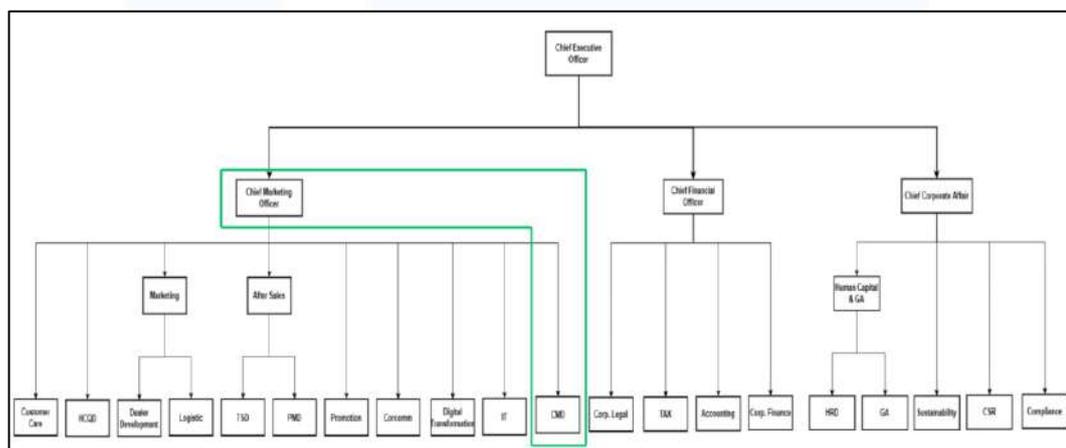


## BAB III

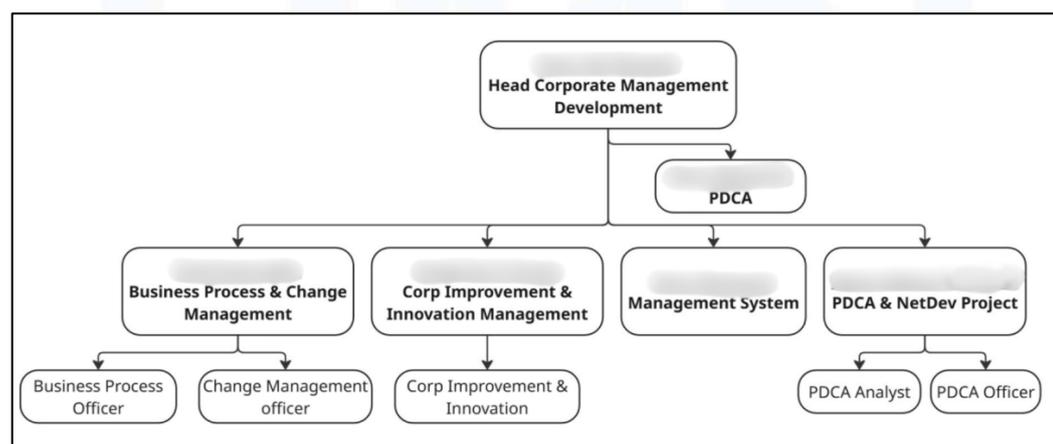
### PELAKSANAAN KERJA MAGANG

#### 3.1 Kedudukan dan Koordinasi

PT XYZ memiliki satu departemen yang menangani proses bisnis dan data-data perusahaan, yaitu *Corporate Management Development* (CMD). Oleh karena itu, mahasiswa dengan peminatan mengarah ke *Data Science*, *Data Engineer*, dan *Data Analyst* akan ditugaskan di divisi PDCA dengan posisi *PDCA Analyst* yang merupakan sub divisi dari divisi *Corporate Management Development* (CMD). Gambar 3.1 memaparkan posisi PDCA pada struktur organisasi perusahaan dan Gambar 3.2 memaparkan alur dari divisi yang bertanggung jawab atas PDCA.



Gambar 3.1 Letak Divisi CMD pada Struktur Organisasi



Gambar 3.2 Struktur Divisi CMD

Dari bagan pada Gambar 3.2, departemen CMD memiliki kedudukan dan struktur yang lebih detail. Dipimpin oleh *Head Department* yang bertanggung jawab atas divisi CMD dan PDCA. Departemen CMD dibagi lagi menjadi beberapa bagian, yaitu *Business Process and Change Management*, *Corporate Improvement and Innovation Management*, *Management System*, dan *PDCA and Network Development Project*. Mahasiswa yang melaksanakan kerja magang di divisi PDCA akan disupervisi oleh Bapak M yang merupakan *Head* dari divisi *PDCA and Network Development Project*. Bapak M bertanggung jawab dalam mengajar, memberikan usulan, dan mendidik serta meningkatkan potensi mahasiswa magang dalam memberikan kontribusi terhadap perusahaan.

### 3.2 Tugas dan Uraian Kerja Magang

Divisi PDCA memiliki peran penting dalam perusahaan dalam memastikan data yang diolah dan digunakan valid dan dapat dijadikan wawasan untuk meningkatkan performa perusahaan. Beberapa tanggung jawab PDCA meliputi melakukan pengecekan terhadap data setiap harinya di setiap departemen dan divisi pada perusahaan, membangun *dashboard* interaktif yang berfungsi untuk memonitor progress dari perusahaan di setiap aspek (keuangan, penjualan jasa dan produk, performa bengkel), mengidentifikasi perubahan pola pada data berdasarkan data historis, dan menyelidiki perubahan yang ada untuk membuat perencanaan strategis. Perangkat lunak yang digunakan untuk melaksanakan kerja magang adalah Power BI dan Python. Pengecekan basis data menggunakan laptop kantor mentor dan dilakukan secara bersama-sama menggunakan query SQL karena mahasiswa magang tidak mendapatkan akses basis data. Tabel 3.1 memaparkan tugas yang dikerjakan selama melakukan kerja magang.

Tabel 3.1 Aktivitas dan Periode Waktu Penugasan

Nomor	Aktivitas	Tanggal	Minggu
Pengenalan lingkungan perusahaan, karyawan, dan beradaptasi terhadap data yang ada			
1	Perkenalan dengan karyawan dan lingkungan perusahaan	13 Januari	3

Nomor	Aktivitas	Tanggal	Minggu
2	Pengenalan terhadap data perusahaan	14-16 Januari	3
3	Menjelaskan wawasan yang dapat diidentifikasi dari data	17 Januari	3
Analisis sentimen dan pembangunan <i>dashboard</i> terhadap data produk dari komentar media sosial			
4	Merencanakan proyek dengan metode CRISP-DM untuk analisis sentimen terhadap data media sosial TikTok mengenai motor Vario 160	20-21 Januari	4
5	Mengumpulkan data sentimen produk dari media sosial	22 Januari	4
6	Melakukan praproses terhadap data yang sudah dikumpulkan	28-31 Januari	5
7	Melakukan klasifikasi terhadap data yang sudah dilakukan praproses	3-4 Februari	6
8	Membangun <i>dashboard</i> terhadap data klasifikasi yang sudah diproses	5 Februari	6
Analisis sentimen dan pembangunan <i>dashboard</i> terhadap bengkel PT XYZ dan bengkel non-PT XYZ			
9	Melakukan perencanaan analisis sentimen terhadap PT XYZ	6 Februari	6
10	Mengumpulkan dan mengeksplorasi data sentimen yang sudah dikumpulkan	7, 10-11 Februari	6 & 7
11	Melakukan praproses pada data dan membangun <i>dashboard</i> untuk hasil klasifikasi sentimen	12-14, 17 Februari	7 & 8
12	Perbaikan terhadap model klasifikasi yang dibangun	18-20 Februari	8
13	Mempresentasikan <i>dashboard</i> yang sudah dibangun dan melanjutkan analisis sentimen terhadap bengkel PT XYZ di	21 Februari	8

Nomor	Aktivitas	Tanggal	Minggu
	daerah Tangerang serta bengkel non PT XYZ		
14	Mengumpulkan data sentimen di bengkel PT XYZ dan bengkel yang sudah ditentukan	24-25 Februari	9
15	Melakukan praproses pada data sentimen bengkel PT XYZ dan bengkel luar PT XYZ yang sudah ditentukan	26-28 Februari, 3-4 Maret	9 & 10
16	Menerapkan metode klasifikasi terhadap data yang sudah diproses	5-7 Maret	10
17	Membangun <i>dashboard</i> dari data yang sudah diklasifikasi dan dipresentasikan	10-14, 17 Maret	11 & 12
Analisis sentimen terhadap sebutan “Vario Jamet” dan pembangunan <i>dashboard</i> monitoring			
18	Melakukan analisis sentimen terhadap data komen Youtube mengenai sebutan negatif yang sering diutarakan	18 Maret	12
19	Merencanakan pengumpulan data menggunakan API Youtube	19 Maret	12
20	Mengumpulkan data video Youtube yang sudah ditentukan	19 Maret	12
21	Melakukan klasifikasi dari data yang sudah di praproses	20-21 Maret	12
22	Merevisi klasifikasi yang dilakukan dan mengubah analisis	24-25 Maret	13
23	Melakukan analisis sarkasme terhadap data komentar	26-28 Maret	13
24	Melakukan revisi terhadap metode klasifikasi yang dilakukan	7 April	15
25	Menyimpulkan hasil analisis sarkasme	7 April	15

Nomor	Aktivitas	Tanggal	Minggu
Pengembangan pembangunan <i>dashboard</i> penjualan motor Vario dengan data lain			
26	Analisis data lanjutan menggunakan <i>dashboard</i> untuk penjualan motor Vario	8 April	15
27	Merencanakan data yang perlu dikumpulkan dari basis data	8-9 April	15
28	Membersihkan data-data yang sudah dikumpulkan	10 April	15
29	Membangun <i>dashboard</i> dari data-data yang sudah dikumpulkan	11, 14-16 April	15 & 16
30	Mempresentasikan dan menyimpulkan hasil dari pembangunan <i>dashboard</i>	25 April	17
31	Menambahkan analisis sentimen untuk menjadi faktor pendukung penurunan penjualan	28-30 April	18
32	Membangun <i>dashboard</i> dengan data-data faktor pendukung	1-2, 5-7 Mei	18 & 19
33	Melanjutkan pembangunan <i>dashboard</i> dan mengumpulkan kembali data-data serta diskusi dengan <i>supervisor</i>	8-9 Mei	19
34	Menganalisis kembali data-data yang menjadi faktor pengaruh dari data penjualan	13-15 Mei	20
35	Mempresentasikan hasil <i>dashboard</i> yang sudah dibangun dan menjelaskan kesimpulan dari proyek	16 Mei	20
36	Menutup proyek <i>dashboard</i>	19 Mei	21
Penutupan pelaksanaan kerja magang			
37	Penutupan pelaksanaan kerja magang	20 Mei	21

(Sumber olahan peneliti, 2025)

Berdasarkan tabel di atas, berikut adalah penjelasan singkat mengenai pekerjaan yang dilakukan selama pelaksanaan kerja magang.

1. Pengenalan lingkungan perusahaan, karyawan, dan beradaptasi dengan data yang ada.

Pengenalan lingkungan perusahaan, karyawan, dan beradaptasi dengan data yang ada dilakukan dengan perkenalan dengan karyawan yang ada di hari pertama lalu dilanjutkan dengan mengenal data faktur dan data servis. Hasil pemahaman data kemudian dipresentasikan dengan *supervisor* untuk mengetahui proyek yang akan dibangun dan melaksanakan deskripsi pekerjaan yang akan dilakukan.

2. Analisis sentimen dan pembangunan *dashboard* terhadap data produk dari komentar media sosial

Pemenuhan deskripsi pekerjaan dilakukan pada tahap ini di mana dilakukan proyek analisis sentimen mengenai data produk, yaitu motor Vario 160 melalui komentar media sosial TikTok dari akun PT XYZ dan PT pusat. Proyek ini berlangsung selama 4 minggu setelah melakukan pengenalan data di minggu pertama. Hasil analisis sentimen digunakan untuk mengetahui perbandingan sentimen antara 2 akun, mengetahui keautentikan respon pelanggan, dan mengetahui sentimen masyarakat mengenai motor Vario 160.

3. Analisis sentimen dan pembangunan *dashboard* terhadap bengkel PT XYZ dan bengkel non-PT XYZ

Proyek analisis sentimen dan pembangunan *dashboard* terhadap bengkel PT XYZ di Jakarta dan Tangerang bersinggungan dengan salah satu proyek besar perusahaan dan proyek ini diminta secara langsung oleh CMO untuk dikerjakan. Proyek ini berlangsung selama 3 minggu dan dipresentasikan. Hasil analisis sentimen dan pembangunan *dashboard* digunakan untuk mengetahui sentimen pelanggan mengenai PT XYZ yang dikunjungi, sentimen bengkel non-PT XYZ, dan melihat perbandingan sentimen antara bengkel PT XYZ dan bengkel non-PT XYZ.

4. Analisis sentimen terhadap sebutan dan pembangunan *dashboard monitoring*

Proyek analisis sentimen terhadap sebutan negatif berlandaskan dari data yang diperoleh melalui *dashboard monitoring* penjualan di PT XYZ yang menunjukkan beberapa pengulangan penurunan penjualan dari tahun ke tahun. Proyek ini berlangsung selama 4 minggu dan hasil pembangunan *dashboard* didiskusikan dengan *supervisor*. Proyek ini bertujuan untuk mengetahui pengaruh sentimen dari sebutan negatif terhadap penjualan motor Vario 160 yang menurun di beberapa kuartal khusus untuk tahun 2024. Hasil dari proyek ini akan digunakan untuk melakukan prediksi terhadap sentimen yang mengandung sarkasme dengan model yang telah dibangun.

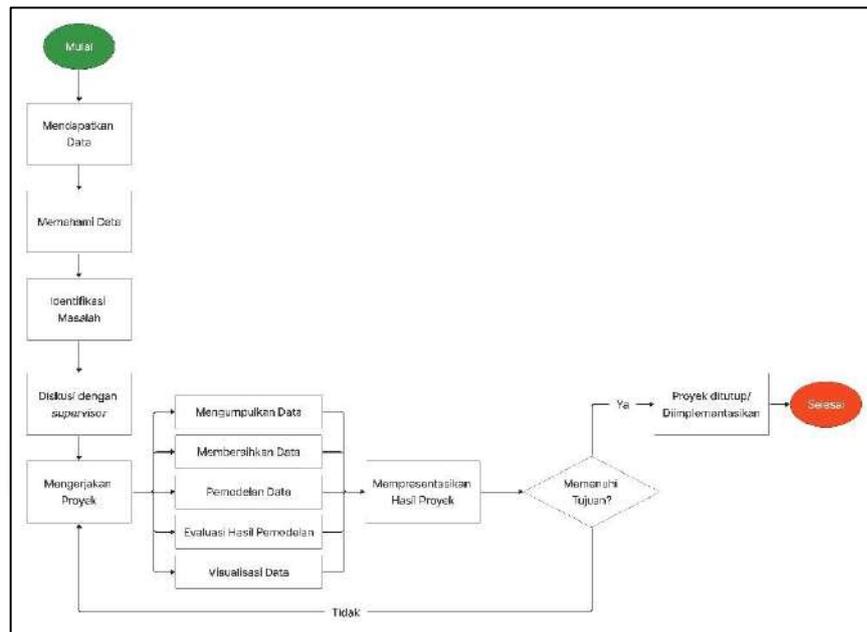
5. Pengembangan pembangunan *dashboard* penjualan motor Vario 160 dengan data lain

Proyek ini merupakan lanjutan dari proyek analisis sentimen sebelumnya. Dengan hasil yang belum cukup menjawab pertanyaan penurunan penjualan di kuartal 4, proyek ini dibentuk untuk menemukan faktor tersebut. Proyek ini melaksanakan deskripsi pekerjaan dalam hal memeriksa basis data untuk memastikan konsistensi data dan mengidentifikasi pola penting.

6. Penutupan pelaksanaan kerja magang

Pelaksanaan kerja magang ditutup dan dilakukan perpisahan dengan karyawan perusahaan.

Divisi PDCA memiliki alur kerja sendiri untuk mengerjakan permintaan *user*. Alur dalam mengerjakan tugas magang adalah melalui *user* yang akan meminta PDCA *Analyst* untuk melakukan analisis data terhadap proyek yang akan dikerjakan oleh departemen *user*. Analisis data berupa mengambil data dari basis data, memverifikasi kebenaran data, memanipulasi (merapikan) data, dan melakukan analisis data dalam bentuk *dashboard* atau grafik. Setelah memperoleh wawasan dari hasil analisis, dilakukan diskusi bersama dengan *intern supervisor* untuk melakukan analisis yang komprehensif. Setelah dilakukan diskusi, revisi, dan membuat kesimpulan, dilakukan presentasi kepada *user* terhadap hasil analisis data yang sudah dilakukan dan melakukan pengambilan keputusan bisnis untuk departemennya. Gambar 3.3 memaparkan alur dalam melakukan pengerjaan tugas.



Gambar 3.3 Alur Pengerjaan Tugas

Tugas dikerjakan dengan mendapatkan data terlebih dahulu dan dipahami untuk mengidentifikasi masalah yang ada pada data. Setelah menemukan masalah, dilakukan diskusi dengan *supervisor* untuk mengetahui proyek yang dapat dijalankan untuk menyelesaikan masalah ini. Setelah mengetahui proyek yang akan dikerjakan, proses pengerjaan proyek dilakukan dengan tahapan mengumpulkan data, membersihkan data, pemodelan data, evaluasi hasil pemodelan, dan visualisasi data. Proses mengumpulkan data hingga evaluasi menggunakan Python dan visualisasi data menggunakan Power BI untuk membangun *dashboard*. Hasil visualisasi data akan dipresentasikan kepada *supervisor* untuk mendapatkan wawasan dan memenuhi tujuan dibentuknya proyek ini. Apabila proyek belum mencapai tujuan, dilakukan analisis lebih dalam lagi dan memulai kembali proses pengerjaan proyek. Apabila proyek sudah mencapai tujuan, maka dibuat simpulan dan melanjutkan dengan proyek lain.

### 3.2.1 Pengenalan Terhadap Data Perusahaan

Pada minggu pertama pelaksanaan kerja magang, kegiatan diawali dengan proses adaptasi terhadap lingkungan kerja yang baru. Penyesuaian ini meliputi adaptasi terhadap lingkungan perusahaan, memahami komunikasi

antardepartemen, pekerjaan yang dilakukan sehari-harinya serta proyek yang sedang berjalan, dan berkenalan dengan karyawan perusahaan. Selain itu, diperkenalkan juga berbagai perangkat dan sistem yang digunakan dalam pengelolaan dan analisis data perusahaan, yaitu Microsoft Power BI, MySQL, KNIME, dan Python untuk analisis data. Namun, untuk pelaksanaan kerja magang, hanya menggunakan Python untuk *coding* dan Power BI untuk membangun *dashboard* serta menggunakan laptop kantor untuk mengakses basis data bersama dengan mentor. Proses adaptasi ini diberlakukan selama seminggu untuk seluruh mahasiswa yang baru melaksanakan kerja magang. Gambar 3.4 memaparkan proses adaptasi dan masa perkenalan lingkungan.



Gambar 3.4 Diskusi dengan *Supervisor*

Fokus utama pada minggu pertama adalah memahami data penjualan dan data pelanggan yang diambil dari basis data perusahaan. Data tersebut mencakup informasi seperti riwayat pembelian, karakteristik konsumen, waktu transaksi, serta segmentasi pelanggan. Langkah awal yang dilakukan adalah memahami struktur data dan memonitor pola distribusi yang ada pada data. Pemahaman menyeluruh terhadap isi basis data menjadi tahap yang penting sebelum untuk mengetahui hubungan setiap tabel dan data yang ada. Gambar

3.5 adalah data faktur yang diambil dari basis data dan Gambar 3.6 adalah hasil pemahaman dari isi data faktur dari sisi demografi.



Gambar 3.5 Data Faktur



Gambar 3.6 Hasil Pemahaman Data dari Sisi Dealer dan Produk

Pemahaman data faktur diminta untuk mendetail dan memahami fungsinya untuk dilakukan analisis yang komprehensif. Data faktur dari sisi *dealer* dan produk (motor) berfungsi untuk mengetahui jumlah motor yang dijual oleh setiap *dealer* beserta karakteristik motor yang terjual.

MULTIMEDIA  
NUSANTARA



Gambar 3.7 Hasil Pemahaman Data dari Sisi Demografi

Gambar 3.7 memaparkan pemahaman data faktor dari sisi demografi atau pelanggan. Hal ini berfungsi untuk melakukan analisis yang lebih mendalam mengenai pelanggan serta karakteristik konsumen yang membeli motor. Dari data tersebut, pelanggan dapat diklasifikasikan dan dapat dianalisis lebih lanjut untuk kebutuhan pemasaran dan penawaran motor.



Gambar 3.8 Hasil Pemahaman Data dari Sisi Pendapatan

Gambar 3.8 memaparkan pemahaman data faktor dari sisi pendapatan. Hal ini berfungsi untuk melakukan analisis yang mendalam mengenai jumlah pendapatan yang didapatkan perusahaan dalam periode waktu tertentu serta menganalisis faktor-faktor yang menyebabkan pendapatan naik atau turun.

