

Nama: Zulfa Nabilah Nurvitasari

*Learning Path: Data Analytics*

Tugas Topic 1 & 2: Esai *Use Case Data Science* & Membuat Halaman Github

1. Buatlah sebuah tulisan mengenai *use case data science* di industri tertentu yang dapat kalian pilih di bawah ini dan bahas dengan detail bagaimana data science diaplikasikan, metode yang digunakan, contoh data yang dipakai, dan impact terhadap bisnis di industri tersebut.

Industri yang Dibahas: Retail

### **Implementasi Data Science dalam Ritel Online: Analisis Customer Retention dan Clustering Customer dengan Metode K-Means**

#### **LATAR BELAKANG**

Di zaman yang sudah maju akan teknologi ini, banyak sekali industri-industri yang mengikuti perkembangan zaman dengan mengikuti tren teknologi guna mengikuti pola hidup masyarakat yang terus berkembang agar bisnis yang dijalankannya dapat tetap bertahan. Salah satu industri yang mengikuti perkembangan zaman adalah sektor ritel. Perkembangan sektor ritel dalam bidang teknologi ini didukung dengan adanya pandemi Covid-19 pada tahun 2020 dimana berbelanja *online* dianggap menjadi solusi di era pandemi tersebut. Oleh karena itu, banyak toko ritel luring yang tutup dan digantikan dengan toko ritel *online*.

Toko ritel *online* pun semakin banyak dan berkembang setiap tahunnya. Berdasarkan data BPS (2022) hingga 15 September 2022, sebanyak 34,10% usaha melakukan kegiatan *e-commerce*. Presentase ini meningkat jika dibandingkan dengan presentase pada 31 Desember 2021 dan 2020, yang secara berturut-turut sebesar 32,23% dan 25,25%. Tidak hanya itu, Direktur Pemberdayaan Informatika, Direktorat Jendral Aplikasi Informatika Kominfo, Septriana Tangkary melalui situs Kominfo menyatakan bahwa pertumbuhan *e-commerce* di Indonesia pada tahun 2019 mencapai 78% tertinggi di dunia. "Indonesia merupakan negara 10 terbesar pertumbuhan *e-commerce* dengan pertumbuhan 78% dan berada di peringkat ke-1," kata Septriana Tangkary di Pamekasan, Jawa Timur, Rabu (27/2/19).

Karena ketatnya persaingan antar toko ritel online, maka mereka pun berlomba-lomba dalam mempromosikan toko ritel *onlinenya* lewat marketing digital. Beberapa strategi yang digunakan dalam promosi adalah memanfaatkan *social media* dengan membuat *content marketing*, promosi melalui *influencer*, *branding* produk, dan memberikan promo atau potongan harga untuk pelanggan. Banyak sekali toko ritel online yang memberikan promo kepada pelanggan dengan tujuan agar pelanggan tertarik untuk berbelanja di toko mereka.

Persaingan harga antar toko ritel *online* yang sangat ketat memunculkan istilah 'perang harga' dan 'bakar duit' untuk sebuah promo. Dari sudut pandang pelanggan, hal ini sangat menguntungkan dirinya. Akan tetapi, di lain sisi, perusahaan *e-commerce* harus melakukan analisis terhadap data para pelanggan sebelum memberi promo atau potongan harga tersebut untuk memastikan bahwa promo yang diberikan kepada pelanggan tidak berpengaruh negatif terhadap perusahaan dan diharapkan dari promo yang diberikan tersebut dapat membawa keuntungan yang besar bagi perusahaan *e-commerce*.

Data yang digunakan untuk analisis berasal dari data transaksi penjualan dan pembelian para konsumen yang menggunakan jasa *e-commerce* tersebut. Karena pengguna *e-commerce* semakin meningkat setiap tahunnya, maka data konsumen yang dimiliki perusahaan juga semakin banyak dan semakin beraneka ragam dan tersimpan secara elektronik dalam suatu *database*. Data tersebut tidak berguna jika tidak diolah. Oleh karena itu, *data science* diperlukan untuk mengolah data terutama data yang besar atau bisa disebut dengan *big data* sehingga informasi yang ada pada data tersebut dapat diketahui dan dapat digunakan sebagai acuan dalam mengambil suatu tindakan. Salah satu contohnya ialah informasi mengenai pertumbuhan pelanggan yang digunakan untuk pertimbangan dalam pemberian promo oleh perusahaan *e-commerce*. Dalam penelitian ini akan membahas implementasi *data science* terhadap suatu ritel online untuk menganalisis *customer retention* dengan metode *cohort analytics* dan pengelompokan pelanggan/*clustering* dengan algoritma K-Means. Data yang digunakan ialah data penjualan dari sebuah perusahaan ritel online yang diambil selama satu tahun terakhir.

