



**ANALISIS DATASET DEPARTMENT PRODUCT PADA
THE LOOK E-COMMERCE
(STUDI KASUS FINAL PROJECT RUANGGURU CAMP)**

KERJA PRAKTIK



UNIVERSITAS
Dinamika

**FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS DINAMIKA
2022**

**ANALISIS DATASET DEPARTMENT PRODUCT PADA
THE LOOK E-COMMERCE
(STUDI KASUS FINAL PROJECT RUANGGURU CAMP)**

Diajukan sebagian salah satu syarat untuk menyelesaikan

Program Sarjana Akuntansi



UNIVERSITAS
Dinamika

Disusun Oleh :
Nama : Karisma Dwi Fernanda
NIM : 19430200006

Program : S1 (Strata Satu)

Jurusan : Akuntansi

FAKULTAS EKONOMI DAN BISNIS

UNIVERSITAS DINAMIKA

2022

LEMBAR PENGESAHAN
ANALISIS DATASET DEPARTMENT PRODUCT PADA
THE LOOK E-COMMERCE
(STUDI KASUS FINAL PROJECT RUANGGURU CAMP)

Laporan Kerja Praktik oleh

Karisma Dwi Fernanda

NIM : 19430200006

Telah dipraksa, diuji dan disetujui

Surabaya, 08 Juli 2022



Pembimbing

Digitally signed by Tony
Soebijono
DN: cn=Tony Soebijono,
o=Universitas Dinamika,
ou=S1 Akuntansi,
email=tonys@dinamika.
ac.id, c=ID
Date: 2022.07.13
08:57:13 +07'00'

Tony Soebijono, S.E., S.H., M.Ak
NIDN. 0703127302

Disetujui :

Penyelia

Universitas
Ruang Raya Indonesia

Sheldi Aini

Mengetahui,

Ketua Program Studi S1 Akuntansi

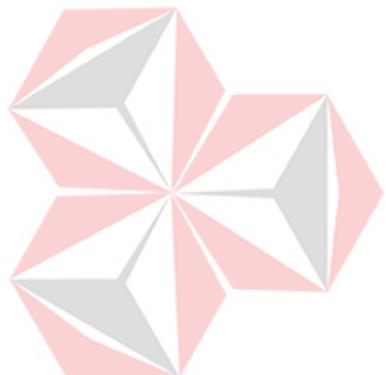
Arifin Puji Widodo Digitally signed by
Arifin Puji Widodo

Date: 2022.07.14
09:15:07 +07'00'

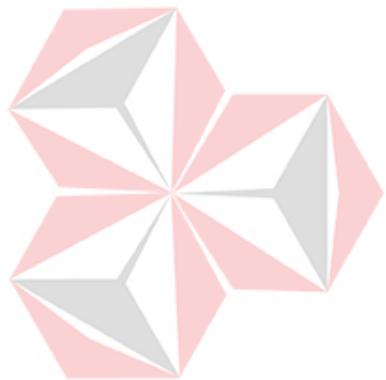
Arifin Puji Widodo, S.E., MSA
NIDN. 0721026801

Patience is needed when you want to success.

- Karisma Dwi Fernanda -



UNIVERSITAS
Dinamika



Laporan Kerja Praktik ini

Saya persembahkan kepada

Keluarga, Dosen Pembimbing, dan

Teman-teman tercintah saya.

UNIVERSITAS
Dinamika

SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, saya:

Nama : Karisma Dwi Fernanda
NIM : 19430200006
Program Studi : S1 Akuntansi
Fakultas : Fakultas Ekonomi dan Bisnis
Jenis Karya : Laporan Kerja Praktik
Judul Karya : **ANALISIS DATASET DEPARTMENT PRODUCT
PADA THE LOOK E-COMMERCE (STUDI KASUS
FINAL PROJECT RUANGGURU CAMP)**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, dialihmediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 08 Juli 2022

Yang menyatakan

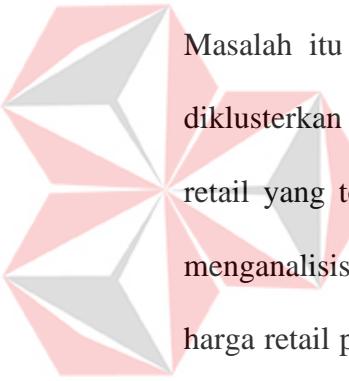


Karisma Dwi Fernanda

NIM: 19430200006

ABSTRAK

PT. Ruang Raya Indonesia (Ruangguru) merupakan perseroan terbatas yang bergerak di bidang pendidikan nonformal. Ruangguru memiliki program yang bernama *data, business analytics & operations bootcamp* yang terdaftar pada kegiatan Kampus Merdeka. Pada akhir program peserta akan diberikan *final project* yang berkaitan dengan data. Peserta diberikan sebuah dataset dari perusahaan fiktif yang bernama “The Look”. The Look merupakan situs pakaian *e-commerce* fiktif yang dikembangkan oleh tim Looker. Dataset berisi informasi tentang pelanggan, produk, pesanan, logistik, acara web, dan kampanye pemasaran digital. Department product sendiri memiliki masalah mengenai minimnya jumlah produk yang terjual.



Masalah itu terjadi salah satunya karena faktor dari harga retail yang belum diklusterkan dengan jelas sehingga *customer* The Look tidak tahu detail dari harga retail yang tergolong murah, mahal, dan standard. Maka dari itu, penulis ingin menganalisis dataset pada department product khususnya dengan mengklusterkan harga retail produk berdasarkan kategori produk. Hasil dari *clustering* harga retail berdasarkan kategori produk menggunakan algoritma K-Means mempunyai nilai $k = 3$ sesuai dengan hasil penentuan jumlah *cluster* terbaik menggunakan *elbow method* sehingga harga retail tersebut dikelompokkan kedalam 3 *cluster*.

Kata kunci : Data, *Final project*, Harga Retail, *Clustering*

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat melaksanakan Kerja Praktik dan menyelesaikan pembuatan laporan dari Kerja Praktik ini. Laporan ini disusun berdasarkan Kerja Praktik dan hasil studi yang dilakukan selama kurang lebih 5 (lima) bulan di PT. Ruang Raya Indonesia. Kerja Praktik ini membahas tentang analisis dataset department product pada the look e-commerce (studi kasus final project Ruangguru CAMP)

Penyelesaian laporan Kerja Praktik ini tidak terlepas dari bantuan berbagai pihak yang telah memberikan banyak masukan, nasihat, saran, kritik, dan dukungan moral maupun materil kepada penulis. Oleh karena itu penulis menyampaikan rasa terima kasih kepada:

1. Ibu dan Bapak tercinta serta keluarga yang selalu mendoakan, mendukung, dan memberikan semangat di setiap langkah dan aktivitas penulis.
2. Bapak Dr. Drs. Antok Supryanto, M.M selaku Dekan Fakultas Ekonomi dan Bisnis Universitas Dinamika.
3. Bapak Arifin Puji Widodo, S.E., MSA selaku Ketua Program Studi S1 Akuntansi Universitas Dinamika yang telah memberikan izin kepada penulis untuk melakukan Kerja Praktik.
4. Bapak Tony Soebijono, S.E., S.H., M.Ak. selaku Dosen Pembimbing yang mengarahkandan memberikan dukungan sehingga terselesainya laporan kerja praktek ini dengan tepat waktu.

5. Ibu Shelda Aini selaku Content Development Manager PT. Ruang Raya Indonesia yang telah memberi kesempatan dalam mengikuti program Ruangguru CAMP ini.
6. Ibu Ina Muthia Kanza selaku Mentor Program Ruangguru CAMP yang telah membimbing mulai awal kegiatan studi independen hingga akhir
7. Teman-teman tercinta yang memberikan bantuan dan dukungan dalam penyusunan laporan ini.
8. Pihak-pihak lain yang tidak dapat disebutkan satu-persatu yang telah memberikan bantuan dan dukungan kepada penulis.

Semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah membantu dan memberikan bimbingan serta nasehat dalam proses Kerja Praktik ini.

Penulis menyadari bahwa Kerja Praktik yang dikerjakan ini masih banyak terdapat kekurangan sehingga kritik yang bersifat membangun dan saran dari semua pihak sangatlah diharapkan agar aplikasi ini dapat diperbaiki menjadi lebih baik lagi. Semoga laporan Kerja Praktik ini dapat diterima dan bermanfaat bagi penulis dan semua pihak.

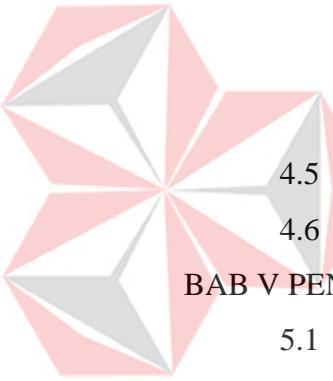
Surabaya, 08 Juli 2022



Karisma Dwi Fernanda

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
1.6 Sistematika Penulisan	5
BAB II.....	7
GAMBARAN UMUM	7
2.1 Latar Belakang Perusahaan	7
2.2 Identitas Perusahaan	8
2.3 Visi Perusahaan	8
2.4 Misi Perusahaan.....	8
2.5 Struktur Organisasi	9
BAB III LANDASAN TEORI.....	13
3.1 The Look.....	13
3.2 <i>Clustering</i>	13
3.3 CRISP-DM	14
3.4 Google Colab (<i>Python</i>).....	17
3.5 pgAdmin 4 (PostgreSQL).....	18
3.6 Tableau	20



BAB IV	23
DESKRIPSI PEKERJAAN.....	23
4.1 <i>Business Understanding</i>	23
4.2 <i>Data Understanding</i>	24
4.2.1 Join Data	25
4.2.2 Import Library pada <i>Python</i>	25
4.2.3 Pemahaman Data	26
4.3 <i>Data Preparation</i>	27
4.3.1 Mengubah Data <i>Type</i>	27
4.3.2 Mengatasi <i>Missing Value</i>	28
4.3.3 Eksplorasi Data	29
4.4 <i>Modeling</i>	33
4.4.1 <i>Model Evaluation</i>	33
4.4.2 <i>Cluster Cardinality</i>	33
4.4.3 <i>Elbow Method</i>	34
4.4.4 <i>Silhouette Coefficient</i>	36
4.5 <i>Evaluation</i>	38
4.6 <i>Deployment</i>	39
BAB V PENUTUP.....	41
5.1 Kesimpulan.....	41
5.2 Saran	41
DAFTAR PUSTAKA	43
LAMPIRAN.....	45

DAFTAR TABEL

	Halaman
Tabel 1. Hasil identifikasi masalah	24
Tabel 2. <i>Cluster</i> harga retail.....	40



DAFTAR GAMBAR

Halaman

Gambar 1. Logo PT. Ruang Raya Indonesia.....	7
Gambar 2. Struktur organisasi perusahaan.....	9
Gambar 3. Tahapan dalam CRISP-DM (al, 2000).....	15
Gambar 4. <i>Create</i> notebok baru	18
Gambar 5. Halaman Utama Google Colab.....	18
Gambar 6. Halaman awal pgAdmin 4.....	19
Gambar 7. Input <i>password</i>	20
Gambar 8. Halaman utama pgAdmin 4	20
Gambar 9. Halaman awal tableau.....	21
Gambar 10. Halaman utama tableau.....	22
Gambar 11. Skema analisis dataset.....	23
Gambar 12. Query Join pada PosrtgreSQL.....	25
Gambar 13. <i>Import library</i>	26
Gambar 14. Jumlah baris dan kolom.....	26
Gambar 15. <i>Missing value</i>	27
Gambar 16. Perubahan data <i>type</i>	28
Gambar 17. Hasil mengatasi <i>missing value</i>	29
Gambar 18. Eksplorasi kolom created at.....	29
Gambar 19. Eksplorasi kolom sold at.....	30
Gambar 20. Eksplorasi kolom product category.....	31
Gambar 21. Korelasi antar variabel	32
Gambar 22. Penentuan <i>cluster</i>	33

Gambar 23. Visualisasi <i>cluster cardinality</i>	34
Gambar 24. Perhitungan <i>inertia</i>	35
Gambar 25. Visualisasi <i>elbow method</i>	35
Gambar 26. Perhitungan koefisien <i>silhouette</i>	37
Gambar 27. Visualisasi <i>silhouette coefficient</i>	37
Gambar 28. Dashboard department product	39



DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Form KP-3 Surat Balasan	45
Lampiran 2. Form KP-5 Acuan Kerja	46
Lampiran 3. Form KP-5 Garis Besar Rencana Kerja Mingguan	47
Lampiran 4. Form KP-6 Log Harian dan Catatan Perubahan Acuan Kerja	49
Lampiran 5. Form KP-7 Kehadiran Kerja Praktik	81
Lampiran 6. Kartu Bimbingan Kerja Praktik	89
Lampiran 7. Biodata Penulis	90



BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk memperkuat komitmen Ruangguru dalam mewujudkan pemerataan pendidikan ke seluruh Indonesia, Ruangguru berkolaborasi dengan Kampus Merdeka dalam mencetak talenta-talenta digital dari Sabang sampai Merauke. Guna membekali dan menyiapkan para pemuda dan pemudi negeri menjadi talenta digital, Ruangguru menyediakan proses pembelajaran yang mendalam berdasarkan kurikulum yang berbasis kompetensi yang diperlukan di dunia bisnis *digital* saat ini, serta mentor terbaik yang berpengalaman dalam menangani proyek bisnis yang nyata.