Kode Ahas	Nama Ahas	Tanggal Serv	No PKB	No Polisi	No Mesin	No Rangka	Type Kendar	Kode Group	Tahun Kenda	Dealer	Penpu	Segment	Jenis Pekerja	Jenis Layanan	Alasan Kr	Al	Kilometer	Jasa/Part nu	Deskripsi	Jas	Nilai	Jasa	

Gambar 3.9 Data Servis

Data Servis		

Gambar 3.10 Hasil Pemahaman Data Servis dari Sisi Bengkel

Setelah memahami data faktur, data servis juga perlu dipahami untuk mengenal pelanggan secara komprehensif. Gambar 3.9 memaparkan data servis yang akan dipahami secara mendalam. Data servis dari sisi bengkel PT XYZ dan *dealer* bermaksud untuk mengenal lebih dalam pelanggan di setiap PT XYZ dan *dealer* dan di mana pelanggan membeli serta servis motor. Gambar 3.10 memaparkan data servis dari sisi PT XYZ dan *dealer*. Analisis ini berfungsi untuk mengidentifikasi wilayah PT XYZ dan *dealer* yang paling sering dikunjungi untuk melakukan servis dan membeli motor oleh pelanggan.

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Gambar 3.11 Hasil Pemahaman Data Servis dari Sisi Servis

Gambar 3.11 memaparkan data servis dari sisi servis dan produk. Analisis ini berfungsi untuk mengetahui waktu servis yang paling sering dilakukan oleh pelanggan, karakteristik pelanggan, dan kebutuhan konsumen yang dikeluhkan kepada PT XYZ dan *dealer*. Selain itu, analisis ini juga diperlukan untuk mengetahui sistem servis dan penjualan yang dilakukan serta efektivitas sistem yang diberlakukan.

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Gambar 3.12 Hasil Pemahaman Data Servis dari Sisi Pelanggan

Gambar 3.12 memaparkan pemahaman data servis dari sisi pelanggan. Analisis ini berfungsi untuk mengetahui kelengkapan data pelanggan dan karakteristik pelanggan yang melakukan servis. Selain itu, pergerakan pelanggan juga dapat diketahui melalui alamat yang menunjukkan tempat tinggal pelanggan dan PT XYZ yang dikunjungi untuk melakukan servis.

Analisis yang dilakukan terhadap kolom-kolom dari data faktur dan data servis dilanjutkan dengan perencanaan proyek yang akan dilakukan selanjutnya melalui diskusi dengan *supervisor*. Hasil diskusi mengarah terhadap analisis sentimen sekaligus menganalisis data faktur untuk mengetahui pengaruh sentimen masyarakat terhadap penjualan motor. Hal ini juga bertujuan untuk meningkatkan pengenalan terhadap motor-motor dan sentimen masyarakat terhadap motor-motor yang dijual. Analisis sentimen dilakukan terhadap 2 akun, yaitu akun PT pusat dan akun PT XYZ. Hal ini bertujuan untuk mengetahui akurasi sentimen melalui persamaan-persamaan sentimen yang dimiliki dari 2 akun.

### **3.2.2 Menerapkan metode CRISP-DM untuk Analisis Sentimen terhadap Data Komentar TikTok mengenai motor Vario 160**

Metode CRIPS-DM digunakan dalam melakukan analisis sentimen untuk mengetahui sentimen masyarakat terhadap motor Vario 160. Metode ini digunakan untuk melakukan tahapan-tahapan analisis data dengan teratur dan

rapih sehingga apabila ada kesalahan atau kesulitan pada satu tahap, dapat diselesaikan dengan teratur. Proyek analisis sentimen terhadap sepeda motor Vario 160 dilakukan melalui platform media sosial TikTok dari akun PT pusat dan PT XYZ. Proyek ini dirancang setelah melakukan diskusi dengan *supervisor* saat minggu perkenalan dan adaptasi lingkungan. Gambar 3.13 menunjukkan draf perencanaan proyek sentimen dengan metode CRISP-DM.

CRISP-DM	Description
Business Understanding	Tidak tahu sentimen atau respon customer terhadap motor-motor yang dipromosikan, dijual, dan dirilis.
Data Understanding	Membutuhkan data sentimen berupa komen pada sosial media yang kemudian dieksplorasi dan diklasifikasikan lalu dibangun model prediktif.
Data Preparation	Text preprocessing: <ul style="list-style-type: none"> <li>• Convert to lowercase</li> <li>• Remove special characters</li> <li>• Tokenization</li> <li>• Remove stopwords</li> <li>• Stemming / Lemmatization</li> </ul>
Data Modelling	Modelling dengan deep learning (BERT / LSTM)
Evaluation	Menggunakan confusion matrix untuk mengetahui apakah hasil overfitting atau tidak
Deployment	Dashboard klasifikasi sentimen

Gambar 3.13 Draft Perencanaan Proyek

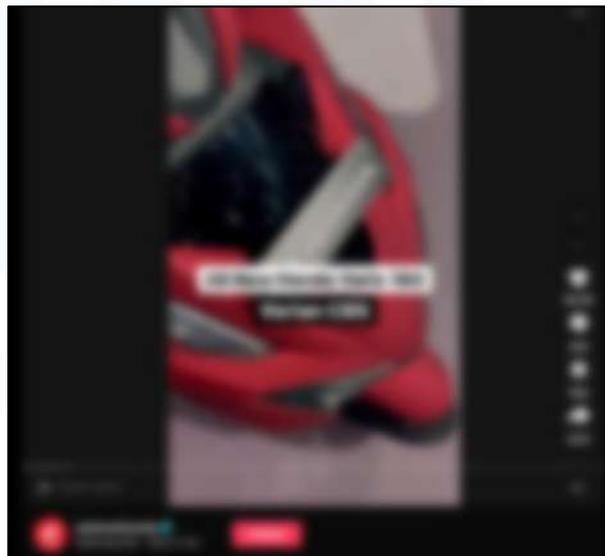
### 3.2.2.1 Business Understanding

*Business understanding* adalah tahap pertama pada CRISP-DM untuk mengidentifikasi masalah yang ditemukan di dalam perusahaan [7]. Masalah yang ditemukan adalah PT XYZ tidak tahu sentimen pelanggan dan seberapa autentik sentimen tersebut terhadap motor-motor yang dipromosikan di sosial media. Oleh karena itu, untuk proyek ini, digunakan motor Vario 160 sebagai uji coba untuk mengukur sentimen dan autentik respon masyarakat. Pemicu adanya analisis sentimen ini adalah PT XYZ memiliki akun media sosial pribadi dan memiliki pemasaran sendiri, sedangkan PT Pusat juga memiliki akun media sosial sendiri dengan pemasaran sendiri. Setelah dilihat secara sekilas, respon

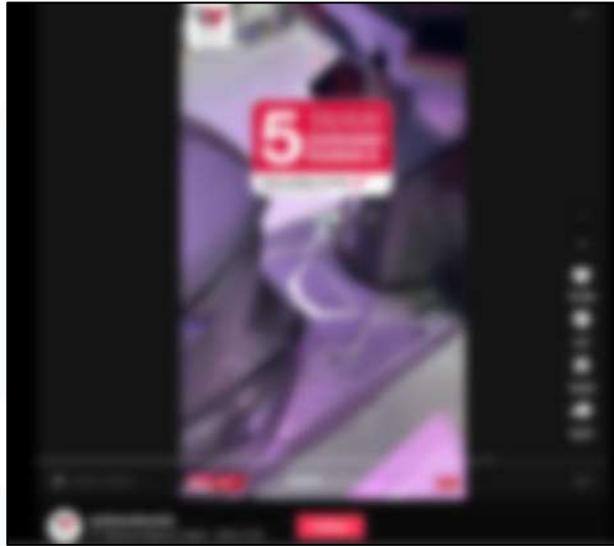
yang diberikan di kedua cukup sama, tetapi tidak cukup untuk menyimpulkan seberapa autentik respon yang diberikan oleh masyarakat. Oleh karena itu, analisis ini memiliki 3 tujuan, yaitu mengetahui sentimen masyarakat, melihat perbandingan sentimen antara akun media sosial PT XYZ dan PT Pusat, dan mengukur seberapa autentik respon masyarakat terhadap motor Vario 160.

### 3.2.2.2 Data Understanding

*Data understanding* adalah tahap kedua dalam metode CRISP-DM yang berfungsi untuk mengetahui data yang diperlukan untuk mengatasi masalah yang ditemukan [7]. Data yang diperlukan untuk mengatasi masalah ini adalah data sentimen masyarakat yang dikumpulkan melalui media sosial TikTok dari akun PT pusat dan PT XYZ. Dari masalah yang ditemukan, pengumpulan data sentimen harus memiliki beberapa kriteria, yaitu video dari kedua akun harus memiliki konten yang sama, yaitu mempromosikan motor Vario 160 dan memiliki jumlah komentar yang setara. Gambar 3.14 dan Gambar 3.15 menunjukkan video-video yang dikumpulkan dari akun TikTok PT pusat dan PT XYZ.



Gambar 3.14 Video Akun A



Gambar 3.15 Video Akun B

Proses pengumpulan data dilakukan dengan metode *scraping* menggunakan situs web yang menyediakan sarana untuk melakukan *scraping*, tetapi hanya terbatas hingga 500 komentar. Data dari akun media sosial dapat terkumpulkan dengan 424 baris dari akun Wahana Artha dan 449 baris dari akun PT XYZ. Jumlah data sudah baik dengan selisih yang tidak banyak dan dapat dianggap setara. Gambar 3.16 memaparkan data yang akan diolah dan Gambar 3.17 memaparkan hasil eksplorasi yang dilakukan.

Date	Text	URL	Likes	Replies	Type	Author	Author Profile
27 Jan, 2025 01:34	menarik bisa jadi di masa depan... PCN... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Siya	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 07:03	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Dani Ad	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 08:33	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Acu... 1856	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 09:53	lalu bingung siapa sih yang jadi... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	1	Comment	Wahana Artha	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 11:28	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 15:04	apa jadinya... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	gunt... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 17:04	kalau... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	AKA... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 19:04	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Nisa... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 19:52	Siapa... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Mon... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 20:30	lalu... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Nisa... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 20:28	lalu... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Mon... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 20:27	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Nisa... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Jan, 2025 20:26	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Nisa... (text partially obscured)	https://www.tiktok.com/@wahanaartha
28 Dec, 2024 14:36	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	30NONE	https://www.tiktok.com/@wahanaartha
27 Dec, 2024 12:40	lalu... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	AURIZ	https://www.tiktok.com/@wahanaartha
25 Dec, 2024 14:55	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	FA	https://www.tiktok.com/@wahanaartha
25 Dec, 2024 10:34	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	sih... (text partially obscured)	https://www.tiktok.com/@wahanaartha
24 Dec, 2024 10:57	OTK... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Ry	https://www.tiktok.com/@wahanaartha
24 Dec, 2024 10:50	tergang... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
24 Dec, 2024 10:14	lalu... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Ry	https://www.tiktok.com/@wahanaartha
24 Dec, 2024 10:13	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
24 Dec, 2024 09:54	lalu... (text partially obscured)	https://www.tiktok.com/@wahanaartha	1	0	Comment	Ry	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 08:59	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 20:46	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Ridha	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:28	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	1	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:16	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Dani... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:39	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Wahana Artha	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:48	menarik... (text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:39	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:30	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha
22 Dec, 2024 09:30	(text partially obscured)	https://www.tiktok.com/@wahanaartha	0	0	Comment	Andi... (text partially obscured)	https://www.tiktok.com/@wahanaartha

Gambar 3.16 Hasil Scraping Data

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 449 entries, 0 to 448
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Date            449 non-null    object
1   Text            449 non-null    object
2   URL             449 non-null    object
3   Likes           449 non-null    int64
4   Replies         449 non-null    int64
5   Type            449 non-null    object
6   Author          449 non-null    object
7   Author Profile  449 non-null    object
8                   448 non-null    object
dtypes: int64(2), object(7)
memory usage: 31.7+ KB
```

Gambar 3.17 Hasil Eksplorasi Data

### 3.2.2.3 Data Preparation

*Data preparation* adalah tahap ketiga dalam metode CRISP-DM yang berfungsi untuk membersihkan data agar dapat diproses dengan baik pada tahap *modeling* [7]. Persiapan data dilakukan setelah pengumpulan data dan dimulai dengan melakukan riset [8], [9], [10] mengenai tahapan praproses yang akan dilakukan terhadap data. Hal ini bertujuan untuk mengetahui tahapan yang tepat untuk melakukan praproses dan mengetahui cara-cara untuk mengatasi *error*. Tahapan praproses yang dilalui adalah *lowercasing* untuk menghilangkan huruf kapital, *dejojize and spacing removing punctuations* untuk menghilangkan tanda baca, *removing numbers* untuk menghilangkan angka, *removing stopwords* untuk menghilangkan konjungsi, dan *encoding emojis* [8], [10], [11] untuk mengartikan *emoji* yang tidak dapat dibaca oleh model. Tujuan dari tahap ini adalah memastikan bahwa data berada dalam kondisi yang bersih untuk menghasilkan akurasi yang tinggi saat diklasifikasikan pada tahap *modeling*. Berikut adalah penjelasan spesifik tahapan praproses yang dilakukan beserta dengan kode yang dijalankan.

1. *Lowercase comment*

Tahap ini dilakukan untuk mengubah huruf kapital menjadi huruf kecil. Tahap ini dilakukan untuk menghilangkan redundan pada data saat

dilanjutkan ke tahap praproses selanjutnya. Contohnya adalah saat membuat komentar “Bagus”, “bagus”, dan “BAGUS”, ketiganya memiliki makna yang sama, tetapi jika tidak diubah menjadi huruf kecil akan diidentifikasi berbeda di tahap praproses selanjutnya. Gambar 3.18 memaparkan kode *lowercasing* dan hasil setelah kode dijalankan.

```
# Lowercasing
df['Text'] = df['Text'].str.lower()
df['Text'].head()

0    vario sih krn multifungsi 🤔 buat beli galon elp...
1
2
3    jadi bingung pngn ambil variou atau pcx... 🤔 ba...
4
Name: Text, dtype: object
```

Gambar 3.18 Kode Lowercasing Comments beserta Hasilnya

## 2. *Encoding and Spacing Encoded Emoji*

Tahap ini dilakukan untuk melakukan *encoding emoji* yang ada pada komentar dan memberikan spasi antara komentar yang sudah di-*encode*. Tahap ini bertujuan untuk mengartikan *emoji* yang ada sehingga dapat dibaca saat tahap *modeling*. *Encoding* berhasil dilakukan dan ditunjukkan pada baris kedua di mana *emoji* tersenyum di-*encode* menjadi "smiling\_face". Gambar 3.19 memaparkan kode dan hasil *encoding* yang sudah dilakukan.

```
# Encoding and Spacing Encoded Emoji
def demojize_with_spacing(text):
    text = emoji.demojize(text, delimiters=(" ", " "))
    return text

df['Text'] = df['Text'].apply(demojize_with_spacing)
df['Text'].head()

0    vario sih krn multifungsi face_with_tears_of_j...
1                                     smiling_face
2                                     smiling_face_with_smiling_eyes
3    jadi bingung pngn ambil variou atau pcx... bea...
4                                     rolling_on_the_floor_laughing
Name: Text, dtype: object
```

Gambar 3.19 Kode Encoding dan Spacing Emoji beserta Hasilnya

## 3. *Removing punctuation*

Tahap ini dilakukan untuk menghilangkan tanda baca yang ada pada komentar. Tahap ini bertujuan untuk menghilangkan tanda baca,

tetapi tanpa menghilangkan tanda baca “\_” dari hasil *encoding emoji* karena hasil *encoding* akan digunakan untuk diklasifikasikan pada tahap *modeling*. Penghilangan tanda baca berhasil dilakukan karena tanda baca “\_” tidak dihilangkan oleh kode. Gambar 3.20 memaparkan kode dan hasil menghilangkan tanda baca yang sudah dilakukan.

```
# Removing Punctuation

import string

df['Text'] = df['Text'].apply(lambda x: ''.join([char for char in x if char not in
                                                string.punctuation or char in['_', ':']]))

df['Text'].head(30)

0    vario sih krn multifungsi face_with_tears_of_j...
1          smiling_face
2          smiling_face_with_smiling_eyes
3    jadi bingung pngn ambil variou atau pcx beamin...
4          rolling_on_the_floor_laughing
5    ga juga bg yg penting punya jaminan bpkb atau ...
6          kalau kur harus punya rumah sendiri kan
7    mantap bgt cashbacknya aku bukan 11 kmrin beli...
```

Gambar 3.20 Kode Removing Punctuation beserta Hasilnya

#### 4. *Removing numbers*

Tahap ini dilakukan untuk menghilangkan angka yang ada pada komentar. Tahap ini bertujuan untuk menghilangkan angka agar tidak redundan pada saat diproses. Menghilangkan angka berhasil dilakukan dan ditunjukkan pada baris ke-7 di mana saat menghilangkan tanda baca, angka 11 ada dan di tahap ini sudah hilang. Gambar 3.21 memaparkan kode dan hasil menghilangkan tanda baca yang sudah dilakukan.

```
# Removing Numbers

df['Text'] = df['Text'].apply(lambda x: ''.join([char for char in x if not char.isdigit()]))
df['Text'].head(30)

0    vario sih krn multifungsi face_with_tears_of_j...
1          smiling_face
2          smiling_face_with_smiling_eyes
3    jadi bingung pngn ambil variou atau pcx beamin...
4          rolling_on_the_floor_laughing
5    ga juga bg yg penting punya jaminan bpkb atau ...
6          kalau kur harus punya rumah sendiri kan
7    mantap bgt cashbacknya aku bukan  kmrin beli c...
```

Gambar 3.21 Kode Removing Numbers beserta Hasilnya

#### 5. *Removing stopwords*

Tahap ini dilakukan untuk menghilangkan konjungsi dan kata-kata yang tidak diperlukan dalam komentar. Tahap ini bertujuan untuk menghilangkan konjungsi dan beberapa kata yang tidak diperlukan agar pemrosesan saat tahap *modeling* tidak lama. Gambar 3.22 memaparkan

kode dan hasil menghilangkan konjungsi dan kata-kata yang tidak diperlukan dalam komentar.

```
# Removing Stopwords

from nltk.corpus import stopwords

stop_words = set(stopwords.words('indonesian'))

df['Text'] = df['Text'].apply(lambda x: ' '.join([word for word in x.split() if word not in stop_words]))
df['Text'].head(30)
```

```
0    vario sih krn multifungsi face_with_tears_of_j...
1                                smiling_face
2                                smiling_face_with_smiling_eyes
3    bingung pngn ambil variou pcx beaming_face_wit...
4                                rolling_on_the_floor_laughing
5    ga bg yg jaminan bpkb sertifikat tanah bisasam...
6                                kur rumah
7    mantap bgt cashbacknya kmrin beli cashback nya...
```

Gambar 3.22 Kode Removing Stopwords beserta Hasilnya

## 6. *Translating Comments*

Tahap ini dilakukan untuk menerjemahkan komentar dari Bahasa Indonesia ke Bahasa Inggris. Tahap ini bertujuan agar komentar dapat dimengerti dengan akurat pada tahap *modeling*. Gambar 3.23 memaparkan kode dan hasil menerjemahkan komentar dari Bahasa Indonesia ke Bahasa Inggris.

```
# Translating Comment

def translate_text(text):
    try:
        return googletranslater(source='id', target='en').translate(text)
    except:
        return text

df['translated_text'] = df['text'].apply(translate_text)
df.head(10)
```

Text	URL	Likes	Replies	Type	Author	Author Profile	Translated Text
ingri si_j_	https://www.tiktok.com/@wahanshonda/video/7337...	1	0	Comment	SQ	https://www.tiktok.com/@zaara99	Vario because multifunctional face_with_tears...
face	https://www.tiktok.com/@wahanshonda/video/7337...	0	0	Comment	Cud Adi	https://www.tiktok.com/@cudadi	smiling_face
eyes	https://www.tiktok.com/@wahanshonda/video/7337...	0	0	Comment	Aon...1985	https://www.tiktok.com/@aon...1985	smiling_face_with_smiling_eyes
ngin wil_	https://www.tiktok.com/@wahanshonda/video/7337...	0	1	Comment	mommy yani	https://www.tiktok.com/@vanano6983	Confused PNGN Take Variou Pcx Beaming Face_wit...

Gambar 3.23 Kode Translating Comments beserta Hasilnya

## 7. *Dropping Missing Comments*

Tahap ini dilakukan untuk menghapus baris yang bernilai kosong. Tahap ini bertujuan untuk menghapus kolom-kolom yang mungkin akan diklasifikasi pada tahap *modeling* walaupun bernilai kosong. Hasil kode ini dapat dibandingkan dengan data pada tahap *data understanding* di mana jumlah baris pada data adalah 449 dan sekarang berjumlah 439

baris. Tahap ini dilakukan di akhir karena ada beberapa komentar yang sebelum dipraproses tidak hilang saat kode ini dijalankan. Gambar 3.24 memaparkan kode dan hasil menghapus komentar yang bernilai kosong.

