

Exploratory Data Analysis Untuk Data Belanja Pelanggan dan Pendapatan Bisnis

Widi Hastomo^{1*}, Adhitio Satyo Bayangkari Karno², Sudjiran³, Dodi Arif⁴, Eka Sally Moreta⁵

¹Program Studi Teknologi Informasi, ITB Ahmad Dahlan Jakarta

^{2,4}Program Studi Manajemen, Universitas Gunadarma

^{3,5}Program Studi Sistem Informasi, STMIK Jakarta STI&K

¹Jl. Ir H. Juanda No.77, Cireundeu, Kec. Ciputat Tim., Kota Tangerang Selatan, Banten 15419, Indonesia

^{2,4}Jl. Margonda Raya No.100, Pondok Cina, Beji, Kota Depok, Jawa Barat 16424, Indonesia

^{3,5}Jl. BRI Radio Dalam No.17, Kebayoran. Baru, Kota Jakarta Selatan, DKI Jakarta 12140, Indonesia

E-mail: widie.has@gmail.com¹, adh1t10@yahoo.com², ontosenosujiran@gmail.com³, dodiarif8@gmail.com⁴, ekamoreta@gmail.com⁵

Abstrak

Info Naskah:

Naskah masuk: 22 Mei 2022

Direvisi: 13 Juli 2022

Diterima: 22 Juli 2022

Pandangan manajemen mekanistik sedang digantikan oleh pandangan yang lebih terukur untuk mendorong bisnis berdasarkan kemampuannya guna mengubah data menjadi pengetahuan dan wawasan. Industri belum memanfaatkan *data driven* secara maksimal untuk menunjang strategi bisnis mereka. Eksperimen ini bertujuan untuk menganalisa data belanja pelanggan dan pendapatan bisnis retail fashion menggunakan *dataset* transaksi yang diperoleh dari salah satu kompetisi di Kaggle.com (*H&M Personalized Fashion Recommendations*). Metode yang digunakan yaitu EDA menggunakan bahasa *python*, hasil dari eksperimen ini yaitu memperlihatkan kuantitas transaksi berdasarkan umur pelanggan, produk dengan penjualan terbanyak dan hanya sekali serta jenis produk yang memberikan pendapatan terbesar dan terkecil. Hasil eksperimen ini diharapkan dapat mendukung kemampuan simulasi, agar industri dapat mengembangkan hasil analitik yang memberikan pandangan ke depan, serta membantu menjembatani kesenjangan antara *knowledge* dan *wisdom* untuk memberikan hasil analitik yang bersifat saran yang berkelanjutan. Tantangan dalam eksperimen ini yaitu jumlah *dataset* yang besar, membutuhkan lingkungan operasi yang mumpuni.

Keywords:

EDA;
python;
customer shopping;
revenue.

Abstract

A more quantifiable perspective is assuming the role of mechanistic management in an effort to enhance business based on its capacity to transform data into knowledge and insight. The industry has not completely supported its business strategy also with driven data. Using a transaction dataset taken from one of the Kaggle.com challenges, this experiment attempts to determine consumer spending patterns and Retail Fashion business revenues (*H&M Personalized Fashion Recommendations*). The results of the experiment are the number of transactions based on customer age, the most sales product and one-time purchased item, and the type of product that generates the highest and smallest income. The approach employed is EDA using the Python language. In order for businesses to generate analytical findings that provide future perspectives and to help identify the gap by delivering analytical results in the form of suggestions that can be perpetuated, the findings of this experiment are intended to support the capabilities of simulation. The challenge in this experiment is the abundance of datasets, which necessitates a suitable operating environment.

*Penulis korespondensi:

Widi Hastomo

E-mail: widie.has@gmail.com

1. Pendahuluan

Era globalisasi yang sangat pesat dan ketat akan kompetisi, setiap organisasi atau bisnis dituntut untuk mampu meningkatkan daya saing. Peningkatan daya saing bisnis di era globalisasi dilakukan dengan merangkul perkembangan teknologi informasi dan komunikasi (TIK). Kemampuan TIK yang semakin canggih mampu mengumpulkan, mengolah, menganalisa data bisnis dalam jumlah besar (*Big Data*) [1] serta mampu menghasilkan informasi tersembunyi yang menunjukkan nilai inti (*core value*) dan karakteristik nyata suatu bisnis berdasarkan perjalanan data suatu organisasi atau bisnis tersebut.

Digitalisasi dan kemajuan dalam teknologi terkait erat dengan inovasi yang signifikan dalam industri ritel. Laju perubahan telah dipercepat secara signifikan selama krisis Covid-19. Misalnya, selama pandemi, volume belanja secara global meningkat dari Februari 2020 hingga April 2021 [2]. Dengan demikian, terjadi lonjakan inovasi dalam model bisnis ritel untuk mengatasi peningkatan harapan pelanggan, adopsi teknologi, integrasi rantai pasokan, tantangan logistik, dan pemasaran digital [3]. Hal ini dapat diartikan bahwa jika sebuah bisnis tidak memanfaatkan teknologi secara optimal, maka bisnis tersebut mendekati masa terpuruk.

Kemampuan TIK saat ini dalam mengolah data dalam jumlah besar dapat mengurangi bias, berbeda dengan analisis tradisional dengan hanya mengandalkan data yang terbatas (*sample*). Dengan analisa *Big Data* (BD) dapat memberikan pola data yang lebih akurat, walaupun diperlukan metode yang lebih rumit. Ketersediaan *library pandas* dalam bahasa *python* membantu mengatasi data dengan kapasitas puluhan hingga ratusan juta *dataset* [4], dibandingkan dengan mengundung *software spreadsheet* dengan kemampuan maksimum hanya 1 jutaan saja.

