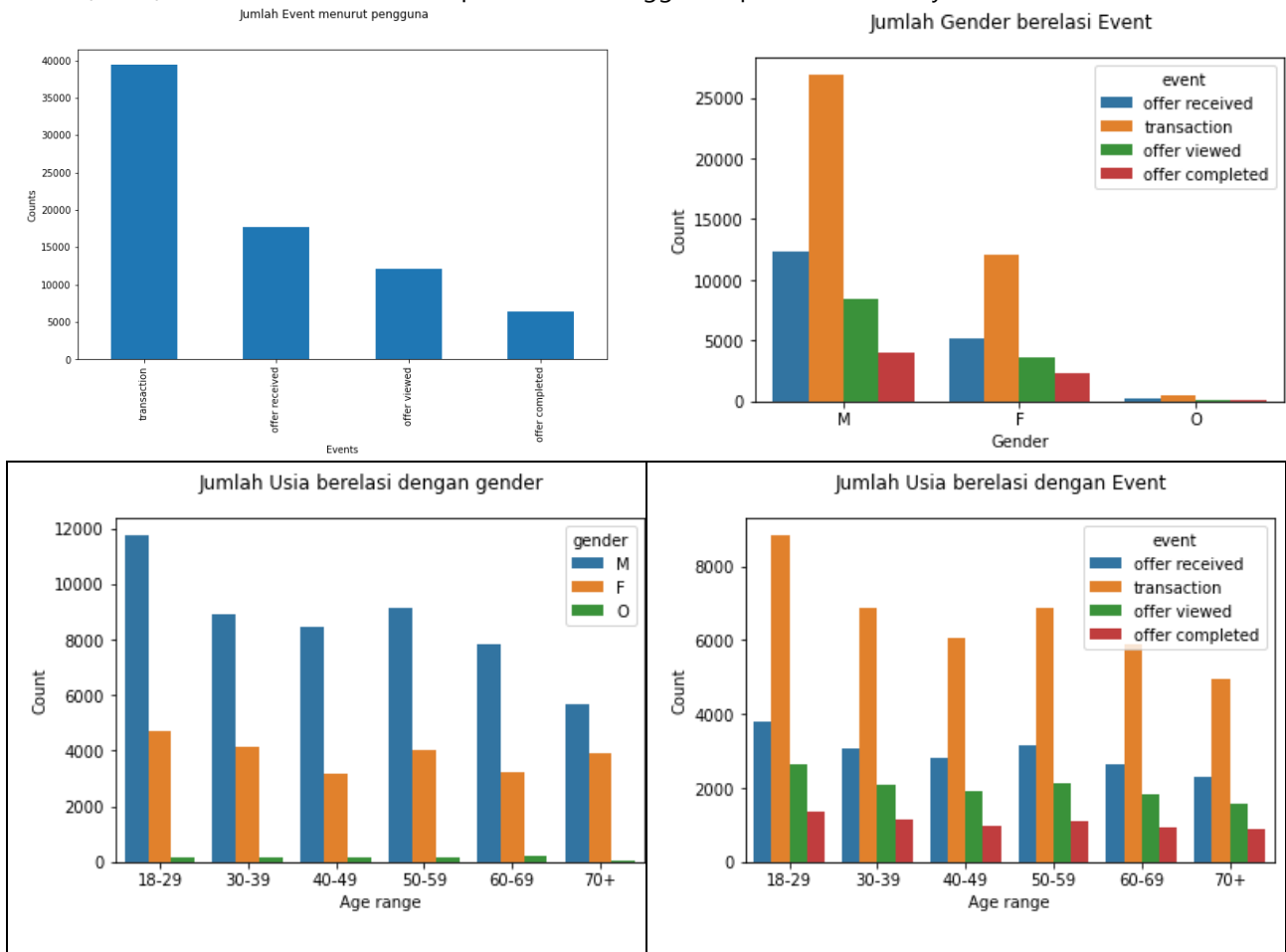
Gambar 5 Analisis Cluster 2

**d) Demografi Pendapatan Menengah ke Tinggi rentang \$70.000 hingga \$90.000**

Segmen pelanggan Starbucks ini berpenghasilan antara \$71.000 hingga \$90.000 dan dikelompokkan sebagai pelanggan berpenghasilan menengah. Demografi ini memiliki 2.194 pelanggan dengan rentang usia 36 hingga