```
# Dropping Missing Comments

df = df.dropna(subset=['Text'])
df = df[~df['Text'].str.strip().eq('')]

df = df.reset_index(drop=True)
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 439 entries, 0 to 438
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Date             439 non-null   object
1   Text             439 non-null   object
2   URL              439 non-null   object
3   Likes           439 non-null   int64
4   Replies         439 non-null   int64
5   Type            439 non-null   object
6   Author          439 non-null   object
7   Author Profile  439 non-null   object
8                   438 non-null   object
9   Translated_Text 439 non-null   object
dtypes: int64(2), object(8)
memory usage: 34.4+ KB
```

Gambar 3.24 Kode Menghapus Nilai Kosong beserta Hasilnya

### 3.2.2.4 Modeling

*Modeling* adalah tahap keempat pada metode CRISP-DM yang berfungsi untuk memproses data yang sudah dibersihkan pada tahap *data preparation* untuk memperoleh hasil yang dapat dievaluasi pada tahap selanjutnya [7]. Setelah melakukan praproses, sentimen yang sudah bersih akan diklasifikasikan menggunakan model NLP. Model pertama yang diuji coba adalah VADER, tetapi 1 kendala ditemukan yaitu *emoji* yang tidak diproses dengan baik walaupun sudah dalam bentuk teks. Hasil klasifikasi sangat buruk dan banyak misklasifikasi karena konteks sentimen tidak sesuai dengan konteks yang diberikan menggunakan *emoji*. Oleh karena itu, diteliti beberapa artikel jurnal untuk menemukan solusi. Salah satu solusi yang ditemukan adalah menggunakan TextBlob yang merupakan *library* Python untuk mengklasifikasi sentimen berdasarkan polaritas kata. Namun, seluruh komentar harus

diterjemahkan ke Bahasa Inggris terlebih dahulu agar dapat diidentifikasi oleh TextBlob. Setelah didiskusikan dengan *supervisor*, diputuskan untuk menggunakan TextBlob untuk mengklasifikasi berdasarkan hasil terjemahan yang sudah dilakukan. Selain itu, untuk mengatasi misklasifikasi *emoji*, *dictionary* dibangun secara manual terhadap *emoticon* dan beberapa kata sifat yang sudah dalam bentuk teks seperti pada Gambar 3.25. Hal ini bertujuan untuk meningkatkan hasil akurasi dan tidak mengurangi jumlah komentar yang ada.

```
emoji_sentiment_dict = {
    "grinning_face": "Positive", "grinning_face_with_big_eyes": "Positive", "grinning_face_with_smiling_eyes": "Positive",
    "beaming_face_with_smiling_eyes": "Positive", "grinning_squinting_face": "Positive", "grinning_face_with_sweat": "Positive",
    "rolling_on_the_floor_laughing": "Positive", "smiling_face_with_smiling_eyes": "Positive", "smiling_face_with_heart-eyes": "Positive",
    "smiling_face_with_hearts": "Positive", "smiling_face_with_sunglasses": "Positive", "thumbs_up": "Positive",
    "pensive_face": "Negative", "crying_face": "Negative", "loudly_crying_face": "Negative", "angry_face": "Negative",
    "enraged_face": "Negative", "thumbs_down": "Negative", "broken_heart": "Negative", "neutral_face": "Neutral",
    "expressionless_face": "Neutral", "face_without_mouth": "Neutral"
}
```

Gambar 3.25 Dictionary Encoded Emoji

Setelah membuat *dictionary*, dibuat kode *function* untuk mengklasifikasikan komentar yang sudah di *demojize* dan diterjemahkan. Gambar 3.26 memaparkan kode *function* yang dijalankan untuk mengklasifikasikan komentar beserta hasilnya.

```
def classify_sentiment(text):
    if not isinstance(text, str):
        return 'Neutral'

    # Classifying Demojized Sentiment
    words = text.split()
    emoji_matches = [word.strip(':') for word in words if word.strip(':') in emoji_sentiment_dict]

    if emoji_matches:
        sentiments = [emoji_sentiment_dict[emo] for emo in emoji_matches]
        return max(set(sentiments), key=sentiments.count)

    # TextBlob Sentiment
    blob = TextBlob(text)
    polarity = blob.sentiment.polarity

    if polarity > 0.1:
        return 'Positive'
    elif polarity < -0.1:
        return 'Negative'
    else:
        return 'Neutral'

df['Sentiment'] = df['Translated_Text'].apply(classify_sentiment)
df['Polarity'] = df['Translated_Text'].apply(lambda x: TextBlob(x).sentiment.polarity)
print(df[['Translated_Text', 'Sentiment', 'Polarity']])
```

	Translated_Text	Sentiment	Polarity
0	Vario because multifunctional face_with_tear_...	Neutral	0.000000
1	smiling_face	Neutral	0.000000
2	smiling_face_with_smiling_eyes	Positive	0.000000
3	Confused PNGN Take Variou Pcx Beaming_face_wit...	Negative	-0.400000
4	rolling_on_the_floor_laughing	Positive	0.000000

Gambar 3.26 Kode Function TextBlob

*Dictionary emoji* dibangun berdasarkan diskusi yang dilakukan dengan *supervisor* dan mentor. Klasifikasi kemudian dilakukan

menggunakan TextBlob berdasarkan polaritas sentimen. Penilaian polaritas berfungsi untuk mengetahui kecenderungan emosional di dalam teks. Contohnya adalah apabila ada sebuah kata sifat, seperti “ramah” atau “buruk” memiliki penilaian polaritas yang dapat digolongkan menjadi 1 atau -1 di mana angka 1 berarti sentimen tersebut cenderung positif dan -1 berarti sentimen tersebut cenderung negatif.

### 3.2.2.5 Evaluation

*Evaluation* adalah tahap kelima dalam metode CRISP-DM yang berfungsi untuk mengevaluasi hasil klasifikasi yang sudah dilakukan menggunakan metode tertentu [7]. Hasil klasifikasi menunjukkan mayoritas sentimen masyarakat terhadap motor Vario 160 adalah netral dari kedua akun. Hasil klasifikasi dipaparkan pada Gambar 3.27 dan Gambar 3.28.

```
df['Sentiment'].value_counts()
Sentiment
Neutral    269
Positive   136
Negative    34
Name: count, dtype: int64
```

Gambar 3.27 Hasil Klasifikasi Akun PT pusat

```
df['Sentiment'].value_counts()
Sentiment
Neutral    259
Positive   135
Negative    24
Name: count, dtype: int64
```

Gambar 3.28 Hasil Klasifikasi Akun PT XYZ

Hasil klasifikasi yang sudah diperoleh akan diuji dengan data yang sebenarnya atau *true label*. Data *true label* dibangun melalui pembuatan *library* secara manual berdasarkan hasil diskusi yang sudah dilakukan dengan *supervisor* dan mentor. Gambar 3.29 memaparkan *library* yang dibangun.

```

classifying.py X
C:\Users\Sheera > OneDrive - Universitas Multimedia Nusantara > Dokumen > Project > classifying.py > ...
1 positive_keywords = ['alhamdulillah', 'economical', 'great', 'bought', 'buy', 'good', 'comfortable', 'cool', 'want',
2                       'dream', 'grinning_face', 'grinning_face_with_big_eyes', 'grinning_face_with_smiling_eyes',
3                       'beaming_face_with_smiling_eyes', 'grinning_squinting_face', 'grinning_face_with_sweat',
4                       'rolling_on_the_floor_laughing', 'smiling_face_with_smiling_eyes', 'smiling_face_with_heart-eyes',
5                       'smiling_face_with_hearts', 'smiling_face_with_sunglasses', 'thumbs_up']
6 negative_keywords = ['esaf', 'esap', 'bad', 'hard', 'pensive_face', 'crying_face', 'loudly_crying_face', 'angry_face',
7                       'enraged_face', 'thumbs_down', 'broken_heart']
8
9 def classify_sentiment(text):
10     text = text.lower()
11     pos_score = sum(word in text for word in positive_keywords)
12     neg_score = sum(word in text for word in negative_keywords)
13
14     if pos_score > neg_score:
15         return 'Positive'
16     elif neg_score > pos_score:
17         return 'Negative'
18     else:
19         return 'Neutral'

```

Gambar 3.29 Library untuk Klasifikasi True Label

Kolom *true label* akan dimasukan ke dataset menggunakan kode pada Gambar 3.30 untuk data akun PT pusat dan Gambar 3.31 untuk data akun PT XYZ beserta dengan hasil klasifikasinya.

```

from classifying import classify_sentiment
df['True_Label'] = df['Text'].apply(classify_sentiment)
df['True_Label'].value_counts()

True_Label
Neutral      338
Positive      74
Negative      27
Name: count, dtype: int64

```

Gambar 3.30 Kode Klasifikasi True Label Akun A beserta Hasilnya

```

from classifying import classify_sentiment
df['True_Label'] = df['Text'].apply(classify_sentiment)
df['True_Label'].value_counts()

True_Label
Neutral      333
Positive      76
Negative       9
Name: count, dtype: int64

```

Gambar 3.31 Kode Klasifikasi True Label Akun B beserta Hasilnya

Penambahan kolom *true label* pada dataset akan diuji menggunakan *accuracy score* dan *classification report*. Hasil pengujian dan kode untuk data akun PT pusat dipaparkan pada Gambar 3.32 dan data akun PT XYZ pada Gambar 3.33.

```
print("Accuracy:", accuracy_score(df['True_Label'], df['predicted']))
print("\nClassification Report:\n", classification_report(df['True_Label'], df['predicted']))
```

Accuracy: 0.7471526195899773

Classification Report:

	precision	recall	f1-score	support
Negative	0.41	0.52	0.46	27
Neutral	0.93	0.74	0.82	338
Positive	0.48	0.88	0.62	74
accuracy			0.75	439
macro avg	0.61	0.71	0.63	439
weighted avg	0.82	0.75	0.76	439

Gambar 3.32 Hasil Uji Akurasi Klasifikasi Akun PT pusat

```
print("Accuracy:", accuracy_score(df['True_Label'], df['predicted']))
print("\nClassification Report:\n", classification_report(df['True_Label'], df['predicted']))
```

Accuracy: 0.7751196172248804

Classification Report:

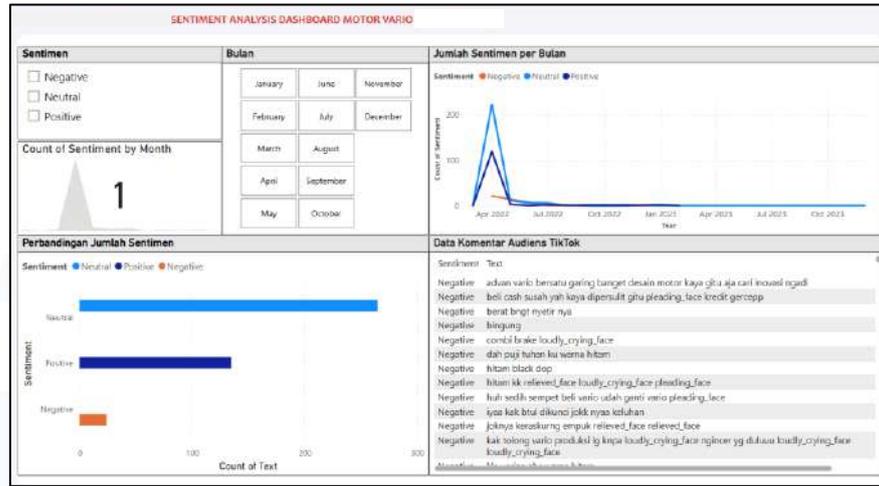
	precision	recall	f1-score	support
Negative	0.38	1.00	0.55	9
Neutral	0.96	0.75	0.84	333
Positive	0.49	0.87	0.63	76
accuracy			0.78	418
macro avg	0.61	0.87	0.67	418
weighted avg	0.86	0.78	0.80	418

Gambar 3.33 Hasil Uji Akurasi Klasifikasi Akun PT XYZ

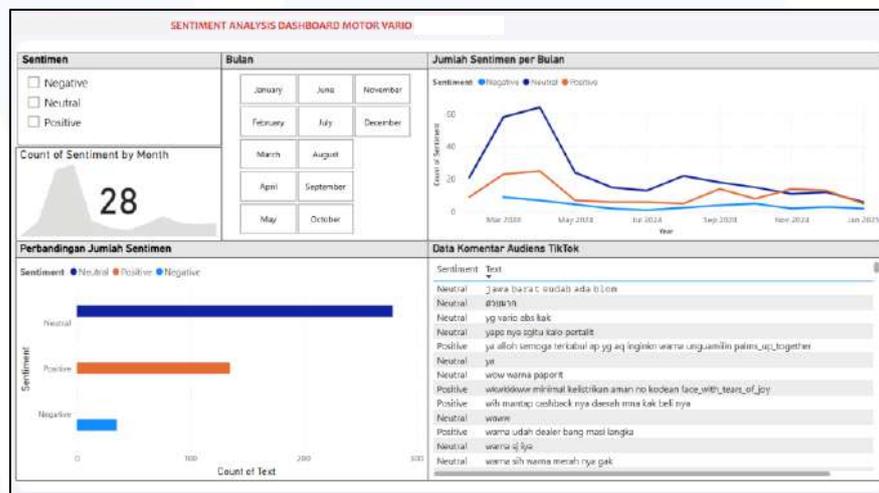
Hasil pengujian untuk data akun PT pusat memperoleh akurasi sebesar 74% dan data akun PT XYZ memperoleh akurasi sebesar 77%. Kedua hasil memiliki akurasi di atas rata-rata, yaitu 70%, tetapi klasifikasi sentimen negatif dan positif memiliki presisi yang kurang baik. Hal ini disebabkan oleh data untuk sentimen negatif dan positif yang tidak seimbang dengan sentimen netral.

### 3.2.2.6 Deployment

*Deployment* adalah tahap terakhir pada metode CRISP-DM yang berfungsi untuk menerapkan hasil evaluasi yang sudah dilakukan [7]. Pembangunan *dashboard* untuk hasil analisis sentimen dilakukan menggunakan Microsoft Power BI. Pembangunan *dashboard* untuk perusahaan memiliki kriteria warna dan ukuran tersendiri yang perlu diikuti. Oleh karena itu, pemilihan warna dan desain tampilan hanya perlu mengikuti ketentuan yang sudah diberikan. Hasil pembangunan *dashboard* untuk hasil analisis pada akun PT pusat dapat dilihat pada Gambar 3.34 dan akun PT XYZ pada Gambar 3.35.

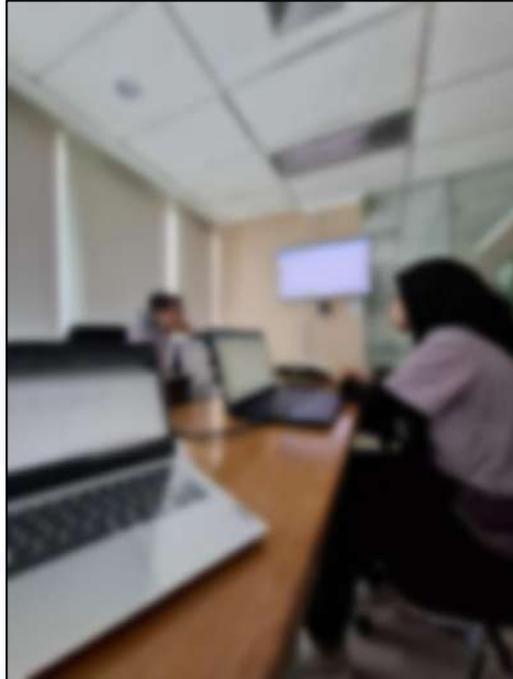


Gambar 3.34 Dashboard Sentimen Akun PT Pusat



Gambar 3.35 Dashboard Sentimen Akun PT XYZ

Hasil pembangunan *dashboard* ini juga dilakukan diskusi dan presentasi kepada *supervisor* yang dipaparkan pada Gambar 3.36.



Gambar 3.36 Presentasi dengan Supervisor

Setelah dilakukan presentasi, diperoleh kesimpulan bahwa proyek ini sudah mencapai 3 tujuan yang ingin dicapai pada saat tahap *business understanding*. Berikut adalah penjelasan setiap tujuan yang sudah terpenuhi.

a. Mengetahui sentimen masyarakat

Dari kedua hasil klasifikasi, dapat disimpulkan bahwa masyarakat memiliki sifat netral terhadap pemasaran motor Vario 160 yang dilakukan. Hal ini ditunjukkan melalui komentar-komentar yang hanya menanyakan mengenai spesifikasi motor dan menyebutkan sebutan “Alhamdulillah sudah beli”. Namun, ada beberapa komentar yang menarik mengenai sentimen negatif terhadap motor, yaitu rangka yang mudah rusak atau patah dan istilah “jamet” yang seringkali membuat pelanggan menolak untuk membeli motor Vario 160.

b. Melihat perbandingan sentimen antara akun media sosial PT XYZ dan PT pusat

Hasil klasifikasi sentimen antara kedua akun tergolong serupa karena jumlah sentimen positif, netral, dan negatif kedua akun hampir

sama. Perbandingan sentimen ingin dilihat karena akun PT pusat yang dimiliki oleh PT pusat memiliki pengikut yang lebih banyak daripada akun yang dimiliki oleh PT XYZ. Namun, setelah melakukan klasifikasi ini, ditemukan bahwa kedua respon cukup sama.

c. Mengukur seberapa autentik respon masyarakat terhadap motor Vario 160

Melanjutkan tujuan kedua yang melihat perbandingan sentimen antara 2 akun, keautentikan respon masyarakat juga akan dianalisis untuk mengetahui keaslian respon masyarakat terhadap motor Vario 160. Kedua akun memiliki respon yang banyak, tetapi keautentikan respon tidak terlalu terlihat karena setiap respon hanya berupa ungkapan syukur dan bukan respon yang benar-benar mengomentari motor Vario 160. Respon masyarakat mayoritas hanya menanyakan harga motor Vario 160 dan mengungkapkan “Alhamdulillah sudah beli” atau “Semoga bisa beli”. Dari pernyataan-pernyataan ini, tidak dapat disimpulkan ketertarikan masyarakat terhadap motor Vario 160.

Presentasi yang dilakukan setelah pembangunan *dashboard* ditutup dan dimulai kembali dengan proyek baru yang berhubungan dengan analisis sentimen karena respon masyarakat dapat diteliti lebih lanjut melalui media atau metode lainnya. Proyek ini ditutup dengan *dashboard* analisis sentimen untuk motor Vario 160.

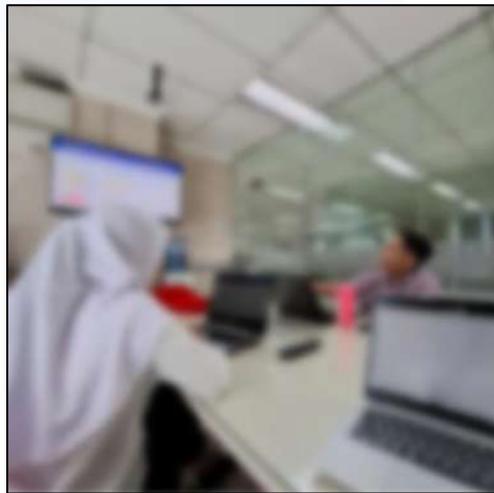
### **3.2.3 Menerapkan Metode CRISP-DM untuk Analisis Sentimen terhadap bengkel PT XYZ di Jakarta dan Tangerang**

Proyek analisis sentimen terhadap bengkel PT XYZ di Jakarta dan Tangerang bertujuan untuk mengidentifikasi ulasan masyarakat dan memantau performa layanan bengkel PT XYZ. Pada awalnya proyek ini hanya sebatas ingin mengetahui sentimen pelanggan terhadap 3 bengkel PT XYZ di Jakarta. Namun, di tengah menjalani proyek, CMO meminta untuk memberhentikan proyek dan melanjutkannya dengan analisis sentimen untuk bengkel-bengkel PT XYZ di daerah Cikupa, Periuk, dan Tigaraksa. Selain itu, diminta juga

untuk melakukan analisis sentimen terhadap bengkel-bengkel non-PT XYZ dengan ulasan terbanyak yang ada di 3 daerah tersebut untuk mengetahui perbandingan sentimen.

### 3.2.3.1 Business Understanding

Proyek ini dimulai dengan mengidentifikasi masalah-masalah yang ditemukan dan ingin diteliti. Sentimen pelanggan terhadap bengkel PT XYZ di Tangerang belum pernah diteliti sehingga performa bengkel PT XYZ tidak pernah diketahui dan dinilai. Masalah tersebut membentuk 3 tujuan proyek analisis sentimen dibuat, yaitu untuk mengetahui sentimen masyarakat yang sudah pernah membeli motor di bengkel PT XYZ daerah Cikupa, Periuk, dan Tigaraksa, mengetahui performa bengkel PT XYZ dari sisi pelanggan, dan mengetahui perbandingan sentimen antara bengkel PT XYZ dan bengkel non-PT XYZ. Gambar 3.37 memaparkan diskusi dengan *supervisor* mengenai proyek ini.

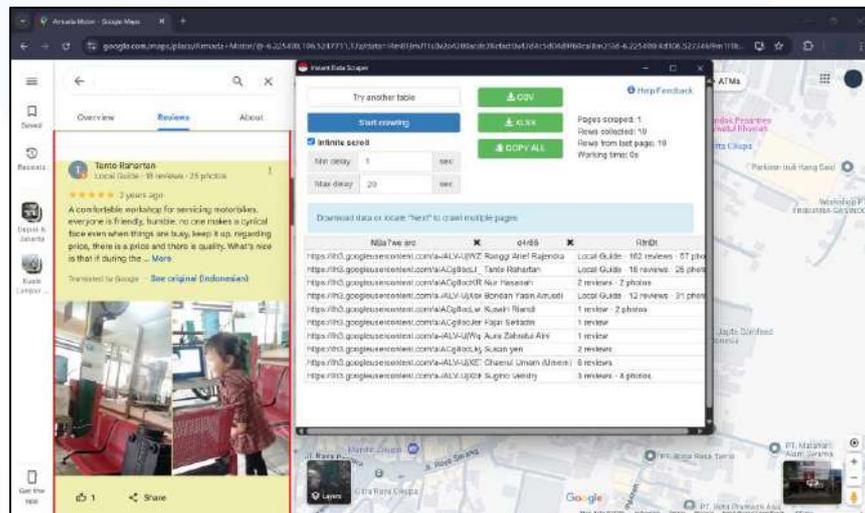


Gambar 3.37 Diskusi dengan Supervisor

### 3.2.3.2 Data Understanding

Data yang dibutuhkan untuk menjalankan proyek ini adalah data ulasan bengkel PT XYZ di Cikupa, Periuk, dan Tigaraksa, serta bengkel non-PT XYZ yang diperoleh dari Google Review. Proyek ini dimulai dengan melakukan *scraping* data menggunakan platform yang sama dengan proyek sentimen motor Vario 160. Namun, ditemukan beberapa

kendala, yaitu keterbatasan kuota untuk mengumpulkan ulasan karena platform berbayar dan ulasan yang tertumpuk dengan balasan ulasan oleh admin bengkel PT XYZ (*reply*) sehingga ulasan tidak bersih. Oleh karena itu, dicari alternatif lain untuk melakukan *scraping* ulasan. Alternatif ditemukan dengan menggunakan *extension* untuk *web browser* Google Chrome di mana *extension* tersebut dapat merekam ulasan yang terlihat di halaman web, tetapi tidak merekam balasan ulasan. Gambar 3.38 memaparkan cara menggunakan *extension* untuk *web browser* Google Chrome dan hasil *scraping* salah satu bengkel PT XYZ dipaparkan pada Gambar 3.39.



Gambar 3.38 Scraping dengan Extension

URL	Name	Address	Reviews	Photos	Response
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	102 reviews	07 photos		Bagikan
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	18 reviews	05 photos		Tanggapi
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	12 reviews	01 photos		Bagikan
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	1 review	2 photos		Tanggapi
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	1 review			Bagikan
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	2 reviews			Bagikan
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	0 reviews			Bagikan
https://www.google.com/maps/place/PT+XYZ+Rajerda	Local Guide	3 reviews	4 photos		Bagikan

Gambar 3.39 Hasil Scraping Data

Data yang sudah di-scraping akan diekplorasi terlebih dahulu menggunakan Python untuk mengetahui informasi umum mengenai data. Gambar 3.40 memaparkan data yang diimpor dan Gambar 3.41 memaparkan informasi berupa jumlah nilai yang hilang dan tipe data.

