

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

3.1.1 Outlet *Online Food Delivery* (OFD)

Pada penelitian ini objek yang akan diteliti adalah mengenai ulasan pada restoran cepat saji. McDonald's adalah perusahaan besar dengan bentuk *Fast Food Restaurant* dimana menjual berbagai macam makanan dengan adat Eropa-Indonesia. Dengan masuknya McDonald's di Indonesia menjadikan menu-menu yang dijual sudah sesuai dengan *pallette* Indonesia. Untuk sekarang, McDonald's telah memiliki banyak sekali outlet. Outlet-outlet McDonald's memiliki penempatan geografis yang tepat, sehingga Summarecon Digital Center juga memiliki satu outlet McDonald's. Walaupun begitu, ulasan dari pembeli tetap tidak dapat dilihat dari pembelian offline. Hal itu dapat terjadi dikarenakan pada pembelian offline mengarah kepada pelayanan cepat saji, sehingga *phase* yang akan dijalankan begitu cepat. Atas alasan tersebut maka pembelian secara *online* menjadi sebuah opsi untuk memberikan komentar sekaligus mendapatkan proses pembelian layaknya pembelian secara *offline*.

GoFood adalah sebuah anak perusahaan dari Gojek yang menjual jasa pengiriman makanan. GoFood menyediakan jasa ojek dimana bertugas untuk membeli produk yang dipesan dan mengantarkan makanan tersebut kepada pembeli. GoFood dikatakan sebagai jasa pelayanan pembelian makanan *online* karena dari segi pembeli, mereka tidak harus pergi menuju tempat dan memesan secara langsung. Di Indonesia, GoFood adalah platform terbesar yang melayani jasa pelayanan pembelian makanan. Banyak pengguna yang telah memberikan komentar terhadap outlet yang dibeli pada platform GoFood. komentar terhadap outlet McDonald's merupakan data dari penelitian ini.

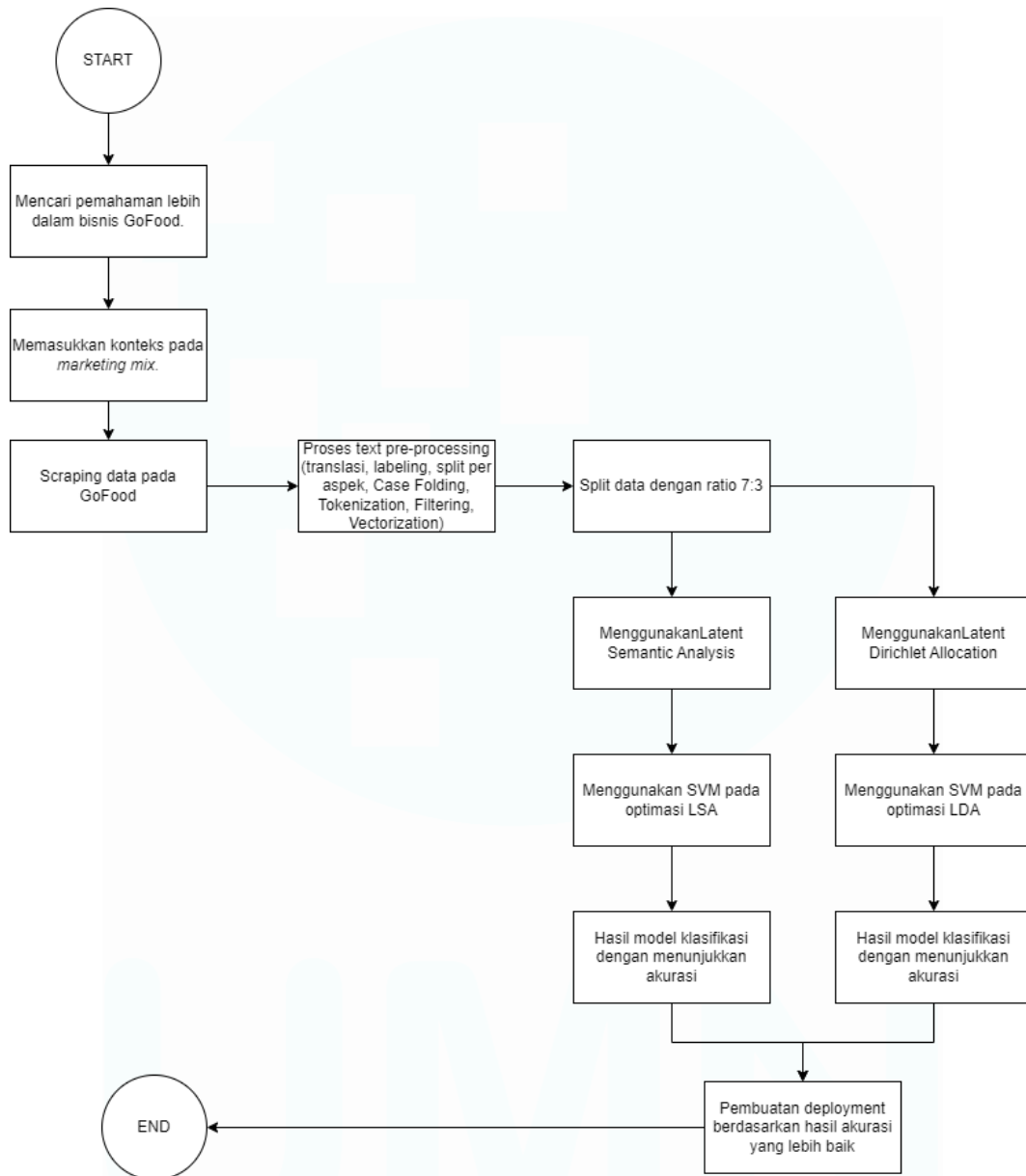
Penilaian dari penelitian ini yaitu aspek. Aspek yang dipilih yaitu *marketing mix*. Pada *marketing mix*, terdapat 7P sebagai poin-poin yang dapat dinilai. 7P