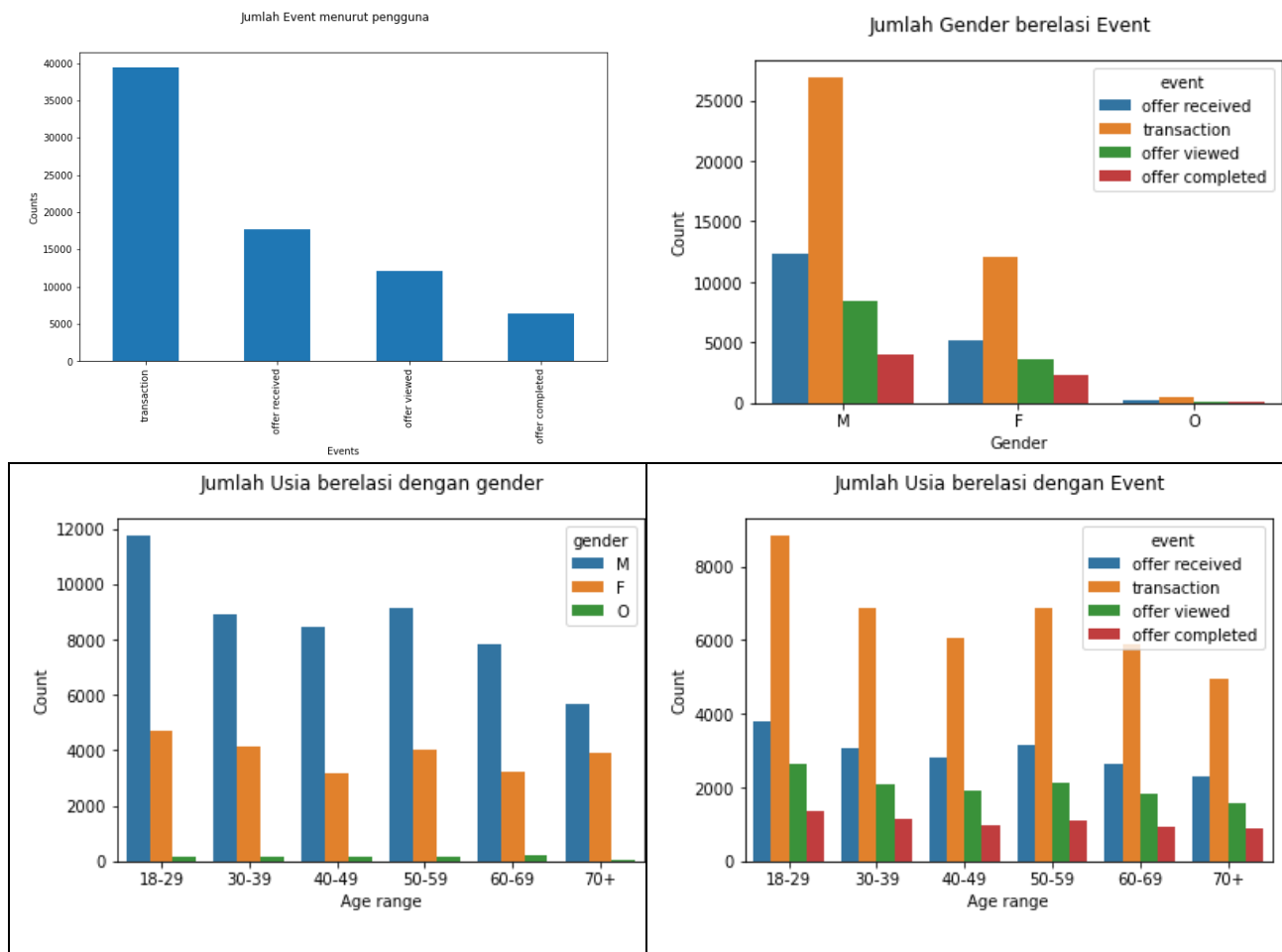
101 dan usia rata-rata 63 tahun. kelompok ini mencakup 36.652 aktivitas yang terdiri dari 13.597 transaksi. 9.814 penawaran diterima dengan 7.236 (74%) dari penawaran yang dilihat dan dari 7.236 penawaran yang dilihat, 6.005 penawaran diselesaikan dengan menghasilkan 83% dari penawaran yang dilihat. Penawaran diskon mendapat respon lebih sebesar 50,5% (3.030) dari penawaran yang diselesaikan dan BOGO sebesar 49,5% (2.975). Penawaran dalam kelompok ini lebih tinggi dari penawaran lainnya 3.301.