```

# Import Data
ad = pd.read_csv('4muda.csv')
ad.head()

```

	NBa7we	src	d4r55	RfnDt	eaLgGf	hCCjke	rsqaWe	wiI7pd	dSlJg	dSlJg 2	znYl0	nM6d2c	DZSIDd	wiI7p	
0	https://lh3.googleusercontent.com/a/ACg8ocL16...	Local Guide	Tanto Rahartan	15 ulasan	25 foto			2 tahun lalu	Bengkel yg nyaman untuk servis motor semua n...			Bagikan	Tanggapan dari pemilik	2 tahun lalu	Trimak bro
1	https://lh3.googleusercontent.com/a/ACg8oclem4...	Fajar Sebadin	1 ulasan					sebulan lalu	Servis buruk bausd cut ilang satu perma servis...			Suka	Tanggapan dari pemilik	sebulan lalu	Maa maaf un kel nya vt
2	https://lh3.googleusercontent.com/a/ACg8ocLwgd...	Kusali Rendi	1 ulasan	2 foto				7 bulan lalu	Terima kasih atas tanggapannya\nSesuai janji...			Bagikan	Tanggapan dari pemilik	7 bulan lalu	Maa un pesa bok service le
3	https://lh3.googleusercontent.com/a/ACg8odKRAK...	Nur Hasanah	2 ulasan	2 foto				2 tahun lalu	Pelayanan nya baik, mekanik nya ramah\nFasalit...			Bagikan	Tanggapan dari pemilik	2 tahun lalu	terimaka

Gambar 3.40 Hasil Impor Data

```

ad.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37 entries, 0 to 36
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   NBa7we      29 non-null    object
1   src         29 non-null    object
2   d4r55       29 non-null    object
3   RfnDt       27 non-null    object
4   eaLgGf      29 non-null    object
5   hCCjke      29 non-null    object
6   rsqaWe      29 non-null    object
7   wiI7pd      18 non-null    object
8   dSlJg       18 non-null    object
9   dSlJg 2     18 non-null    object
10  znYl0       13 non-null    object
11  nM6d2c      19 non-null    object
12  DZSIDd      19 non-null    object
13  wiI7pd 2    19 non-null    object
14  znYl0 2     17 non-null    object
dtypes: object(14)
memory usage: 4.2+ KB

```

Gambar 3.41 Informasi Data

### 3.2.3.3 Data Preparation

Persiapan data yang akan diolah di tahap *modeling* akan dilakukan praproses. Urutan tahapan praproses untuk sentimen sama dengan proyek sentimen motor Vario 160 dengan sedikit tambahan

dalam memanipulasi kolom. Berikut adalah penjelasan tahapan praproses secara terperinci.

### 1. Drop Columns

Pada tahap eksplorasi data, ditemukan banyak kolom-kolom yang tidak diperlukan di dalam proses analisis. Oleh karena itu, dilakukan penghapusan kolom yang tidak diperlukan dengan kode dan hasil yang dipaparkan pada Gambar 3.42.

```
# Drop Columns
ad = ad.drop(columns = ['Rfndt', 'ealGof', 'hCCjke', 'dS1jg', 'zNv10', 'dS1jg_2', 'zNv10_2', 'nFEdZc', 'DZSIDd', 'wi17pd_2', 'zNv10_2'], axis = 1)
ad.head()
```

	Nba7we src	d4r55	rsqaWe	wi17pd
0	https://lh3.googleusercontent.com/a/ACg8ocLI_6...	Tanto Rahartan	2 tahun lalu	Bengkel yg nyaman untuk servis motor , semua n...
1	https://lh3.googleusercontent.com/a/ACg8ocIem4...	Fajar Setiadin	sebulan lalu	Servis buruk baid cvt ilang satu pertma servis...
2	https://lh3.googleusercontent.com/a/ACg8ocLwgD...	Kusairi Riandi	7 bulan lalu	Terima kasih atas tanggapannya.\nSesuai janji ...
3	https://lh3.googleusercontent.com/a/ACg8ocKRAK...	Nur Hasanah	2 tahun lalu	Pelayanan nya baik, mekanik nya ramah\nFasall...
4	https://lh3.googleusercontent.com/a-/ALV-UjXsk...	Bondan Yasin Arrusdi	setahun lalu	Seal oli mesin cvt jebol,dari pertigaan lampu ...

Gambar 3.42 Kode Menghapus Kolom beserta Hasilnya

### 2. Rename Columns

Kolom-kolom yang diperlukan akan diubah menjadi nama yang seharusnya. Contohnya adalah kolom 'Nba7we src' diubah menjadi 'URL' karena mengandung *link* pelanggan yang mengkommentar Gambar 3.43 memaparkan kode dan hasil menjalankannya.

```
# Renaming Columns
ad.rename(columns={'Nba7we src': 'URL',
                  'd4r55': 'Nama Pelanggan',
                  'rsqaWe': 'Waktu Ulasan',
                  'wi17pd': 'Ulasan'}, inplace=True)
ad.head(30)
```

	URL	Nama Pelanggan	Waktu Ulasan	Ulasan
0	https://lh3.googleusercontent.com/a/ACg8ocLI_6...	Tanto Rahartan	2 tahun lalu	Bengkel yg nyaman untuk servis motor , semua n...
1	https://lh3.googleusercontent.com/a/ACg8ocIem4...	Fajar Setiadin	sebulan lalu	Servis buruk baid cvt ilang satu pertma servis...
2	https://lh3.googleusercontent.com/a/ACg8ocLwgD...	Kusairi Riandi	7 bulan lalu	Terima kasih atas tanggapannya.\nSesuai janji ...
3	https://lh3.googleusercontent.com/a/ACg8ocKRAK...	Nur Hasanah	2 tahun lalu	Pelayanan nya baik, mekanik nya ramah\nFasall...
4	https://lh3.googleusercontent.com/a-/ALV-UjXsk...	Bondan Yasin Arrusdi	setahun lalu	Seal oli mesin cvt jebol,dari pertigaan lampu ...
5	https://lh3.googleusercontent.com/a-/ALV-UjXE7...	Chaerul Umam (Umem)	3 bulan lalu	tempat mengkhawatirkan\ntidak direkomendasikan...

Gambar 3.43 Kode Mengganti Nama Kolom beserta Hasilnya

### 3. Mengubah Waktu menjadi Tahun

Pengumpulan data dengan *extension* hanya mengembalikan waktu pelanggan berkomentar dengan format 2 tahun lalu, sebulan lalu, dan sebagainya. Penulisan format tanggal tidak spesifik menjadi format

‘DD/MM/YY’. Oleh karena itu, dengan pernyataan seperti “2 tahun lalu”, akan diubah menjadi tahun dengan menghitung 2025 dikurangi dengan waktu ulasan dinyatakan. Contohnya adalah 2 tahun lalu menjadi tahun 2023. Gambar 3.44 memaparkan kode *function* untuk mengganti format tanggal dan Gambar 3.45 adalah hasil kode dijalankan.

```
# Function to convert time to year
def convert_to_year(date_string):
    today = datetime.today()
    delta = None

    if "tahun" in date_string:
        if "setahun" in date_string:
            delta = timedelta(days=365)
        else:
            num_years = int(date_string.split()[0])
            delta = timedelta(days=365 * num_years)
    elif "bulan" in date_string:
        if "sebulan" in date_string:
            delta = timedelta(days=30) # 1 month
        else:
            num_months = int(date_string.split()[0])
            delta = timedelta(days=30 * num_months)
    elif "minggu" in date_string:
        if "seminggu" in date_string:
            delta = timedelta(weeks=1)
        else:
            num_weeks = int(date_string.split()[0])
            delta = timedelta(weeks=num_weeks)
    elif "hari" in date_string:
        num_days = int(date_string.split()[0])
        delta = timedelta(days=num_days)

    if delta is None:
        return "Unknown"

    past_date = today - delta
    past_year = past_date.year

    return str(past_year)
```

Gambar 3.44 Kode Function Mengubah Format Waktu

	Waktu Ulasan	Ulasan Tahun
0	2 tahun lalu	Bengkel yg nyaman untuk servis motor , semua n... 2023
1	sebulan lalu	Servis buruk baud cvt ilang satu pertma servis... 2025
2	7 bulan lalu	Terima kasih atas tanggapannya.\nSesuai janji ... 2024
3	2 tahun lalu	Pelayanan nya baik, mekanik nya ramah\nFasalit... 2023
4	setahun lalu	Seal oli mesin cvt jebol,dari pertigaan lampu ... 2024
5	3 bulan lalu	tempat mengkhawatirkan\ntidak direkomendasikan... 2024

Gambar 3.45 Hasil Menjalankan Kode

#### 4. Lowercasing Data

Tahap praproses khusus untuk komentar dimulai dengan mengubah seluruh kata menjadi huruf kecil. Gambar 3.46 memaparkan kode untuk mengubah seluruh huruf menjadi huruf kecil beserta hasilnya.

```
# Lowercasing Data

ad['Ulasan'] = ad['Ulasan'].str.lower()
ad['Ulasan'].head()

0    bengkel yg nyaman untuk servis motor , semua n...
1    servis buruk baud cvt ilang satu pertma servis...
2    terima kasih atas tanggapannya.\nsesuai janji ...
3    pelayanan nya baik, mekanik nya ramah\nfasalit...
4    seal oli mesin cvt jebol,dari pertigaan lampu ...
Name: Ulasan, dtype: object
```

Gambar 3.46 Kode Lowercasing Data beserta Hasilnya

### 5. Demojize Data

Untuk mempermudah klasifikasi *emoji*, dilakukan *demojize* untuk mengganti *emoji* dalam format UTF-8 menjadi teks, seperti thumbs\_up. Kemudian dilakukan penghapusan tanda baca yang diperlukan, kecuali “\_” dan “:” karena kedua tanda baca menjadi penyambung *emoji* yang sudah *demojize*. Gambar 3.47 memaparkan kode untuk melakukan *demojize* beserta dengan hasilnya.

```
# Demojize Data

def demojize_with_spacing(text):
    text = emoji.demojize(text, delimiters=(" ", " "))
    return text

ad['Ulasan'] = ad['Ulasan'].astype(str).apply(demojize_with_spacing)
ad['Ulasan'].head()

0    bengkel yg nyaman untuk servis motor , semua n...
1    servis buruk baud cvt ilang satu pertma servis...
2    terima kasih atas tanggapannya.\nsesuai janji ...
3    pelayanan nya baik, mekanik nya ramah\nfasalit...
4    seal oli mesin cvt jebol,dari pertigaan lampu ...
Name: Ulasan, dtype: object
```

Gambar 3.47 Kode Demojize beserta Hasilnya

### 6. Removing Punctuation

Data dengan tanda baca akan dihilangkan dengan kode ini, tetapi tidak menghilangkan hasil *demojize* dengan mengecualikan tanda baca “:” dan “\_”. Gambar 3.48 memaparkan kode menghilangkan tanda baca beserta hasilnya.

```

# Removing Punctuation
import string

ad['Ulasan'] = ad['Ulasan'].apply(lambda x: ''.join([char for char in x if char not in string.punctuation or char in ('_', ':')]))
ad['Ulasan'].head(30)

0    bengkel yg nyaman untuk servis motor semua ny...
1    servis buruk baud cvt ilang satu pertma servis...
2    terima kasih atas tanggapannya\nsesuai janji s...
3    pelayanan nya baik mekanik nya ramah\nfasalita...
4    seal oli mesin cvt jeboldari pertigaan lampu m...

```

Gambar 3.48 Kode Menghilangkan Tanda Baca beserta Hasilnya

## 7. *Removing Numbers*

Manipulasi data dilanjutkan dengan menghilangkan angka untuk mempermudah pemrosesan data pada tahap selanjutnya. Gambar 3.49 memaparkan kode dan hasil setelah kode dijalankan.

```

# Removing Numbers

ad['Ulasan'] = ad['Ulasan'].apply(lambda x: ''.join([char for char in x if not char.isdigit()]))
ad['Ulasan'].head(30)

0    bengkel yg nyaman untuk servis motor semua ny...
1    servis buruk baud cvt ilang satu pertma servis...
2    terima kasih atas tanggapannya\nsesuai janji s...
3    pelayanan nya baik mekanik nya ramah\nfasalita...
4    seal oli mesin cvt jeboldari pertigaan lampu m...

```

Gambar 3.49 Kode Menghilangkan Angka beserta Hasilnya

## 8. *Removing Stopwords*

Dilanjutkan dengan menghilangkan konjungsi dan kata penghubung untuk mempermudah pemrosesan pada tahap selanjutnya. Gambar 3.50 memaparkan kode beserta hasil menjalankan kode.

```

# Removing Stopwords

stop_words = set(stopwords.words('indonesian'))

ad['Ulasan'] = ad['Ulasan'].apply(lambda x: ' '.join([word for word in x.split() if word not in stop_words]))
ad['Ulasan'].head(30)

0    bengkel yg nyaman servis motor ny ramah humble...
1    servis buruk baud cvt ilang pertma servis ga m...
2    terima kasih tanggapannya sesuai janji rating ...
3    pelayanan nya mekanik nya ramah fasalitas nya ...
4    seal oli mesin cvt jeboldari pertigaan lampu m...

```

Gambar 3.50 Kode Menghilangkan Stopwords beserta Hasilnya

## 9. *Translating Text*

Tahap ini perlu dilakukan karena pada tahap modeling akan menggunakan TextBlob untuk melakukan klasifikasi. Oleh karena itu, kalimat ulasan diterjemahkan menjadi Bahasa Inggris. Gambar 3.51 memaparkan kode beserta hasil menjalankan kode.

```
# Translating Text

def translate_text(text):
    try:
        return translator.translate(text, src='id', dest='en').text
    except:
        return text

ad['Translated_Text'] = ad['Ulasan'].apply(translate_text)

ad[['Ulasan', 'Translated_Text']].head(10)
```

	Ulasan	Translated_Text
0	bengkel yg nyaman servis motor ny ramah humble...	Comfortable Workshop Motorcycle Services Frien...
1	servis buruk baud cvt ilang pertma servis ga m...	Bad service Bolt CVT The first service does no...
2	terima kasih tanggapannya sesuai janji rating ...	Thank you for the response according to the pr...
3	pelayanan nya mekanik nya ramah fasalitas nya ...	the mechanical service is friendly fasality, g...
4	seal oli mesin cvt jeboldari pertigaan lampu m...	SEAL OLI MACHINE CVT Jebold from the T -juncti...

Gambar 3.51 Kode Menerjemahkan Ulasan beserta Hasilnya

### 10. *Dropping Missing Value*

Setelah seluruh tahap praproses untuk ulasan dilakukan, menghapus nilai yang hilang baru dilakukan. Tahap ini dilakukan di akhir karena ada beberapa praproses yang menghilangkan seluruh kata sehingga menghasilkan baris dengan nilai yang hilang. Oleh karena itu, setelah seluruh tahapan praproses ulasan dilakukan, penghapusan nilai yang hilang baru dilakukan. Gambar 3.52 memaparkan kode beserta hasil menjalankan kode.

```
# Dropping Missing Comments

ad = ad.dropna(subset=['Ulasan'])
ad = ad[~ad['Ulasan'].str.strip().eq('')]

ad = ad.reset_index(drop=True)
ad.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37 entries, 0 to 36
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   URL              29 non-null     object
1   Nama Pelanggan  29 non-null     object
2   Waktu Ulasan    29 non-null     object
3   Ulasan          37 non-null     object
4   Translated_Text 37 non-null     object
5   Sentiment       37 non-null     object
6   Polarity        37 non-null     float64
7   Subjectivity    37 non-null     float64
dtypes: float64(2), object(6)
memory usage: 2.4+ KB
```

Gambar 3.52 Kode Menghapus Nilai yang Hilang

### 11. *Add New Column*

Tahap ini berfungsi untuk menambahkan kolom nomor ID yang berfungsi untuk menjadi *unique key* yang dapat difilter saat dibangun

dashboard. Nomor ID diambil melalui kode bengkel PT XYZ yang sudah ada pada master data PT XYZ dan dimasukkan ke dalam data. Gambar 3.53 memaparkan kode dan hasil menjalankan kode.