Library matplotlib, seaborn dan lainnya didalam *python* mampu menampilkan visualisasi data dengan tampilan menarik, dinamis [5] dan memudahkan para pemangku kepentingan untuk mengambil keputusan cepat dan tepat. Eksperimen ini menggunakan bahasa *python*, selain bersifat *open source*, mudah diperoleh, juga memiliki kemampuan untuk mengolah data besar, banyak digunakan dalam analisa data [6]–[16], *data mining* [17] dan *data science* [18]. Analisa data mampu memberikan gambaran profil bisnis dalam selang waktu tertentu. Merupakan tahap awal yang umum dilakukan sebelum masuk ke dalam tahap proses analisa yang lebih dalam [19]. Dengan analisa data dapat diketahui kelompok data *outlier*, data kosong (tidak lengkap), dan berbagai ringkasan statistik lainnya [20].

Penelitian terkait analisis pembelian konsumen dan prediksi perilaku konsumen telah dilakukan oleh [21] menggunakan model *Deep Neural Network (DNN)*. Penelitian terkait yang menggunakan *Artificial Intelligence (AI) Natural Language Processing (NLP)* telah meningkatkan kinerja bisnis melalui analisis sentimen [22]. Penelitian yang telah dilakukan oleh [23] analisis urutan perilaku konsumen, minat konsumsi dan memprediksi perilaku pembayaran menggunakan *Multivariate Real-Time Sequence Analysis*. Penelitian yang telah dilakukan oleh [24] untuk mengembangkan kecerdasan bisnis dan peningkatan kepuasan tentang *E-commerce* menggunakan

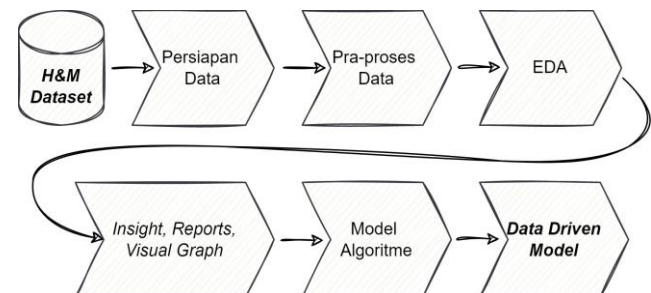
algoritme *Machine Learning (ML)*. Penelitian yang bertujuan untuk mengatasi masalah kelebihan informasi dan kehilangan informasi dalam pemilihan platform belanja *online e-commerce*, sistem rekomendasi yang dipersonalisasi menggunakan teknologi penyaringan informasi menggunakan ML [25].

Penelitian terkait sentimen pelanggan dalam menentukan analisis belanja menggunakan metode ML NLP [26]. Penelitian yang dilakukan oleh [27] yang bertujuan untuk memberikan model yang efisien untuk mengoptimalkan sistem penjualan produk di perusahaan farmasi menggunakan metode *clustering* dan berdasarkan indikator dan algoritme ML. Penelitian yang dilakukan oleh [28] terkait rekomendasi produk untuk konsumen pada layanan *e-commerce*, untuk membantu memahami pelanggan. Penelitian terkait proses bisnis digital, pemasaran, meningkatkan wujud dan fokus kepada kebutuhan pelanggan menggunakan ML dengan luaran visualisasi grafik [29]. Penelitian yang dilakukan oleh [30] yang bertujuan untuk meningkatkan akurasi dan efektivitas rekomendasi yang dipersonalisasi, sehingga meningkatkan penjualan produk dan kepuasan pengguna, dengan metode ML.

Berdasarkan dari paparan sebelumnya, penelitian ini bertujuan untuk memanfaatkan data guna menciptakan siklus melalui umpan balik dari proses EDA, serta meta data pelanggan dan produk yang telah dibeli yang menjadi pendapatan bisnis dalam industri tersebut. Eksperimen ini diharapkan mampu membuka wawasan baru, menciptakan pengetahuan dan data dapat diperkaya dengan konteks, agar dapat diintegrasikan untuk membentuk kearifan dalam menentukan strategi bisnis kedepannya.

2. Metode

Ilustrasi dari metode eksperimen ini terlihat pada Gambar 1. Sebelum ke tahap *Exploratory Data Analysis (EDA)*, dilakukan proses persiapan data dan pra-proses data. Hasil dari EDA berupa wawasan, laporan dalam bentuk grafik secara visual, selanjutnya membuat model algoritme dan hasil akhirnya berupa *data driven model*.



Gambar 1. Ilustrasi model eksperimen

2.1 Persiapan Data

Jumlah data yang digunakan dalam proses eksperimen ini mencapai 31.788.324 baris transaksi dalam durasi dua tahun (2018-2020) bisnis retail *fashion* dari *H&M group*. Eksperimen ini menggunakan *dataset* yang diperoleh dari keikutsertaan dalam kompetisi Kaggle [31] "*H&M*

```
Jumlah data = 105542
articles.columns = Index(['article_id', 'product_code', 'prod_name', 'product_type_no',
'product_type_name', 'product_group_name', 'graphical_appearance_no',
'graphical_appearance_name', 'colour_group_code', 'colour_group_name',
'perceived_colour_value_id', 'perceived_colour_value_name',
'perceived_colour_master_id', 'perceived_colour_master_name',
'department_no', 'department_name', 'index_code', 'index_name',
'index_group_no', 'index_group_name', 'section_no', 'section_name',
'garment_group_no', 'garment_group_name', 'detail_desc'],
dtype='object')
```

