



Analisis *Customer Retention* Dalam Ritel Online Data *United States E-Commerce Records 2020*

Andy Hermawan¹, Muhamad Fauzi Hakim², B Hilda Nida Alistiqlal³,
Bagas Dio Hanggoro⁴

¹Universitas Indraprasta PGRI, Jakarta

²⁻⁴Purwadhika Digital Technology School, Jakarta

Email: andy.hermawan@unindra.ac.id¹, fauzi.paradise@gmail.com², hildanida01@gmail.com³,
bagasdioh@gmail.com⁴

Abstract. *This study investigates the level of customer loyalty in e-commerce in the United States in 2020. Using e-commerce customer data from 2020 available on Kaggle, this research analyzes cohort retention to understand the number of customers from each cohort who continue to make purchases in subsequent months. Additionally, this study evaluates the implementation of data analysis to assess the effectiveness of promotional strategies and their impact on customer loyalty and retention. The results show that retention percentages decline over time, with the number of active customers decreasing after registration. However, some cohorts exhibit better retention performance. For instance, the April cohort saw a significant increase from 18.46% to 29.23%, and the May cohort experienced the highest increase from 14.29% to 30.16%. On the other hand, the January cohort showed the most drastic retention drop from 100% to 10.42% after the first month, with the largest monthly decline from 23.96% in April to 7.29% in May. These findings provide practical guidance for e-commerce companies in developing more effective strategies to improve customer retention. Further analysis is recommended to understand the specific causes of these retention trends and to develop more holistic strategies to support better business decisions and increase long-term profitability.*

Keywords: : e-commerce, customer retention, cohort retention, data analysis, promotional strategies, customer loyalty, United States.

Abstrak. Penelitian ini menyelidiki tingkat loyalitas pelanggan e-commerce di Amerika Serikat pada tahun 2020. Menggunakan data pelanggan e-commerce dari tahun 2020 yang tersedia di Kaggle, penelitian ini menganalisis retensi kohort untuk memahami jumlah pelanggan dari setiap kohort yang terus melakukan pembelian pada bulan-bulan berikutnya. Selain itu, penelitian ini juga mengevaluasi implementasi analisis data untuk menilai efektivitas strategi promosi dan dampaknya terhadap loyalitas serta retensi pelanggan. Hasil penelitian menunjukkan bahwa persentase retensi menurun seiring waktu, dengan jumlah pelanggan yang tetap aktif berkurang setelah pendaftaran. Namun, beberapa kohort menunjukkan performa retensi yang lebih baik. Misalnya, Cohort April mengalami peningkatan signifikan dari 18.46% ke 29.23%, dan Cohort Mei mengalami peningkatan tertinggi dari 14.29% ke 30.16%. Di sisi lain, Cohort Januari menunjukkan penurunan retensi paling drastis dari 100% ke 10.42% setelah bulan pertama, dengan penurunan bulanan terbesar dari 23.96% di bulan April menjadi 7.29% di bulan Mei. Temuan ini memberikan panduan praktis bagi perusahaan e-commerce dalam mengembangkan strategi yang lebih efektif untuk meningkatkan retensi pelanggan. Analisis lebih lanjut direkomendasikan untuk memahami penyebab spesifik dari tren retensi ini dan mengembangkan strategi yang lebih holistik guna mendukung keputusan bisnis yang lebih baik serta meningkatkan keuntungan jangka panjang.

Kata kunci: e-commerce, retensi pelanggan, cohort retention, analisis data, strategi promosi, loyalitas pelanggan, Amerika Serikat.