```

row_name = len(ad)
ad["ID"] = 1035
ad.head()

```

	URL	Nama Pelanggan	Waktu Ulasan	Ulasan	Translated_Text	Sentiment	Polarity	Subjectivity	ID
0	https://h3.googleusercontent.com/a/ACg8ocU6...	Ganto Roherton	2 tahun lalu	bengkel yg nyaman servis motor ny ramah humble...	Comfortable Workshop Motorcycle Services Fries...	Positive	0.135000	0.660000	1035
1	https://h3.googleusercontent.com/a/ACg8ocUem4...	Fajar Selindin	sobulan lalu	servis buruk baid cvt ilang pertama servis ga m...	Bad service Bolt CVT The first service does no...	Negative	-0.225000	0.500000	1035
2	https://h3.googleusercontent.com/a/ACg8ocUwgd...	Kusaini Ritandi	7 bulan lalu	terima kasih tanggapannya sesuai jeng rating ...	Thank you for the response according to the pr...	Neutral	0.000000	0.000000	1035
3	https://h3.googleusercontent.com/a/ACg8ocKRAK...	Nur Hasanah	2 tahun lalu	pelayanan nya mekanik nya ramah fasilitas nya ...	the mechanical service is friendly facility, g...	Positive	0.295833	0.783333	1035
4	https://h3.googleusercontent.com/a/ALV-UjYok...	Bondan Yasin Arrudi	satuhun lalu	seal oli mesin cvt jebolde ri perigaan lampu m...	SEAL OIL MACHINE CVT Jebold from the T-junct...	Positive	0.165667	0.416667	1035

Gambar 3.53 Kode Menambahkan Kolom Baru beserta Hasilnya

## 12. Merge Excel

Data-data ulasan yang sudah dipraproses akan digabungkan menjadi 1 file agar dapat digunakan pada tahap modeling untuk diklasifikasikan. Penggabungan file dilakukan di akhir karena di awal praproses, dapat dilihat bahwa kolom-kolom yang di *scraping* memiliki nama yang berbeda dan pada file lainnya, setiap nama kolom berbeda. Oleh karena itu, dilakukan praproses mandiri terlebih dahulu lalu digabungkan menjadi satu. Total file berjumlah 11 untuk bengkel PT XYZ dan bengkel non-PT XYZ, yaitu 5 file bengkel PT XYZ dan 6 file bengkel non-PT XYZ. Gambar 3.54 memaparkan kode untuk menggabungkan file dan Gambar 3.55 memaparkan hasil penggabungan yang sudah dilakukan.

```

folder_path = "C:/Users/Sheera/OneDrive - Universitas Multimedia Nusantara/Dokumen/Project/Sentiment
file_list = [f for f in os.listdir(folder_path) if f.endswith('.xlsx')]

df_list = []

for file in file_list:
    file_path = os.path.join(folder_path, file)
    df = pd.read_excel(file_path)
    df_list.append(df)

merged_df = pd.concat(df_list, ignore_index=True)
merged_df.to_excel('Complete_Sentiment.xlsx', index=False)

print("Successfully merged", len(file_list), "files.")
Successfully merged 11 files.

```

Gambar 3.54 Kode untuk Menggabungkan File

```
df = pd.read_excel('Complete_Sentiment.xlsx')
df['ID'].value_counts()

ID
17835    385
8045     385
10935    375
8747     226
JM       173
YM       160
ZP       150
GT       105
1035     36
MM        35
Name: count, dtype: int64
```

Gambar 3.55 Hasil Penggabungan Data

### 3.2.3.4 Modeling

Tahap pemodelan dilakukan menggunakan kode yang sama seperti proyek sentimen X. Pertama dibuat *dictionary* yang sudah mengklasifikasikan *emoji* menjadi sentimen tertentu kemudian diklasifikasikan kembali menggunakan *TextBlob* untuk mengetahui polaritas dan subjektivitas kalimat ulasan. Gambar 3.56 memaparkan kode untuk membuat *function* yang akan mengklasifikasikan kalimat ulasan dan Gambar 3.57 memaparkan hasil yang didapatkan dari menjalankan kode.

```
emoji_sentiment_dict = {
    "grinning_face": "Positive", "grinning_face_with_big_eyes": "Positive", "grinning_face_with_smiling_eyes": "Positive",
    "beaming_face_with_smiling_eyes": "Positive", "grinning_squinting_face": "Positive", "grinning_face_with_sweat": "Positive",
    "rolling_on_the_floor_laughing": "Positive", "smiling_face_with_smiling_eyes": "Positive", "smiling_face_with_heart_eyes": "Positive",
    "smiling_face_with_hearts": "Positive", "smiling_face_with_sunglasses": "Positive", "thumbs_up": "Positive",
    "pensive_face": "Negative", "crying_face": "Negative", "loudly_crying_face": "Negative", "angry_face": "Negative",
    "enraged_face": "Negative", "thumbs_down": "Negative", "broken_heart": "Negative", "neutral_face": "Neutral",
    "expressionless_face": "Neutral", "face_without_mouth": "Neutral"
}

def classify_sentiment(text):
    if not isinstance(text, str):
        return 'Neutral'

    # Classifying Desejized Sentiment
    words = text.split()
    emoji_matches = [word.strip('.') for word in words if word.strip('.') in emoji_sentiment_dict]

    if emoji_matches:
        sentiments = [emoji_sentiment_dict[emo] for emo in emoji_matches]
        return max(set(sentiments), key=sentiments.count)

    # TextBlob Sentiment
    blob = TextBlob(text)
    polarity = blob.sentiment.polarity

    if polarity > 0.1:
        return 'Positive'
    elif polarity < -0.1:
        return 'Negative'
    else:
        return 'Neutral'

ad['Sentiment'] = ad['Translated_Text'].apply(classify_sentiment)
ad['Polarity'] = ad['Translated_Text'].apply(lambda x: TextBlob(x).sentiment.polarity)
ad['Subjectivity'] = ad['Translated_Text'].apply(lambda x: TextBlob(x).sentiment.subjectivity)

ad[['Translated_Text', 'Sentiment', 'Polarity', 'Subjectivity']]
```

Gambar 3.56 Kode Function Klasifikasi

	Translated_Text	Sentiment	Polarity	Subjectivity
0	Comfortable Workshop Motorcycle Services Frien...	Positive	0.135000	0.660000
1	Bad service Bolt CVT The first service does no...	Negative	-0.225000	0.500000
2	Thank you for the response according to the pr...	Neutral	0.000000	0.000000
3	the mechanical service is friendly fasality, g...	Positive	0.295833	0.783333
4	SEAL OLI MACHINE CVT Jebold from the T -juncti...	Positive	0.166667	0.416667
5	worrying about recommended my motorbike heavy ...	Negative	-0.200000	0.500000

Gambar 3.57 Hasil Menjalankan Kode Function

### 3.2.3.5 Evaluation

Evaluasi dilakukan dengan uji akurasi klasifikasi menggunakan data *true label* yang dibuat menggunakan *library* yang dibangun dari hasil diskusi dengan *supervisor*. Gambar 3.58 memaparkan *library* yang sudah dibangun.

```

classifying.py
C:\Users> Sheera > OneDrive - Universitas Multimedia Nusantara > Dokumen > Project > Sentiment > classifying.py > ...
1 positive_keywords = ['puas', 'memuaskan', 'ramah', 'nyaman', 'dingin', 'coffee', 'minuman', 'bersih', 'bagus',
2   'pelayanan', 'oke', 'mantap', 'membantu', 'cepat', 'thumbs up', 'gratis', 'ruang', 'tunggu',
3   'baik', 'kopi', 'sukses', 'lengkap', 'service', 'profesional', 'mekanik', 'teliti', 'progress',
4   'best', 'cepat', 'ok']
5 negative_keywords = ['kecewa', 'mahal', 'panas', 'mengecewakan', 'parah', 'jelek', 'rusak', 'buruk', 'baut', 'berat',
6   'lama']
7
8 def classify_sentiment(text):
9   text = text.lower()
10  pos_score = sum(word in text for word in positive_keywords)
11  neg_score = sum(word in text for word in negative_keywords)
12
13  if pos_score > neg_score:
14    return 'Positive'
15  elif neg_score > pos_score:
16    return 'Negative'
17  else:
18    return 'Neutral'

```

Gambar 3.58 Library Klasifikasi Data True Label

Kata-kata yang digolongkan menjadi positif dan negatif diambil berdasarkan kata-kata yang sering disebutkan di dalam komentar. Untuk mengetahui kata-kata yang sering disebutkan, digunakan Word Cloud untuk memvisualisasikan kata-kata tersebut. Gambar 3.59 memaparkan visualisasi Word Cloud.





akurasi yang didapatkan adalah 79% dan dinilai cukup baik. Namun, nilai presisi untuk sentimen negatif belum menunjukkan angka yang cukup tinggi, yaitu 27%. Hal ini disebabkan oleh jumlah data sentimen negatif yang tidak seimbang dengan sentimen positif dan netral.

```
print("Accuracy:", accuracy_score(df['True_Label'], df['predicted']))
print("\nClassification Report:\n", classification_report(df['True_Label'], df['predicted']))

Accuracy: 0.7950372208436725

Classification Report:
              precision    recall  f1-score   support

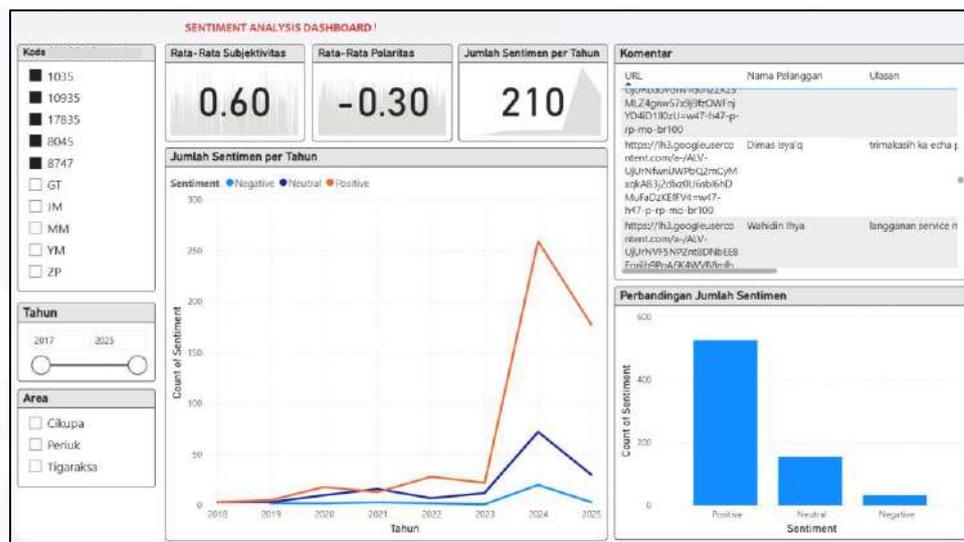
 Negative     0.27     0.67     0.38         33
  Neutral     0.55     0.63     0.59        429
  Positive     0.91     0.84     0.87       1553

 accuracy              0.80         2015
 macro avg              0.61         2015
 weighted avg           0.81         2015
```

Gambar 3.62 Nilai Akurasi dan Classification Report

### 3.2.3.6 Deployment

Data sentimen bengkel PT XYZ dan bengkel non-PT XYZ yang sudah digabungkan digunakan untuk membangun *dashboard*. Gambar 3.63 memaparkan *dashboard* analisis sentimen yang difilter untuk bengkel PT XYZ dan Gambar 3.64 memaparkan *dashboard* yang difilter untuk bengkel non-PT XYZ.



Gambar 3.63 Dashboard Bengkel PT XYZ



Gambar 3.64 Dashboard Bengkel Non-PT XYZ

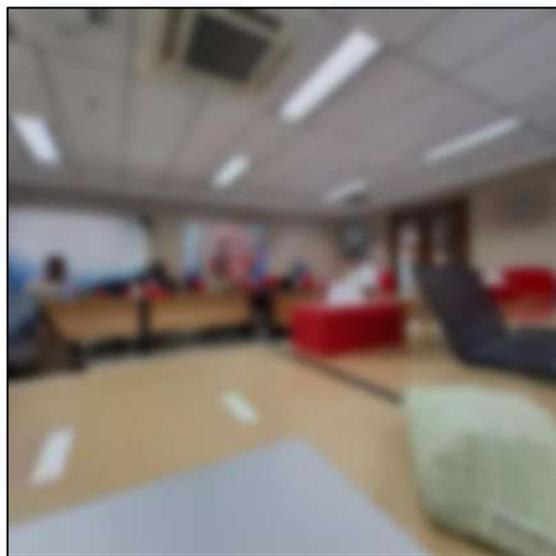
Kesimpulan dari analisis sentimen untuk *dashboard* yang difilter untuk bengkel non-PT XYZ adalah pelanggan lebih suka mengunjungi bengkel non-PT XYZ karena biaya yang lebih murah, servis yang lebih cepat, dan harga suku cadang yang lebih murah. Terkadang bengkel-bengkel ini menyediakan servis yang jauh lebih lengkap daripada bengkel PT XYZ yang ada sehingga pelanggan lebih cenderung sering pergi kesana. Terlepas dari sentimen positif, sentimen negatif yang diperoleh cenderung sama dengan bengkel PT XYZ. Sebagian besar masyarakat merasa kesal karena pelayanan yang terkadang buruk dan tidak ramah, serta seringkali dibiarkan menunggu motor yang diservis cukup lama. Finalisasi pembangunan *dashboard* dipresentasikan kepada direktur untuk menjelaskan wawasan yang didapatkan.

Wawasan yang diperoleh dari pembangunan *dashboard* tersebut adalah seluruh bengkel PT XYZ memiliki sentimen yang positif dari masyarakat baik dalam hal melakukan servis maupun melayani pelanggan. Pelayanan pelanggan yang diberikan, antara lain menyediakan ruang tunggu yang dingin beserta tempat dan kursi yang bersih, menyediakan makanan ringan dan minuman seperti air, kopi, dan teh, serta kemudahan proses pembuatan janji menggunakan aplikasi agar

dapat menentukan jadwal untuk melakukan servis. Selain itu, pelayanan menggunakan aplikasi sebagai pengingat untuk melakukan servis sangat membantu karena diberikan notifikasi untuk melakukan servis dalam periode waktu tertentu.

Terlepas dari sentimen positif yang didapatkan, terdapat sentimen negatif yang dibutuhkan pengawasan lebih melihat polaritas yang bernilai negatif, yaitu beberapa terdapat beberapa bengkel PT XYZ yang memiliki pelayanan buruk karena tidak ramah dan sopan. Selain itu, terdapat beberapa pelanggan yang risih dan tidak suka saat diajarkan menggunakan aplikasi perusahaan untuk membuat janji temu. Hal ini dapat diatasi dengan kembali mengingatkan kepada bengkel PT XYZ untuk meningkatkan pelayanan yang lebih baik, serta menggunakan pendekatan yang lebih ramah dan nyaman saat memperkenalkan aplikasi untuk membuat janji servis serta menjelaskan alasan peralihan sistem dari pembuatan janji manual menjadi lebih modern.

Pembangunan *dashboard* yang sudah dilakukan akan dipresentasikan bersama dengan penjelasan proyek besar perusahaan. Gambar 3.65 memaparkan *meeting* yang dilakukan untuk melakukan presentasi.



Gambar 3.65 Presentasi Hasil Proyek dengan CMO

Penjelasan saat presentasi dilakukan secara terperinci untuk menjelaskan setiap komponen yang ada pada *dashboard*, baik secara teori maupun ulasan yang ada. Berikut adalah penjelasan yang dibuat saat melakukan presentasi.

- a. *Dashboard* pertama menunjukkan hasil analisis sentimen pelanggan terhadap bengkel PT XYZ di Cikupa, Periuk, dan Tigaraksa dengan rata-rata 210 komentar per tahun. Rata-rata subjektivitas mencapai 0,60, yang berarti sebagian besar komentar mengandung opini pribadi atau ekspresi emosi. Sementara itu, rata-rata polaritas sebesar -0,30 mencerminkan bahwa sentimen cenderung ke arah negatif, meskipun distribusi bar chart menunjukkan bahwa komentar positif mendominasi jumlah komentar keseluruhan, diikuti oleh netral dan sedikit komentar negatif. Grafik tren sentimen per tahun memperlihatkan lonjakan tajam komentar positif pada tahun 2024, yang kemudian menurun di 2025.
- b. Sebaliknya, *dashboard* kedua menampilkan hasil analisis pada bengkel non-PT XYZ dengan rata-rata 21 komentar. Rata-rata subjektivitas jauh lebih rendah, yaitu 0,25, yang berarti bahwa komentar yang dianalisis cenderung lebih objektif atau berbasis fakta. Polaritas yang bernilai positif menandakan bahwa komentar lebih banyak bernada positif secara keseluruhan. Distribusi sentimen memperlihatkan dominasi komentar positif dan sedikit komentar negatif. Namun, jumlah data yang lebih kecil menyebabkan analisis yang tidak seimbang. Komentar seperti “mantap”, “mantaap”, dan deskripsi teknis ringan seperti “caliper piston rcb” mencerminkan opini yang lebih singkat, langsung, dan cenderung kurang emosi dibandingkan dengan komentar dalam *dashboard* bengkel PT XYZ.
- c. Dapat disimpulkan kedua data memiliki beberapa perbedaan yang cukup kritis dalam melakukan pembangunan *dashboard* dan mengukur nilai sentimen menggunakan polaritas dan subjektivitas, yaitu skala data. *Dashboard* yang di-filter untuk bengkel PT XYZ memiliki data yang lebih banyak dengan emosi pengguna yang lebih ekspresif,

meskipun cenderung negatif dari sisi polaritas sehingga perlu dimonitor lebih lanjut untuk pelayanan yang lebih baik. Sementara itu, *dashboard* yang di-*filter* untuk bengkel non-PT XYZ, walaupun memiliki polaritas positif, mencerminkan komentar yang cenderung datar, singkat, dan kurang emosional. Secara teori [12], tingginya subjektivitas pada *dashboard* bengkel PT XYZ berarti keterlibatan emosional pelanggan yang kuat, sedangkan polaritas negatif menandakan adanya kritik atau ketidakpuasan yang tersirat.

### **3.2.4 Menerapkan metode CRISP-DM untuk Analisis Sentimen terhadap Komentar Youtube mengenai “Vario Jamet”**

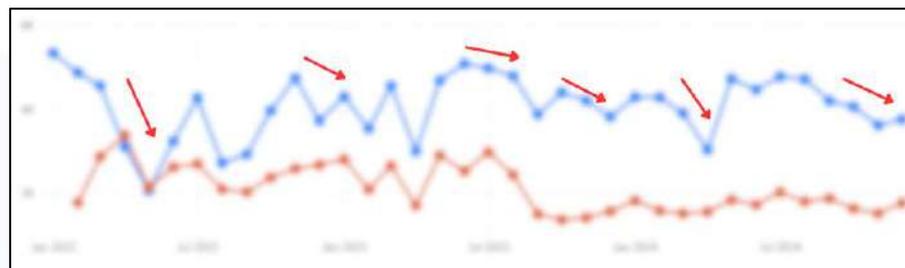
Proyek analisis sentimen mengenai “Vario Jamet” dimulai dari sebutan yang seringkali muncul di media sosial karena adanya tren memodifikasi motor menjadi “Thailook”. Gambar 3.66 memaparkan contoh motor “Thailook”. Analisis sentimen dimulai karena sebagian besar masyarakat yang berkomentar di media sosial mengatakan bahwa sebutan “Vario Jamet” bukan untuk motornya, melainkan untuk orang yang melakukan modifikasi terhadap motor. Setelah ditelusuri lebih lanjut, istilah “Vario Jamet” merupakan sebutan untuk orang-orang yang suka melakukan modifikasi pada motor dengan mengganti knalpot, ban, dan cat pada badan motor menyerupai “Thailook”. Namun, yang menjadi permasalahan adalah masyarakat yang melakukan modifikasi ini seringkali menjadi penyebab isu saat berlalu lintas, seperti suara knalpot yang berisik, modifikasi yang berlebihan, warna yang tidak selaras, dan warna lampu yang nyentrik. Oleh karena itu, sebutan ini memungkinkan menjadi faktor penyebab turunnya penjualan motor Vario. Permasalahan ini secara spesifik akan dijelaskan pada tahap *business understanding*.



Gambar 3.66 Motor Thailook [13]

### 3.2.4.1 Business Understanding

Proyek dimulai dengan masalah yang ditemukan dari menganalisis data penjualan motor Vario pada *dashboard* yang sudah dibangun dan ditemukan penurunan penjualan pada kuartal 2 dan kuartal 4 selama 3 tahun. Namun, untuk proyek ini, dispesifikan hanya untuk tahun 2024 agar pengumpulan data lebih mudah dan terkini. Gambar 3.67 memaparkan grafik garis yang mengalami penurunan penjualan pada tahun 2024.



Gambar 3.67 Grafik Penjualan Motor Vario

Hal ini menjadi masalah baru yang perlu dilakukan pendekatan dari sisi sentimen. Salah satu sebutan yang seringkali diutarakan oleh masyarakat adalah “Vario Jamet” dan dikhawatirkan dapat menurunkan penjualan motor Vario akibat sentimen ini. Oleh karena itu, dimulai proyek analisis sentimen terhadap komentar masyarakat mengenai “Vario Jamet” yang diambil dari YouTube. Platform ini dipilih karena memiliki ulasan dan komentar yang banyak mengenai “Vario Jamet”

melalui video-video ulasan motor dan bongkar rangka. Untuk memulai proyek ini, dilakukan diskusi dengan *supervisor* untuk menjelaskan proyek yang akan dilakukan. Gambar 3.68 memaparkan diskusi yang dilakukan di kantor.



Gambar 3.68 Diskusi dengan Supervisor

Diskusi dengan *supervisor* juga menghasilkan tujuan diadakan proyek ini, yaitu mengerti tujuan sebutan “Vario Jamet”, mengetahui pengaruh “Vario Jamet” terhadap penurunan penjualan, dan

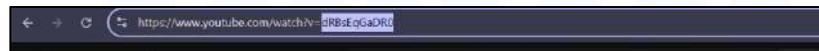
#### **3.2.4.2 Data Understanding**

Pengumpulan data dimulai dengan memilih video yang akan dilakukan *scraping*. Pemilihan video memiliki kriteria yang harus dipenuhi, yaitu berada pada lini waktu yang sama dengan data yang akan dianalisis, memiliki komentar yang banyak, dan merupakan akun yang sering mengulas mesin dan rangka motor. Video YouTube yang digunakan berasal dari salah satu kreator konten YouTube otomotif yang sudah melakukan *review* motor dan rangka bertahun-tahun [14]. Gambar 3.69 memaparkan video yang akan dilakukan *scraping*.

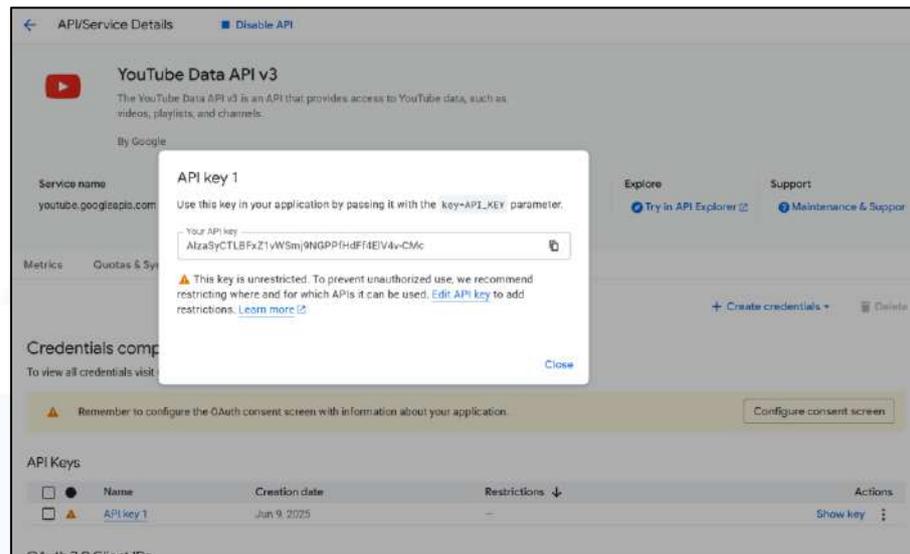


Gambar 3.69 Video Youtube yang Dilakukan Scraping

Video yang sudah dipilih akan dilakukan *scraping* menggunakan API YouTube yang didapatkan dari YouTube. Metode ini bersumber dari video YouTube yang menjelaskan tahapan melakukan *scraping* hingga mengekspor file [15]. Tahap pertama adalah membuka Google Colab dan mengaktifkan Gemini untuk membantu membuat kode *scraping* menggunakan YouTube API dan kunci API pribadi. API video diperoleh dari pranala video setelah tanda “=” dan API pribadi diambil dari YouTube yang diaktifkan melalui *credentials* untuk YouTube. Gambar 3.70 memaparkan kunci API video YouTube dan Gambar 3.71 memaparkan kunci API pribadi yang didapatkan dari YouTube.



Gambar 3.70 API Video yang akan Di-scraping



Gambar 3.71 Kunci API Youtube untuk Scraping

Kunci API yang sudah didapatkan akan dimasukkan ke dalam kode yang sudah dibuat oleh Gemini pada Google Colab. Gambar 3.72 memaparkan kode untuk membuat *function scraping* setiap komentar yang ada pada video.