merupakan tambahan dari 4P dan *service-based*. Untuk penelitian ini akan menggunakan 4P saja karena *Place, Promotion, Price, dan Product* sudah cukup merepresentatifkan poin jual perusahaan[15]. 3 Poin yang tidak digunakan beserta alasannya antara lain :

1. *People*. *People* yang dimaksud adalah orang-orang yang ada pada McDonald's seperti pekerja, *manager*, dan sebagainya. Poin *People* tidak digunakan dikarenakan orang-orang yang bergabung pada perusahaan McDonald's telah distandarisasi berdasarkan persyaratan kerja. Selain itu, McDonald's juga memberikan masa *probation*, yaitu masa untuk melatih calon pekerja McDonald's sesuai jangka waktu yang diperlukan. Alasan lainnya yaitu data yang digunakan tidak memiliki poin *People*, karena orang yang memberikan layanan kepada pembeli adalah ojek, yaitu produk dari GoFood. Orang di McDonald's tidak memberikan efek secara langsung kepada pelanggan.
2. *Process*. *Process* yang dimaksud yaitu tahapan pembelian dari pembeli masuk ke outlet hingga selesai menikmati. Pada saat pembelian *online*, proses tidak dapat dinilai. Dikarenakan, pengguna mengalami proses yang berbeda ketika pembelian *online* dan *offline*. Proses pembelian *offline* dapat memberikan nilai ke McDonald's, namun pembelian secara *offline* tidak dapat memberikan nilai ke McDonald's.
3. *Physical Evidence*. *Physical Evidence* yang dimaksud yaitu bentuk fisik bahwa perusahaan ini memiliki bentuk fisik seperti kantor pusat. Alasan poin ini tidak digunakan adalah dari sudut pandang pembeli tidak terlalu diperlukan dan mempengaruhi komentar terhadap perusahaan.

Kesimpulannya yaitu penelitian ini dilakukan pada McDonald's Summarecon Digital Center dimana data yang akan digunakan berasal dari platform GoFood. Penelitian ini dilakukan berdasarkan aspek-aspek, yaitu *Product, Place, Price, Promotion* (4P). Pengambilan data akan dilakukan pada tahun 2024 dengan estimasi jumlah data minimal 1000 data ulasan.

3.2 Alur Penelitian



Gambar 3. 1 Flowchart alur penelitian secara keseluruhan berdasarkan *framework* CRISP-DM.

Alur penelitian ini akan menggunakan *framework* CRISP-DM. Seluruh proses dari alur penelitian terdapat pada Gambar 3. 1, dimana seluruh proses berdasarkan artikel jurnal[21] dan diadaptasi sesuai kebutuhan penelitian, sehingga tahapan alur akan menjadi seperti berikut:

1. *Business Understanding*

Dalam alur ini telah ditentukan bahwa bisnis yang dijalankan McDonald's akan diukur dalam 4P yaitu *Product*, *Place*, *Price*, dan *Promotion*. Semua

aspek seharusnya telah diberikan komentar oleh *customer* sehingga bisnis McDonald's sudah berjalan. Respons komentar *customer* didapat dari McDonald's yang melakukan 4P dengan caranya sendiri, seperti:

1. *Product*

Produk yang dijual oleh McDonald's beraneka ragam seperti Muffin, Burger, Chicken, nasi, kentang goreng, soda, dan sebagainya. Produk yang dijual dikombinasi dengan tren, contohnya kimchi ayam goreng. Yang akan dinilai dari produk adalah enak dan kesegaran.