Sebagai organisasi yang bergerak di dunia pendidikan, sejak awal, Ruangguru dipercaya pentingnya investasi pada pengembangan sumber daya manusia. Ruangguru mencoba menghidupkan nilai tersebut melalui berbagai program dan inisiatif yang dilakukan (baik dalam operasi bisnis maupun kegiatan non-komersial organisasi), termasuk membuat produk pembelajaran bagi para profesional di Indonesia dan termasuk bagi para mahasiswa.

Ruangguru berkomitmen untuk meningkatkan kualitas sumber daya manusia dan menyiapkan mahasiswa agar siap untuk bersaing di dunia kerja. Melalui program Studi Independen, Ruangguru menyediakan kesempatan bagi mahasiswa untuk meningkatkan kompetensi melalui proses pembelajaran yang praktis dan mendalam. Mahasiswa akan dibekali oleh pelatihan *online* berbasis kompetensi yang difasilitasi oleh Ruangguru agar menjembatani pengetahuan pada industri kerja yang tidak termasuk dalam kurikulum pada perguruan tinggi.

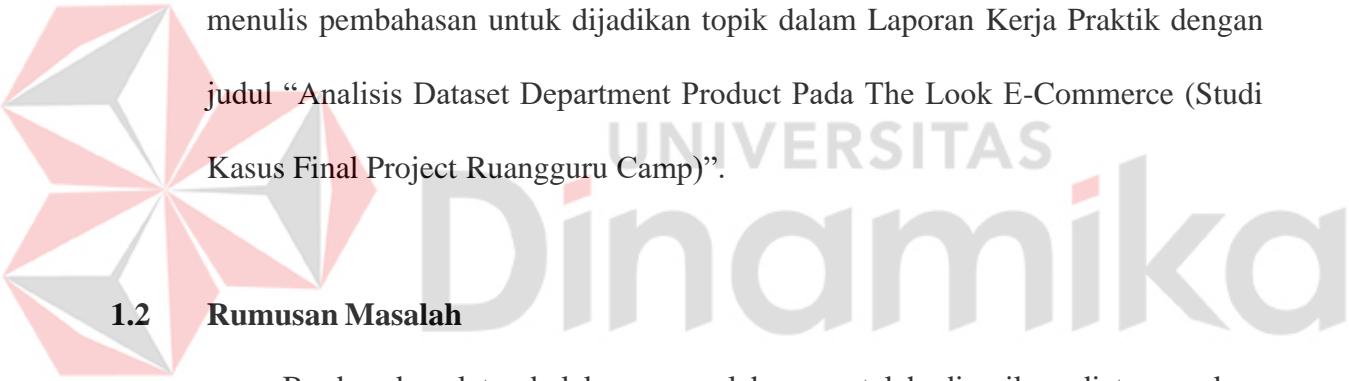
Tidak hanya itu, Ruangguru juga memberikan akses gratis pada mahasiswa peserta Studi Independen di Ruangguru untuk belajar dan memilih pelatihan yang dibutuhkan untuk persiapan masuk di dunia era digital sesuai dengan aspirasi berkarir masing-masing individu.

Pada program Studi Independen Ruangguru bersama Kampus Merdeka, materi ajar yang akan diberikan kepada mahasiswa akan disusun oleh tim ahli pengembangan konten dari Ruangguru bersama para instruktur berpengalaman dan didukung oleh teknologi canggih untuk mempermudah proses pembelajaran.

Ruangguru memiliki program yang bernama *data, business analytics & operations bootcamp* yang terdaftar pada kegiatan Kampus Merdeka. Aktivitas dari program tersebut adalah diadakan kelas secara *online* untuk pemahaman materi mengenai ruang lingkup data yang dilakukan setiap hari kerja yaitu senin – jum’at. Selanjutnya, pada akhir program peserta akan diberikan *final project* yang berkaitan dengan data.

Final project merupakan kegiatan akhir yang dilakukan peserta Merdeka Belajar Kampus Merdeka (MBKM) di Ruangguru yang dimana aktivitas dari kegiatan ini adalah membuat *insight* yang bisa membantu perusahaan melakukan optimasi perusahaan. Peserta diberikan sebuah dataset dari perusahaan fiktif yang bernama “The Look”. The look merupakan situs pakaian *e-commerce* fiktif yang dikembangkan oleh tim Looker. Dataset berisi informasi tentang pelanggan, produk, pesanan, logistik, acara web, dan kampanye pemasaran digital. Isi dari kumpulan data ini adalah sintetis, dan diberikan kepada praktisi industri untuk tujuan penemuan, pengujian, dan evaluasi produk.

Perusahaan The Look memiliki beberapa department antara lain department human resource, product, dan sales. Penulis hanya bertanggung jawab pada satu department yaitu pada department product. Department product sendiri memiliki masalah mengenai minimnya jumlah produk yang terjual. Masalah itu terjadi salah satunya karena faktor dari harga retail yang belum diklusterkan dengan jelas sehingga *customer* The Look tidak tahu detail dari harga retail yang tergolong murah, mahal, dan standard. Maka dari itu, penulis ingin menganalisis dataset pada department product khususnya dengan mengklusterkan harga retail produk berdasarkan kategori produk.

Berdasarkan latar belakang yang telah diuraikan diatas, maka dengan ini menulis pembahasan untuk dijadikan topik dalam Laporan Kerja Praktik dengan judul “Analisis Dataset Department Product Pada The Look E-Commerce (Studi Kasus Final Project Ruangguru Camp)”. 

1.2

Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan diatas, maka dapat dikemukakan rumusan masalah adalah bagaimana cara menganalisis dataset department product pada The Look E-Commerce?

1.3 Batasan Masalah

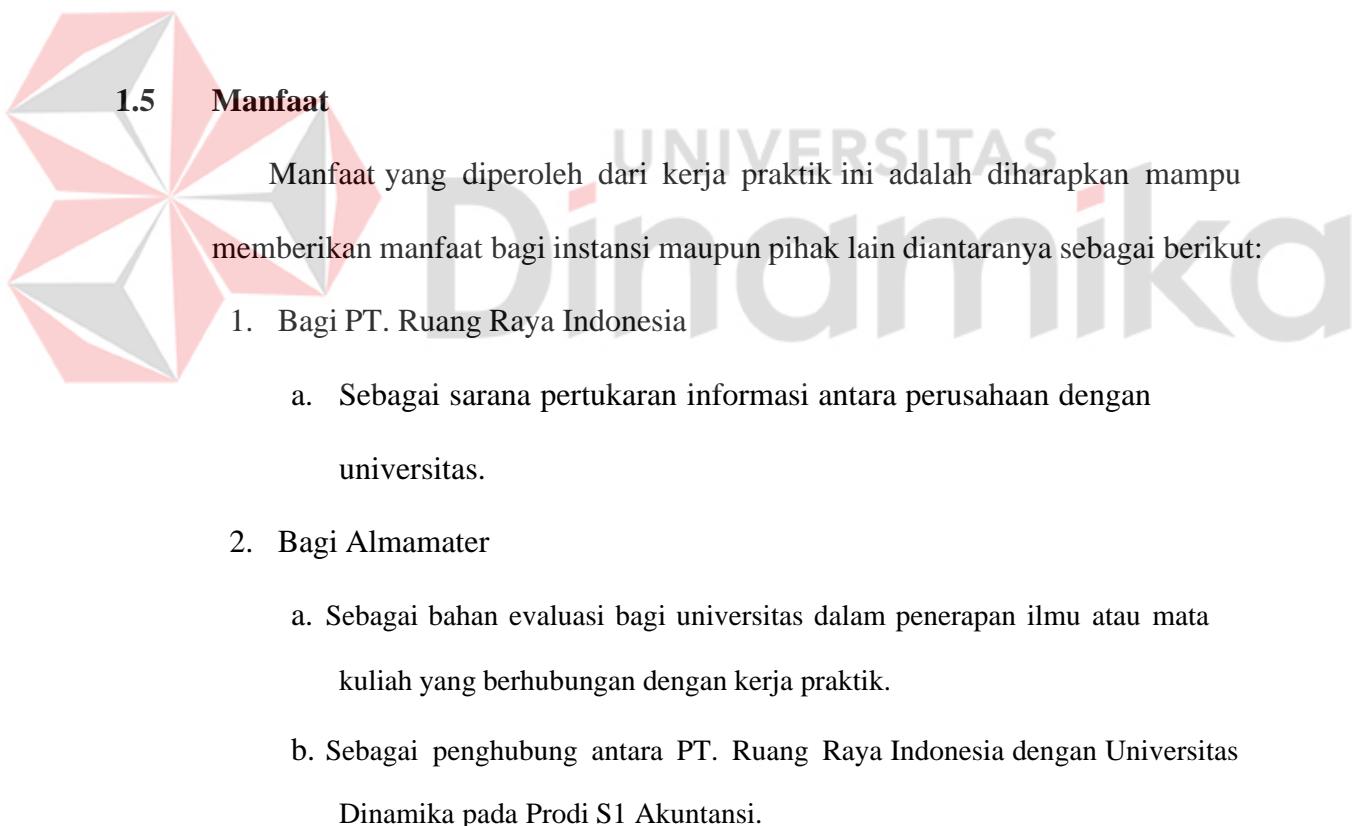
Berdasarkan rumusan masalah yang telah diuraikan, agar penulisan lebih fokus dan terarah, maka penelitian ini akan dibatasi oleh hal sebagai berikut:

1. Penulis membahas analisis dataset department product pada perusahaan The Look e-commerce yang berlangsung pada periode tahun 2019 – 2022.

2. Dataset yang digunakan *final project* department product ada 2 (dua) antara lain data inventory items (data primer) dan product (data sekunder).
3. Penulis hanya menggunakan model *clustering* untuk mengklusterkan harga retail produk

1.4 Tujuan

Berdasarkan dari latar belakang dan rumusan masalah, dapat disesuaikan bahwa tujuan dari kerja praktik ini adalah untuk menganalisis dataset department product pada The Look *e-commerce* dengan mengklusterkan harga retail produk menjadi tiga cluster yaitu harga murah, mahal, dan standar berdasarkan kategori produk.



3. Bagi Penulis

- a. Sebagai media belajar untuk penerapan mata kuliah yang selama ini dipelajari ke dunia kerja.
- b. Sebagai penambah ilmu baru bagi penulis.