```
# Replace with your YouTube API Key
DEVELOPER_KEY = "AlzaSyCTLBFxZ1vWSmj9NGPPfHdF4EIV4v-CMc"

# Replace with the video ID of the YouTube video
VIDEO_ID = "dR8sEqGaDR0"

def get_comments(video_id, part="snippet", max_results=500):
    youtube = build("youtube", "v3", developerKey=DEVELOPER_KEY)
    comments = []
    nextPageToken = None

    while len(comments) < max_results:
        try:
            # Retrieve comment thread using the youtube.commentThreads().list() method
            response = youtube.commentThreads().list(
                part=part,
                videoId=video_id,
                textFormat="plainText",
                maxResults=min(100, max_results - len(comments)), # Get up to 100 or remaining
                pageToken=nextPageToken
            ).execute()

            for item in response["items"]:
                comment_text = item["snippet"]["topLevelComment"]["snippet"]["textDisplay"]
                likes = item["snippet"]["topLevelComment"]["snippet"]["likeCount"]
                published_at = item["snippet"]["topLevelComment"]["snippet"]["publishedAt"] # Extract published date
                comments.append({"comment": comment_text, "num_of_likes": likes, "published_at": published_at})

            nextPageToken = response.get("nextPageToken")
            if not nextPageToken:
                break # No more pages

        except HttpError as error:
            print(f"An HTTP error occurred: {error}")
            return None

    return comments
```

Gambar 3.72 Kode untuk Scraping Komentar pada Video

Komentar pada video YouTube akan di-*scraping* menggunakan *function* tersebut lalu diekspor menggunakan *function* yang dipaparkan pada Gambar 3.73.

```
def main():
    # Get comments from the video
    comments = get_comments(VIDEO_ID)

    if comments:
        # Create a pandas dataframe from the comments list
        df = pd.DataFrame(comments)

        # Sort dataframe by number of likes in descending order
        df = df.sort_values(by=['num_of_likes'], ascending=False)

        # Print a preview of the first 10 rows
        print(df.head(10))

        # Export dataframe to a CSV file named "comments.csv"
        df.to_csv("comments.csv", index=False)
    else:
        print("Error: Could not retrieve comments from video.")

if __name__ == "__main__":
    main()
```

Gambar 3.73 Function Ekspor Dataset

Data komentar yang sudah di-*scraping* diekspor menjadi file *Excel* yang dapat diunduh dan diolah. Data tersebut dilakukan praproses dan diolah menggunakan *Python* sama seperti tahapan pada proyek-proyek sebelumnya. Setelah diimpor, dilakukan eksplorasi dan dilanjutkan dengan praproses untuk membersihkan data. Gambar 3.74 memaparkan data yang sudah di-*scraping*.

	A	B	C	D	E
1	comment	num_of_li	published_at		
2	Bukan karena motor nya tapi pengguna nya, padahal dulu pengennya vario, tapi pas liat	227	2024-11-14T08:09:45Z		
3	satu kelemahan nya, build quality nya.	121	2024-11-14T08:19:45Z		
4	Vario selama gak di list ini:	116	2024-11-14T08:59:59Z		
5	Kemampuan otak manusia dalam memodifikasi bisa anda nilai sendiri arahnya	75	2024-11-14T08:02:08Z		
6	jamet itu modifikasi yg nggak berkonsep, ya misalnya modifikasi yg banyak part yg nggak j	35	2024-11-14T09:51:45Z		
7	Selagi mereka gk pake knalpot bacot mereka yang bikin budi ( budeg dikit ) + lepas spakbo	32	2024-11-14T08:13:28Z		
8	mau aja kalian percaya istilah bginianðŸ”„, semua motor kalo modifnya masuk ya keren ke	25	2024-11-14T11:04:38Z		
9	5:26	22	2024-11-14T10:19:19Z		
10	Motor apapun , kalo udah modif aneh-aneh, knalpot berisik, dan diceper-ceper ya jametle	19	2024-11-14T08:32:39Z		
11	Uda metik, 150cc lagi pakai knalpot racing = JAMET	19	2024-11-14T08:34:02Z		
12	Tpi jujur vario 125/150 keren bgt design nya. ga kegedean ga kekecilan jg. Dlu smpet pngel	17	2024-11-14T10:28:29Z		
13	Setuju sangat. Yang bukan motor jamet itu honda spacy. Ada komunitasnya & modifnya ju	17	2024-11-14T12:01:15Z		
14	banyak fenyeponk nya	11	2024-11-14T13:08:11Z		
15	Perasaan ane doang apa gimana yah user vario kebanyakan arogan, entah itu bapak bap	11	2024-11-14T09:10:26Z		
16	Motor-motor Honda itu ada aja segmennya	11	2024-11-14T09:50:14Z		
17	Saya pake vario 125 led cbs dari 2018 sampai sekarang, sama sekali gk pernah	11	2024-11-14T16:16:08Z		
18	Motornya sih kagak salah cuman user nya aja yang ngeselin	10	2024-11-14T08:55:59Z		
19	Udah di invasi sama jamet ngaberss, helm njs, knalpot suara mencreet kencing kaga.	9	2024-11-14T10:54:27Z		
20	cc kecil 150 cc sok2an make kenalpot racing..emang aga laen anak muda di negeri ini	9	2024-11-14T09:09:52Z		
21	Gw punya 125 led old, dari beli sampe sekarang masih standar aja, gak suka modif lebih	8	2024-11-15T02:11:57Z		
22	Vario paling enak yang techno 125 old sama beat karbu body plastik masih tebal, rangka d	8	2024-11-14T14:00:34Z		
23	Semua akan jamet pada waktunya, ðŸ”„	7	2024-11-14T10:28:59Z		
24	Stop lamp teks memang jamet sih itu stoplamp apa jam masjid	6	2024-11-14T10:51:11Z		
25	Sekarang Ngabers Vario itu Ibarat User Mobil Diesel Klo Mereka Ngebacotnya "M3	5	2024-11-14T09:06:01Z		
26	Gw mlh baru ambil CBR 150 2005 mulus mumpung murah belum di goreng.	5	2024-11-14T08:39:30Z		
27	Sebenarnya modif sah2 aja kalo eksekusinya bener, tapi masalahnya sebagian besar ngabi	5	2024-11-14T10:10:56Z		
28	Pengguna hongda bilek ketika punya beat/vario ciri khas nya yaitu spion dilepas, velg jadi	5	2024-11-14T11:00:45Z		
29	Sebenarnya kebanyakan modifikasi vario itu oke oke aja. Tapi gw ada dua hal yang gak	5	2024-11-14T08:48:36Z		
30	3:05 over power, ya gw setuju min, gw beli vario trus 3 bulan kemudian gw jual, buat gw yi	5	2024-11-14T08:45:27Z		

Gambar 3.74 Dataset Hasil Scraping Komentar

Jumlah komentar yang berhasil untuk di-*scraping* adalah 397 baris beserta dengan waktu menulis komentar dan jamnya.

### 3.2.4.3 Data Preparation

Persiapan data dilakukan dengan melakukan klasifikasi sarkasme secara manual. Klasifikasi ini bertujuan untuk mengetahui kepada siapa kalimat sarkasme dituju. Dalam kasus ini, sebutan “Vario Jamet” belum diketahui untuk orang, motor, keduanya, atau bukan keduanya. Oleh karena itu, dilakukan klasifikasi terlebih dahulu secara manual agar konteks dari komentar dapat dimengerti. Klasifikasi ini juga diperlukan hasil yang akurat karena akan digunakan untuk melakukan prediksi sentimen menggunakan input dan output di tahap evaluasi.

Pelabelan manual memang membutuhkan waktu yang cukup lama, tetapi menurut beberapa artikel, pelabelan ini penting agar model dapat dilatih dan mempelajari data dengan baik untuk memprediksi sentimen [16], [17], [18]. Selain itu, beberapa artikel juga mengatakan bahwa klasifikasi dapat dilakukan dengan lebih efisien waktu menggunakan *ChatGPT* dan survei terhadap beberapa orang yang

mengerti mengenai topik yang akan dianalisis. Hasil pencarian metode ini didiskusikan dengan *supervisor* dan diputuskan untuk melakukan pelabelan dengan *ChatGPT* dan pelabelan dengan mentor. Hasil pelabelan diakumulasikan dan dituliskan menjadi “Orang”, “Motor”, “Keduanya”, dan “Tidak Keduanya”. Gambar 3.75 memaparkan dataset hasil pelabelan yang sudah diakumulasikan.

comment	published_sarkas
Ngabers vario emg JAMET semuanya !!	2024-11-1 Orang
Yg jamet org2nya, gw liat2 yg di video keren2 aja, kecuali bannya cacingan + knalpot tembak muka. Kalo gw asal knalpot ge	2024-11-1 Orang
jujur aja,gue 2 bulan lalu beli cash vario 125 new,hbis itu modif pasang biled(cuma buat penerangan,ane motoran biasa,gak	2024-11-1 Keduanya
Gw pake Vario KZR, modif ga banyak, itupun cuma looks dan Laser D2, bahkan lampu utama masih dibiarin bohlam, karen	2024-11-1 Orang
Kalau ngeliat modifan vario rasanya paling bener modifan (PLAT KT)	2024-11-1 Orang
Bahas HONDA CB atau GL100 bang.	2024-11-1 Orang
Ya sekarang banyak sih,gw pemakai vegar r balok std jreng,dan gw sering dikatakan sama user" vario dengan alasan yang r	2024-11-1 Orang
alhamdulillah gua udah jual vario	2024-11-1 Motor
Gw paling gak suka motor matic pake knalpot racing (mau 100 - 250cc)	2024-11-1 Motor
Iya, pengen bnget nyiram bensin lalu di bakar, vario itu udah bagus standaran bgtu, malah di kasih ban cacing dn knalpot ml	2024-12-1 Orang
Vario selama gak di list ini:	2024-11-1 Orang
Vario 2015 trbaik sih elegan dapet, sexy dapet, misterius dapet, sporty dapet.	2024-11-1 Tidak Keduanya
Gak sir mke knalpot samlong"an mending sih knalpot yg da brand kenal kyk Shinjiro atau arm	2024-11-1 Orang
Kegunaan kenalpot kek gitu cuman buat banjir doang biar gak ke masukan air tuh knalpot.	2024-11-1 Orang
Paling sebel liat vario n aerox pake knalpot bronk..	2024-11-1 Orang
Yang gua kesel lebih ke user nya sih, apalagi kalo yang dah pake knalpot samlong kalo gak std racing, kalo udah di jalan nge	2024-11-1 Orang
Bagus aja gitu kalau di liat nya apa lagi yang "Mbeerr"..... Anying anying gakak gua	2024-11-1 Tidak Keduanya
Ya lu liat aja tiap hari di jalan pasti ada yang begitu. Lampu biled led warna warni dangdut bis malem, bore up knalpot bron	2024-11-1 Orang
Motor-motor Honda itu ada aja segmennya	2024-11-1 Keduanya

Gambar 3.75 Hasil Pelabelan dengan ChatGPT dan Manual

Dataset yang sudah dilabel akan digunakan untuk melakukan praproses. Praproses pada proyek ini menggunakan *function* yang merangkap seluruh metode praproses teks. Pertama, membuat kode untuk memanggil *stopwords* dan *stemmer* Bahasa Indonesia dan Inggris. Kode ini akan digunakan saat menjalankan *function* praproses dalam menghapus konjungsi atau kata sambung dan menghilangkan imbuhan pada kata. Kemudian dilanjutkan dengan membuat kode *function* *detect\_language* yang berfungsi untuk mendeteksi bahasa yang digunakan saat membuat komentar dan membuat kolom baru untuk memasukkan bahasa yang digunakan. Gambar 3.76 memaparkan kode pemanggilan *stemmer* dan *function* *detect\_language*.

```

# Stopwords dan stemmer untuk Bahasa Inggris dan Bahasa Indonesia
stopwords_id = set(StopWordRemoverFactory().get_stop_words())
stopwords_en = set(stopwords.words('english'))

stemmer_id = SastrawiStemmerFactory().create_stemmer()
stemmer_en = PorterStemmer()

# Function Language Detection
def detect_language(text):
    if not isinstance(text, str) or not text.strip():
        return 'id'
    try:
        clean_text = re.sub(r'^[a-zA-Z\s]', '', text)
        if not clean_text.strip():
            return 'id'
        lang = detect(clean_text)
        return lang if lang in ['en', 'id'] else 'id'
    except:
        return 'id'

```

Gambar 3.76 Kode Memanggil Stemmer dan Function Detect Language

Setelah membuat *function* deteksi bahasa, dilanjutkan dengan membuat *function* `preprocess_text` untuk melakukan praproses pada teks. Tahapan praproses yang dilakukan di *function* adalah memberikan spasi, menghilangkan tanda baca, mengubah huruf menjadi huruf kecil, tokenisasi, memilih *stopwords* dan *stemmer* berdasarkan bahasa, dan menghapus *stopwords* dan melakukan *stemming*. Gambar 3.77 memaparkan *function* untuk melakukan praproses.

```

# Function untuk Text Preprocessing
def preprocess_text(text, lang):
    if not isinstance(text, str) or not text.strip():
        return ''

    # Lowercasing, Spacing, dan Removing Punctuation
    text = text.replace("_x000D_", " ").replace("\n", " ").replace("\r", " ")
    text = re.sub(r"http\S+", "", text)
    text = re.sub(r"[^\w\s]", "", text)
    text = re.sub(r"\s+", " ", text).lower().strip()

    # Tokenisasi
    tokens = word_tokenize(text)

    # Pilih stopwords dan stemmer berdasarkan bahasa
    stopwords_set = stopwords_en if lang == 'en' else stopwords_id
    stemmer = stemmer_en if lang == 'en' else stemmer_id

    # Hapus stopwords & stemming
    final_tokens = []
    for word in tokens:
        if word.isalpha() and word not in stopwords_set:
            if lang == 'en':
                final_tokens.append(stemmer.stem(word))
            else:
                final_tokens.append(stemmer.stem(word)) # stemmer_id sudah string

    return " ".join(final_tokens)

```

Gambar 3.77 Function Praproses Data

Setelah *function* dibuat, dataset akan dipraproses dengan menjalankan kode *function* Gambar 3.78 memaparkan kode untuk menjalankan *function* yang dibuat beserta hasil praproses.

```
df['lang'] = df['comment'].apply(detect_language)
df['cleaned'] = df.apply(lambda row: preprocess_text(row['comment'], row['lang']), axis=1)
df['cleaned'].head(30)

0          ngabers vario emg jamet semua
1  yg jamet gw yg video aja ban cacing knalpot te...
2  jujur aja gue bulan lalu beli cash vario new h...
3  gw pake vario kZR modif ga banyak itu cuma loo...
4  kalau ngeliat modifan vario rasa paling bener ...
5          bahas honda cb bang
```

Gambar 3.78 Kode Menjalankan Function Praproses beserta Hasilnya

Tahap praproses dilanjutkan dengan melakukan *encoding* terhadap tujuan sarkasme. Gambar 3.79 memaparkan kode *encoding* dan Gambar 3.80 memaparkan hasilnya.

```
variable_map = {'Orang': 0, 'Motor': 1, 'Keduanya': 2, 'Tidak Keduanya': 3}
df['classified'] = df['sarkas'].map(variable_map)
```

Gambar 3.79 Kode Encoding

	comment	published_at	sarkas	lang	cleaned	classified
0	Ngabers vario emg JAMET semuanya !!	2024-11-15T14:55:13Z	Orang	id	ngabers vario emg jamet semua	0
1	Yg jamet org2nya, gw liat2 yg di video keren2 ...	2024-11-14T13:19:45Z	Orang	id	yg jamet gw yg video ajo ban cacing knalpot te...	0
2	Jujur aja, gue 2 bulan lalu beli cash vario 125...	2024-11-15T17:24:00Z	Keduanya	id	jujur aja gue bulan lalu beli cash vario new h...	2
3	Gw pake Vario KZR, modif ga banyak, itupun cum...	2024-11-15T15:57:52Z	Orang	id	gw pake vario kZR modif ga banyak itu cuma loo...	0
4	Kalau ngeliat modifan vario rasanya paling ben...	2024-11-17T05:15:32Z	Orang	id	kalau ngeliat modifan vario rasa paling bener ...	0

Gambar 3.80 Hasil Encoding

### 3.2.4.4 Modeling

Pemodelan dilakukan dengan *Sentence Bidirectional Encoder Representations* atau SBERT. SBERT adalah model yang dapat mengubah kalimat menjadi numerik yang merepresentasikan arti kalimat secara kontekstual [19]. Penggunaan SBERT berfungsi untuk mengerti konteks kalimat yang dibuat sehingga model dapat menilai dan mengklasifikasikan teks dengan benar. Pemodelan dengan SBERT akan dilakukan bersama dengan algoritma SVM untuk memprediksi input kalimat untuk menguji keakurasian model. Gambar 3.81 memaparkan kode yang digunakan untuk melatih data terhadap model SBERT dengan mengubah kalimat menjadi angka.