PENDAHULUAN

Perusahaan ritel *online* atau *e-commerce* telah menjadi salah satu sektor yang berkembang pesat, terutama di Amerika Serikat (AS). Pertumbuhan teknologi yang pesat, didukung oleh akses internet yang semakin mudah dan perangkat mobilitas yang semakin canggih, serta perubahan perilaku konsumen yang mengutamakan kenyamanan dan kecepatan, telah mengubah cara orang berbelanja. Konsumen kini lebih memilih berbelanja *online* karena

kemudahan dalam membandingkan harga, membaca ulasan produk, dan melakukan transaksi dari mana saja tanpa perlu mengunjungi toko fisik. Selain itu, perkembangan metode pembayaran digital dan sistem logistik yang efisien turut mendorong pertumbuhan *e-commerce* di AS. Pandemi COVID-19 juga berperan dalam mempercepat adopsi *e-commerce*, karena banyak konsumen beralih ke belanja *online* untuk memenuhi kebutuhan sehari-hari selama masa pembatasan sosial (Zelalem, dkk., 2022).

Untuk memahami pola pembelian dan retensi pelanggan di industri *e-commerce*, analisis data menjadi sangat penting. Salah satu sumber data yang kaya adalah dataset dari Kaggle yang berisi catatan transaksi *e-commerce* di AS pada tahun 2020. Dataset ini mencakup informasi penting seperti tanggal pesanan, mode pengiriman, identifikasi pelanggan, kategori produk, penjualan, jumlah barang yang dibeli, diskon dan keuntungan. Kategori produk yang dijual dalam *e-commerce* ini meliputi teknologi, furnitur, dan perlengkapan kantor. Selain itu, analisis ini dapat membantu dalam mengidentifikasi tren penjualan, mengevaluasi efektivitas kampanye pemasaran, dan mengoptimalkan strategi bisnis untuk meningkatkan kepuasan dan loyalitas pelanggan (Gonzalez & Rabbi, 2023).

Persaingan harga yang ketat telah memunculkan istilah 'perang harga' dan 'bakar duit' untuk promosi. Dari perspektif pelanggan, hal ini sangat menguntungkan karena mereka mendapatkan harga yang lebih rendah. Namun, dari perspektif perusahaan, keputusan ini harus dianalisis secara cermat untuk memastikan dampak positif terhadap performa penjualan. Data mengenai transaksi penjualan dan pembelian yang berjumlah besar tersimpan secara elektronik dalam database perusahaan dan dapat dimanfaatkan untuk menghitung pertumbuhan pelanggan. Analisis performa penjualan yang tercatat setiap tahun dalam suatu database sangat penting untuk memahami dampaknya secara menyeluruh terhadap bisnis. Selain itu, riset pemasaran dapat membantu perusahaan dalam beberapa cara, seperti meningkatkan kualitas pengambilan keputusan, menelusuri masalah, mengidentifikasi peluang pasar baru, dan mengembangkan strategi pemasaran yang lebih efektif (Aparicio, dkk., 2024).

Riset pemasaran yang baik membantu perusahaan fokus pada hal terpenting, yaitu mempertahankan konsumen yang ada. Ini juga membantu mereka memahami pasar dengan lebih baik, memperingatkan tentang tren pasar, dan mengukur nilai persepsi (*perceived value*) atas produk dan jasa. Melalui riset pemasaran, perusahaan dapat memperoleh wawasan berharga yang mendukung strategi bisnis mereka secara keseluruhan dan membantu meningkatkan tingkat kepuasan konsumen serta kualitas pengambilan keputusan (Fedushko & Ustyianovych, 2022). Selain itu, riset pemasaran memungkinkan perusahaan untuk menyesuaikan penawaran mereka dengan kebutuhan dan preferensi pelanggan,

mengidentifikasi peluang untuk inovasi produk, dan mengembangkan kampanye pemasaran yang lebih efektif dan terarah, terlebih saat kondisi dimana adanya hal baru yang tiba-tiba terjadi seperti wabah COVID-19 yang tidak terprediksi terjadi sehingga perlu dilakukan riset pemasaran yang lebih terkini (Fatihudin & Firmansyah., 2019).