Gambar 5 Analisis Cluster 3

**e) Demografi Penghasilan Tinggi lebih besar atau sama dengan \$91.000**

Segmen pelanggan Starbucks ini berpenghasilan antara \$91.000 hingga \$110.000 dan \$111.000+ dan dikelompokkan sebagai pelanggan berpenghasilan tinggi. Demografi ini memiliki 2.194 pelanggan dengan usia mulai dari 36 hingga 101 dengan usia rata-rata 63. kelompok ini menyumbang 36.652 aktivitas terdiri dari 13.597 transaksi. 9.814 penawaran diterima dengan 7.236 (74%) dari penawaran yang dilihat dan dari 7.236 penawaran yang dilihat, 6.005 penawaran diselesaikan dengan menghasilkan 83% dari penawaran yang dilihat. Penawaran diskon mendapat tanggapan sebesar 50,5% (3.030) dari penawaran yang diselesaikan dan BOGO sebesar 49,5% (2.975). Penawaran informasi dalam grup ini lebih tinggi dari penawaran lainnya 3.301 dengan jumlah total yang dibelanjakan sekitar \$405.055.76 per transaksi dimana wanita melakukan sebagian besar transaksi sebesar \$238.828,13 dan pria \$162.276,96. Tidak seperti kelompok lain, wanita lebih banyak daripada laki-laki dengan lebih banyak transaksi diselesaikan oleh wanita. Mayoritas pelanggan berusia rentang 50 hingga 59 yang melakukan lebih banyak transaksi, kemudian pelanggan dengan rentang usia 70+ yang sedikit lebih tinggi dari kelompok usia 60 - 69. Dalam kelompok ini tidak ada pelanggan antara 18 - 29 dan sejumlah kecil pelanggan dalam kisaran 30 - 39.



Gambar 5 Analisis Cluster 4

### 3. Modelling

Pada bagian ini akan uraikan hasil prediksi bagaimana pelanggan Starbucks akan menanggapi penawaran dengan menggunakan tiga metode klasifikasi yang berbeda seperti *Regresi Logistik*, *KNeighbors Classifier* dan *Random Forest Classifier*. Berikut hasil ketiga model yang diusulkan dapat dilihat pada tabel 1.