1.6 Sistematika Penulisan

Guna untuk memudahkan pembaca dan penulis dalam melakukan pemahaman secara menyeluruh terhadap permasalahan yang sedang dibahas, maka terdapat sistematika penulisan yang dibagi ke dalam beberapa bab, antara lain:

BAB I PENDAHULUAN

Merupakan bab yang berisi tentang gambaran umum dari permasalahan yang akan dibahas. Dalam pendahuluan ini terdiri dari beberapa sub bab, yaitu latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II GAMBARAN UMUM INSTANSI

Merupakan bab yang berisi tentang gambaran umum dari perusahaan yang telah diteliti. Dalam gambaran umum perusahaan ini terdiri dari beberapa sub bab, yaitu sejarah, visi misi, logo, dan struktur organisasi.

BAB III LANDASAN TEORI

Dalam landasan teori ini yang dianggap berhubungan dengan kerja praktik yang dilakukan dan menjadi acuan dalam menyelesaikan masalah.

BAB IV DESKRIPSI PEKERJAAN

Dalam deskripsi pekerjaan ini akan dijelaskan tentang kegiatan yang dilakukan oleh penulis selama kerja praktik dan juga pembahasan mengenai langkah – langkah yang digunakan dalam melakukan peramalan penjualan properti rumah.

BAB V PENUTUP

Dalam bab ini berisi tentang kesimpulan dari serangkaian pembahasan,

keterbatasan penulis dan saran



BAB II

GAMBARAN UMUM

2.1 Latar Belakang Perusahaan

PT. Ruang Raya Indonesia merupakan perseroan terbatas yang bergerak di bidang pendidikan non-formal yang didirikan menurut dan berdasarkan hukum yang berlaku di Indonesia serta telah memperoleh Izin Pendirian Satuan Pendidikan Nonformal dan Izin Operasional Lembaga Kursus Pelatihan dengan Nomor 3/A.5a/31.74.01/-1.851.332/2018. Perusahaan ini didirikan sejak tahun 2014 oleh Belva Devara dan Iman Usman, yang keduanya berhasil masuk dalam jajaran pengusaha sukses di bawah 30 tahun melalui *Forbes 30 under 30* untuk sektor teknologi konsumen di Asia. Di tahun 2019, mereka mendapat penghargaan sebagai *Emerging Entrepreneur* dari *Ernst & Young*.



Gambar 1. Logo PT. Ruang Raya Indonesia

2.2 Identitas Perusahaan

Nama Instansi : PT. Ruang Raya Indonesia
Alamat : Jl. Dr. Saharjo No.161, Manggarai Selatan, Tebet, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12860
No. Telepon : 081578200000, Fax (021)40008000
Website : www.ruangguru.com
Email : info@ruangguru.com

2.3 Visi Perusahaan

Ruangguru memiliki visi untuk memberikan layanan belajar bermutu dengan harga terjangkau, memaksimalkan mutu dan mewujudkan lapangan kerja bagi tenaga pendidik, serta memudahkan pelajar, tenaga pendidik, maupun orang tua dalam kegiatannya melalui teknologi Pendidikan (Ruangguru, n.d.)

2.4

Misi Perusahaan

Untuk mewujudkan visi yang dituju, Ruangguru memiliki misi untuk mewujudkan visinya, diantara lain :

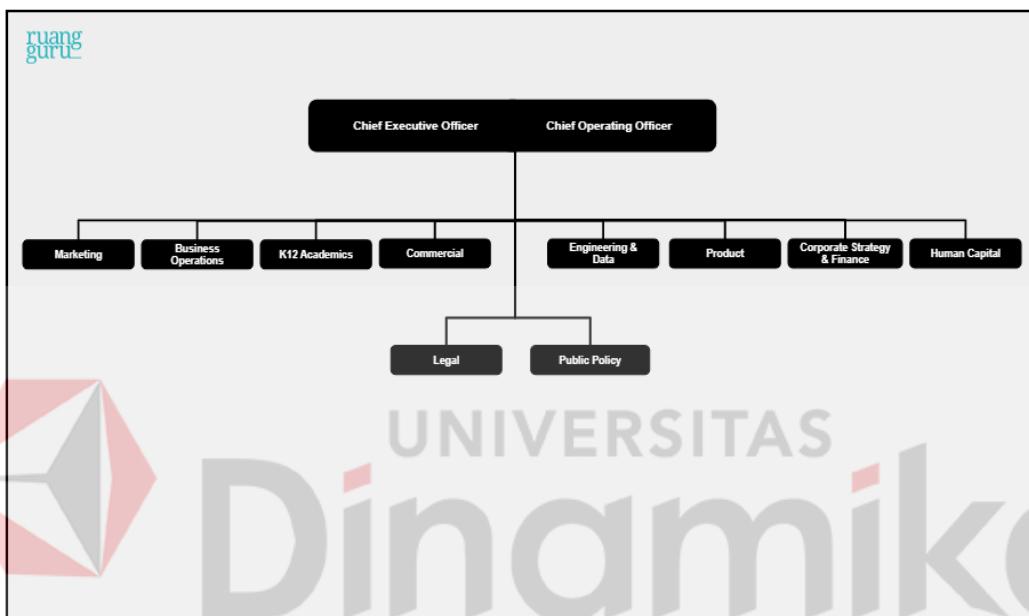
1. Menyediakan dan memperluas akses terhadap Pendidikan berkualitas melalui teknologi untuk semua siswa, kapan saja dan di mana saja.
2. Menyediakan layanan Pendidikan dan materi pembelajaran dari guru-guru terbaik Indonesia, yang bisa diakses oleh seluruh siswa di mana saja dengan biaya yang terjangkau

3. Meningkatkan kualitas guru dengan menciptakan lapangan pekerjaan dan penghasilan bagi guru di Indonesia

2.5 Struktur Organisasi

Dalam melakukan proses aktivitas, PT. Ruang Raya Indonesia memiliki struktural organisasi untuk menjalankannya. dapat dilihat pada Gambar 2.

Gambar 2. Bagan struktur organisasi PT. Ruang Raya Indonesia



Gambar 2. Struktur organisasi perusahaan

Pada struktur organisasi diatas, dapat dilihat bahwa perusahaan ini dipimpin oleh seorang CEO (Chief Executive Officer) dan COO (Chief Operating Officer). Semua departemen akan bertanggung jawab untuk melaporkan pekerjaan pada CEO dan COO Ruangguru. Adapun tugas yang dilakukan oleh setiap departemen adalah:

- 
1. CEO (Chief Executive Officer)
 - a. Menciptakan, mengomunikasikan, dan mengimplementasikan visi misi dan gambaran umum tujuan organisasi atau perusahaan kepada seluruh karyawan
 - b. Menetapkan kebijakan yang berkaitan dengan keberlangsungan perusahaan
 2. COO (Chief Operating Officer)
 - a. memastikan perusahaan berjalan sesuai dengan regulasi yang ada
 - b. memastikan perusahaan menjalankan strateginya dengan tepat
 - c. membuat kebijakan untuk efektivitas kerja
 3. Marketing
 - a. Mengenalkan produk kepada masyarakat
 - b. Menjalin hubungan baik antara perusahaan dan konsumen
 - c. Menginformasikan evaluasi dari produk yang dijual
 4. Business Operations
 - a. Mengelola pengembangan strategi
 - b. Melaksanakan tujuan operasional
 - c. Memantau kinerja bisnis
 5. K12 Academics
 - a. Menyusun kurikulum pembelajaran online
 - b. Menyesuaikan Pendidikan pengajaran kepada pelanggan
 6. Commercial
 - a. Mengembangkan relasi dengan klien
 - b. Menyusun rencana pertumbuhan bisnis

c. Mengembangkan panduan penetapan harga

d. Mengawasi vendor dan kontraktor

7. Engineering & Data

- a. Membangun sistem aplikasi yang digunakan oleh karyawan dan pelanggan
- b. Memelihara keberlangsungan penggunaan aplikasi
- c. Membangun dan menguji arsitektur data
- d. Mengembangkan algoritma yang dapat mengubah data menjadi informasi

8. Product

a. Menjalankan *customer research*

b. Menetapkan *product vision*

c. Mengembangkan produk

d. Menjaga *product growth*

9. Corporate strategy & Finance

a. Merumuskan visi dan misi organisasi

b. Menetapkan uni startegi bisnis

c. Alokasi sumberdaya

d. Mengidentifikasi kesempatan pertumbuhan bisnis

e. Mengelola keuangan perusahaan

10. Human Capital

a. Merekut karyawan baru

b. Mengadakan *On boarding*

c. Mengakan training and development

d. Menilai kinerja karyawan



BAB III

LANDASAN TEORI

3.1 The Look

The Look merupakan situs pakaian *e-commerce* fiktif yang dikembangkan oleh tim Looker. Dataset dari The Look berisi informasi tentang pelanggan, produk, pesanan, logistik, acara web, dan kampanye pemasaran digital. Isi dari kumpulan data ini adalah sintetis, dan diberikan kepada praktisi industri untuk tujuan penemuan, pengujian, dan evaluasi produk.

Perusahaan ini terdiri dari beberapa departemen antara lain Human Resource, Product, dan Sales. Department Human Resource memiliki tanggung jawab untuk mengurus segala sesuatu yang berhubungan dengan tenaga kerja. Department Product memiliki tanggung jawab dalam perencanaan *product*, melakukan riset *product* dan mengembangkan *product*. Departemen Sales memiliki tanggung jawab menjual dan memastikan kepuasan pelanggan setelah membeli dan menggunakan produk.

3.2 Clustering

Clustering merupakan sebuah teknik yang digunakan untuk menemukan sub-kelompok atau *cluster*. Prinsip dari *clustering* adalah data – data yang termasuk dalam *cluster* yang sama “diharapkan” memiliki kesamaan satu dengan yang lainnya, sedangkan data – data yang ada dalam *cluster* yang berbeda cukup berbeda dari satu *cluster* dengan *cluster* lainnya.

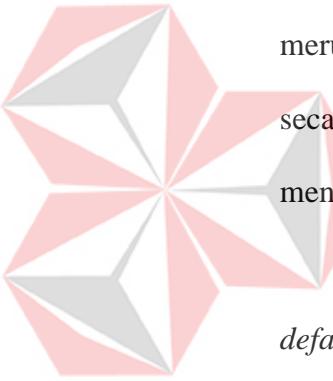
(Dubes, 1998) menjelaskan definisi *clustering* sebagai berikut :

1. Data dalam *cluster* yang sama, harus “semirip” mungkin
2. Data dalam *cluster* yang berbeda, harus sangat berbeda
3. Pengukuran kesamaan dan ketaksamaan harus jelas dan praktis

Secara umum, terdapat empat pendekatan *clustering* yang sering digunakan antara lain *centroid-based clustering*, *density-based clustering*, *distribution-based clustering*, dan *hierarchial-clustering*.

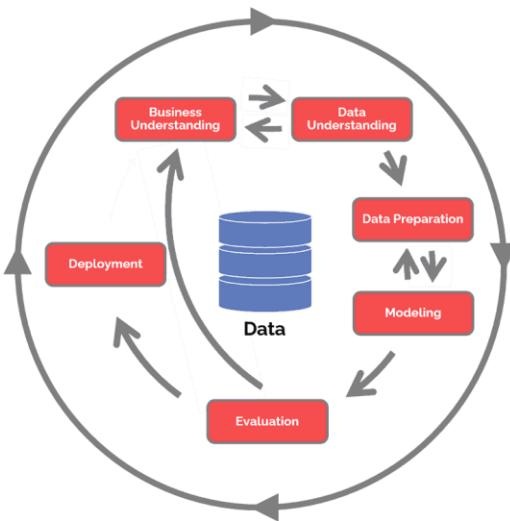
3.3 CRISP-DM

Cross-Industry Standard Process for Data Mining atau CRISP-DM



merupakan salah satu model proses datamining yang masih digunakan secara luas di kalangan industri karena keunggulannya dalam menyelesaikan banyak persoalan dalam proyek data mining.