Perubahan kondisi perekonomian dan tren teknologi dalam berbelanja, seperti penurunan daya beli masyarakat dan tren di sejumlah kota besar yang beralih ke pemanfaatan belanja *online*, juga memainkan peran penting dalam pertumbuhan *e-commerce*. Pada tahun 2020, ritel *online* berkembang sangat pesat, dipicu oleh kondisi pandemi COVID-19, dimana berbelanja online dianggap sebagai solusi di era kebiasaan baru. Toko ritel pun kemudian digantikan oleh toko ritel *online*. Kondisi ini membuat perusahaan ritel harus bertahan dan mengikuti tren teknologi serta pola yang berkembang di masyarakat untuk tetap menjalankan usahanya. Beberapa strategi yang digunakan termasuk penggunaan pemasaran digital dengan memanfaatkan media sosial dan influencer, branding produk, dan memberikan potongan harga atau promo kepada pelanggan. Banyak perusahaan memberikan harga promo kepada pelanggan dengan tujuan untuk menarik mereka agar membeli produk tersebut. Selain itu, peningkatan pengalaman pengguna di platform online melalui layanan pelanggan yang responsif dan pengiriman yang cepat juga menjadi kunci keberhasilan dalam mempertahankan dan menarik pelanggan baru (Donthu & Gustafsson., 2020)

Penelitian ini akan mengeksplorasi tingkat loyalitas pelanggan *e-commerce* di AS pada tahun 2020. Tujuan utama penelitian ini adalah untuk menganalisis retensi kohort berdasarkan data pelanggan *e-commerce* di Amerika Serikat tahun 2020. Dengan cara ini, kita dapat memahami jumlah pelanggan dari setiap kohort yang terus melakukan pembelian pada bulan-bulan berikutnya. Penelitian ini juga akan mengevaluasi implementasi analisis data untuk menilai efektivitas strategi promosi dan dampaknya terhadap loyalitas serta retensi pelanggan. Menggunakan dataset transaksi *e-commerce* dari AS tahun 2020 yang tersedia di Kaggle mengenai data pelanggan, penelitian ini bertujuan untuk memberikan wawasan yang dapat membantu perusahaan membuat keputusan yang lebih baik dan meningkatkan keuntungan jangka panjang.

KAJIAN TEORITIS

Retensi atau tingkat loyalitas pelanggan memiliki dampak signifikan bagi bisnis. Retensi dalam konteks bisnis merujuk pada kecenderungan pelanggan untuk terus menggunakan produk atau layanan dari penyedia layanan mereka (Ranaweera & Neely, 2003). Retensi pelanggan mencerminkan pengalaman interaksi mereka dengan perusahaan terkait dan

memainkan peran penting bagi perusahaan tersebut dalam menciptakan keunggulan kompetitif yang berkelanjutan di tengah persaingan industri (Andreini, dkk., 2018). Perusahaan di berbagai sektor semakin menaruh perhatian kepada tingkat retensi pelanggan mereka. Hal ini dilakukan guna mendeteksi sejak dini risiko pelanggan mereka yang ingin beralih atau berhenti menggunakan layanan perusahaan terkait (Ascarza, 2018).

Tingkat retensi yang dikelola sejak dini dapat membantu perusahaan terkait untuk membuat rencana peningkatan layanan mereka yang sesuai meskipun dengan sumber daya terbatas (Hung, dkk., 2020). Pengelolaan retensi sejak dini dapat dilakukan dengan perbaikan layanan secara berkala. Hal ini dapat meningkatkan tingkat retensi pelanggan yang lebih tinggi di masa mendatang. Loyalitas pelanggan disebabkan karena sikap atau perilaku mereka. Hal ini dapat dipengaruhi oleh tingkat kualitas produk yang baik, reputasi bagus perusahaan di pasar, lokasi toko yang memadai dan tingkat harga (Xhema, dkk., 2018). Oleh sebab itu, langkah terbaik bagi suatu perusahaan untuk meningkatkan loyalitas pelanggan mereka adalah dengan lebih fokus pada kualitas produk dan/ layanan mereka serta reputasi perusahaan. Alih-alih menurunkan biaya dengan menempatkan toko mereka di daerah terpencil. Hal ini dapat mempengaruhi perilaku pelanggan mereka ke arah negatif. Karena lokasi toko mempengaruhi perilaku pelanggan, semakin jauh lokasi toko, maka semakin banyak pelanggan akan beralih ke pesaing (Xhema, dkk., 2018).