Gambar 3. Ekstraksi file “articles.csv”

```
Jumlah data = 1371980
customers.columns = Index(['customer_id', 'FN', 'Active', 'club_member_status',
'fashion_news_frequency', 'age', 'postal_code'],
dtype='object')
```

	customer_id	FN	Active	club_member_status	fashion_news_frequency	age	postal_code
0	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	NaN	NaN	ACTIVE	NONE	49.0	52043ee2162cf5aa7ee79974281641c6f11a68d276429a...
1	0000423b00ade91418ccea3b26c6af3dd342b51fd051e...	NaN	NaN	ACTIVE	NONE	25.0	2973abc54daa8a5f8ccfe9362140c63247c5ee03f1d93...
2	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	NaN	NaN	ACTIVE	NONE	24.0	64f17e6a330a85798e4998f62d0930d14db8db1c054af6...
3	00005ca1c9ed5f5146b52ac8639a40ca9d57aeff4d1bd2...	NaN	NaN	ACTIVE	NONE	54.0	5d36574f52495e81f019b680c843c443bd343d5ca5b1c2...
4	00006413d8573cd20ed7128e53b7b13819fe5cfc2d801f...	1.0	1.0	ACTIVE	Regularly	52.0	25fa5ddee9aac01b35208d01736e57942317d756b32ddd...
...
1371975	ffffb78b6eaac697a8a5dfbf2bfa8113ee5b403e474...	NaN	NaN	ACTIVE	NONE	24.0	7aa399f7e669990daba2d92c577b52237380662f36480b...
1371976	ffffc5046a6143d29a04fb8c424ca49a76e5cdf4fab5...	NaN	NaN	ACTIVE	NONE	21.0	3f47f1279beb72215f4de557d950e0fa73789d24acb5e...
1371977	ffffc35913a0bee60e8741cb2b4e78b8a98ee5f2e6a1...	1.0	1.0	ACTIVE	Regularly	21.0	4563fc79215672cd6a863f2b4b5f6b8f98f2d96ed590e...
1371978	ffffd7744cebct3aca44ae7049d2a94b87074c3d4ffe38...	1.0	1.0	ACTIVE	Regularly	18.0	8892c18e9bc3dca6aa4000cb8094fc4b51ee8db2ed14d7...
1371979	ffffd9ac14e89946416d80e791d064701994755c3ab686...	NaN	NaN	PRE-CREATE	NONE	65.0	0a1a03306fb2f62164c2a439b38c0caa64b40deaae8687...

1371980 rows x 7 columns

Gambar 4. Ekstraksi File “customer.csv”

Jumlah data = 1371980

	customer_id	prediction
0	00000dbacae5abe5e23885899a1fa44253a17956c6d1c3...	0706016001 0706016002 0372860001 0610776002 07...
1	0000423b00ade91418ccea3b26c6af3dd342b51fd051e...	0706016001 0706016002 0372860001 0610776002 07...
2	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	0706016001 0706016002 0372860001 0610776002 07...
3	00005ca1c9ed5f5146b52ac8639a40ca9d57aeff4d1bd2...	0706016001 0706016002 0372860001 0610776002 07...
4	00006413d8573cd20ed7128e53b7b13819fe5cfc2d801f...	0706016001 0706016002 0372860001 0610776002 07...

Gambar 5. File “sample_submission.csv”

Personalized Fashion Recommendations”. H&M Group adalah brand bisnis produk pakaian (fashion) dengan 53 pasar online dan 4.850 toko fisik diseluruh dunia.

Dataset terdiri dari 4 file berekstensi csv (“articles.csv”, “customer.csv”, “sample_submission.csv” dan “transactions_train.csv”) dan 1 folder “images” berisi file gambar produk berekstensi jpg. File “articles.csv” adalah informasi mengenai detail metadata tiap produk pakaian yang dijual. File ini berisi 105.542 baris data dengan 20 kolom fitur detail kriteria tiap produk (Gambar 3). Ekstraksi file “customer.csv” berjumlah 1.371.980 dengan 7 kolom detail metadata customer, diperlihatkan pada Gambar 4.

File “sample_submission.csv” menunjukkan format file yang harus di kirim (submit) pada kompetisi. File ini berisi prediksi 1.371.980 data customer yang diperkirakan membeli jenis produk selanjutnya dalam 12 periode waktu berikutnya, diperlihatkan Gambar 5.

File “transactions_train.csv”, 31.788.324 baris data dengan 5 fitur kolom metadata berisi informasi transaksi pembelian yang berlangsung dalam kurun waktu 2 tahun (20-9-2018 sampai dengan 22-9-2020), pada Gambar 6. Folder “images” yang berisi 86 folder dengan tiap folder berisi gambar pakaian dan fashion jenis tertentu, pada Gambar 7.