Menurut (Mariscal, 2010) menyatakan bahwa CRISP-DM sebagai *defacto* menjadi standar untuk pengembangan proyek data mining. Dalam proses CRISP-DM memberikan gambaran mengenai siklus hidup proyek data mining. CRISP-DM memiliki 6 (enam) tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* seperti yang ditunjukkan pada gambar 3 (al, 2000).



Gambar 3. Tahapan dalam CRISP-DM (al, 2000)

Masing – masing tahapan dijelaskan sebagai berikut :

1. *Business Understanding*

Merupakan tahap pertama CRISP-DM. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun dan mendapatkan data serta bagaimana untuk mencocokkan tujuan atas pemodelan sehingga model terabik dapat dibangun.

2. *Data Understanding*

Merupakan tahap kedua CRISP-DM. Pada tahap ini secara garis besar untuk memeriksa data sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan pondasi *analytics* untuk sebuah penelitian dengan membuat ringkasan (*summary*) dan mengidentifikasi masalah dalam data.

3. *Data Preparation*

Merupakan tahap ketiga CRISP-DM. Pada tahap ini secara garis besar untuk memperbaiki masalah dalam data. Aktivitas yang dilakukan pada

tahap ini biasanya seperti melakukan transformasi dan melakukan pembersihan data agar siap untuk diolah pada tahap *modeling*.

4. *Modeling*

Merupakan tahap keempat CRISP-DM. Pada tahap ini secara garis besar membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistik dan menentukan algoritma data mining yang diterapkan. Beberapa modeling yang biasa dilakukan adalah *classification*, *time series forecasting*, *clustering*, *finding relation*, *prediction*, dan sebagainya.

5. *Evaluation*

Merupakan tahap kelima CRISP-DM. pada tahap ini secara garis besar melakukan interpretasi hasil dari data mining yang dihasilkan dalam proses pemodelan tahap sebelumnya. Biasanya dalam evaluasi ini dilakukan terhadap model yang diterapkan dengan tujuan model yang ditentukan sesuai dengan tujuan dalam tahap pertama.

6. *Deployment*

Merupakan tahap keenam CRISP-DM. Pada tahap ini secara garis besar melakukan rencana penggunaan model. Rencana *deployment* tidak ada model yang statis tetapi model tersebut dibangun dari data yang diwakili pada waktu tertentu sehingga perubahan waktu dapat menyebabkan perubahan atas karakter data. Pemodelan harus dipantau dan mungkin diganti dengan model yang telah diperbaiki.

3.4 Google Colab (*Python*)

Google colab merupakan *coding environment* bahasa pemograman *python* dengan format “*notebook*”. *Software* ini pada dasarnya seperti *jupyter notebook* gratis berbentuk *cloud* yang dijalankan menggunakan browser seperti google chrome dan sejenisnya.

Dengan penggunannya google colab memiliki manfaat diantaranya memudahkan untuk menjalankan program pada komputer dengan spesifikasi tinggi (GPU Tesla, RAM 12GB, Disk 300GB yang masih bisa tersambung dengan google drive, akses internet cepat untuk mendownload file besar), dapat berkolaborasi dengan orang lain dengan membagi *codingan* secara *online*, dan mudah *merunning deep learning program* via *handphone*. Berikut adalah beberapa tahapan untuk membuat *project* pada google colab.

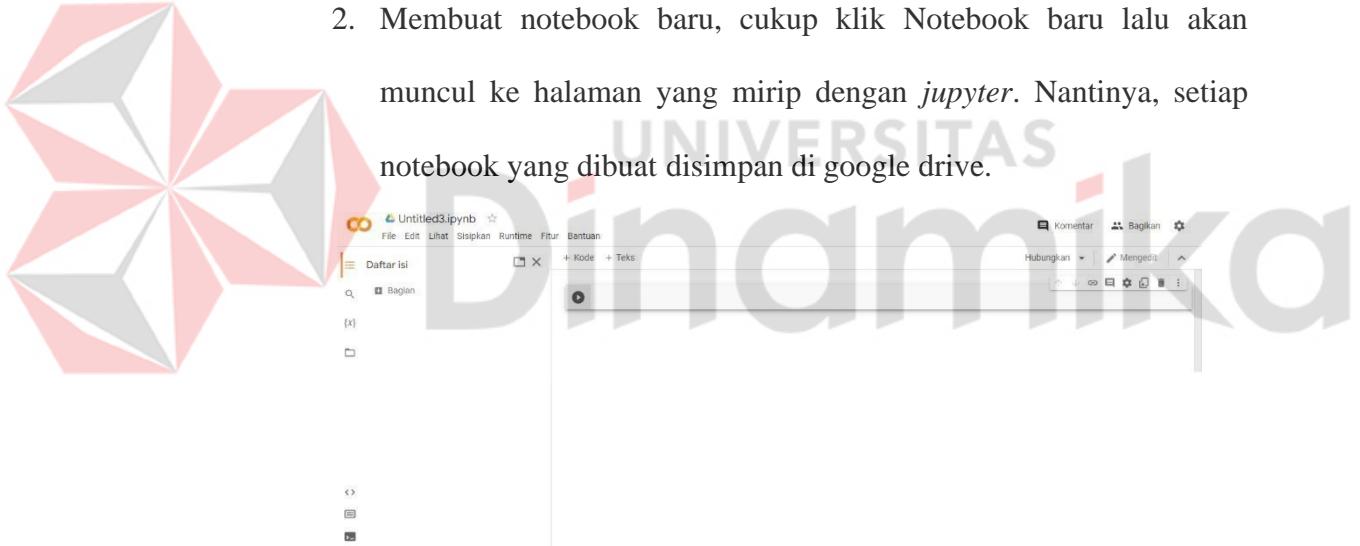
1. Langkah awal yang dibutuhkan adalah akun google dan silahkan membuka link <https://colab.research.google.com/>. Setelah itu disuguhkan dengan tampilan berikut :

Filter notebook				
Judul	Terakhir dibuka	Pertama kali dibuka	More	
CO Selamat Datang di Colaboratory	00.01	12 April		
Final Project_Product.ipynb	13 Juni	10 Juni		
Untitled2.ipynb	13 Juni	13 Juni		
Marketing - Data User- Final Project RG	13 Juni	13 Juni		
Salinan dari Course 2_Meeting in Class 24_ver 2.ipynb	12 Juni	12 Juni		

[Notebook baru](#) [Batal](#)

Gambar 4. *Create* noteboek baru

2. Membuat notebook baru, cukup klik Notebook baru lalu akan muncul ke halaman yang mirip dengan *jupyter*. Nantinya, setiap notebook yang dibuat disimpan di google drive.



Gambar 5. Halaman utama google colab

3.5 pgAdmin 4 (PostgreSQL)

PostgreSQL merupakan *object-relational database management system* berbasis *open source*. Secara *database* relasional, fungsi utama PostgreSQL adalah tempat untuk menyimpan dan mengelola data melalui

perintah atau query SQL. Dengan query tersebut, maka PostgreSQL bisa digunakan untuk :

1. Membuat atau memanipulasi tabel
2. Memanipulasi isian data atau *value* dari tabel
3. Mengelola transaksi pada *database*

pgAdmin merupakan tool GUI (*graphical user interface*) standar yang ada dalam paket PostgreSQL dan mendukung semua operasi dan fitur dari PostgreSQL.

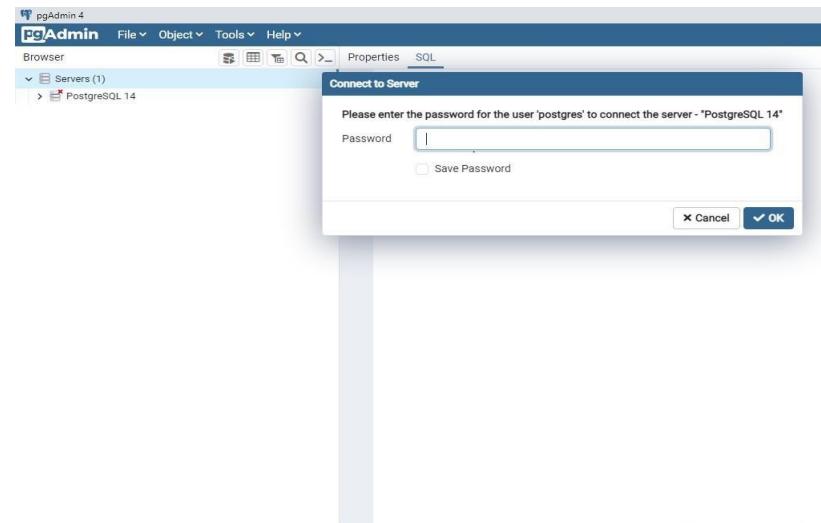
Berikut adalah beberapa tahapan untuk membuat *project* pada PostgreSQL.

1. Langkah awal adalah dengan membuka pgAdmin 4, dengan menekan tombol *Windows* dan ketikkan “pgAdmin 4” lalu klik.



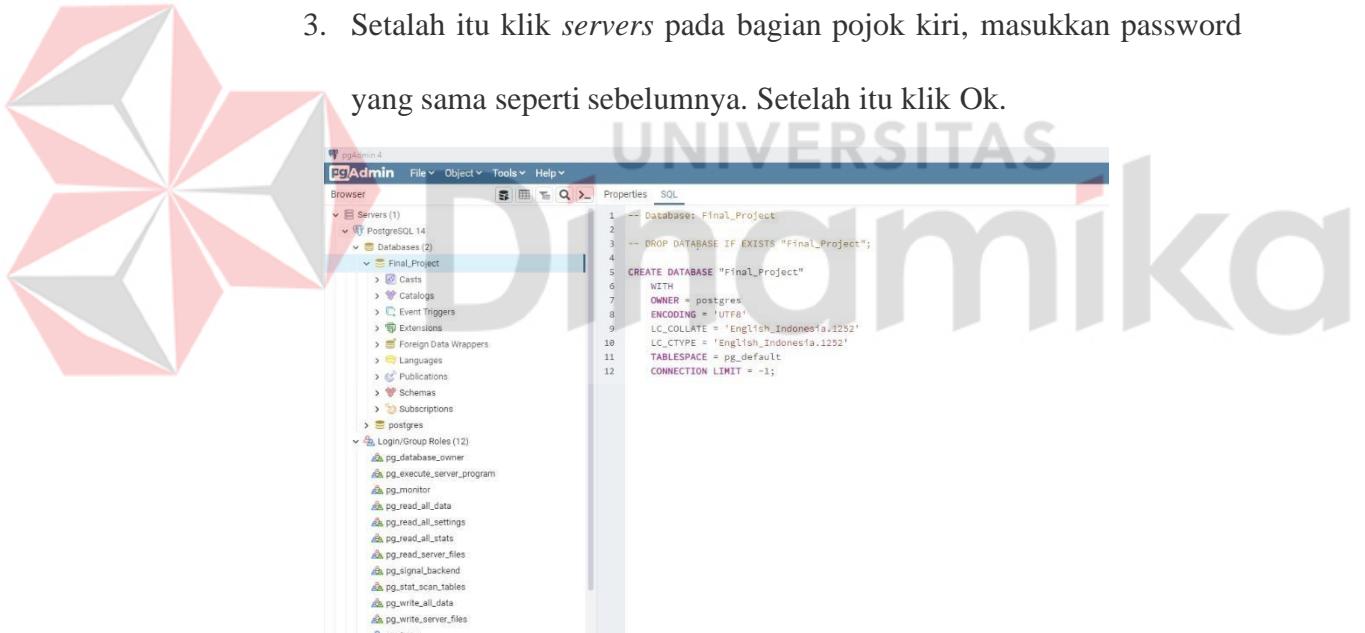
Gambar 6. Halaman awal pgAdmin 4

- Masukkan password yang sudah terhubung dengan *database server*, setelah itu klik Ok



Gambar 7. Input password

- Setalah itu klik *servers* pada bagian pojok kiri, masukkan password yang sama seperti sebelumnya. Setelah itu klik Ok.