## LANDASAN TEORI

### Data

Data adalah kumpulan fakta yang diperoleh dari hasil pengamatan atau pengukuran. Berdasarkan bentuknya, data dibedakan menjadi dua kategori, yaitu:

- a) Data terorganisir (*organized data*) merupakan data yang terstruktur atau berurutan dalam baris atau kolom, dimana setiap baris mewakili satu nilai dan setiap kolom mewakili karakteristik nilai tersebut.
- b) Data tidak terorganisir (*unorganized data*) merupakan data yang bentuknya bebas seperti *teks raw audio*/ sinyal yang harus diubah menjadi terstruktur.

### Big Data

*Big data* adalah kumpulan data yang sangat besar dan kompleks. Oleh karena itu, sangat susah mengolah data tersebut dengan teknik manajemen data tradisional sehingga diperlukan *data science* untuk mengolah data tersebut sehingga informasi yang terkandung dalam data dapat diekstraksi dengan baik. *Big data* memiliki karakteristik yang terdiri dari:

- Volume  
*Big data* memerlukan infrastruktur yang elastis dan pengolahan yang terdistribusi
- Velocity  
Pada *e-commerce*, biasanya tingkat kecepatan data yang keluar dan masuk sekitar satu petabyte/detik. Hal ini lebih cepat dari pengolahan data dan perlindungan data di perusahaan *e-commerce*. Hal ini dapat memburuk ketika terdapat jutaan konsumen yang masuk ke situs *e-commerce* secara bersamaan.
- Variety  
*Big data* merupakan kumpulan data yang diambil dari berbagai macam sumber yang terorganisir, tidak terstruktur, dan tidak terorganisir. Data dapat berupa gambar, video, dokumen dan yang lainnya. Oleh karena itu, data beberapa kali diatur agar dapat diolah sehingga memberikan informasi yang diperlukan untuk analitik data.

- Value

*Big data* memiliki informasi yang sangat besar dan sangat berguna bagi perusahaan *e-commerce* yang didapatkan melalui pengolahan dan analisis yang tepat.

### **Data Science**

*Data Science* adalah ilmu pengetahuan multidisiplin yang mempelajari tentang data atau dapat juga diartikan sebagai proses dalam menghasilkan informasi atau pengetahuan dari data (*data insight*). Untuk menghasilkan produk data dengan *data science*, maka harus melalui tiga fase terlebih dahulu, yaitu fase desain data, pengumpulan data, dan analisis data. *Data Science* memiliki beberapa cabang ilmu, diantaranya:

a) *Traditional Software*

*Traditional Software* merupakan cabang ilmu hasil dari irisan cabang ilmu komputer dengan SME (*Subject Matter Expertise*). SME adalah pengetahuan yang digunakan untuk mengembangkan sistem yang dapat membantu proses bisnis pada sebuah instansi

b) *Traditional Research*

*Traditional Research* merupakan cabang ilmu hasil dari irisan cabang ilmu matematika dan statistika dengan SME (*Subject Matter Expertise*).

c) *Machine Learning*

*Machine Learning* adalah cabang ilmu kecerdasan buatan (*Artificial Intelligence*) yang mempelajari tentang cara memberikan kemampuan belajar pada sebuah mesin (computer) menggunakan algoritma tertentu.

Dalam prosesnya, *data science* memiliki enam tahapan yang harus dilewati, yaitu:

- 1) Menentukan tujuan
- 2) Pengambilan data. Data yang sudah dikumpulkan biasanya masih mentah yang perlu di transformasi dan dipoles terlebih dahulu sebelum digunakan.
- 3) *Preprocessing data*. Karena data yang didapatkan masih mentah, maka diperlukan pendeteksian dan perbaikan berbagai jenis kesalahan pada data.
- 4) Eksplorasi data. Tahap ini bertujuan untuk mendapatkan pemahaman mendalam tentang data. Hal ini dapat dilakukan dengan melihat pola, mencari korelasi, dan penyimpanan pada teknik visual dan deskriptif.
- 5) Pemodelan data. Pemodelan data dilakukan untuk mendapatkan informasi pada data dan juga melakukan prediksi.
- 6) Mengomunikasikan dan memvisualisasikan hasil. Pada tahap terakhir, yang perlu dilakukan ialah menyimpulkan hasil serta mengkomunikasikannya dalam visualisasi yang baik.