```

# Load data dan split
texts = df['cleaned'].tolist()
labels = df['classified'].tolist()
train_texts, val_texts, train_labels, val_labels = train_test_split(texts, labels, test_size=0.2, random_state=42)

# Siapkan data dengan InputExample
train_data = [InputExample(texts=[text, text], label=label) for text, label in zip(train_texts, train_labels)]

# DataLoader
train_loader = DataLoader(train_data, shuffle=True, batch_size=16)

# Load model SBERT
model = SentenceTransformer('paraphrase-multilingual-MiniLM-L12-v2')

# Menunjukkan loss function
loss = losses.SoftmaxLoss(
    model=model,
    sentence_embedding_dimension=model.get_sentence_embedding_dimension(),
    num_labels=len(set(labels))
)

# Fine-tuning model
model.fit(
    train_objectives=[(train_loader, loss)],
    epochs=3, # Bisa ubah jumlah epoch
    warmup_steps=10,
    output_path='sbert_model_sederhana'
)

```

Gambar 3.81 Kode Melatih SBERT dengan Data

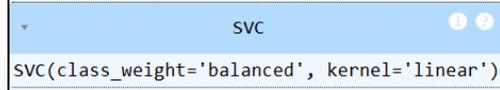
Data yang sudah dilatih disimpan ke 'sbert\_model\_sederhana' dan digunakan untuk melakukan prediksi bersama dengan algoritma SVM. Algoritma SVM digunakan karena memiliki kecocokan dengan model SBERT dalam melakukan klasifikasi dan prediksi kalimat. *Support Vector Machine* (SVM) adalah algoritma *supervised machine learning* yang digunakan untuk melakukan klasifikasi dan prediksi [20]. Dalam hal ini, dilakukan klasifikasi dan prediksi. Gambar 3.82 memaparkan kode untuk melatih kalimat menggunakan model SVM.

```

# Encode kalimat pelatihan
train_embeddings = model.encode(train_texts.tolist(), convert_to_tensor=False)

# Latih SVM
svm = SVC(kernel='linear', class_weight='balanced')
svm.fit(train_embeddings, y_train)

```



Gambar 3.82 Training Data dengan SVM

Data yang sudah dilatih akan digunakan untuk dilakukan penyeimbangan kelas karena label 1 (orang) dan 2 (keduanya) memiliki jumlah yang sedikit sehingga model tidak belajar dengan maksimal pola kalimat yang berlabel tersebut. Oleh karena itu, digunakan SMOTE

untuk menyeimbangkan kelas. SMOTE digunakan karena memiliki performa yang baik dalam menyeimbangkan kelas dan mengurangi *overfitting* data yang dapat mengurangi akurasi [20]. Hasil SMOTE ditunjukkan di bagian bawah di mana kelas yang tidak seimbang menjadi seimbang. Gambar 3.83 memaparkan kode memanggil model SBERT dan menerapkan SMOTE pada data.

```
# Memanggil klasifikasi yang sudah di encode
y_encoded = df['classified']

# Memanggil Model yang sudah dilatih dengan SBERT
model = SentenceTransformer('fine_tuned_sbert_model')
X_embeddings = model.encode(df['cleaned'], show_progress_bar=True)

# Melakukan penyeimbangan kelas dengan SMOTE
smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X_embeddings, y_encoded)

print("Before SMOTE:", dict(pd.Series(y_encoded).value_counts()))
print("After SMOTE:", dict(pd.Series(y_resampled).value_counts()))

Batches: 100% ██████████ 13/13 [00:04<00:00, 5.62it/s]
Before SMOTE: {0: 198, 3: 119, 1: 56, 2: 19}
After SMOTE: {0: 198, 2: 198, 1: 198, 3: 198}
```

Gambar 3.83 Kode Memanggil Model SBERT dan SMOTE

Data yang sudah seimbang dibagi menjadi 20% data tes dan 80% data latih menggunakan `train_test_split`. Kemudian data dilatih dengan model SVM dan diuji menggunakan *accuracy score* dan *confusion matrix*. Gambar 3.84 memaparkan kode yang digunakan untuk melatih data.

```

# Split data 80-20
X_train, X_test, y_train, y_test = train_test_split(
    X_resampled, y_resampled, test_size=0.2, random_state=42
)

# Melatih Model SVM
svm = SVC()
svm.fit(X_train, y_train)

# Predict & Evaluate
y_pred = svm.predict(X_test)

# Accuracy Score
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", accuracy)

# Classification Report
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# Confusion Matrix
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))

```

Gambar 3.84 Kode SVM, Accuracy Score, dan Confusion Matrix

### 3.2.4.5 Evaluation

Pemodelan yang dilakukan pada tahap sebelumnya diperoleh nilai akurasi dan *confusion matrix* untuk mengevaluasi performa model yang sudah dilatih. Gambar 3.85 memaparkan hasil evaluasi pemodelan yang dilakukan.

```

Akurasi: 0.8616352201257862

Classification Report:
              precision    recall  f1-score   support

     0         0.75         0.81         0.78         37
     1         0.92         0.83         0.87         41
     2         0.90         1.00         0.95         37
     3         0.88         0.82         0.85         44

 accuracy          0.86         0.86         0.86         159
 macro avg         0.86         0.86         0.86         159
 weighted avg      0.86         0.86         0.86         159

Confusion Matrix:
[[30  1  3  3]
 [ 4 34  1  2]
 [ 0  0 37  0]
 [ 6  2  0 36]]

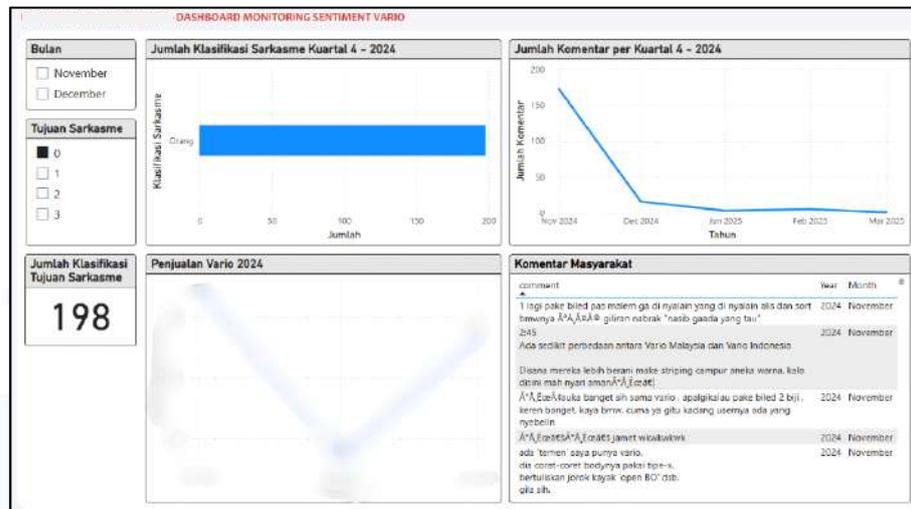
```

Gambar 3.85 Nilai Akurasi dan Confusion Matrix

Nilai akurasi menunjukkan nilai yang cukup tinggi, yaitu 86% dengan hasil *confusion matrix* yang cukup maksimal dengan nilai presisi untuk setiap kelas di atas 70%. Namun, perlu diperhatikan lebih untuk kelas 0 dan diperbanyak datanya untuk melatih model.

#### **3.2.4.6 Deployment**

Hasil klasifikasi menunjukkan bahwa memang sarkasme yang ditujukan adalah kepada motor yang melakukan modifikasi dan mengganggu masyarakat lain yang sedang berlalu lintas. Sebagian besar mengeluh karena suara knalpot yang berisik saat di jalan dengan bunyi yang nyaring. Namun, ada beberapa masyarakat juga yang akhirnya tidak ingin membeli motor Vario bahkan membencinya karena pendapat masyarakat lainnya mengenai “Vario Jamet”. Hipotesis awal saat melakukan analisis ini adalah karena sebutan “Vario Jamet”, maka penjualan motor Vario akan menurun. Saat dibandingkan dengan data penjualan, analisis ini tidak sejalan dengan data penjualan karena saat sarkas sebagian besar sarkas yang diberikan adalah kepada orangnya bukan motornya. Oleh karena itu, analisis sarkasme diberhentikan sampai disini dan dilanjutkan mencari faktor lain yang mempengaruhi penurunan penjualan pada kuartal 2 dan kuartal 4. Gambar 3.86 memaparkan *dashboard* yang sudah dibangun.



Gambar 3.86 Dashboard Sentimen Motor Vario

Hasil proyek ini digunakan untuk menjadi landasan dalam melakukan pemasaran motor Vario. Contohnya adalah dapat mempromosikan motor Vario sebagai motor “Vario Jamet” dan dipasarkan bagi masyarakat yang suka melakukan modifikasi pada motor. Untuk menarik pelanggan yang sering melakukan modifikasi terhadap motor, dapat dibuat konten pemasaran yang sering menyebutkan sebutan “Vario Jamet”. Selain itu, dari sisi servis dan modifikasi, bengkel PT XYZ dapat menyediakan paketan untuk melakukan servis mesin dan rangka sekaligus perawatan modifikasi yang sudah dilakukan. Dari sisi penjualan motor, *dealer* dapat menyediakan paketan untuk melakukan modifikasi saat pelanggan akan membeli motor.

Model SVM yang sudah dilatih menggunakan SBERT akan digunakan untuk melakukan prediksi terhadap komentar yang diberikan. Uji coba melakukan prediksi sangat sederhana menggunakan Python input dan output. Gambar 3.87 memaparkan kode yang digunakan beserta hasilnya. Hasil uji coba cukup akurat karena dapat mengklasifikasikan komentar “motor vario produk gagal” menjadi sarkasme untuk motor dan bukan orang, keduanya, dan tidak keduanya.

```

# Load model yang sudah di-fine-tune
model = SentenceTransformer('fine_tuned_sbert_model')

# Kalimat yang ingin diklasifikasikan
new_sentence = "vario produk gagal"
embedding = model.encode([new_sentence])

# Contoh prediksi dengan SVM (jika kamu latih SVM terpisah)
prediction = svm.predict(embedding)
print("Prediksi kelas:", prediction)

Prediksi kelas: [1]

```

Gambar 3.87 Hasil Prediksi Sentimen

### 3.2.5 Analisis Faktor Penurunan Penjualan Motor Vario

Landasan dari kelanjutan analisis ini adalah mencari faktor yang mempengaruhi penurunan penjualan motor Vario. Hal ini bertujuan untuk menggali lebih dalam pengaruh eksternal yang mempengaruhi terjadinya penurunan penjualan motor Vario. Beberapa faktor yang seringkali diutarakan oleh departemen PDCA saat menganalisis *dashboard* yang memonitor penjualan dan *dashboard customer care* adalah hari raya idul fitri, data promosi dan diskon, dan pengeluaran rangka motor Vario. Oleh karena itu, keempat faktor ini akan dibuat menjadi 3 *dashboard* yang berbeda untuk menganalisis lebih lanjut faktor yang paling mempengaruhi penurunan penjualan motor Vario pada kuartal 2 dan kuartal 4.

Pembangunan *dashboard* dilakukan dengan mengumpulkan data dari basis data menggunakan *query*. Gambar 3.88, Gambar 3.89, Gambar 3.90, dan Gambar 3.91 memaparkan *query* yang digunakan untuk mendapatkan data dari basis data menggunakan laptop mentor.

```

SELECT
  m.mhnfaktur_tgl,
  m.nomor_mesin,
  m.cust_code,
  m.cust_jenis,
  m.jual_typebyr,
  m.dp_actual,
  m.tenor_actual,
  m.cicilan_actual,
  CASE
    WHEN LOWER(m.mtr_type_desc) LIKE '%adv150%' OR LOWER(m.mtr_type_desc)
    LIKE '%adv160%' THEN 'ADV Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%beat%' AND LOWER(m.mtr_type_desc) NOT
    LIKE '%beat street%' THEN 'BeAT Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%beat street%' THEN 'Beat Street'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cb 150r%' OR LOWER(m.mtr_type_desc)
    LIKE '%cb150r%' OR LOWER(m.mtr_type_desc) LIKE '%cb150%' THEN 'CB 150R'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cb 150x%' OR LOWER(m.mtr_type_desc)
    LIKE '%cb150x%' THEN 'CB150X'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cbr 150%' OR LOWER(m.mtr_type_desc)
    LIKE '%cbr150%' THEN 'CBR 150 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cbr%' AND LOWER(m.mtr_type_desc) NOT
    LIKE '%cbr 150%' THEN 'CBR Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%crf150%' THEN 'CRF 150 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%crf250%' OR LOWER(m.mtr_type_desc)
    LIKE '%crf 250%' THEN 'CRF 250 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%ct125%' THEN 'CT125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%em1%' THEN 'EM1 E'
    WHEN LOWER(m.mtr_type_desc) LIKE '%forza%' THEN 'FORZA'
    WHEN LOWER(m.mtr_type_desc) LIKE '%genio%' THEN 'GENIO'
    WHEN LOWER(m.mtr_type_desc) LIKE '%monkey%' THEN 'MONKEY 125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%pcx%' THEN 'PCX'
    WHEN LOWER(m.mtr_type_desc) LIKE '%revo%' THEN 'Revo Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%scoopy%' THEN 'Scoopy'
    WHEN LOWER(m.mtr_type_desc) LIKE '%sonic%' THEN 'Sonic 150 R'
    WHEN LOWER(m.mtr_type_desc) LIKE '%st125%' THEN 'ST125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%stylo%' THEN 'Stylo 160'
    WHEN LOWER(m.mtr_type_desc) LIKE '%super cub%' THEN 'Super Cub C125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%supra%' THEN 'Supra X Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario 125%' THEN 'Vario 125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario 160%' THEN 'Vario 160'
    WHEN LOWER(m.mtr_type_desc) LIKE '%verza%' THEN 'Verza Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%blade%' THEN 'Blade Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario%' AND LOWER(m.mtr_type_desc)
    NOT LIKE '%vario 125%' AND LOWER(m.mtr_type_desc) NOT LIKE '%vario 160%'
  
```

Gambar 3.88 Query untuk Mendapatkan Faktur Mei 2024-2025

```

THEN 'Vario Series'
  WHEN LOWER(m.mtr_type_desc) LIKE '%spacy%' THEN 'Spacy'
  WHEN LOWER(m.mtr_type_desc) LIKE '%mega pro%' THEN 'Mega'
  ELSE 'Big Bike'
END AS Grup_Series,
  inv.motor_tahun AS mtr_tahun,
  inv.mtr_hrgjual,
  inv.jual_discount,
  c.cust_sumber,
  c.cust_agama,
  c.agama_stnk,
  c.test_ride,
  c.penghasilan
FROM mohon_fakturs m
JOIN invoice_details inv ON inv.nomor_mesin = m.nomor_mesin
JOIN customers c ON c.cust_code = m.cust_code
WHERE YEAR(m.mhnfaktur_tgl) = 2024
UNION ALL

```

Gambar 3.89 Lanjutan Query Sebelumnya

```

SELECT
  m.mhnfaktur_tgl,
  m.nomor_mesin,
  m.cust_code,
  m.cust_jenis,
  m.jual_tipebyr,
  m.dp_actual,
  m.tenor_actual,
  m.cicilan_actual,
  CASE
    WHEN LOWER(m.mtr_type_desc) LIKE '%adv150%' OR LOWER(m.mtr_type_desc)
    LIKE '%adv160%' THEN 'ADV Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%beat%' AND LOWER(m.mtr_type_desc) NOT
    LIKE '%beat street%' THEN 'BEAT Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%beat street%' THEN 'Beat Street'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cb 150r%' OR LOWER(m.mtr_type_desc)
    LIKE '%cb150r%' OR LOWER(m.mtr_type_desc) LIKE '%cb150%' THEN 'CB 150R'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cb 150x%' OR LOWER(m.mtr_type_desc)
    LIKE '%cb150x%' THEN 'CB150X'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cbr 150%' OR LOWER(m.mtr_type_desc)
    LIKE '%cbr150%' THEN 'CBR 150 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%cbr%' AND LOWER(m.mtr_type_desc) NOT
    LIKE '%cbr 150%' THEN 'CBR Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%crf150%' THEN 'CRF 150 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%crf250%' OR LOWER(m.mtr_type_desc)
    LIKE '%crf 250%' THEN 'CRF 250 Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%ct125%' THEN 'CT125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%em1%' THEN 'EM1 E'
    WHEN LOWER(m.mtr_type_desc) LIKE '%forza%' THEN 'FORZA'
    WHEN LOWER(m.mtr_type_desc) LIKE '%genio%' THEN 'GENIO'
    WHEN LOWER(m.mtr_type_desc) LIKE '%monkey%' THEN 'MONKEY 125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%pcx%' THEN 'PCX'
    WHEN LOWER(m.mtr_type_desc) LIKE '%revo%' THEN 'Revo Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%scoopy%' THEN 'Scoopy'
    WHEN LOWER(m.mtr_type_desc) LIKE '%sonic%' THEN 'Sonic 150 R'
    WHEN LOWER(m.mtr_type_desc) LIKE '%st125%' THEN 'ST125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%stylo%' THEN 'Stylo 160'
    WHEN LOWER(m.mtr_type_desc) LIKE '%super cub%' THEN 'Super Cub C125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%supra%' THEN 'Supra X Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario 125%' THEN 'Vario 125'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario 160%' THEN 'Vario 160'
    WHEN LOWER(m.mtr_type_desc) LIKE '%verza%' THEN 'Verza Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%blade%' THEN 'Blade Series'
    WHEN LOWER(m.mtr_type_desc) LIKE '%vario%' AND LOWER(m.mtr_type_desc)
    NOT LIKE '%vario 125%' AND LOWER(m.mtr_type_desc) NOT LIKE '%vario 160%'
    THEN 'Vario Series'
  
```

Gambar 3.90 Query untuk Mendapatkan Data Faktor Sebelum Mei 2024