Pelanggan cenderung memiliki niat yang lebih kuat untuk tetap menggunakan layanan suatu perusahaan jika mereka memulainya dengan promosi bebas biaya bulanan dibandingkan dengan promosi bebas biaya pendaftaran (Kim, 2019). Hal tersebut menunjukkan bahwa nilai yang dirasakan mempunyai pengaruh positif yang signifikan terhadap tingkat pelanggan suatu perusahaan. Hal ini bisa menjadi pertimbangan strategi awal perusahaan. Di samping itu, diketahui pula pemasaran media sosial tidak memberikan dampak signifikan terhadap nilai yang dirasakan tetapi memberikan dampak positif dan signifikan terhadap retensi pelanggan (Hanaysha, 2018). Hal ini menunjukkan bahwa langkah atau strategi awal perusahaan memiliki dampak positif yang signifikan terhadap retensi atau loyalitas pelanggan mereka.

Ada hubungan yang kuat antara biaya pindah dan perilaku retensi pelanggan. Kompleksitas rencana layanan yang mencerminkan harga dan penggunaan berhubungan positif dengan perilaku retensi pelanggan (Seo, dkk., 2008). Di samping itu, terdapat hubungan yang kuat antara biaya peralihan dan perilaku retensi pelanggan. Kompleksitas rencana layanan, mencerminkan harga dan penggunaan layanan yang berhubungan positif dengan perilaku retensi pelanggan (Seo, dkk., 2008). Tamuliene & Gabryte (2014) menemukan adanya korelasi positif yang signifikan secara statistik antara faktor-faktor dalam model teoritis: kepuasan

pelanggan, kualitas hubungan, biaya peralihan, dan retensi pelanggan. Kepuasan pelanggan secara signifikan mempengaruhi retensi pelanggan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan *analisis cohort*. Metode analisis ini digunakan untuk memahami retensi pelanggan dengan membagi mereka ke dalam kelompok atau *cohort* berdasarkan waktu mereka pertama kali mulai menggunakan layanan atau produk. Analisis ini memberikan wawasan tentang bagaimana retensi pelanggan berubah dari waktu ke waktu dan membantu mengidentifikasi faktor-faktor yang mempengaruhi retensi. Dalam metode ini terdapat sejumlah komponen utama, diantaranya: (1) *cohort* yang merupakan pengelompokan pengguna berdasarkan karakteristik waktu tertentu, seperti tanggal pendaftaran atau tanggal pertama kali melakukan pembelian (Johnson, 2019); (2) *retention rate* merupakan persentase pengguna dalam cohort yang tetap aktif atau melakukan tindakan tertentu selama periode waktu tertentu (Brown, 2018); (3) periode analisis yang merupakan interval waktu yang digunakan untuk mengukur retensi, misalnya harian, mingguan, bulanan (Anderson & Lee, 2017).

Cohort retention analysis dapat dilakukan dengan menggunakan berbagai pendekatan statistik dan analisis data seperti regresi logistik, survival analysis, dan *time-series analysis* untuk mengidentifikasi faktor-faktor yang mempengaruhi retensi dan membuat prediksi berdasarkan pola retensi yang diamati (Meyer, 2016). Visualisasi data menggunakan *heatmap* dan *lineplot* dapat memberikan wawasan yang lebih dalam mengenai retensi pelanggan dari waktu ke waktu (Anderson & Lee, 2017). *Heatmap* menunjukkan representasi grafis dari data di mana nilai-nilai individu dalam matriks direpresentasikan sebagai warna (Brown, 2018). Warna yang lebih gelap biasanya menunjukkan retensi yang lebih tinggi, sementara warna yang lebih terang menunjukkan retensi yang lebih rendah (Tamuliene & Gabryte, 2014). *Heatmap* sangat berguna untuk melihat perubahan retensi pelanggan dari bulan ke bulan dalam satu pandangan komprehensif (Syaqirah & Faizurrahman, 2014). Penggunaan *lineplot* menunjukkan penggambaran tren yang dengan menggunakan garis yang menunjukkan tingkat retensi dari waktu ke waktu (Meyer, 2016). *Lineplot* sangat berguna untuk melihat bagaimana retensi pelanggan berubah dari waktu ke waktu untuk setiap cohort secara individual (Smith, 2020).