Jumlah data= 31788324
 periode waktu = 2018-09-20 - 2020-09-22

	t_dat	customer_id	article_id	price	sales_channel_id
0	2018-09-20	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	663713001	0.050831	2
1	2018-09-20	000058a12d5b43e67d225668fa1f8d618c13dc232df0ca...	541518023	0.030492	2
2	2018-09-20	00007d2de826758b65a93dd24ce629ed66842531df6699...	505221004	0.015237	2
3	2018-09-20	00007d2de826758b65a93dd24ce629ed66842531df6699...	685687003	0.016932	2
4	2018-09-20	00007d2de826758b65a93dd24ce629ed66842531df6699...	685687004	0.016932	2

Gambar 6. File “transactions_train.csv”



Gambar 7. Produk pakaian dan fashion H&M group

```
print(articles.isnull().sum())
```

article_id	0	perceived_colour_master_name	0
product_code	0	department_no	0
prod_name	0	department_name	0
product_type_no	0	index_code	0
product_type_name	0	index_name	0
product_group_name	0	index_group_no	0
graphical_appearance_no	0	index_group_name	0
graphical_appearance_name	0	section_no	0
colour_group_code	0	section_name	0
colour_group_name	0	garment_group_no	0
perceived_colour_value_id	0	garment_group_name	0
perceived_colour_value_name	0	detail_desc	416
perceived_colour_master_id	0	dtype: int64	

Gambar 8. Ringkasan statistik “articles.csv”

```
print(customers.isnull().sum())
```

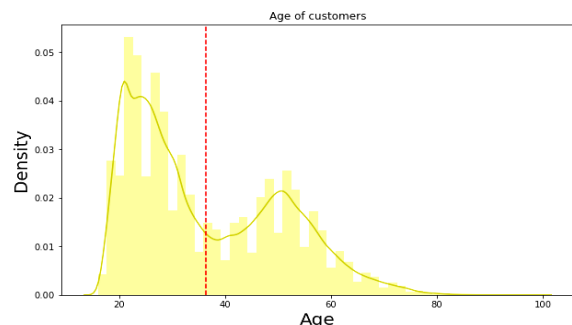
customer_id	0
FN	895050
Active	907576
club_member_status	6062
fashion_news_frequency	16009
age	15861
postal_code	0
dtype: int64	

```
customers.describe()
```

	FN	Active	age
count	476930.0	464404.0	1.356119e+06
mean	1.0	1.0	3.638696e+01
std	0.0	0.0	1.431363e+01
min	1.0	1.0	1.600000e+01
25%	1.0	1.0	2.400000e+01
50%	1.0	1.0	3.200000e+01
75%	1.0	1.0	4.900000e+01
max	1.0	1.0	9.900000e+01

Gambar 9. Ringkasan statistik “customers.csv”

average age of employees = 36.386964565794



Gambar 10. Sebaran density belanja berdasarkan umur customer.

2.2 Pra-prosesing Data

Untuk memudahkan memahami pola dasar data dipergunakan berbagai fungsi statistik dan visualisasi grafik. Cara pendekatan ini dapat juga dipergunakan untuk melihat data outlier dan anomaly data, tahap ini juga dipergunakan sebagai awal masuk ke tahap cleansing data, pencarian dan penanganan terhadap data kosong [5]. Untuk mengetahui jenis data atribut ordinal dan nominal diperlukan ringkasan atau descriptive statistic (mean, median, modus, quartil, min, max, standard deviation) terhadap ke tiga file (“articles.csv”, “customer.csv”, dan “transactions_train.csv”), pada Gambar 8, Gambar 9 dan Gambar 10.

Dari gambar 8, dapat diketahui bahwa data “articles.csv” tidak memiliki data kosong (null) dan kode produk berjumlah 1055420e+05. Rerata umur pelanggan yang telah berbelanja merupakan usia produktif, yaitu 36 tahun, hal ini terlihat dari ekstraksi dataset pada gambar 6.

```
tx_train.describe()
```

	article_id	price	sales_channel_id
count	3.178832e+07	3.178832e+07	3.178832e+07
mean	6.962272e+08	2.782927e-02	1.704028e+00
std	1.334480e+08	1.918113e-02	4.564786e-01
min	1.087750e+08	1.694915e-05	1.000000e+00
25%	6.328030e+08	1.581356e-02	1.000000e+00
50%	7.145820e+08	2.540678e-02	2.000000e+00
75%	7.865240e+08	3.388136e-02	2.000000e+00
max	9.562170e+08	5.915254e-01	2.000000e+00

```
tx_train.isnull().sum()
```

t_dat	0
customer_id	0
article_id	0
price	0
sales_channel_id	0
dtype: int64	

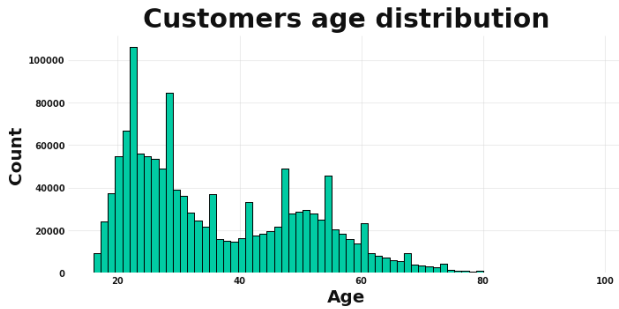
Gambar 11. Ringkasan statistik “transactions_train.csv”

Data “transactions_train.csv” tidak berisi data kosong (null), gambar 11 dapat diketahui bahwa sales dengan channel 2 dapat menjual produk lebih dari 50% dibandingkan sales channel 1.

3. Hasil dan Pembahasan

3.1 Kuantitas Pembelian Dan Pendapatan Perusahaan Berdasarkan Umur Pelanggan

Gambar 12-a dan 12-b memperlihatkan kerapatan belanja pelanggan tertinggi yang berada dalam rentang pada kisaran umur 20-30 tahun. Dengan melakukan plotting umur pelanggan dan jumlah kuantitas pembelian dapat di visualisasikan dalam bentuk vertical bar chart, memberikan informasi bahwa pelanggan berumur 20-30 tahun telah melakukan pembelian dengan kuantitas terbesar (42.5%), dan juga memberikan pendapatan tertinggi perusahaan (41.9%).