2. *Place*

Hal yang menjadi kunci dalam poin Place atau tempat adalah kenyamanan, kebersihan, dan strategis. Pastinya outlet yang dibuka McDonald's strategis namun untuk kasus ini perlu perubahan. Data yang diambil berdasarkan kasus *customer* melakukan pembelian online sehingga poin akan berubah. Place akan menjadi lama atau tidaknya dalam pengantaran dan higienis dalam pembungkusan.

3. *Price*

Harga telah dilakukan oleh McDonald's dengan memberikan harga beserta pajak dan potongan lainnya yang diberikan dari GoFood. Pembelian makanan pada *Online Food Delivery* (OFD) kebanyakan memasang harga lebih tinggi dikarenakan dari sisi outlet perlunya seperti *packaging fee*.

4. *Promotion*

Promosi yang dilakukan dengan menawarkan produk yang *trendy* dan sering dibeli. Namun, McDonald's tidak membedakan satu produk dengan lainnya sehingga semuanya terpampang sebagai menu. Promosi dilakukan oleh perusahaan dan untuk penelitian ini di GoFood dengan voucher dari aplikasi GoJek.

2. *Data Understanding*

Data Understanding merupakan tindakan untuk memahami data yang akan digunakan dan dikumpulkan dalam satu file dengan format .csv. Data akan dilakukan *scraping* dengan memahami sumber data lebih

mendalam. *Scraping* akan dilakukan secara otomatis dan menggunakan pemantauan jaringan untuk mendeteksi *Application Programming Interface* (API). Setelah mendapatkan API yang sesuai, dilakukan proses *scraping* pada API tersebut. *Scraping* dilaksanakan pada tanggal 14 April 2024, namun data terbaru ada pada tanggal 11 April 2024. Tidak ada alasan spesifik terkait pada tanggal 14 April 2024. Hal ini dikarenakan data pada penelitian ini dapat diperbarui terus menerus hingga terakhir pada tanggal 14 April 2024.

```
In [6]: url = "https://gofood.co.id/api/outlets/ba99cfaf-7fda-4f98-a9b9-4ee969526d2a/reviews?page=2&page_size=20"
page = requests.get(url)
print(page)
<Response [403]>
```

Gambar 3. 2 Hasil penggunaan *library BeautifulSoup* dan *requests* untuk automatic *scraping* dengan hasil *Response 403* berarti *Forbidden*

Dari Gambar 3. 2 sudah terlihat bahwa percobaan otomatisasi *scraping* pada Python telah dilakukan yaitu menggunakan *library BeautifulSoup*. Hal ini dikarenakan adanya penjagaan dengan *response 403*. *Response 403* adalah respons yang diberikan dengan error *Forbidden*, yang berarti dilarang untuk mengakses API tersebut. Percobaan otomatis pada *scraping* akan dilanjutkan dengan cara pemantauan jaringan website.

Proses *Scraping* secara otomatis dilakukan menggunakan tools sebagai *Robotic Process Automation (RPA)*. Ketika mendapatkan API yang diinginkan, RPA dibutuhkan untuk *scrape* data dari API tersebut dan dilakukan *stacking* data pada excel. Proses yang menggunakan RPA memerlukan tindakan repetitif, cepat, dan efisien, sehingga RPA adalah alat yang tepat.

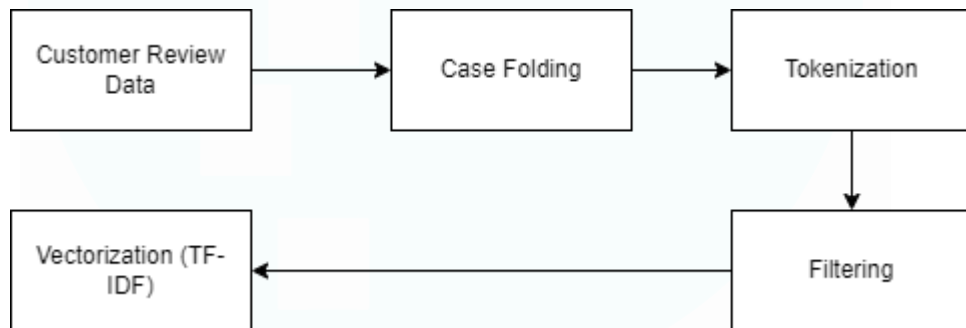
Data yang digunakan adalah 4 kolom dan minimal 1000 baris. Data tersebut adalah utuh yang didapat dari API kecuali pada kolom *createdAt*, dimana format tanggal dirubah menjadi *user friendly*. 4 kolom tersebut akan digunakan dalam pengolahan selanjutnya.