HASIL DAN PEMBAHASAN

Pada bagian ini, penelitian ini memaparkan langkah-langkah analisis yang dilakukan.

Ask the Questions

Fokus dari penelitian ini adalah analisis terhadap kinerja penjualan. Beberapa pertanyaan yang harus dijawab oleh data adalah terkait dengan kenaikan penjualan dari waktu ke waktu. Penelitian ini akan menggunakan jumlah transaksi per bulan untuk menentukan apakah tren penjualan cenderung naik atau turun. Selain itu, pertumbuhan pelanggan juga akan dianalisis untuk melihat apakah jumlah pelanggan bertambah atau tidak. Hal ini dilakukan melalui analisis *customer retention* dengan *metode cohort* untuk mengukur seberapa baik perusahaan dalam mempertahankan pelanggan dari waktu ke waktu. Sebelum melakukan analisis lebih lanjut, berikut adalah deskripsi data secara umum yang disajikan dalam tabel 1.

Tabel 1. Daftar Variabel Awal Penelitian

No	Nama Kolom	Deskripsi
1	Order Date	Represents Dates
2	Row ID	Represents Row ID
3	Order ID	Represents Order ID
4	Ship Mode	Represents Ship Mode
5	Customer ID	Represents Customer ID
6	Segment	Represents Segments
7	Country	This Column Represent Country
8	City	Represent Cities
9	State	This column represent States
10	Postal Code	Represents Postal Code
11	Region	Represent Region
12	Product ID	Represents Product ID
13	Category	Represents Category
14	Sub-Category	Represents Sub-Category
15	Product Name	Represent Product Name
16	Sales	Represent Sales
17	Quantity	Represent Quantity
18	Discount	Represent Discounts
19	Profit	Represent Profit

Obtaining Data

Data yang digunakan dalam penelitian ini data publik yang diambil dari pangkalan data Kaggle. Proses analisis dilakukan dengan menggunakan alat Google Collaboratory. Sebelum masuk ke *exploring data*, beberapa tahapan yang dilakukan adalah:

a) Mengatur koneksi data dan import data ke Google Collaboratory

Pengaturan koneksi dari Google Drive ke Google Collaboratory dilakukan dengan menggunakan *link e-mail* akun google yang terdaftar. Proses data *cleaning* atau pembersihan data menggunakan *Python* menghasilkan data yang bersih dan terstruktur dengan baik. Tujuan

utama proses ini adalah untuk memastikan bahwa data yang digunakan dalam menganalisis atau pemodelan adalah akurat, konsisten, dan lengkap (Kotsiantis, dkk., 2006). Data yang telah diproses dapat digunakan untuk melakukan analisis lebih lanjut, seperti analisis statistik, visualisasi data dan interpretasi hasil analisis (McKinney., 2013).

b) *Data understanding*

Pemahaman data dilakukan dengan menggunakan berbagai fungsi dan teknik seperti analisis deskriptif, pengujian hipotesis, dan eksplorasi visual. Fungsi-fungsi ini memungkinkan peneliti untuk mengidentifikasi pola, tren, dan anomali dalam data yang dapat memberikan wawasan penting sebelum melangkah ke tahap analisis lebih lanjut (Hadley & Garrett., 2016):

1) Distribusi data

Untuk memahami distribusi dari data yang menjadi perhatian dapat dilakukan dengan sintaks `df.describe()`. Fungsi ini menampilkan ringkasan statistik dari data numerik, termasuk nilai rata-rata, standar deviasi, nilai minimum dan maksimum, serta kuartil (Shmueli, ,dkk., 2016).