### **Data Mining**

*Data mining* adalah proses untuk mengekstraksi dan mengidentifikasi informasi dan pengetahuan yang terkait dari berbagai *database* besar. disiplin ilmu gabungan dari statistik, matematika, kecerdasan buatan, dan *machine learning*. Berdasarkan tugas yang dilakukan, data mining dibagi menjadi enam kelompok, yaitu deskripsi, prediksi, asosiasi, estimasi, klasifikasi dan pengklasteran.

### ***Clustering dengan metode K-Means***

Pengklasteran adalah proses mengidentifikasi dan pengelompokan data pengamatan atau objek-objek berdasarkan kemiripan. Pengklasteran merupakan salah satu kegunaan dari *data science*. Salah satu algoritma *clustering* yang paling banyak digunakan adalah K-Means. Algoritma K-Means melakukan *clustering* dengan metode partisi dan melakukan pemodelan tanpa supervise berdasarkan persamaan karakteristik.

### ***Cohort analytics***

*Cohort* adalah beberapa individu yang masuk ke dalam sebuah sistem pada waktu yang bersamaan (kelompok entitas). *Cohort* biasanya dikelompokkan berdasarkan tahun atau ukuran periode waktu tertentu. *Cohort* analisis adalah analisis yang menjelaskan mengenai perbedaan antara dua dimensi temporal. Keanggotaan dalam *Cohort* ditentukan oleh waktu interval yang bergantung pada pertimbangan analisis dan kondisi yang ingin diteliti.

### ***E-Commerce***

Pengertian *e-commerce* secara umum merujuk pada semua bentuk transaksi yang berkaitan dengan aktifitas komersial, baik dalam organisasi maupun individual yang berdasarkan pemrosesan transaksi data yang didigitalisasikan.

## **HASIL DAN PEMBAHASAN**

Berdasarkan enam tahapan dalam *data science*, maka hasil dan pembahasan penelitian dijelaskan sebagai berikut:

### **1) Menentukan tujuan**

Tujuan penelitian ini yaitu untuk menganalisis kinerja penjualan dengan menggunakan jumlah transaksi per bulan. Selain itu, ingin dilihat juga pertumbuhan pelanggan melalui analisis *customer retention* dengan metode *cohort*.

### **2) Pengambilan data**

Data yang digunakan dalam penelitian ini merupakan data public yang bersumber dari pangkalan data Kaggle.

### **3) Preprocessing data**

Sebelum melakukan eksploratori, data perlu dipreprocessing terlebih dahulu agar siap digunakan. Tahap *preprocessing* yang dilakukan adalah *data cleaning* seperti mengecek dan mengatasi data hilang, pengecekan dan penanganan data duplikat serta tahap *data manipulation*.

### **4) Eksplorasi data**

Setelah data bersih dari data hilang dan duplikasi, maka langkah selanjutnya ialah eksplorasi data. Beberapa eksploratori data yang dilakukan ialah:

#### **a) Analisis pendapatan per-bulan**

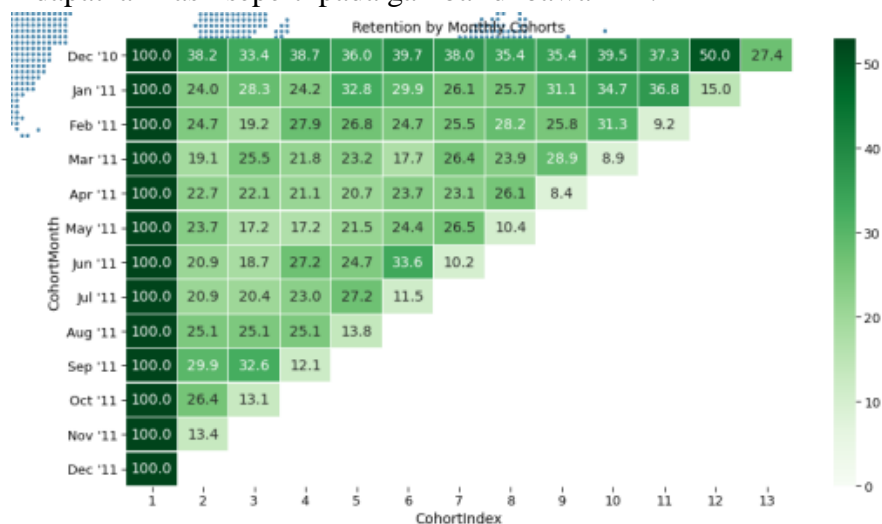
Hasil perhitungan volume penjualan pada tahun 2011 dapat dilihat melalui diagram batang berikut.



Dari gambar tersebut, kita dapat melihat bahwa selama tahun 2011 volume penjualan dari bulan Januari hingga Agustus cenderung stabil meskipun terdapat kenaikan atau penurunan sedikit pada jangka waktu tersebut. Sedangkan, volume penjualan pada bulan September hingga November mengalami peningkatan setiap bulannya walaupun pada akhirnya pada bulan Desember mengalami penurunan. Dapat diketahui pula bahwa volume penjualan tertinggi terjadi pada bulan November, sedangkan volume penjualan terendah terjadi pada bulan Desember.

#### b) Analisis pertumbuhan pelanggan

Analisis pertumbuhan pelanggan dapat dilakukan dengan menganalisis *customer retention*. Didapatkan hasil seperti pada gambar di bawah ini:



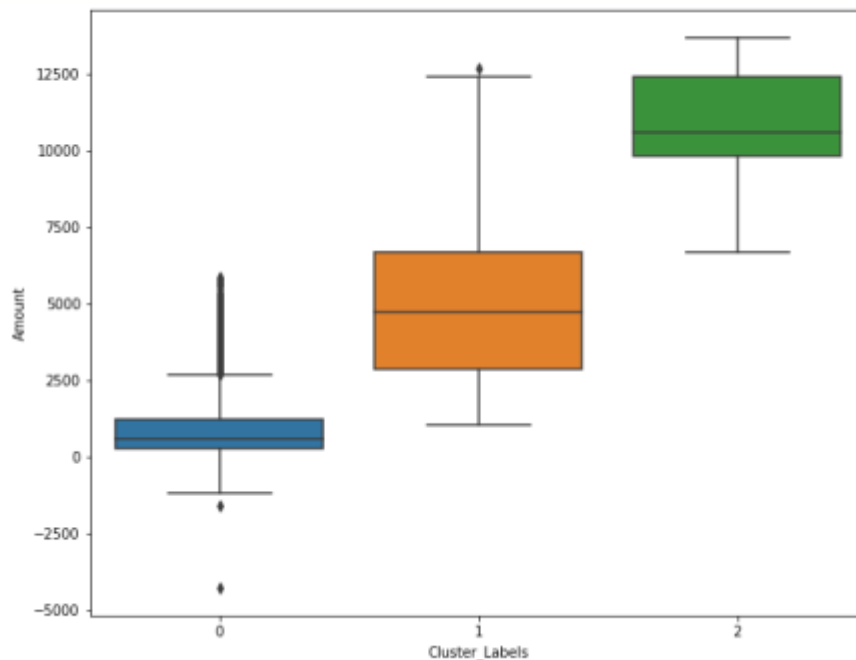
Berdasarkan gambar di atas, dapat diketahui bahwa pertumbuhan pelanggan mengalami penurunan sebesar 27,4% - 50%, hal ini menandakan bahwa target presentase yang ideal tidak tercapai seperti pada bulan pertama atau naik pada tahun 2011.

### 5) Pemodelan data

Langkah pemodelan data menggunakan K-Means *Clustering*. Jumlah kluster terbaik ditentukan oleh metode Elbow Criterion. Dengan model yang telah didapatkan, maka kejadian yang akan datang dapat diprediksi sehingga dapat dilakukan pencegahan jika akan terjadi kejadian yang tidak diinginkan.