Gambar 8. Halaman utama pgAdmin 4

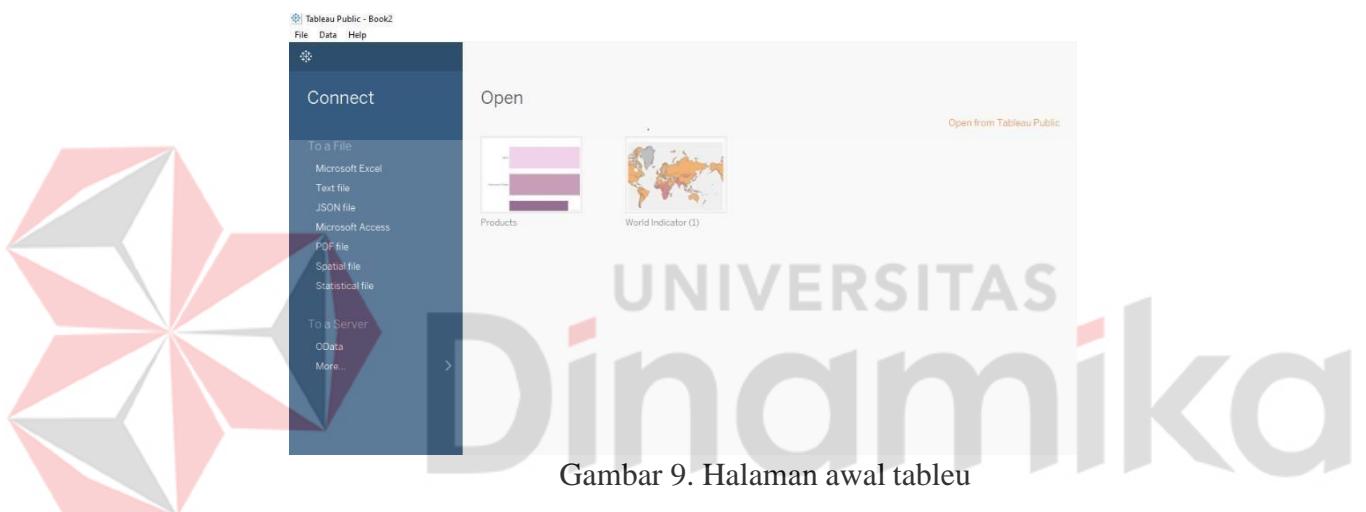
3.6 Tableu

Tableu merupakan platform *business intelligence* untuk membantu dalam melihat dan memahami data. Hal tersebut karena tableau memiliki beberapa keunggulan dalam memvisualisasikan data :

1. Mempersingkat waktu untuk mendapatkan *insight*
2. Meningkatkan akurasi untuk mendapatkan *insight*
3. Meningkatkan *engagement*

Tableu memiliki beberapa *product* antara lain tableau *desktop*, tableau *reader*, tableau *public*, tableau *online*, dan tableau *server*. Berikut adalah beberapa tahapan untuk membuat *project* pada tableau.

1. Langkah awal adalah dengan membuka tableau, dengan menekan tombol *Windows* dan ketikkan “tableu” lalu klik.



2. Pilih “*to a file*” atau “*to a server*” yang sesuai dengan tipe file yang ingin diunggah. Selanjutnya akan disuguhkan tampilan seperti gambar 10.

The screenshot shows the Tableau Public interface. On the left, the 'Connections' pane shows a single connection named 'Data_Tableau'. The 'Files' pane lists several CSV files: 'data_all.csv', 'data_gabungan.csv', 'Data_Tableau.csv', 'department_product.csv', 'distribution_centers.csv', 'employees.csv', 'events.csv', 'inventory_items.csv', 'order_items.csv', 'orders.csv', 'products.csv', and 'users.csv'. A 'New Union' option is also present. The main workspace is titled 'Data_Tableau' and contains a single data source named 'Data_Tableau.csv'. The data source preview shows columns: 'Id' (Type: Alphanumeric, Field Name: Data_Tableau.csv, Physical Table: Data_Tableau.csv, Remote: No), 'Product Id' (Type: Alphanumeric, Field Name: Data_Tableau.csv, Physical Table: Data_Tableau.csv, Remote: No), 'Created At' (Type: Date/Time, Field Name: Data_Tableau.csv, Physical Table: Data_Tableau.csv, Remote: No), 'Sold At' (Type: Alphanumeric, Field Name: Data_Tableau.csv, Physical Table: Data_Tableau.csv, Remote: No), and 'Cost' (Type: Number, Field Name: Data_Tableau.csv, Physical Table: Data_Tableau.csv, Remote: No). Buttons for 'Update Now' and 'Update Automatically' are visible. A 'Need more data?' section with a 'Drag tables here to relate them' message and a 'Learn more' link is also present.

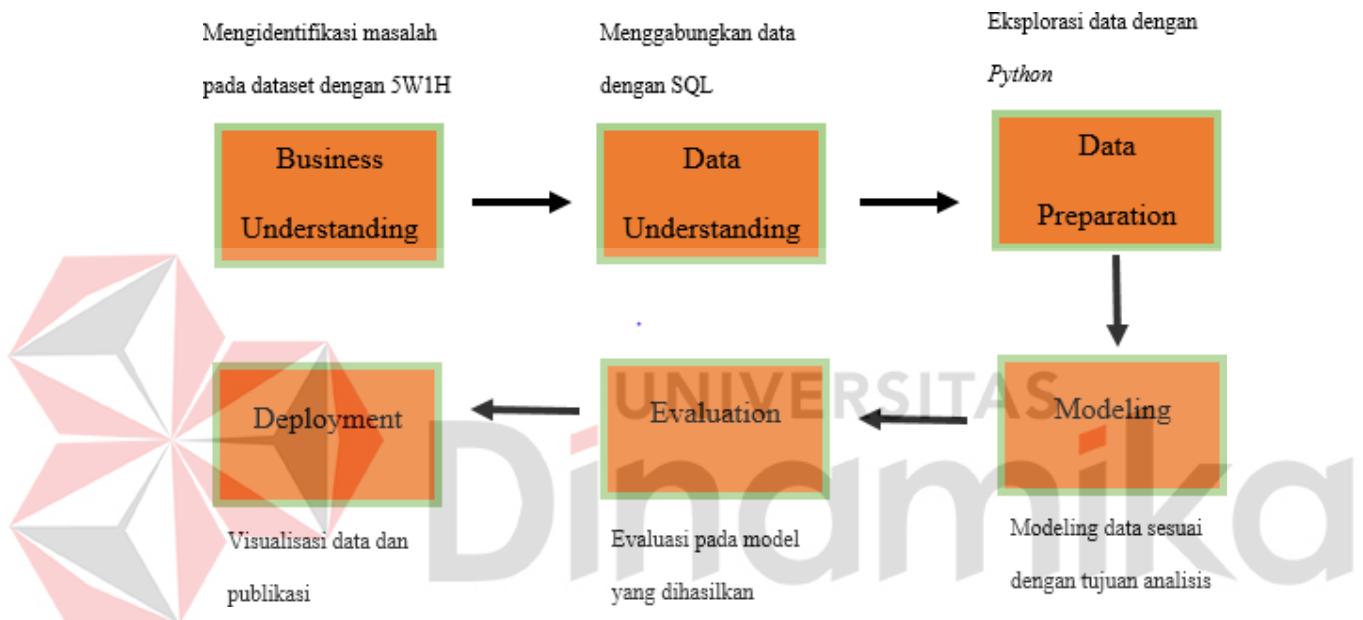
Gambar 10. Halaman utama tableau



BAB IV

DESKRIPSI PEKERJAAN

Dalam menganalisis dataset pada department product yang digunakan pada penelitian ini yaitu menggunakan metode CRISP – DM (*Cross-Industry Standard Process for Data Mining*), dengan skema atau tahapan pada Gambar 11.



Gambar 11. Skema analisis dataset

4.1 *Business Understanding*

Sesuai dengan gambar 11, tahap awal yang dilakukan dalam penelitian ini yaitu *business understanding*. Tahap *business understanding* dimulai dengan menganalisis masalah pada dataset department product. berikut analisis masalah dan alternatif solusi yang dapat dilihat pada Tabel 1.

Tabel 1. Hasil identifikasi masalah

Masalah	Alternatif Solusi
Department product memiliki masalah mengenai minimnya jumlah produk yang terjual. Masalah itu terjadi karena faktor dari harga retail yang belum diklusterkan dengan jelas sehingga pelanggan tidak tahu detail harga yang tergolong murah, standard, dan mahal.	Mengklusterkan harga retail produk menjadi tiga cluster antara lain harga retail yang tergolong murah, mahal, dan standard berdasarkan produk kategori.

4.2 *Data Understanding*

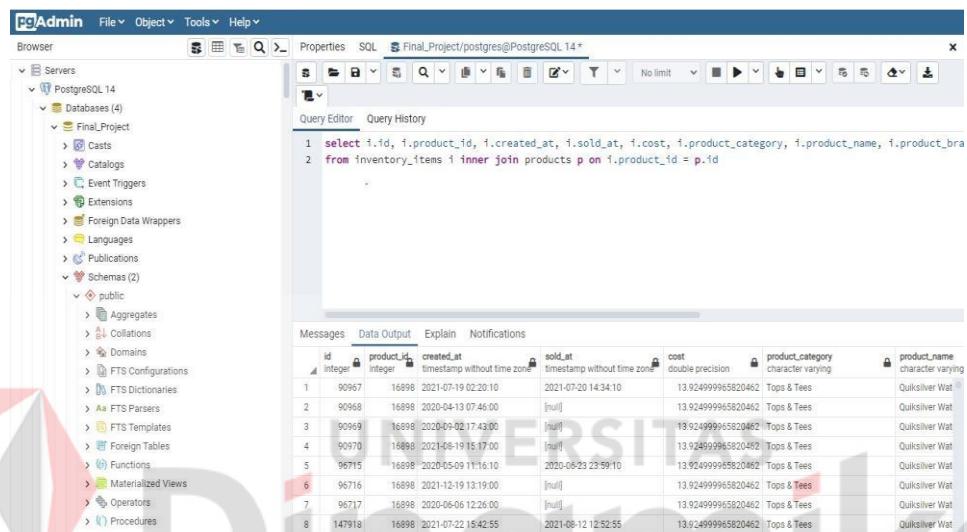


Tahap kedua yaitu tahap *data understanding*, pada tahap ini dilakukan penggabungan data terlebih dahulu melalui PostgreSQL, setelah itu memahami data dengan menggunakan Google Colab (*Python*). Department product menggunakan dua dataset yang terdiri dari data inventory items (data primer) dan data product (data sekunder) sebagai analisis.

Data inventory items terdapat informasi mengenai id, product id, created at, sold at, cost, product category, product name, product brand, product retail price, product department, product sku, dan product distribution center id. Data product terdapat informasi mengenai id, cost, category, name, brand, retail price, department, sku, dan distribution center id. Variabel yang digunakan pada penelitian ini yaitu product retail price dan product category.