2) Tipe data

Untuk memahami tipe data dari setiap kolom dalam dataset dapat digunakan sintaks `df.info()`. Fungsi ini memberikan ringkasan informasi mengenai jumlah entri, tipe data, dan jumlah nilai non-null untuk setiap kolom, yang sangat berguna untuk inspeksi awal dataset (Shmueli, ,dkk., 2017).

c) *Data cleaning*

Proses pelibatan identifikasi dan koreksi data yang tidak akurat, hilang, atau tidak konsisten untuk meningkatkan kualitas data yang akan digunakan dalam analisis selanjutnya. Proses ini mencakup penghapusan data duplikat, penanganan nilai yang hilang, dan standarisasi format data untuk memastikan keandalan dan validitas hasil analisis (Carlo., 2018)

1) Mengecek *missing value*

Untuk mendeteksi nilai yang hilang dalam dataset, dapat digunakan sintaks `df.isnull().sum()`. Fungsi ini akan menghitung jumlah nilai yang hilang di setiap kolom, yang penting untuk langkah awal dalam proses pembersihan data guna memastikan kualitas data yang tinggi untuk analisis lebih lanjut (Joel, ,dkk., 2022).

2) *Me-remove missing value*

Salah satu cara untuk mengatasi nilai yang hilang adalah dengan menghapus data yang kosong tersebut menggunakan sintaks `df.dropna()`, kemudian memeriksa bentuk data baru dengan `df.shape`. Pendekatan ini memastikan bahwa hanya data lengkap yang

digunakan dalam analisis, meskipun ada risiko kehilangan informasi berharga jika banyak data yang dihapus (Van Buuren & Groothuis., 2011).

3) Mengecek data duplikat

Pengecekan data duplikasi juga merupakan bagian penting dari proses pembersihan data. Proses ini dapat dilakukan dengan menggunakan sintaks `df.duplicated().sum()`, yang menghitung jumlah baris duplikat dalam dataset. Identifikasi dan penghapusan data duplikat membantu memastikan keakuratan dan integritas data yang akan digunakan dalam analisis lebih lanjut (Zhang., 2016).

4) Menghapus data yang duplikat

Setelah mengidentifikasi data yang duplikat, langkah selanjutnya adalah menghapus data duplikat tersebut dengan menggunakan sintaks `df.drop_duplicates(keep='first', inplace=True)`, dan kemudian memeriksa bentuk data baru dengan `df.shape`. Pendekatan ini memastikan bahwa hanya satu entri dari data duplikat yang dipertahankan, yang penting untuk menjaga integritas dan kualitas data (Rahman & Islam., 2013).

Exploring Data

Setelah dataset bersih dari missing value dan duplikasi maka langkah berikutnya adalah proses exploring data. Beberapa exploring yang dilakukan adalah:

a) Mengidentifikasi pertumbuhan pelanggan

Analisis untuk mengidentifikasi pelanggan ini dapat dilakukan dengan menganalisis *customer retention*. Untuk dapat mengetahui *customer retention* dapat menggunakan sintaks heatmap pada Gambar 1.

```
1 plt.figure(figsize=(12,8))
2 sns.heatmap(cohort_retention_final,
3             annot=True,
4             cbar=True,
5             linewidths=0.5,
6             cmap='YlOrBu',
7             cbar_kws={'orientation':'horizontal',
8                     annot_kws={'size': 18}})
9 plt.set_title('Monthly Customer Retention Cohort 2020')
10 plt.xlabel('Cohort Month')
11 plt.ylabel('Month Distance')
12 plt.show()
```

gambar 1. Sintaks heatmap program untuk customer retention

Dari hitungan di gambar 1 dapat dibuat visualisasi sebagaimana terlihat pada Gambar 2.