Gambar 12-a. Sebaran jumlah belanja berdasarkan umur pelanggan

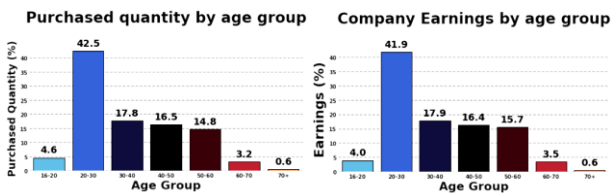
```

article_id  sold_qty
0 7080180001 50287
1 7080180002 35043
2 3728800001 31718
3 6107790002 30199
4 7598710002 26329

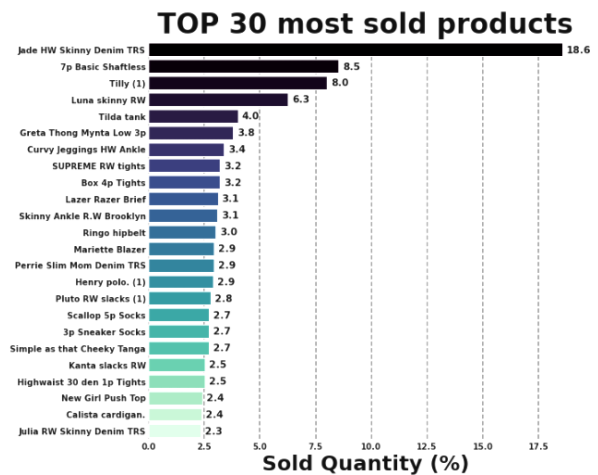
df_sold_qty["sold_qty"].describe()

count    184547.000000
mean      384.857735
std       791.266060
min        1.000000
25%       14.000000
50%        65.000000
75%       286.000000
max      58287.000000
Name: sold_qty, dtype: float64
    
```

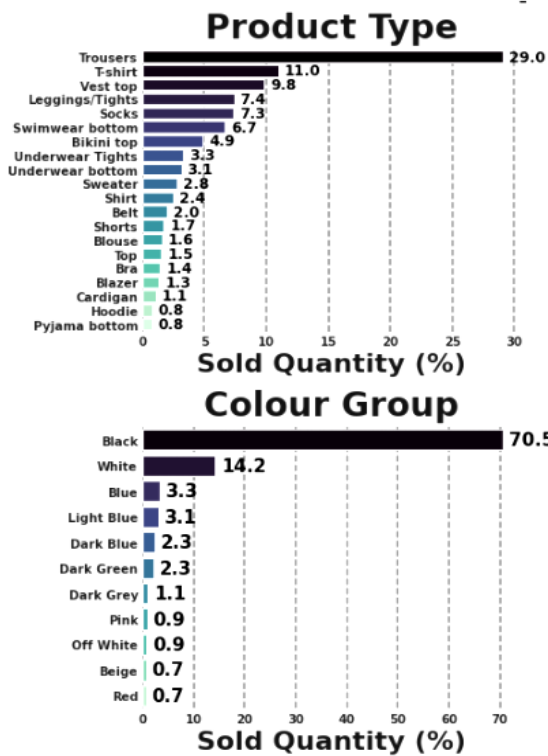
Gambar 13. Ringkasan statistik kuantitas penjualan



Gambar 12-b. Kuantitas belanja dan pendapatan perusahaan berdasarkan umur pelanggan



Gambar 14. Jenis produk yang terjual terbanyak (Top 30)



Gambar 15. Karakteristik produk penjualan terbanyak (Top 100)

3.2 Produk dengan penjualan terbanyak

Dengan mengekstrak kuantitas penjualan per artikel menggunakan metode `value_counts` di kolom `"article_id"` dari data `"transactions_train.csv"`, maka akan diperoleh ringkasan statistik kuantitas penjualan. Dari 104.547 jenis artikel transaksi (yang telah terjual), ada beberapa jenis yang terjual hanya sekali, jenis produk terjual dengan kuantitas 14 (25%), jenis produk terjual sebanyak 65 (50%), jenis produk terjual sebanyak 286 (75%), dan jenis produk yang terjual maksimum sebanyak 50.287, pada Gambar 13.

Dari 100 nama produk yang mempunyai penjualan sangat besar, dilakukan *plotting* 30 artikel tertinggi (gambar 14), dapat diketahui bahwa terdapat 4 jenis produk yang terjual paling banyak dengan hampir 40% dari seluruh penjualan, terdiri dari `"Jade HW Skinny Denim TRS"` dengan hampir 19%, `"7p Basic Shaftless"` sebanyak 8.5%, `"Tilly (1)"` 8% dan `"Luna skinny RW"` sebanyak 6.3% dari seluruh penjualan. Karakteristik produk lain (selain nama produk), akan diperoleh *plot* yang sangat efektif dengan mempertimbangkan dari 100 produk penjualan terbanyak, pada Gambar 15. Hampir 30% tipe produk terjual bertipe `"trouser"`, 38% ber-indeks pakaian wanita (`Ladieswear`), 30% ber-indeks `"Lingerie/Tights"`, lebih dari 70% berwarna hitam, hampir 40% termasuk kelompok produk `"Garment Lower body"`, dan `"Garment Upper body"` sekitar 32%.

3.3 Produk dengan penjualan sekali

Setelah mengetahui jenis produk dengan penjualan terbesar, perlu juga untuk mengetahui jenis produk yang sulit dijual atau hanya sekali terjual. Terdapat 4.491 item terjual hanya sekali (Gambar 16). Dari seluruh transaksi terdapat sekitar 100.000 item berbeda, ini berarti hanya sekitar 5% produk yang terjual hanya sekali.