4.2.1 Join Data

Join data dengan PostgreSQL digunakan untuk menggabungkan dataset inventory items dengan product. Pada join data ini menggunakan *inner join* dimana hanya menampilkan data – data yang mempunyai nilai sama antar tabel pada sebuah *database*. Berikut query untuk penggabungan data yang dapat dilihat pada Gambar 12.



```
1 select i.id, i.product_id, i.created_at, i.sold_at, i.cost, i.product_category, i.product_name, i.product_brand
2 from inventory_items i inner join products p on i.product_id = p.id
```

	id	product_id	created_at	sold_at	cost	product_category	product_name
1	90967	16898	2021-07-19 02:20:10	2021-07-20 14:34:10	13.92499995820462	Tops & Tees	Quiksilver Wat
2	90968	16898	2020-04-13 07:46:00	[null]	13.92499995820462	Tops & Tees	Quiksilver Wat
3	90969	16898	2020-09-02 17:43:00	[null]	13.92499995820462	Tops & Tees	Quiksilver Wat
4	90970	16898	2021-08-19 15:17:00	[null]	13.92499995820462	Tops & Tees	Quiksilver Wat
5	96715	16898	2020-05-09 11:16:10	2020-05-23 23:59:10	13.92499995820462	Tops & Tees	Quiksilver Wat
6	96716	16898	2021-12-19 13:19:00	[null]	13.92499995820462	Tops & Tees	Quiksilver Wat
7	96717	16898	2020-06-06 12:26:00	[null]	13.92499995820462	Tops & Tees	Quiksilver Wat
8	147918	16898	2021-07-22 15:42:55	2021-08-12 12:52:55	13.92499995820462	Tops & Tees	Quiksilver Wat

Gambar 12. Query join pada PosrtgreSQL

4.2.2 Import Library pada Python

Pada tahap ini melakukan *import library* pada *python* yang sesuai dengan kebutuhan. Untuk *import library* dapat dilihat pada Gambar 13.

```
#import package
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib.pyplot import rcParams
from datetime import datetime
import warnings
warnings.filterwarnings('ignore')
```

Gambar 13. *Import library*

4.2.3 Pemahaman Data

Pada tahap ini mencari info seperti jumlah baris, kolom, *missing value*, dan *type* data pada data department product yang telah berhasil digabungkan di PostgreSQL. Berikut hasil dari info jumlah baris, kolom, *missing value*, data *type* yang terlihat seperti Gambar 14 dan 15.

```
product=pd.read_csv('/content/department_product.csv')
product.shape
(487394, 12)
```

Gambar 14. Jumlah baris dan kolom

Dari gambar 14 dapat diketahui bahwa dataset department product memiliki jumlah 487394 baris dan 12 kolom.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 487394 entries, 0 to 487393
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   id               487394 non-null   int64  
 1   product_id       487394 non-null   int64  
 2   created_at       487394 non-null   object  
 3   sold_at          180508 non-null   object  
 4   cost              487394 non-null   float64 
 5   product_category 487394 non-null   object  
 6   product_name     487359 non-null   object  
 7   product_brand    487020 non-null   object  
 8   product_retail_price 487394 non-null   float64 
 9   product_department 487394 non-null   object  
 10  sku              487394 non-null   object  
 11  distribution_center_id 487394 non-null   int64  
dtypes: float64(2), int64(3), object(7)
memory usage: 44.6+ MB

```

Gambar 15. *Missing value*

Dari gambar 15 dapat diketahui bahwa dataset department product terdapat kolom yang memiliki *missing value* antara lain sold at, product name, dan product brand. Type data dari semua kolom rata – rata bertipe *integer, object, dan float*.

4.3 Data Preparation

Pada tahap ini penulis akan melakukan eksplorasi data sesuai kebutuhan analisis penulis. Adapun beberapa eksplorasi yang dilakukan oleh penulis, antara lain :

4.3.1 Mengubah Data Type

Langkah pertama yang harus dilakukan untuk eksplorasi data adalah dengan mengubah data *type* sesuai dengan kebutuhan. Penulis mengubah data *type* pada kolom id, product id, created at, dan sold at.

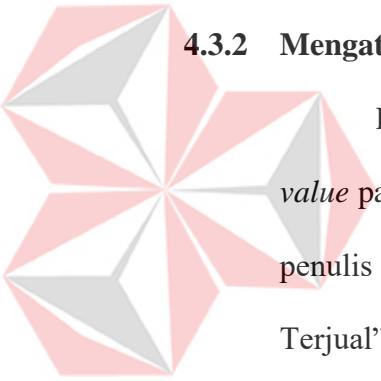
Kolom id dan product id awalnya memiliki data type *integer* tetapi penulis mengubahnya menjadi *object* untuk mempermudah langkah dalam analisis selanjutnya. Selain itu, pada pada kolom created at dan sold at

diubah menjadi *datetime*. Hasil dari perubahan pada data *type* dapat dilihat pada Gambar 16.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 487394 entries, 0 to 487393
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   id                487394 non-null   object  
 1   product_id        487394 non-null   object  
 2   created_at        487394 non-null   datetime64[ns]
 3   sold_at           180508 non-null   datetime64[ns]
 4   cost              487394 non-null   float64 
 5   product_category  487394 non-null   object  
 6   product_name      487359 non-null   object  
 7   product_brand     487020 non-null   object  
 8   product_retail_price 487394 non-null   float64 
 9   product_department 487394 non-null   object  
 10  sku               487394 non-null   object  
 11  distribution_center_id 487394 non-null   int64  
dtypes: datetime64[ns](2), float64(2), int64(1), object(7)
memory usage: 44.6+ MB
```

Gambar 16. Perubahan data *type*

4.3.2 Mengatasi *Missing Value*



Pada tahap *data understanding* diketahui bahwa terdapat *missing value* pada kolom sold at, product name, dan product brand. Pada tahap ini penulis akan mengisi kolom sold at dengan keterangan “Produk Belum Terjual” sedangkan kolom product name dan product brand dihapus karena tidak berpengaruh pada analisis selanjutnya. Hasil dari mengatasi *missing value* dapat dilihat pada gambar 17.

```
id                      0
product_id               0
created_at               0
sold_at                 0
cost                     0
product_category         0
product_name              .
product_brand              .
product_retail_price      0
product_department         0
sku                      0
distribution_center_id    0
dtype: int64
```

Gambar 17. Hasil mengatasi *missing value*

4.3.3 Eksplorasi Data

Dalam tahap ini penulis akan mengeksplor data yang menghasilkan *insight* sederhana seperti mengetahui jumlah produk yang dibuat pada waktu tertentu, jumlah produk yang terjual pada waktu tertentu, jumlah produk kategori yang dijual oleh The Look, dan korelasi antar variabel. Hasil dari eksplorasi data dapat dilihat pada Gambar 18 sampai dengan 21.

2020-07-24 02:17:00	5
2021-05-24 13:37:00	5
2022-04-10 18:04:00	5
2020-07-08 03:59:00	5
2021-11-19 05:16:00	5
..	
2022-02-04 08:00:00	1
2022-05-25 17:29:00	1
2020-03-22 00:08:00	1
2021-06-27 18:57:54	1
2020-06-22 15:50:00	1

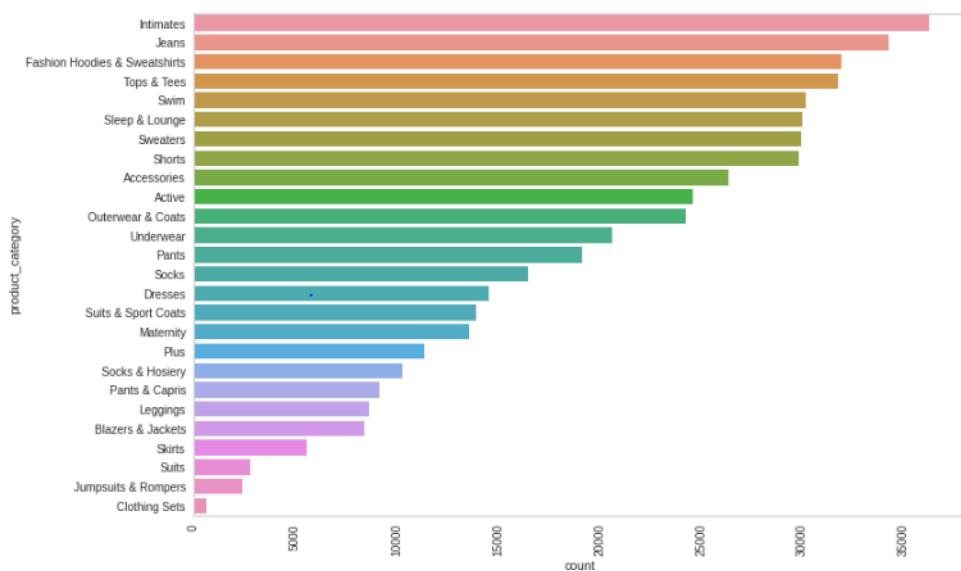
Gambar 18. Eksplor kolom created at

Dari gambar 18 menghasilkan *insight* sederhana mengenai jumlah produk yang dibuat contohnya seperti pada tanggal 24 Juli 2020 jumlah produk yang dibuat sebesar 5. Selanjutnya, penulis melakukan eksplorasi pada kolom sold at karena ingin mengetahui jumlah produk yang terjual dalam waktu tertentu yang dapat dilihat pada Gambar 19.

Produk_belum_terjual	306632
2022-03-10 07:49:24	3
2022-05-29 16:43:56	2
2021-04-01 13:23:22	2
2022-04-14 05:13:58	2
...	
2022-05-09 09:02:21	1
2022-05-08 16:15:30	1
2020-01-28 01:18:42	1
2022-02-28 08:46:26	1
2021-06-27 04:00:11	1

Gambar 19. Eksplorasi kolom sold at

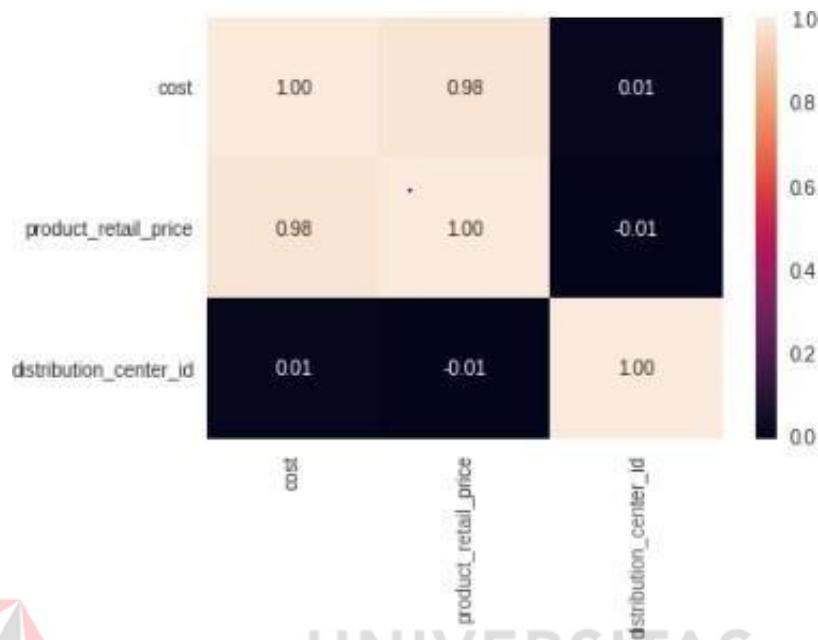
Dari gambar 19 menghasilkan *insight* sederhana mengenai jumlah produk yang belum terjual sebanyak 306632 produk. Selanjutnya, penulis melakukan eksplorasi pada kolom product category dimana ingin mengetahui kategori produk apa saja yang dijual oleh The Look. Kategori produk apa saja yang dijual bisa dilihat pada Gambar 20.



Gambar 20. Eksplorasi kolom product category

Dari gambar 20 menghasilkan *insight* sederhana mengenai produk kategori apa saja yang dijual oleh The Look. Produk kategori yang dijual oleh The Look sebanyak 26 kategori antara lain intimates, jeans, fashion hoodies & sweatshirts, tops & tees, swim, sleep & lounge, sweaters, shorts, accessories, active, outwear & coats, underwear, pants, socks, dresses, suits & sport coats, maternity, plus, socks & hoseiery, pants & capris, laeggings, blazers & jackets, skirts, suits, jumpsuits & rompers, dan closthing sets. Produk kategori yang paling banyak dijual adalah kategori intimates sedangkan yang paling sedikit dijual adalah kategori clothing sets.