Gambar 2. Visualisasi Heatmap *Customer Retention*

atau dapat juga menggunakan sintaks line plot pada gambar 3. sebagai berikut ini

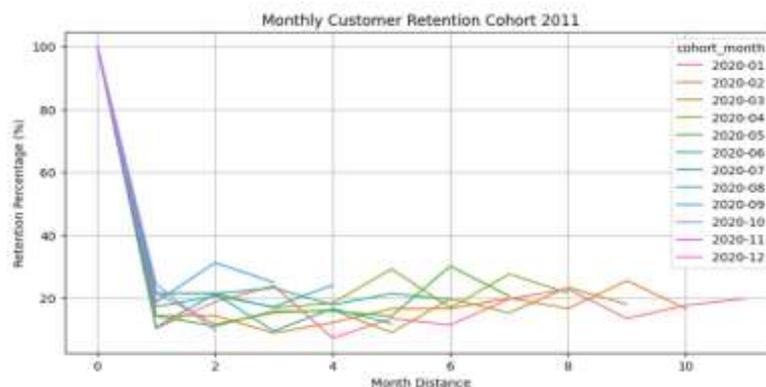
```

1 #Create Lineplot
2
3 plt.figure(figsize=(10,5))
4 ax=sns.lineplot(df_plot,
5                 x='period_distance',
6                 y='percentage_retention',
7                 hue='cohort_month')
8 ax.set_title('Monthly Customer Retention Cohort 2011')
9 ax.set_ylabel('Retention Percentage (%)')
10 ax.set_xlabel('Month Distance')
11 plt.grid()
12 plt.show()

```

Gambar 3. Sintaks Line plot program untuk customer retention

Dari hitungan di gambar 3 dapat dibuat visualisasi Line plot sebagaimana terlihat pada Gambar 4.



Gambar 4. Visualisasi *Lineplot Customer Retention*

Dari gambar 2 dan 4 di atas, dapat dilihat mengenai tren retensi bahwa persentase retensi menurun seiring waktu, dimana jumlah pelanggan yang tetap aktif berkurang seiring berjalannya waktu setelah pendaftaran. *Cohort* terbaik mengalami peningkatan yang signifikan pada *Cohort* April sebesar 18.46% ke 29.23%, atau bisa juga hampir 11% dari bulan agustus ke bulan september. Selain itu pada *Cohort* Mei memiliki peningkatan paling tinggi sebesar 14.29% ke 30.16% atau hampir 16% dari bulan Oktober ke bulan November. Untuk

Cohort Terburuk mengalami Penurunan paling drastis dari awal pendaftaran ke bulan pertama yaitu pada *Cohort* Januari sebesar 100% ke 10.42% mengalami penurunan hampir 90%. Pada *Cohort* Januari juga mengalami penurunan per bulan paling besar yaitu Bulan April ke Bulan Mei sebesar 23.96% ke 7.29% mengalami penurunan hampir 17%.

KESIMPULAN DAN SARAN

Dari hasil visualisasi analisis di atas, dapat disimpulkan bahwa persentase retensi pelanggan menurun seiring waktu, dengan jumlah pelanggan yang tetap aktif berkurang setelah pendaftaran. Meskipun demikian, terdapat beberapa kohort yang menunjukkan performa retensi yang lebih baik. Beberapa kohort, seperti yang berasal dari April dan Mei 2020, memiliki tingkat retensi yang sedikit lebih baik pada bulan-bulan berikutnya dibandingkan dengan kohort lainnya. Ini mungkin mengindikasikan faktor-faktor spesifik selama bulan-bulan tersebut yang mempengaruhi retensi secara positif. Penelitian ini merekomendasikan analisis lebih lanjut untuk dapat membantu untuk memahami penyebab spesifik dari tren ini dan mengembangkan strategi untuk meningkatkan retensi pelanggan di masa mendatang.