```

    WHEN LOWER(m.mtr_type_desc) LIKE '%spacy%' THEN 'Spacy'
    WHEN LOWER(m.mtr_type_desc) LIKE '%mega pro%' THEN 'Mega'
    ELSE 'Big Bike'
  END AS Grup_Series,
  oinv.motor_tahun AS mtr_tahun,
  oinv.mtr_hrgjual,
  oinv.jual_discount,
  c.cust_sumber,
  c.cust_agama,
  c.agama_stnk,
  c.test_ride,
  c.penghasilan
FROM mohon_fakturs m
JOIN old_invoice_details oinv ON oinv.nomor_mesin = m.nomor_mesin
JOIN customers c ON c.cust_code = m.cust_code
WHERE YEAR(m.mhnfaktur_tgl) = 2024
  AND NOT EXISTS (
    SELECT 1 FROM invoice_details inv WHERE inv.nomor_mesin = m.nomor_mesin
  );

```

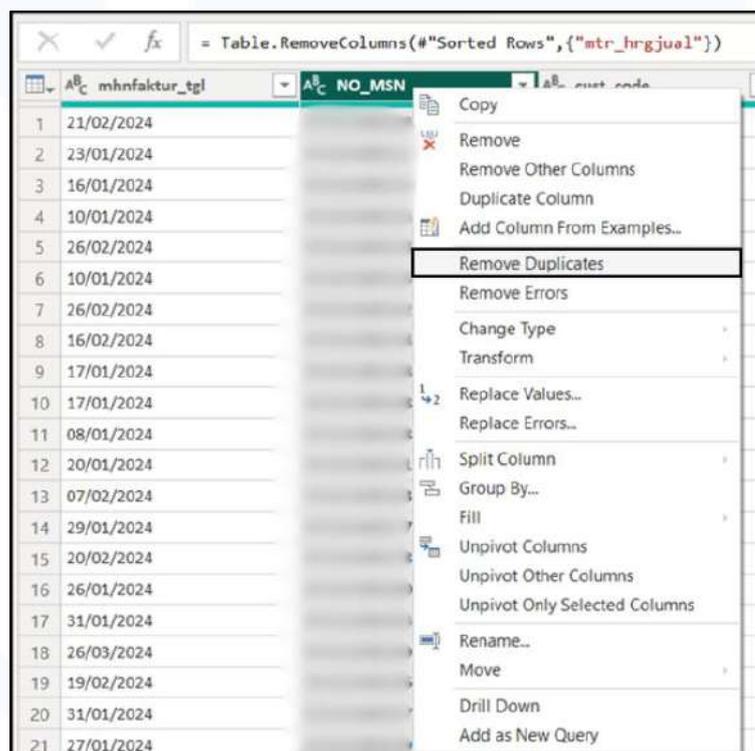
Gambar 3.91 Lanjutan Query Sebelumnya

Data faktur yang dihasilkan dari kueri akan diekspor dalam format CSV dan dimasukkan ke Power BI untuk divisualisasikan. Data faktur akan dihubungkan dengan data pelanggan yang didapatkan dari pemahaman data di minggu pertama. Penghubungan kedua data ini berfungsi untuk menyambungkan data pelanggan dengan data faktur agar dapat dibuat

visualisasi dengan data yang sudah terhubung. Kedua data ini kemudian dihubungkan dengan relasi *one to one* melalui nomor mesin. Namun, ketika dihubungkan, terdapat banyak duplikat karena harga motor yang berbeda, tetapi di dalam satu transaksi yang sama. Gambar 3.92 memaparkan contoh penduplikasian data karena harga motor. Oleh karena itu, kolom harga motor dihapus karena tidak diperlukan dalam visualisasi dan analisis serta data yang duplikat akan dihilangkan. Gambar 3.93 memaparkan data yang akan dihilangkan duplikasinya.

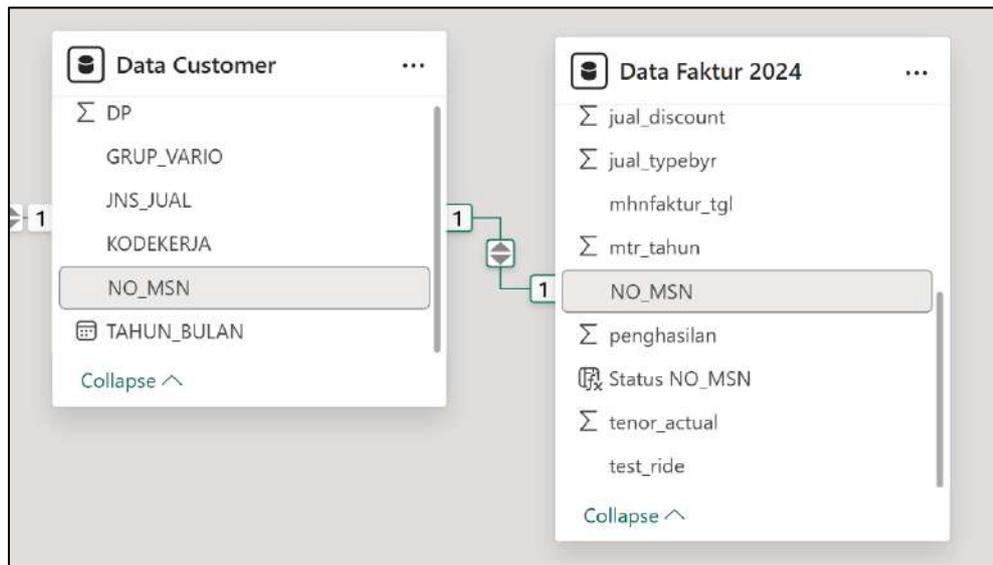
17/01/2024	E	148	C	190
17/01/2024	Ei	148	C	190

Gambar 3.92 Data Duplikat karena Harga Motor



Gambar 3.93 Menghilangkan Duplikasi

Data yang sudah bersih dapat dihubungkan dengan relasi *one to one*. Hasil relasi dipaparkan pada Gambar 3.94.



Gambar 3.94 Relasi *One to One* Data Faktur dan Data Pelanggan

### 3.2.5.1 Dashboard Data Penjualan saat Hari Raya Idul Fitri

*Dashboard* data penjualan saat hari Raya Idul Fitri adalah *dashboard* yang berfungsi untuk memonitor penjualan motor Vario saat hari Raya Idul Fitri pada tahun 2024. Pembangunan *dashboard* ini dilandasi atas penurunan yang terus terjadi di kuartal 2 dari tahun ke tahun. Penurunan ini diasumsikan karena masyarakat lebih memilih untuk menyimpan uangnya untuk berlibur daripada membeli motor, tetapi saat sudah memperoleh THR, masyarakat baru membeli motor karena terjadi kenaikan penjualan di kuartal 3. Oleh karena itu, diperkirakan faktor ini merupakan penyebab penurunan penjualan di kuartal 2.

Pembangunan *dashboard* ini membutuhkan data penjualan dan data pelanggan yang dilengkapi dengan identitas pribadi berupa agama. Agama akan dijadikan sebagai filter untuk mengetahui pelanggan yang beragama Islam dan membeli motor Vario. Pembangunan *dashboard* dimulai dengan membangun relasi antara data penjualan, data pelanggan, dan data faktur motor. Gambar 3.95 memaparkan *dashboard* yang sudah dibangun.



Gambar 3.95 Dashboard Monitoring Hari Raya Idul Fitri

Grafik garis di bagian atas memaparkan penjualan motor Vario pada kuartal 3 yang ditandai dengan panah merah. Terdapat penurunan pada bulan April dan kenaikan di Bulan Mei. Grafik garis di bagian bawah memaparkan agama pelanggan di mana garis yang terletak cukup tinggi adalah agama Islam dan garis yang terletak di bawah adalah agama lainnya. Penurunan dan kenaikan penjualan memiliki pola yang sama dengan kenaikan dan penurunan pelanggan beragama Islam. Dengan ini, *dashboard* menunjukkan bahwa asumsi yang diberikan oleh divisi *customer care* benar karena terjadi penurunan saat hari Raya Idul Fitri dan kenaikan saat mendapatkan THR. Faktor ini dapat dinyatakan sebagai salah satu faktor yang paling berpengaruh dalam penurunan penjualan di kuartal 2 atau menjelang Hari Raya Idul Fitri.

### 3.2.5.2 *Dashboard* Perbandingan Data Penjualan dengan Data

#### Promosi dan Data Diskon

*Dashboard* perbandingan data penjualan dengan data promosi dan data data diskon adalah *dashboard* yang berfungsi untuk memonitor promosi dan diskon yang dilakukan serta pengaruhnya terhadap data penjualan. Pembangunan *dashboard* ini dilandasi oleh asumsi dari *supervisor* di mana promosi dan diskon yang dilakukan di kuartal 2 sedikit. Untuk membangun *dashboard* ini, dibutuhkan data penjualan,

data promosi, dan data diskon. Data diskon dan promosi diambil melalui basis data, tetapi secara langsung di-*query* oleh mentor karena basis data yang belum rapih sehingga dapat memperlambat proses pengumpulan data. Gambar 3.96 memaparkan hasil pembangunan *dashboard* data penjualan dengan data promosi dan data diskon.



Gambar 3.96 Dashboard Data Penjualan dengan Data Promosi dan Data Diskon

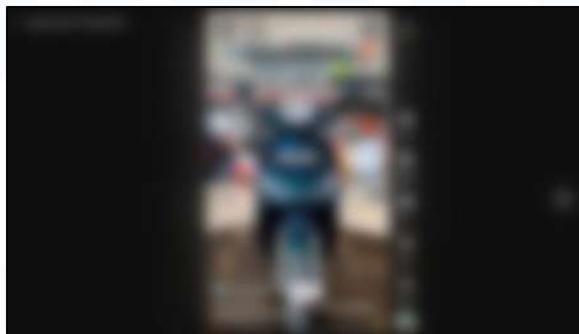
Grafik garis di kanan atas memaparkan penjualan motor Vario pada kuartal 2, grafik garis di kiri bawah memaparkan frekuensi frekuensi penjualan dengan diskon, dan grafik di kanan bawah memaparkan frekuensi promosi yang dilakukan selama 2024. Melalui visualisasi, dapat disimpulkan bahwa data pameran atau promosi tidak memiliki pengaruh terhadap penurunan penjualan. Data pameran memiliki hubungan yang berlawanan dengan data penjualan. Pada Bulan April, di mana pameran memiliki frekuensi yang banyak, penjualan menurun, sedangkan Mei di mana frekuensi pameran mulai menurun, penjualan melonjak tinggi. Data diskon memiliki sedikit keterkaitan dengan data penjualan karena memiliki pola yang cukup sama di mana pada bulan Maret, April, dan Mei di mana penjualan menurun lalu melonjak. Namun, data ini tidak dapat disimpulkan berpengaruh

terhadap penjualan motor karena pada kuartal lain polanya tidak mengikuti grafik penjualan.

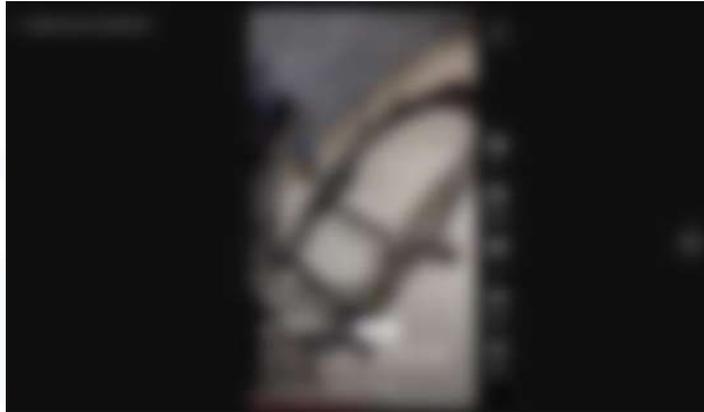
### 3.2.5.3 *Dashboard Analisis Sentimen terhadap Rangka motor Vario*

Dashboard ini dibangun menggunakan data klasifikasi sentimen yang dilakukan terhadap komentar Youtube mengenai rangka pada motor Vario. Masyarakat yang memiliki sentimen negatif karena mudah berkarat dan patah, tetapi belum pernah diteliti seberapa negatifnya sentimen ini terhadap penjualan motor Vario. Oleh karena itu, dilakukan analisis sentimen terhadap rangka motor Vario.

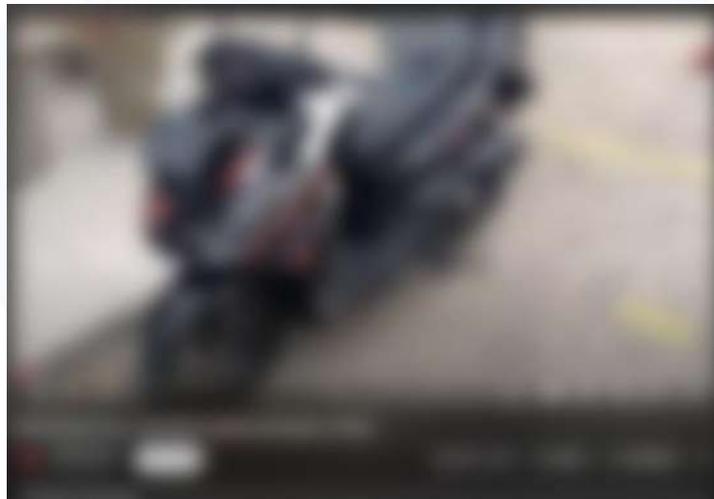
Analisis sentimen dilakukan dengan mengumpulkan data melalui beberapa video Youtube Shorts dan Video yang memiliki konten mengenai rangka motor Vario. Rangka motor Vario merupakan rangka keluaran tahun 2022 sehingga banyak video yang sudah dipublikasikan dari tahun-tahun tersebut. Oleh karena itu, video yang dicari akan difilter sekitar satu tahun atau beberapa bulan yang lalu agar komentar dapat difilter khusus pada kuartal 4 tahun 2024. Gambar 3.97, Gambar 3.98, Gambar 3.99, dan Gambar 3.100 merupakan video Youtube yang akan di-*scraping* komentarnya untuk diklasifikasikan secara manual. Klasifikasi secara manual memiliki alasan yang sama dengan proyek sebelumnya, yaitu perlunya dilakukan pemahaman konteks dan mengerti penyebab banyaknya sentimen tertentu di dalam kalimat.



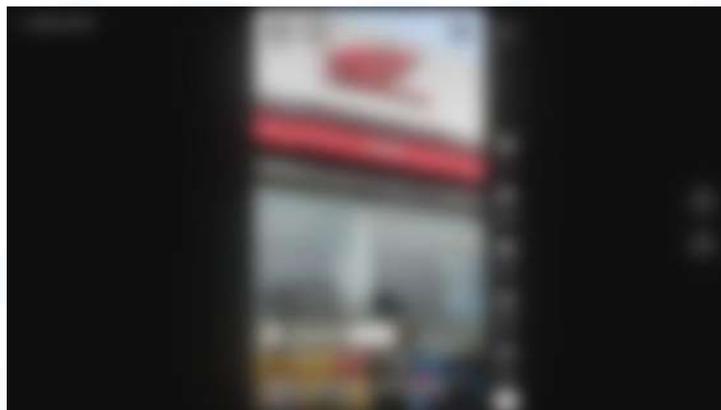
Gambar 3.97 Video 1 [21]



Gambar 3.98 Video 2 [22]



Gambar 3.99 Video 3 [23]



Gambar 3.100 Video 4 [24]

Pengumpulan data dilakukan sama seperti proyek analisis sentimen “Jamet”. Namun, untuk mempercepat pelabelan data, dilakukan pelabelan dengan membuat *library* Python dengan kata-kata

yang sering disebutkan di dalam komentar lalu digolongkan ke kata positif dan negatif. Kata-kata yang tidak tergolong di dalam keduanya akan diklasifikasikan secara netral. Gambar 3.101 memaparkan *library* yang sudah dibangun.

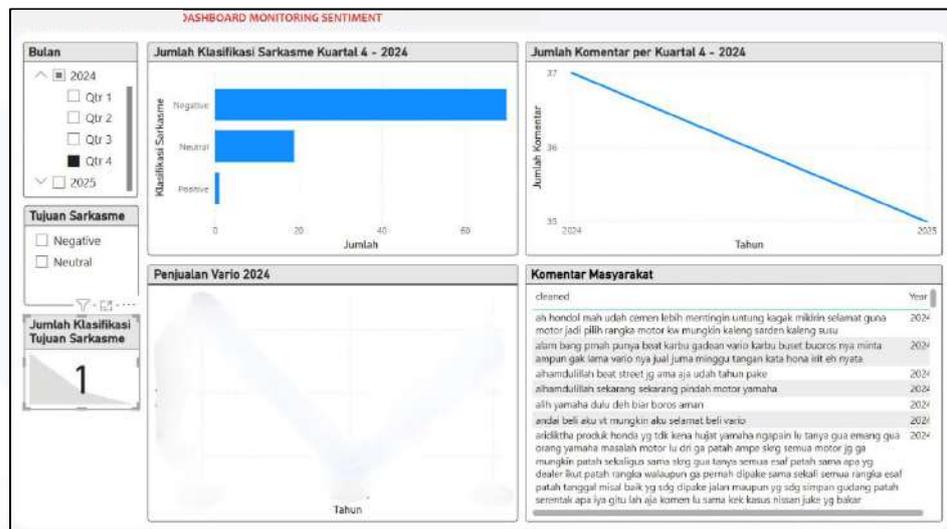
```

classifying.py
C:\Users> Sheera > OneDrive - Universitas Multimedia Nusantara > Dokumen > Project > List Project > classifying.py > ...
1 positive_keywords = ['bagus', 'suka', 'enak', 'nyaman', 'lelit']
2 negative_keywords = ['jelek', 'tipis', 'yamaha', 'krupuk', 'patah', 'boros', 'lama', 'beat', 'aerox', 'garansi', 'ringan', 'rangka',
3 'ragu', 'khawatir', 'sangat', 'esaf', 'karat', 'sampah', 'beban', 'sampah', 'ga', 'ak']
4
5 def classify_sentiment(text):
6     text = text.lower()
7
8     # Jika mengandung kata negatif langsung Negative
9     if any(word in text for word in negative_keywords):
10        return 'Negative'
11
12    # Jika tidak ada kata negatif tapi ada kata positif Positive
13    elif any(word in text for word in positive_keywords):
14        return 'Positive'
15
16    # Jika tidak ada kata positif maupun negatif Neutral
17    else:
18        return 'Neutral'
19

```

Gambar 3.101 Library Klasifikasi Sentimen

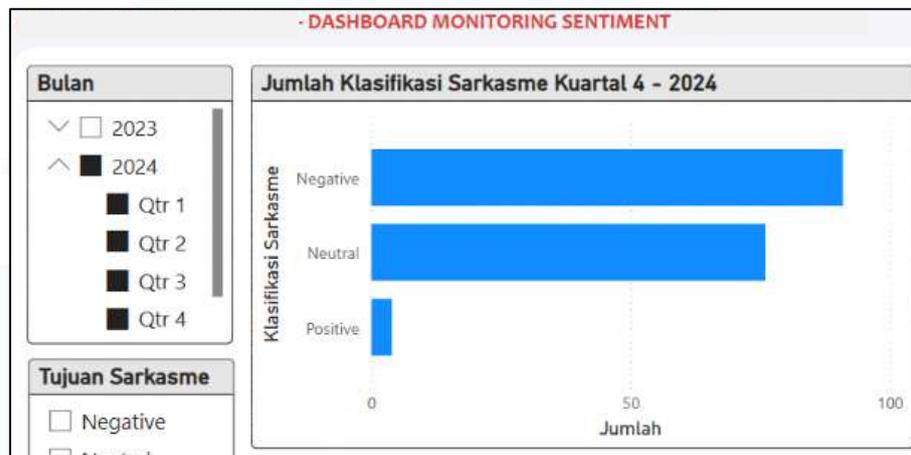
Setelah klasifikasi dilakukan dengan *library*, dilakukan pengecekan kembali terhadap data yang sudah diklasifikasikan. Data yang sudah diperiksa digunakan untuk membangun *dashboard* yang dipaparkan pada Gambar 3.102.



Gambar 3.102 Dashboard Monitoring Sentimen Rangka Motor Vario

*Dashboard* menunjukkan bahwa sentimen terhadap rangka motor Vario mayoritas negatif pada kuartal keempat. Namun, ketiga filter untuk tahun 2024 dinyalakan untuk seluruh kuartal, mayoritas sentimen

masyarakat untuk rangka motor Vario adalah negatif. Gambar 3.103 memaparkan filter 2024 dinyalakan untuk seluruh kuartal.



Gambar 3.103 Sentimen Rangka Motor Y 2024

*Dashboard* ini menunjukkan bahwa sentimen terhadap rangka motor Vario yang berpengaruh besar terhadap penurunan penjualan motor Vario. Mayoritas pelanggan yang sudah membeli motor Vario mengatakan bahwa rangka motor Vario sangat mudah rusak dan berkarat. Pemakaian baru 2-3 bulan rangka sudah berwarna coklat dan cat terkelupas. Selain itu, rangka sangat mudah patah karena tipis dan diibaratkan seperti “ember” dan “kaleng” atau biasa disebut “kerupuk” karena mudah hancur. Video-video ini juga akhirnya membuat masyarakat yang tadinya ingin membeli motor Vario menjadi beralih ke motor merk lain dikarenakan harganya yang mahal, tetapi diberikan rangka yang mudah rusak dan garansi perbaikan yang terkadang harus tetap membayar. Faktor ini menjadi salah satu faktor yang perlu ditindaklanjuti ke bagian pengembangan produk dan *sales* untuk meningkatkan kualitas dan penjualan produk.

### 3.3 Kendala yang Ditemukan

Pelaksanaan kerja magang pastinya tidak terlepas dari kendala-kendala yang dialami. Berikut adalah kendala-kendala yang seringkali dialami selama pengerjaan proyek.

1. Koneksi WIFI yang memblokir beberapa situs web sehingga tidak dapat mengakses beberapa platform yang perlu digunakan untuk mengerjakan proyek. Salah satu situs yang paling sering dikunjungi dalam pengerjaan proyek adalah Youtube dan situs ini termasuk sebagai situs yang diblokir. Hal ini diketahui saat melakukan *scraping* data untuk mengumpulkan sentimen.
2. Kesulitan memanipulasi data menggunakan DAX query pada Power BI. DAX query merupakan salah satu bahan pengajaran pada saat mata kuliah Big Data Analytics, tetapi penggunaan query ini sangat jarang digunakan dan apabila digunakan hanya *copy-paste*.
3. Kesulitan mendapatkan data perusahaan karena tidak memiliki akses terhadap basis data kantor dan harus meminta data atau meng-*query* sendiri di laptop mentor. Pada proyek terakhir, permintaan data sedikit sulit karena harus meminta data kepada mentor agar dapat membangun *dashboard*.
4. Kesulitan dalam memeriksa relasi dan keakurasian data karena basis data belum rapih dan tidak memiliki pemetaan yang benar untuk mengakses data tertentu. Kendala ini paling dirasakan saat membuat proyek terakhir karena data faktur memiliki banyak komponen dan banyak duplikasi karena data yang digabungkan memiliki isi yang berbeda-beda.

### 3.4 Solusi atas Kendala yang Ditemukan

Kendala-kendala yang dialami pastinya diperlukan solusi untuk tetap mengerjakan proyek yang sudah ditugaskan. Berikut adalah solusi-solusi yang ditemukan untuk mengatasi kendala yang ditemukan.

1. Mengatasi kendala koneksi WIFI dengan menggunakan hotspot dari handphone. Hal ini hanya dilakukan saat melakukan *scraping* data untuk mengerjakan proyek.
2. Mempelajari DAX *query* sendiri dengan bantuan *ChatGPT* dan mengerti istilah-istilah untuk membuat kolom baru dan menggabungkan data dari tabel yang berbeda.

3. Mengerjakan proyek lain terlebih dahulu yang tidak perlu menarik data dari basis data perusahaan, seperti *scraping* data yang tidak menggunakan basis data perusahaan dan melakukan praproses sambil menunggu data perusahaan.
4. Memeriksa tabel lain di dalam basis data yang memiliki keterkaitan dengan tabel. Salah satu cara yang dilakukan adalah pada proyek terakhir dengan memeriksa harga jual motor dari faktur dan harga jual motor dari data pelanggan. Apabila isi data berbeda, perlu ditelusuri lebih lanjut dan dicari penyebabnya dengan melibatkan tabel lain juga.