Pada Gambar 17 memberikan informasi bahwa dari produk dengan penjualan minimum, 60% diantaranya adalah produk dengan *item* untuk produk anak-anak.

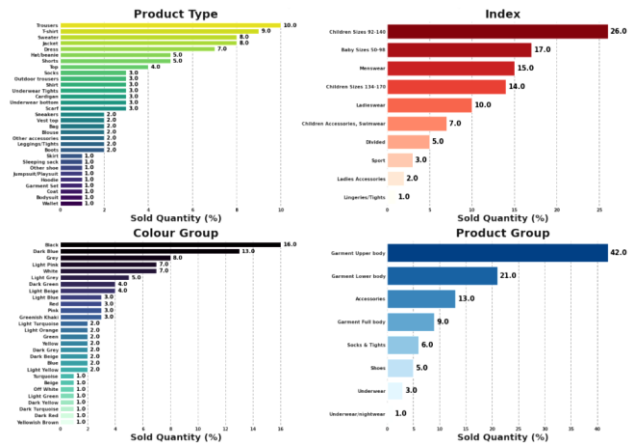
3.4 Jenis Produk Yang Memberikan Pendapatan Terbesar

Setelah melakukan analisa dari tiap produk dengan kuantitas penjualan, selanjutnya dari data yang ada akan di *plot* grafik yang menunjukkan total pendapatan (*earning*) yang diperoleh dari tiap penjualan produk. Pendapatan dihitung dengan mengalikan harga dengan total kuantitas jenis produk yang terjual. Karena alasan *privacy*, maka harga telah ditransformasikan dalam bentuk skala (*scaled*) sehingga tidak dinyatakan dalam bentuk mata uang (*currency*). Dari gambar 18 dapat diketahui pendapatan terbesar yang di peroleh suatu produk adalah 1.631 dengan total pendapatan perusahaan sebesar 884.645.

```
df_sold_qty["sold_qty"].where(lambda x: x==1).dropna()
```

```
100056 1.0
100057 1.0
100058 1.0
...
104542 1.0
104543 1.0
104544 1.0
Name: sold_qty, Length: 4491, dtype: float64
```

Gambar 16. Produk dengan penjualan sekali



Gambar 17. Karakteristik produk terjual hanya sekali

```
article_id  earning
0  708016001  1631.732102
1  708016002  1136.321085
2  588601006  939.268593
3  448509014  781.476390
4  708016003  692.195915

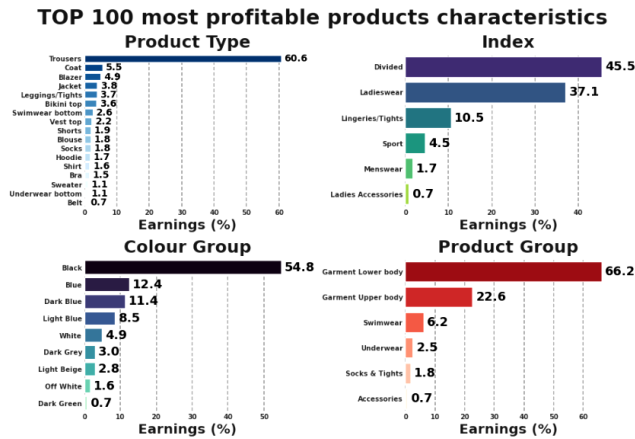
print("Number of different sold articles:", len(df_prices["earning"]))
print("Total Earnings:", df_prices["earning"].sum())
```

Number of different sold articles: 104547
Total Earnings: 884645.974858846

Gambar 18. Besar pendapatan perusahaan



Gambar 19. 50 produk yang memberikan keuntungan terbesar

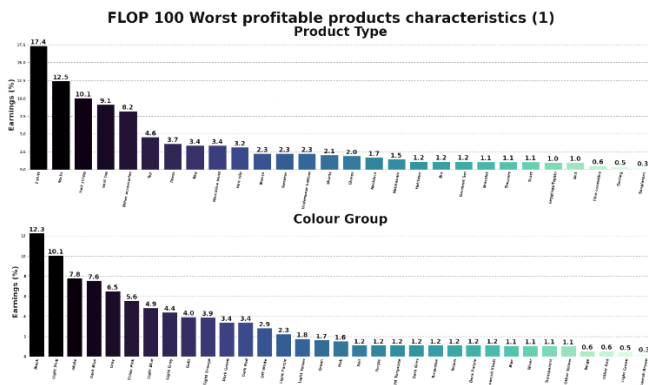


Gambar 20. Karakteristik Top 100 produk yang memberikan keuntungan untuk perusahaan

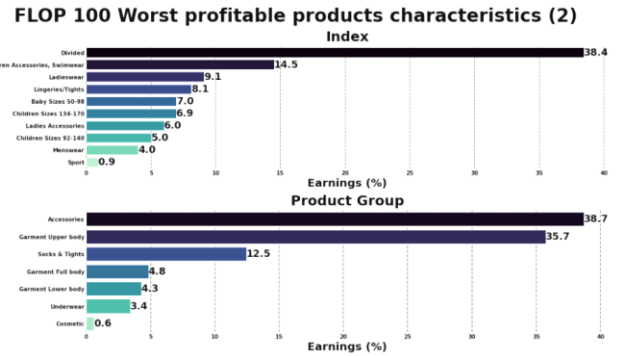
Jenis produk “*Jade HW Skinny Denim TRS*” memberikan keuntungan terbesar yaitu 18% dari 50 produk yang tergolong dapat memberikan banyak keuntungan untuk perusahaan (gambar 17). Wawasan informasi yang diperoleh (Gambar 20) yaitu lebih dari 60% dari 100 produk yang menguntungkan adalah produk berjenis “*Trousers*”, sekitar 50% produk dengan indek koleksi remaja, 37% adalah produk pakaian wanita (*Ladieswear*), 55% produk dengan warna hitam, dan 66,2% produk berhubungan dengan pakaian bagian bawah (*lower body*).