Penelitian ini juga merekomendasikan sejumlah strategi bisnis yang dapat dipertimbangkan dari hasil analisis untuk mempertahankan tingkat retensi pelanggan. Pertama, membentuk strategi *early engagement* seperti, program *onboarding* untuk melibatkan pelanggan secara efektif dalam dua bulan pertama. Hal ini dilakukan untuk pengalaman terarah untuk memastikan pelanggan memahami nilai produk/layanan. Kedua, melakukan analisis segmen dimana identifikasi *cohort* dengan retensi tinggi. Hal ini dilakukan untuk mengidentifikasi karakteristik umum atau strategi yang berhasil. Ketiga, strategi keterlibatan jangka panjang seperti dengan membentuk komunitas. Adanya komunitas di sekitar produk/layanan mendorong pelanggan dapat berinteraksi, berbagi pengalaman, dan mendapatkan dukungan dari rekan-rekan. Komunitas yang kuat dapat meningkatkan loyalitas pelanggan.

DAFTAR REFERENSI

- Anderson, M., & Lee, S. (2017). Analyzing cohort data in digital marketing. *Journal of Marketing Analytics*, 4(3), 189-204.
- Aparicio, D., Metzman, Z., & Rigobon, R. (2024). The pricing strategies of online grocery retailers. *Quantitative Marketing and Economics*, 22(1), 1-21.
- Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1), 80–98.

- Brown, A. (2018). The impact of retention strategies in e-commerce. *Journal of Business Research*, 62(2), 133-145.
- Carlo, S., & Batini, M. (2018). *Data and information quality: Dimensions, principles and techniques*. Springer.
- Donthu, N., & Gustafsson, A. (2020). Effects of COVID-19 on business and research. *Journal of Business Research*, 117, 284-289.
- Fatihudin, D., & Firmansyah, A. (2019). *Pemasaran jasa (strategi, mengukur kepuasan dan loyalitas pelanggan)*. Deepublish.
- Fedushko, S., & Ustyianovych, T. (2022). E-commerce customers behavior research using cohort analysis: A case study of COVID-19. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), 12.
- Gonzalez, M., & Rabbi, F. (2023). Evaluating the impact of big data analytics on personalized e-commerce shopping experiences and customer retention strategies. *Journal of Computational Social Dynamics*, 8(2), 13-25.
- Hadley, W., & Garrett, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.
- Hanaysha, J. R. (2018). Customer retention and the mediating role of perceived value in the retail industry. *World Journal of Entrepreneurship, Management and Sustainable Development*, 14(1), 2–24.
- Hung, P. H., Doi, K., & Inoi, H. (2020). User retention tendency of bus routes based on user behavior transition in an area with low mode share of public transport.
- Joel, L. O., Doorsamy, W., & Paul, B. S. (2022). A review of missing data handling techniques for machine learning. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 5(3), 971-1005.
- Johnson, K. (2019). Customer cohort analysis in practice. *Journal of Consumer Behavior*, 15(1), 29-37.
- Kim, J. (2019). The impact of different price promotions on customer retention. *Journal of Retailing and Consumer Services*, 46(October), 95–102.
- Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science*, 1(2), 111-117.
- McKinney, W. (2013). *Python for data analysis*. O'Reilly Media, Inc.
- Meyer, D. (2016). Advanced techniques in cohort analysis. *Journal of Data Science*, 14(2), 223-237.
- Rahman, M. G., & Islam, M. Z. (2013). Missing value imputation using decision trees and decision forests by splitting and merging records: Two novel techniques. *Knowledge-Based Systems*, 53, 51-65.

- Seo, D. B., Ranganathan, C., & Babad, Y. (2008). Two-level model of customer retention in the US mobile telecommunications service market.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). Data mining for business analytics: Concepts, techniques, and applications in R. John Wiley & Sons.
- Tamuliene, V., & Gabryte, R. (2014). Factors influencing customer retention: A theoretical model. *Business and Management Review*, 5(1), 23-31.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45, 1-67.
- Xhema, J., Groumpos, P. P., & Metin, H. (2018). Switching-costs, corporate image and product quality effect on customer loyalty: Kosovo retail market. *IFAC-PapersOnLine*, 51(30), 51-55.
- Zelalem Jembre, Y., Jung, W. Y., Attique, M., Paul, R., & Kim, B. (2022). Mobile broadband performance evaluation: Analysis of national reports. *Electronics*, 11(3), 485.
- Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, 4(1).