Tabel 1 Perbandingan Hasil Prediksi

Model	precision	recall	f1-score	accuracy
Logistic Regression	0.85	0.91	0.88	0.87
K neighbors Classifier	0.91	0.85	0.88	0.88
Random Forest Classifier	0.93	0.90	0.91	0.91

Berdasarkan tabel 1 dapat dilihat pendekatan model *Random Forest Classifier* memiliki skor akurasi prediksi terbaik sebesar 91%, kemudian *K neighbors Classifier* dan *Logistic Regression*, namun model *Logistic Regression* menghasilkan nilai tertinggi untuk *recall* sebesar 91%. Jika biaya untuk mengirimkan penawaran kepada pelanggan tidak terlalu tinggi, perusahaan Starbucks sangat mungkin untuk menargetkan sebanyak mungkin calon responden, yang berarti diperlukan *f1-score* yang tinggi. Di sisi lain, jika biaya pengiriman penawaran terlalu tinggi, ini berarti Starbucks tidak ingin mengirimkan penawaran kepada pelanggan yang mungkin mengabaikan penawaran tersebut. Dalam hal ini, nilai *precision* lebih tinggi daripada *recall*, maka model *Random Forest Classifier* akan menjadi pilihan.

## Diskusi

Berdasarkan analisis tersebut, ada beberapa hal yang perlu didiskusikan yaitu penawaran yang tidak direncanakan, perusahaan Starbucks kemungkinan "kehilangan" pendapatan hingga \$49.032 dalam sebulan atau sekitar \$588.384 pendapatan dalam setahun. Jadi target pemasaran promo sangat penting dan memainkan peran besar. Pelanggan wanita cenderung berbelanja lebih banyak daripada pelanggan Pria, dengan rata-rata pembelanjaan sebesar \$16,3 per transaksi dibandingkan pria sebesar \$10,4. Pelanggan wanita juga memiliki kecenderungan untuk menyelesaikan penawaran bahkan tanpa melihatnya terlebih dahulu, sehingga perlu mempertimbangkan dalam mengirimkan penawaran kepada mereka. Secara keseluruhan, pelanggan yang menyelesaikan penawaran tanpa melihat penawarannya terlebih dahulu memiliki pendapatan rata-rata yang lebih tinggi, terutama pada penawaran diskon dimana mereka yang menyelesaikan penawaran tanpa melihatnya.

Perusahaan Starbucks juga perlu berhati-hati dalam mengirimkan penawaran terutama penawaran BOGO yang memberikan kerugian sebesar \$31.230 dalam percobaan ini. Satu hal yang bisa dilakukan adalah berhenti memberikan penawaran BOGO kepada pelanggan dengan rata-rata pembelian > 2 cup per transaksi, karena tanpa memberikan penawaran pelanggan cenderung membeli > 2 cup juga sehingga penawaran BOGO sepertinya tidak penting bagi mereka. Kemudian perlu mengirim lebih sedikit penawaran kepada pelanggan wanita, terutama penawaran diskon. Dari data tersebut dapat dilihat bahwa rata-rata pengeluaran pelanggan wanita adalah sebesar \$16,3. Jadi sangat mungkin untuk meningkatkan pembelanjaan minimum untuk penawaran yang di kirimkan kepada mereka.

## KESIMPULAN

Penelitian ini menyajikan analisis bisnis dengan data science yang merupakan salah satu studi kasus pemasaran produk perusahaan Starbucks. Beberapa tahapan diuraikan seperti analisis bisnis, pemahaman data, analisis data eksplorasi, pengelompokan pelanggan hingga membangun model untuk memprediksi bagaimana pelanggan akan merespons penawaran yang berbeda menggunakan tiga metode klasifikasi yang berbeda seperti *Regresi Logistik*, *KNeighbors Classifier* dan *Random Forest Classifier*. Dari hasil analisis data, menghasilkan lima kelompok pelanggan yang telah diuraikan dalam bentuk visualisasi, sedangkan model prediksi menghasilkan akurasi terbaik sebesar 91% pada model *Random Forest Classifier*, kemudian *KNeighbors Classifier* sebesar 88% sedangkan model *Regresi Logistik* sebesar 87%.

## Supplementary Materials (optional)

Sumber dataset tersedia di <https://www.kaggle.com/blacktile/starbucks-app-customer-reward-program-data>

## Kontribusi Penulis

Semua Penulis memiliki kontribusi yang sama dalam makalah ini Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

## Konflik kepentingan

Para penulis menyatakan tidak ada konflik kepentingan.