3.5 Jenis Produk Yang Memberikan Pendapatan Terendah

Setelah melakukan analisa produk yang dapat memberikan keuntungan tertinggi, perlu juga untuk melihat jenis-jenis produk yang terjual dan sedikit memberikan keuntungan untuk perusahaan. Dari gambar 21, menunjukkan *bar chart* karakteristik jenis dan warna produk yang kurang memberikan keuntungan pada perusahaan. Informasi yang diperoleh adalah 17,4% dengan jenis “*t-shirt*” dan 12,5% berjenis kaus kaki (“*socks*”), dengan 12% dominasi warna hitam (*black*) dan 10% berwarna *pink* dan juga terdapat *accessories* lainnya seperti ikat rambut (“*hair bands*”), penjepit (“*clips*”) dan lainnya.



Gambar 21. Karakteristik jenis dan warna 100 produk yang tidak menguntungkan



Gambar 22. Karakteristik indek dan kelompok produk 100 produk yang tidak menguntungkan

Informasi yang diperoleh dari gambar 22, prosentase 38,7% untuk kelompok produk *assesories*, dan sekitar 35,7% untuk jenis produk bayi (“*babies*”). Kategori kosmetik mendapatkan penjualan produk yang paling rendah, yaitu 0,6%.

4. Kesimpulan

EDA menggunakan bahasa python telah mampu menunjukkan dan menyelaraskan bagaimana sebuah data dapat dimanfaatkan untuk organisasi guna menciptakan siklus yang baik melalui umpan balik dari proses data bisnis retail. Hasil simulasi dapat membuka pengetahuan dan wawasan baru yang terukur, mengubah dari data transaksi bisnis menjadi informasi dan memberikan makna dalam menghasilkan informasi serta pengetahuan yang dapat dijadikan rujukan dalam menentukan strategi bisnis retail.

Analisa data belanja pelanggan pada gilirannya dapat dimanfaatkan untuk mengoptimalkan operasi mereka secara berkelanjutan, serta menghasilkan prioritas dan tindakan yang sesuai dengan arah kebijakan organisasi. Setiap pelanggan dapat diberikan rekomendasi yang sesuai berdasarkan dari riwayat transaksi sebelumnya guna meningkatkan pengalaman berbelanja. Organisasi perlu menginvestasikan waktu dalam mengelola hubungan dengan pelanggan untuk mempertahankan sumber pendapatan ini.

Analisa pendapatan bisnis dapat dijadikan acuan dalam menentukan strategi perusahaan kedepannya. Memahami aliran pendapatan memungkinkan untuk menyadari pola pendapatan yang masuk, dan oleh karena itu dapat dengan cepat mengamati pergerakan yang tidak biasa atau perubahan tren pendapatan, dan mengidentifikasi penyebabnya. Penelitian selanjutnya dapat memanfaatkan teknologi dan *tools* Big Data semisal Hadoop Distributed File System (HDFS), Map Reduce (MR), Apache HBASE, Apache Cassandra, Mongo DB dan lainnya.

Ucapan Terimakasih

Penelitian ini terlaksana atas dukungan dari LP3M ITB Ahmad Dahlan Jakarta sehingga berjalan dengan baik.

Daftar Pustaka

[1] H. L. Taboada GL, “Exploratory Data Analysis and Data Envelopment Analysis of Urban Rail Transit,” *Electronics*, vol. 9, no. 8, p. 1270, 2020, doi:

- <https://doi.org/10.3390/electronics9081270>.
- [2] V. Mostaghel, R., Oghazi, P., Parida, V., & Sohrabpour, "Digitalization driven retail business model innovation: Evaluation of past and avenues for future research trends," *J. Bus. Res.*, vol. 146, pp. 134–45, 2022.
- [3] A. Gavrilă, S. G., & de Lucas Ancillo, "Spanish SMEs' digitalization enablers: E-Receipt applications to the offline retail market," *Technol. Forecast. Soc. Change*, vol. 162, 2021.
- [4] W. McKinney, *Pandas: Powerful Python data analysis toolkit. Release 1.1.4*, The Pandas.
- [5] Leinhardt G and Leinhardt S., "Chapter 3: Exploratory Data Analysis: New Tools for the Analysis of Empirical Data," in *Review of Research in Education*, 2019, pp. 85–157. doi: 10.3102/0091732X008001085.
- [6] W. Hastomo, "Gesture Recognition For Pencak Silat Tapak Suci Real-Time Animation," *J. Comput. Sci. Inf.*, vol. 2, no. 13, pp. 91–102, 2020, doi: <http://dx.doi.org/10.21609/jiki.v13i2.855>.
- [7] A. S. B. Karno, W. Hastomo, & Arif, D., and E. S. Moreta, "Optimasi Portofolio Dan Prediksi Cryptocurrency Menggunakan Deep Learning Dalam Bahasa Python," vol. 4, no. September, 2020.
- [8] A. Satyo, B. Karno, W. Hastomo, Y. Efendi, and R. Irawati, "Arsitektur Alexnet Convolution Neural Network (CNN) Untuk Mendeteksi Covid-19 Image Chest-Xray," pp. 482–485, 2021.
- [9] W. Hastomo and A. Satyo, "Kemampuan Long Short Term Memory Machine," vol. 4, no. September, pp. 229–236, 2020.
- [10] W. Hastomo, "Convolution Neural Network Arsitektur Mobilenet-V2 Untuk Mendeteksi Tumor Otak," vol. 5, no. Gambar 1, 2021.
- [11] A. S. B. Karno, W. Hastomo, and I. S. K. Wardhana, "Prediksi Jangka Panjang Covid-19 Indonesia Menggunakan Deep Learning Long-Term," *Semin. Nas. Teknol. Inf. dan Komun.*, pp. 483–490, 2020.
- [12] W. Hastomo, A. S. Bayangkari Karno, N. Kalbuana, A. Meiriki, and Sutarno, "Characteristic Parameters of Epoch Deep Learning to Predict Covid-19 Data in Indonesia," *J. Phys. Conf. Ser.*, vol. 1933, no. 1, p. 012050, 2021, doi: 10.1088/1742-6596/1933/1/012050.
- [13] W. Hastomo, "Klasifikasi Covid-19 Chest X-Ray Dengan Tiga Arsitektur Cnn (Resnet-152, Inceptionresnet-V2, Mobilenet-V2)," vol. 5, no. D1, 2021.
- [14] W. Hastomo and A. Satyo, "Long Short Term Memory Machine Learning Untuk Memprediksi Akurasi Nilai Tukar IDR Terhadap USD," vol. 3, 2019.
- [15] W. Hastomo and S. Bayangkari, "Diagnosa Covid-19 Chest X-Ray Dengan Convolution Neural Network Arsitektur Resnet-152," vol. 2, no. 1, pp. 26–33, 2021.
- [16] W. Hastomo, A. S. B. Karno, N. Kalbuana, E. Nisfiani, and L. ETP, "Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi Covid-19," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 7, no. 2, pp. 133–140, 2021, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/47411>
- [17] I. Kiguchi, M., Saeed, W., & Medi, "Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest," *Appl. Soft Comput.*, vol. 118, no. 108491, 2022.
- [18] C. Ozgur, T. Coliaiu, G. Rogers, Z. Hughes, and B. Myer-Tyson, "C. Ozgur, T. Coliaiu, G. Rogers, Z. Hughes & B. Myer-Tyson "MatLab vs. Python vs. R" *Journal of Data Science Vol. 15 No. 3 pp. 355-372,*" *J. data Sci. JDS*, vol. Vol 15, pp. 355–372, Jun. 2017.
- [19] A. M. Aqajari, S. A. H., Naeini, E. K., Mehrabadi, M. A., Labbaf, S., Dutt, N., & Rahmani, "pyEDA: An open-source python toolkit for pre-processing and feature extraction of electrodermal activity," *Procedia Comput. Sci.*, vol. 184, pp. 99–106, 2021.
- [20] U. Mukhiya, S. K., & Ahmed, *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd, 2020.
- [21] W. Zhang, Y., Wang, A., & Hu, "Deep Learning-Based Consumer Behavior Analysis and Application Research," *Wirel. Commun. Mob. Comput.*, 2022.
- [22] C. Ahmed, A. A. A., Agarwal, S., Kurniawan, I. G. A., Anantadjaya, S. P., & Krishnan, "Business boosting through sentiment analysis using Artificial Intelligence approach," *Int. J. Syst. Assur. Eng. Manag.*, vol. 13, no. 1, pp. 699–709, 2022.
- [23] L. Guo, B. Zhang, and X. Zhao, "A Consumer Behavior Prediction Model Based on Multivariate Real-Time Sequence Analysis," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6688750.
- [24] N. N. Moon, I. M. Talha, and I. Salehin, "An advanced intelligence system in customer online shopping behavior and satisfaction analysis," *Curr. Res. Behav. Sci.*, vol. 2, p. 100051, 2021, doi: <https://doi.org/10.1016/j.crbeha.2021.100051>.
- [25] L. Xu and X. Sang, "E-Commerce Online Shopping Platform Recommendation Model Based on Integrated Personalized Recommendation," *Sci. Program.*, vol. 2022, p. 4823828, 2022, doi: 10.1155/2022/4823828.
- [26] M. Lodha, S. J., & Damle, "Sentiment and Statistical Analysis of Customer Reviews for Strategic decision on Positioning and Marketing," *Int. Conf. Decis. Aid Sci. Appl.*, pp. 100–107, 2022.
- [27] M. Hamzehi and S. Hosseini, "Business intelligence using machine learning algorithms," *Multimed. Tools Appl.*, 2022, doi: 10.1007/s11042-022-13132-3.
- [28] A. Nguyen, F. Krause, D. Hagenmayer, and M. Färber, "Quantifying Explanations of Neural Networks in E-Commerce Based on LRP BT - Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track," 2021, pp. 251–267.
- [29] N. Scheidthauer, J. Knoll, and R. Gross, "Visualizing Customer Journeys: How to Illustrate the Entire Customer Interaction Universe of a Commercial Website in Real Time BT - Innovations for Community Services," 2021, pp. 167–182.
- [30] X. Lin, "Big Data Analysis of Personalized Recommendation in," vol. 653, no. Ssha, pp. 768–771, 2022.
- [31] "H&M Personalized Fashion Recommendations," *kaggle.com*. <https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/data>