Penulis selanjutnya akan mencari hubungan korelasi antar variabel apakah memiliki hubungan yang kuat atau tidak. Hasil korelasi dapat dilihat pada Gambar 21.



Gambar 21. Korelasi antar variabel

Dari gambar 21 menghasilkan *insight* sederhana mengenai hubungan antar variabel. Hubungan antar variabel dikatakan kuat apabila hasilnya mendekati 1.00. berikut variabel yang memiliki hubungan kuat antara lain cost dengan product retail price.

4.4 *Modeling*

Pada tahap ini yang dilakukan adalah *modeling* pada data dengan menggunakan model *clustering* dengan algoritma K-Means. Langkah awal dalam model *clustering* yaitu menetukan nilai k, dimana penulis membuat *cluster* dengan nilai k=3 yang artinya bahwa harga retail akan dikelompokkan menjadi 3 yaitu harga murah, mahal, dan standard.

```
from time import time
K = 3
kmeans_3 = KMeans(n_clusters=K, random_state=11)

start = time()
kmeans_3.fit(product_new1)
print(f"Done fitting kMeans in {time()-start:.3f}s")
```

Done fitting kMeans in 2.542s

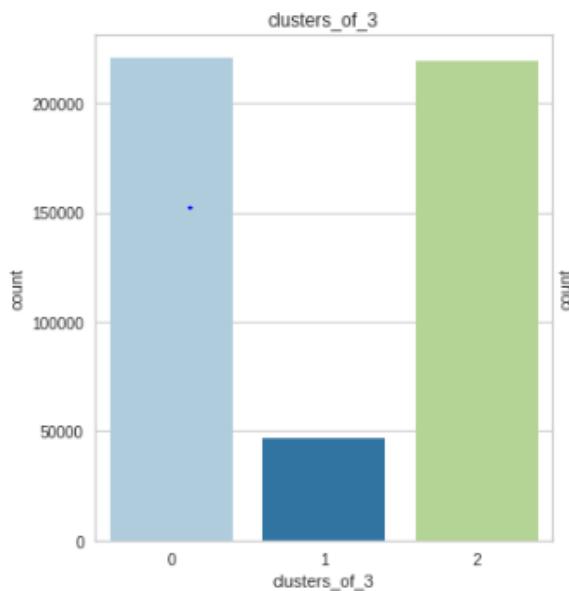
Gambar 22. Penentuan *cluster*

4.4.1 *Model Evaluation*

Pada tahap evaluasi model, penulis mencoba menganalisis dengan menggunakan *cluster cardinality*, *elbow method*, dan *silhouette coefficient* dari masing – masing *cluster* di model tersebut dengan k = 3.

4.4.2 *Cluster Cardinality*

Pada tahap evaluasi model dengan menggunakan *cluster cardinality* dapat menghasilkan *insight* berupa perbedaan antara ketiga *cluster*. Hasil dari perbedaan dapat dilihat pada Gambar 23.



Gambar 23. Visualisasi *cluster cardinality*

Dari visualisasi *cluster cardinality* diatas dapat diketahui bahwa warna biru muda merupakan *cluster 0* (harga retail murah), warna biru tua merupakan *cluster 1* (harga retail mahal), dan warna hijau merupakan *cluster 2* (harga retail standard). Dari model dengan 3 *cluster* ($k = 3$) terdapat perbedaan cukup jauh antara *cluster 0* dan *2* dengan *cluster 1*.

4.4.3 *Elbow Method*

Pada tahap evaluasi model dengan menggunakan *elbow method* dapat menghasilkan *insight* mengenai jumlah *cluster* yang optimal. Evaluasi ini perlu menghitung *inertia* dari seluruh *cluster*. *Inertia* merupakan jumlah kuadrat jarak sampel ke pusat *cluster* terdekat. Penulis mencoba beberapa nilai k (jumlah *cluster*) untuk dihitung intertianya. Berikut hasil perhitungan *inertia* yang dapat dilihat pada Gambar 24.

```

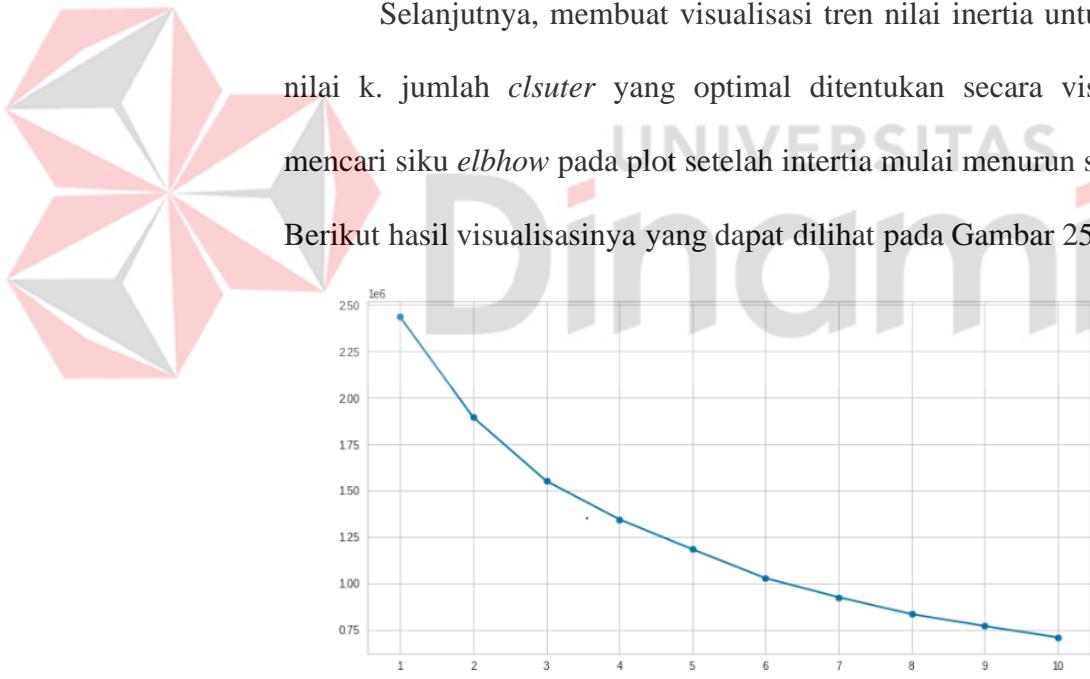
Clustering with n_clusters=1
Done clustering in 0.547s
Clustering with n_clusters=2
Done clustering in 2.612s
Clustering with n_clusters=3
Done clustering in 6.267s
Clustering with n_clusters=4
Done clustering in 6.805s
Clustering with n_clusters=5
Done clustering in 6.203s
Clustering with n_clusters=6
Done clustering in 7.246s
Clustering with n_clusters=7
Done clustering in 8.880s
Clustering with n_clusters=8
Done clustering in 13.016s
Clustering with n_clusters=9
Done clustering in 15.257s
Clustering with n_clusters=10
Done clustering in 11.743s

```

Gambar 24. Perhitungan *inertia*

Selanjutnya, membuat visualisasi tren nilai inertia untuk masing – nilai k. jumlah *clsuter* yang optimal ditentukan secara visual dengan mencari siku *elbow* pada plot setelah *intertia* mulai menurun secara linier.

Berikut hasil visualisasinya yang dapat dilihat pada Gambar 25.



Gambar 25. Visualisasi *elbow method*

Dari visualisasi *elbow method* diatas dapat diketahui bahwa jumlah *cluster* yang optimal ditentukan secara visual dengan mencari bentuk siku *elbow* pada plot setelah *inertia* mulai menurun secara linier. Jadi, jumlah *cluster* yang paling optimal adalah 3.

4.4.4 *Silhouette Coefficient*

Pada tahap evaluasi model dengan menggunakan *silhouette coefficient* dapat menghasilkan *insight* mengenai nilai koefisien yang berada pada rentang antara -1 sampai 1. Apabila koefisien *silhouette* semakin mendekati 1 maka bisa dikatakan *cluster* yang dihasilkan sudah optimal. Sebelumnya, penulis melakukan perhitungan koefisien *silhouette* yang menggunakan persamaan :

$$\frac{b - a}{\max(a, b)}$$

Dengan nilai b adalah jarak antar elemen dalam satu *cluster* dan a adalah jarak suatu elemen dengan elemen lain yang tidak dalam satu *cluster*.

Berikut hasil perhitungan koefisien *silhouette* yang dapat dilihat pada

Gambar 26.



UNIVERSITAS
Dinamika

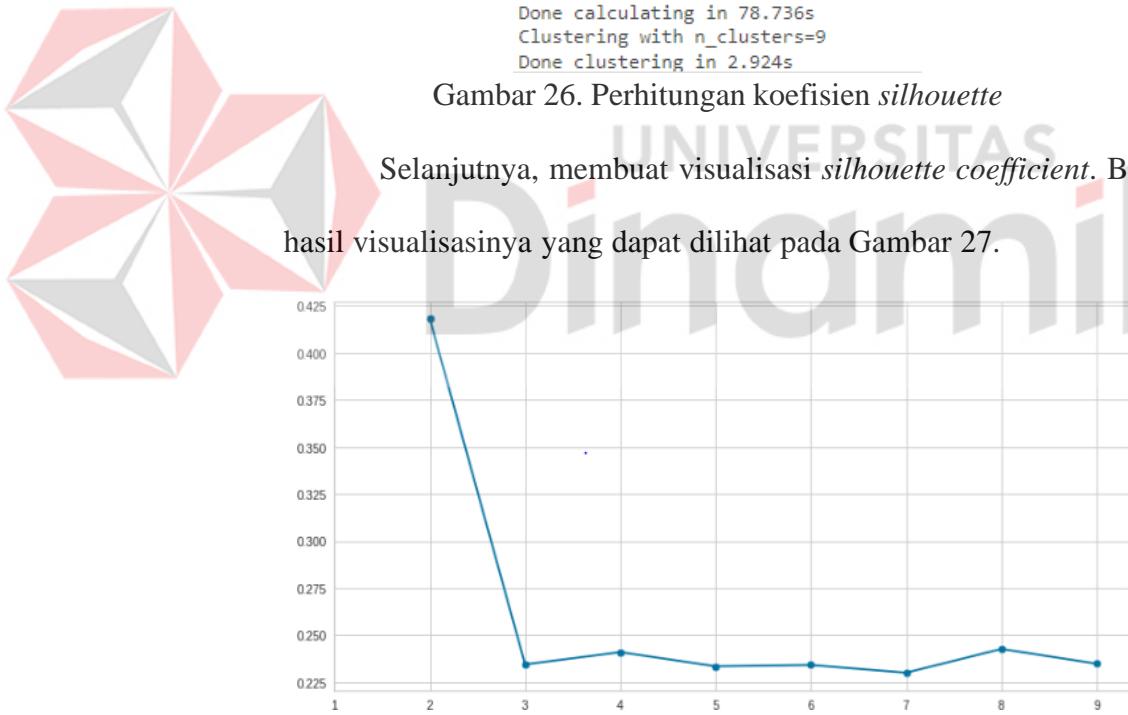
```

Clustering with n_clusters=2
Done clustering in 1.128s
Calculating silhouette coefficient..
Done calculating in 93.324s
Clustering with n_clusters=3
Done clustering in 1.031s
Calculating silhouette coefficient..
Done calculating in 87.876s
Clustering with n_clusters=4
Done clustering in 1.419s
Calculating silhouette coefficient..
Done calculating in 87.067s
Clustering with n_clusters=5
Done clustering in 2.136s
Calculating silhouette coefficient..
Done calculating in 79.998s
Clustering with n_clusters=6
Done clustering in 1.936s
Calculating silhouette coefficient..
Done calculating in 80.066s
Clustering with n_clusters=7
Done clustering in 2.327s
Calculating silhouette coefficient..
Done calculating in 79.795s
Clustering with n_clusters=8
Done clustering in 2.505s
Calculating silhouette coefficient..
Done calculating in 78.736s
Clustering with n_clusters=9
Done clustering in 2.924s

```

Gambar 26. Perhitungan koefisien *silhouette*

Selanjutnya, membuat visualisasi *silhouette coefficient*. Berikut hasil visualisasinya yang dapat dilihat pada Gambar 27.