## 6) Mengomunikasikan dan memvisualisasikan hasil

Hasil visualisasi dari klaster yang terbentuk dengan frekuensi penjualan dapat dilihat pada gambar berikut:



Informasi yang dapat diambil dari box plot di atas adalah pelanggan yang berada di klaster dengan label 2 ialah pelanggan yang memiliki jumlah transaksi paling tinggi jika dibandingkan dengan pelanggan lainnya. Pelanggan di klaster 2 merupakan pelanggan tetap yang loyal. Sedangkan, pelanggan yang berada di klaster dengan label 0 ialah pelanggan yang memiliki jumlah transaksi paling rendah di antara pelanggan lainnya. Untuk pelanggan yang berada di klaster label 1 merupakan pelanggan yang hanya beberapa kali melakukan transaksi.

## KESIMPULAN

Dari data penjualan tersebut, dapat diketahui bahwa perusahaan tersebut mengalami pertumbuhan pelanggan dengan rata-rata 37,42% sepanjang tahun 2011. Volume penjualan tertinggi terjadi pada bulan November, sedangkan volume penjualan terendah terjadi pada bulan Desember. Dengan diketahuinya informasi tersebut, sebaiknya perusahaan meninjau kembali penyebab kenaikan dan penurunan volume penjualan tersebut sehingga diharapkan untuk kedepannya volume perusahaan tetap stabil, bahkan naik dan tidak mengalami penurunan kembali. Selain itu agar pelanggan yang berada di klaster dengan label 1 dapat menjadi pelanggan yang setia dan loyal, sebaiknya para pelanggan tersebut diberi promo atau potongan harga agar mereka lebih sering melakukan transaksi di *e-commerce* tersebut. Dari pernyataan-pernyataan tersebut, dapat disimpulkan bahwa informasi yang terkandung dalam big data dapat diperoleh dengan mengolahnya menggunakan data science. Informasi yang didapatkan tersebut dapat digunakan oleh perusahaan sebagai bahan evaluasi untuk meningkatkan kualitas serta kuantitas perusahaan.

## REFERENSI

- Cielen, D., Meysman, A. and Ali, M. 2016. *Introducing data science: big data, machine learning, and more, using Python tools*.
- Fadli, A. 2020. Konsep Dasar Data Science. *Jurnal Ilmu Komputer*. pp. 1–7.
- Gao, Y. 2021. Big Data Analysis on E-commerce Platform. *Proceedings of the 2021 International Conference on Economic Development and Business Culture (ICEDBC 2021)*. 182(Icedbc), pp. 191–195. doi: 10.2991/aebmr.k.210712.031.  
[https://www.kominfo.go.id/content/detail/16770/kemkominfo-pertumbuhan-e-commerce-indonesiacapai-78-persen/0/sorotan\\_media](https://www.kominfo.go.id/content/detail/16770/kemkominfo-pertumbuhan-e-commerce-indonesiacapai-78-persen/0/sorotan_media). Diakses pada 18 Maret 2023.
- Oktora, R., dkk. 2022. Statistik E-Commerce. pp. 9–41.
- Riswandi. 2019. Transaksi On-Line (E-Commerce): Peluang dan Tantangan Dalam Perspektif Ekonomi Islam. *Angewandte Chemie International Edition*. 6(11). 951–952. 13(April). pp. 15–38.
- Sari, I. P. 2021. Implementasi Data Science dalam Ritel Online: Analisis Customer Retention dan Clustering Customer dengan Metode K-Means. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*. 5(1), pp. 417–425.
- Setyawan, D. A. 2013. Data dan Metode Pengumpulan Data Penelitian. *Metodologi Penelitian*. pp. 9–17.
- Singh, R., Verma, N. and Gupta, M. 2020. ROLE OF E-COMMERCE IN BIG DATA. 11(12), pp. 1770–1777.
- Yuli Mardi. 2019. Data Mining: Klasifikasi Menggunakan Algoritma C4. 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ). *Jurnal Edik Informatika, Jurnal Edik Informatika*. 2(2). pp. 213–219.
2. Buatlah sebuah github page dengan contoh sebagai berikut: <https://chriskhanhtran.github.io/minimal-portfolio/Links to an external site.>
- URL Github: <https://github.com/ZulfaNabilahNurvitasari>
- URL LinkedIn: [www.linkedin.com/in/zulfanabilahnurvitasari](https://www.linkedin.com/in/zulfanabilahnurvitasari)