## REFERENCES

- [1] S. S. Alrumiah and M. Hadwan, "Implementing big data analytics in e-commerce: Vendor and customer view," *IEEE Access*, vol. 9, pp. 37281–37286, 2021, doi: 10.1109/ACCESS.2021.3063615.
- [2] T. Istiqomatin, A. Setiadi, and T. Ekowati, "Effect of Marketing Mix on Consumer Purchase Decisions to Buy Cavendish Banana at Modern Markets in Semarang," *J. Agric. Socioecon. Bus.*, vol. 4, no. 2, pp. 120–

132, 2021.

- [3] J. R. Saura, P. Palos-Sanchez, and B. R. Herráez, "Digital Marketing for Sustainable Growth: Business Models and Online Campaigns Using Sustainable Strategies," *Sustain.* 2020, Vol. 12, Page 1003, vol. 12, no. 3, p. 1003, Jan. 2020, doi: 10.3390/SU12031003.
- [4] A. Reyes-menendez, "Consumer Behavior in the Digital Age," *J. Spat. Organ. Dyn.*, vol. VIII, no. 3, pp. 190–196, 2020.
- [5] K. Maheswari, P. Packia, and A. Priya, "Predicting Customer Behavior in Online Shopping Using SVM Classifier," *IEEE Int. Conf. Intell. Tech. Control. SIGNAL Process.*, 2017.
- [6] V. Vijayakumar and V. Neelanarayanan, *Data Mining Approach for Intelligent Customer Behavior Analysis for a Retail Store*, vol. 49. Cham: Springer International Publishing, 2016.
- [7] Y. Zhang, E. T. Bradlow, and D. S. Small, "Predicting customer value using clumpiness: From RFM to RFMC," *Mark. Sci.*, vol. 34, no. 2, pp. 195–208, 2015, doi: 10.1287/mksc.2014.0873.
- [8] C. Wang *et al.*, "Behavior-interior-aware user preference analysis based on social networks," *Complexity*, vol. 2018, 2018, doi: 10.1155/2018/7371209.
- [9] N. Tsapatsoulis and C. Djouvas, "Opinion mining from social media short texts: Does collective intelligence beat deep learning?," *Front. Robot. AI*, vol. 6, no. JAN, p. 138, Jan. 2019, doi: 10.3389/frobt.2018.00138.
- [10] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *J. Retail. Consum. Serv.*, vol. 61, no. April, p. 102566, 2021, doi: 10.1016/j.jretconser.2021.102566.
- [11] P. Rita, T. Oliveira, and A. Farisa, "The impact of e-service quality and customer satisfaction on customer behavior in online shopping," *Heliyon*, vol. 5, no. 10, p. e02690, Oct. 2019, doi: 10.1016/j.heliyon.2019.e02690.
- [12] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services," *Appl. Comput. Informatics*, vol. 15, no. 2, pp. 94–101, 2019, doi: 10.1016/j.aci.2018.05.004.
- [13] J. Qiu, Z. Lin, and Y. Li, "Predicting customer purchase behavior in the e-commerce context," *Electron. Commer. Res.*, vol. 15, no. 4, pp. 427–452, 2015, doi: 10.1007/s10660-015-9191-6.
- [14] S. Bansal, C. Gupta, and A. Sinha, "Clickstream & behavioral analysis with context awareness for e-commercial applications," *2017 10th Int. Conf. Contemp. Comput. IC3 2017*, vol. 2018-Janua, no. August, pp. 1–6, 2018, doi: 10.1109/IC3.2017.8284328.
- [15] M. Gull and A. Pervaiz, "Customer behavior analysis towards online shopping using data mining," *5th Int. Multi-Topic ICT Conf. Technol. Futur. Gener. IMTIC 2018 - Proc.*, pp. 1–5, 2018, doi: 10.1109/IMTIC.2018.8467262.