Gambar 27. Visualisasi *silhouette coefficient*

Dari visualisasi *silhouette* diatas dapat diketahui bahwa jumlah cluster dengan nilai *silhouette coefficient* yang paling besar adalah 1.

4.5 *Evaluation*

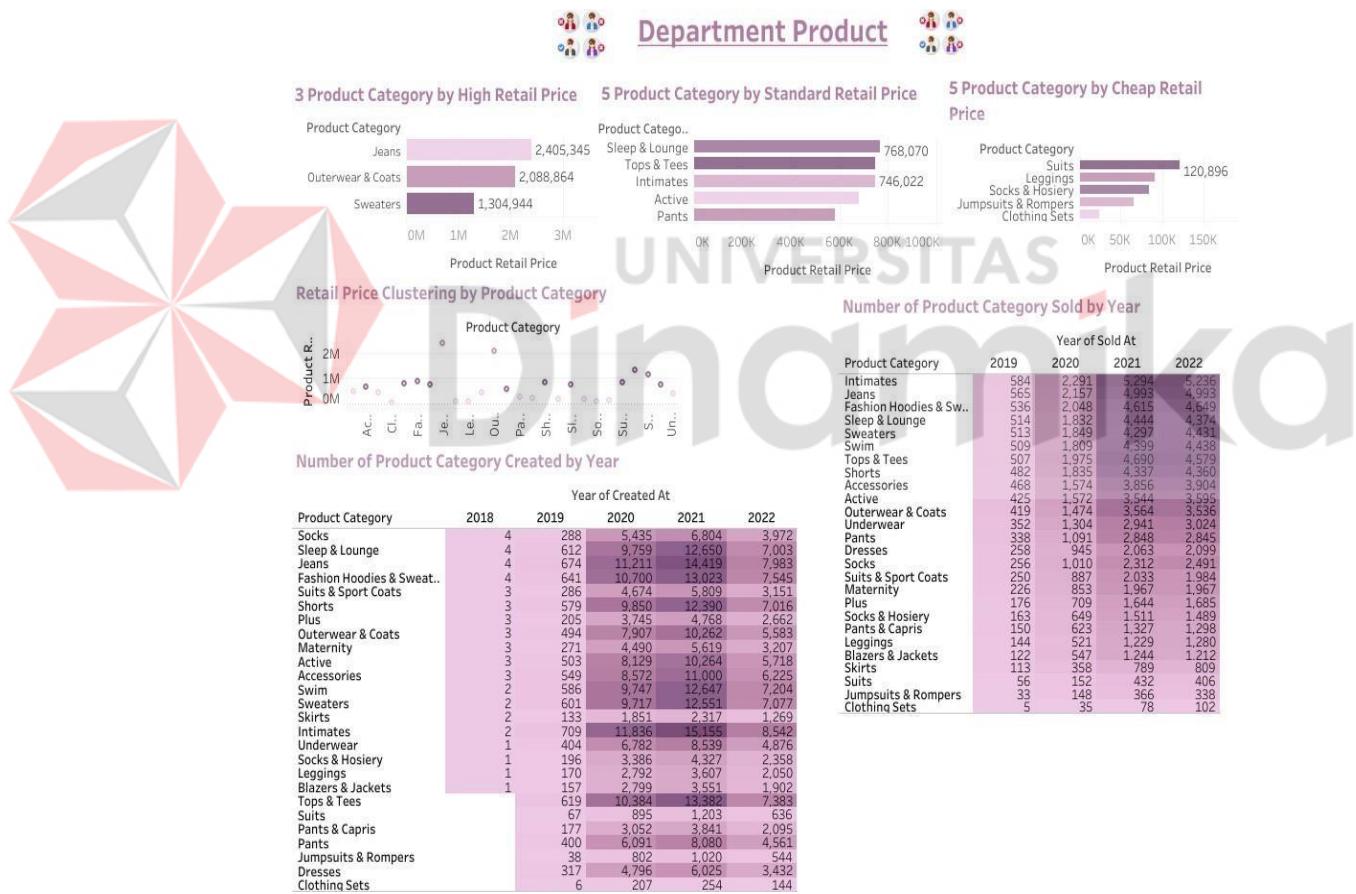
Pada tahap evaluasi akan dijelaskan mengenai hasil dari pengkajian dimana terdapat 1 model simulasi yaitu K-Means *clustering*. Untuk menentukan simulasi yang memiliki *insight* yang optimal dilihat dari 3 (tiga) indeks validitas dengan kriteria *relative*, yaitu *cluster cardinality*, *elbow method*, dan *silhouette coefficient*.

Pada indeks *cluster cardinality* terdapat perbedaan yang cukup jauh antara *cluster 0* dan *2* dengan *cluster 1*. Pada indeks *elbow method* simulasi terbaik ditunjukkan dengan siku *elbow* pada plot setelah *inertia* mulai menurun secara linier. Pada indeks *silhouette coefficient* simulasi yang terbaik ditunjukkan dengan nilai *silhouette* yang semakin mendekati 1.

Berdasarkan evaluasi *insight* yang telah dilakukan, dapat disimpulkan bahwa indeks yang memiliki *performance* optimal merupakan indeks 2 dikarenakan nilai validitas yang sesuai yaitu memiliki jumlah *cluster 3*.

4.6 Deployment

Pada tahap ini penulis akan membuat *dashboard* dengan tableau yang menghasilkan beberapa *insight* seperti merekomendasikan kategori produk yang memiliki harga murah, mahal, standard, hasil *clustering*, jumlah kategori produk yang telah dibuat pada tahun 2018 – 2022, dan jumlah kategori produk yang berhasil terjual pada tahun 2019 – 2022. Kemudian penulis mempublikasikannya melalui tableau *public*. Berikut hasil *dashboard* yang telah dibuat dapat dilihat pada Gambar 28.



Gambar 28. Dashboard department product

Selain itu, penulis akan memperlihatkan hasil *insight* dari harga retail produk berdasarkan kategori produk yang memiliki harga retail murah, mahal, dan standard. Hasil *insight* tersebut dapat dilihat pada Tabel 2.

Tabel 2. *Cluster* harga retail

Cluster Harga Retail	Kategori Produk	Jumlah
Harga Retail Murah	Accessories, Blazer & Jackets, Clothing Sets, Jumpsuits & Rompers, Leggings, Maternity, Pants & Capris, Plus, Skirts, Socks, Socks & Hosiery, Suits, dan Underwear	13
Harga Retail Mahal	Jeans, dan Outwear & Coats	2
Harga Retail Standard	Active, Dresses, Fashion Hoodies & Sweatshirts, Intimates, Pants, Shorts, Sleep & Lounge, Suits & Sport Coats, Sweaters, Swim, dan Tops & Tees.	11

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan kegiatan Kerja Praktik yang sudah dilaksanakan oleh penulis selama 5 bulan di PT. Ruang Raya Indonesia mengenai analisis dataset pada department product, maka dapat ditarik kesimpulan, antara lain:

1. Hasil dari *clustering* harga retail berdasarkan kategori produk menggunakan algoritma K-Means mempunyai nilai $k = 3$ sesuai dengan hasil penentuan jumlah *cluster* terbaik menggunakan *elbow method* sehingga harga retail tersebut dikelompokkan kedalam 3 *cluster*.
2. **Cluster 1** tergolong harga retail murah dengan jumlah 13 kategori produk antara lain: Accessories, Blazer & Jackets, Clothing Sets, Jumpsuits & Rompers, Leggings, Maternity, Pants & Capris, Plus, Skirts, Socks, Socks & Hosiery, Suits, dan Underwear.

Cluster 2 tergolong harga retail mahal dengan jumlah 2 kategori produk antara lain: Jeans, dan Outwear & Coats.

Cluster 3 tergolong harga retail standard dengan jumlah 11 kategori produk antara lain: Active, Dresses, Fashion Hoodies & Sweatshirts, Intimates, Pants, Shorts, Sleep & Lounge, Suits & Sport Coats, Sweaters, Swim, dan Tops & Tees.

5.2 Saran

Analisis dataset dengan menggunakan metode *clustering* tidak sepenuhnya sempurna dan memiliki beberapa kekurangan, sehingga

terdapat beberapa saran yang dapat digunakan untuk pengembangan analisis kedepannya. Saran yang dimaksud antara lain: Apabila metode *clustering* dirasa kurang akurat maka pada penelitian selanjutnya dapat menambahkan metode yang lain sebagai pembanding agar kebijakan yang dibuat lebih akurat.



DAFTAR PUSTAKA

- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>
- Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, 10(2), 133–146. <https://doi.org/10.34010/jamika.v10i2.2683>
- Latupeirissa, S. J., Lewaherilla, N., & Hiariey, A. (2022). *PENGELOMPOKAN KABUPATEN / KOTA DI PROVINSI MALUKU BERDASARKAN DATA KEMISKINAN TAHUN 2021 MENGGUNAKAN METODE K- MEANS CLUSTER (Grouping District / City in Maluku Province Based On Poverty Data for 2021 Using The K-Means Cluster Method)*. 4, 15–22.
- Lisnawati, A. (2019). Model exponential smoothing holt-winter dan model SARIMA untuk peramalan tingkat hunian hotel di provinsi DIY. *UNY Journal*, 6–25.
- Poerwanto, B., & Fa'rifah, R. Y. (2016). Analisis Cluster K-Means dalam Pengelompokan Kemampuan Mahasiswa. *Jurnal Scientific Pinisi*, 2(2), 92–96.
- Print, I., Rahmah, M., Candra, A., & Sembiring, R. W. (2022). *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Identifikasi Predikat Hasil Pengelompokan Data Kualitas Udara dengan Menggunakan Affinity Propagation dan Silhouette Coefficient*. 2.

Rafi Muttaqin, M., Iman Hermanto, T., Agus Sunandar, M., Studi Teknik Informatika, P., & Tinggi Teknologi Wastukancana, S. (2022). *Penerapan K-Means Clustering Dan Cross-Industry Standard Process for Data Mining (Crisp-Dm) Untuk Mengelompokan Penjualan Kue.* 19(1), 38–53. <https://journal.unpak.ac.id/index.php/komputasi>

Ramadanti, E., & Muslih, M. (2022). Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(1), 1–7. <https://doi.org/10.36341/rabit.v7i1.2155>

Ramdhani, F., Hoyyi, A., & Mukid, A. M. (2015). Pengelompokan Provinsi Di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster. *Jurnal GAUSSIAN*, 4(4), 875–884. <http://ejournal-s1.undip.ac.id/index.php/gaussian>

Yuliana Sari, R., Oktavianto, H., & Wahyu Sulistyo, H. (2022). Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia K-Means Algorithm With Elbow Method To Grouping District/City in Central Java Based on Components of Human D. *Jurnal Smart Teknologi*, 3(2), 2774–1702. <http://jurnal.unmuhjember.ac.id/index.php/JST>

Zaifullah, Z., & Yulianto, T. (2022). Analisis Cluster Untuk Pengelompokan Prestasi Mahasiswa Angkatan 2013 Fakultas MIPA Universitas Islam Madura. *Zeta - Math Journal*, 7(1), 1–10. <https://doi.org/10.31102/zeta.2022.7.1.1-10>