Data yang akan didapat terdapat banyak kolom, namun yang menjadi pusat penelitian ini yaitu *review* dan *rating*. *Data Understanding*

akan melakukan pelabelan untuk sentimen. Pelabelan dilakukan atas data yang tersedia, yaitu teks dengan melakukan otomatisasi. Labeling akan dilakukan dengan translasi kepada bahasa Inggris dan *labeling* menggunakan TextBlob. Namun, *rating* akan menjadi pertimbangan jika labeling yang dilakukan pada bahasa Inggris terlalu banyak Netral[25].

3. Data Preparation

Data Preparation akan dilakukan setelah *Data Understanding*. Dikarenakan penelitian ini mengarah kepada teks, sehingga dilakukannya *text pre-processing*.



Gambar 3. 3 Potongan flowchart untuk proses *Data Preparation* berdasarkan Gambar 3. 1

Tahapan melakukan *text pre-processing* terlampir pada Gambar 3.3. Tahap *text-preprocessing* dimulai dari *customer review data* dari .xlsx. Proses dimulai dengan *Case Folding* dengan proses *lower case* teks, menghilangkan numerik, simbol dan emoji, dan spasi berlebihan, termasuk dengan stopwords akan dihapus. Setelah *Case Folding* dilakukan, proses dilanjutkan dengan tokenisasi. Setelah tokenisasi dilakukan, proses vektorisasi akan dilakukan menggunakan TF-IDF.

4. Modelling

Modelling akan dilakukan setelah melakukan *Data Preparation*. *Modelling* menggunakan metode analisis pada algoritma. Metode yang akan digunakan pada penelitian ini adalah metode klasifikasi sentimen *Support Vector Machine*, *Latent Semantic Analysis*, dan *Latent Dirichlet Allocation*.

5. *Evaluation*

Evaluation akan dilakukan setelah *Modelling*. Tahap *Evaluation* akan menggunakan *Confusion Matrix*, sebagai evaluasi dari nilai performa dan akurasi model yang telah dibuat. 4 bentuk tersebut akan menjadi penentu apakah model yang telah dibuat baik atau tidak. Evaluasinya dalam 4 bentuk, yaitu Akurasi, Presisi, *Recall*, F1-score yang didapat dari *Confusion Matrix*.

6. *Deployment*

Deployment akan dilakukan setelah *evaluation* mendapatkan hasil yang optimal. *Deployment* merupakan tahapan dimana memberikan produk pada sebuah analisa. Untuk penelitian ini, *deployment* diterapkan dengan membuat aplikasi sederhana mengambil inputan komentar dan memberikan sentimen berdasarkan inputan tersebut. Aplikasi akan dibuat dalam bentuk *web service*, yaitu *Application Programming Interface* dalam pada local.

3.3 Metode Penelitian

Metode penelitian akan menjelaskan mengenai metode penyelesaian. Metode yang dipakai untuk melakukan *Sentiment Analysis* adalah metode sentimen analisis dengan tugas yaitu *Sentiment Classification* dengan pendekatan *Machine Learning* dan menggunakan level aspek[9]. Dikarenakan menggunakan level aspek, maka *Sentiment Analysis* berubah menjadi *Aspect-Based Sentiment Analysis*. Untuk pendekatan pada penelitian ini dapat menggunakan *Machine Learning* tradisional seperti *Multinomial Naïve Bayes*, *Support Vector Machine* (SVM), dan *Linear Regression*(LR)[10][11].

Seperti yang telah dijelaskan pada penelitian terdahulu[9], SVM dan *Naïve Bayes* banyak digunakan untuk *Sentiment Analysis*. *Linear Regression* juga pernah digunakan dalam analisis sentimen[11]. Berikut tabel perbandingan ketiga model klasifikasi tersebut :

Table 3. 1 Perbandingan Model SVM, Naïve Bayes, dan LR.

Kategori	SVM	Naïve Bayes	LR
Proses pembelajaran klasifikasi	Menemukan <i>hyperplane</i> yang memisahkan kelas yang ada dengan margin terbesar.	Melakukan perhitungan probabilitas suatu peristiwa dengan kemungkinan peristiwa sama yang telah terjadi sebelumnya.	Mengetahui <i>independent variable</i> dan menemukan <i>intercept</i> yang errornya paling minimal.
Penggunaan	Biasa digunakan untuk permasalahan klasifikasi	Biasa digunakan untuk permasalahan klasifikasi.	Biasa digunakan untuk permasalahan regresi atau prediksi.
Kemampuan kerja pada ukuran data yang besar	SVM kurang dapat bekerja baik jika menggunakan dataset yang besar.	Naïve Bayes kurang dapat bekerja baik jika menggunakan dataset yang besar.	LR dapat bekerja baik jika menggunakan dataset yang besar.

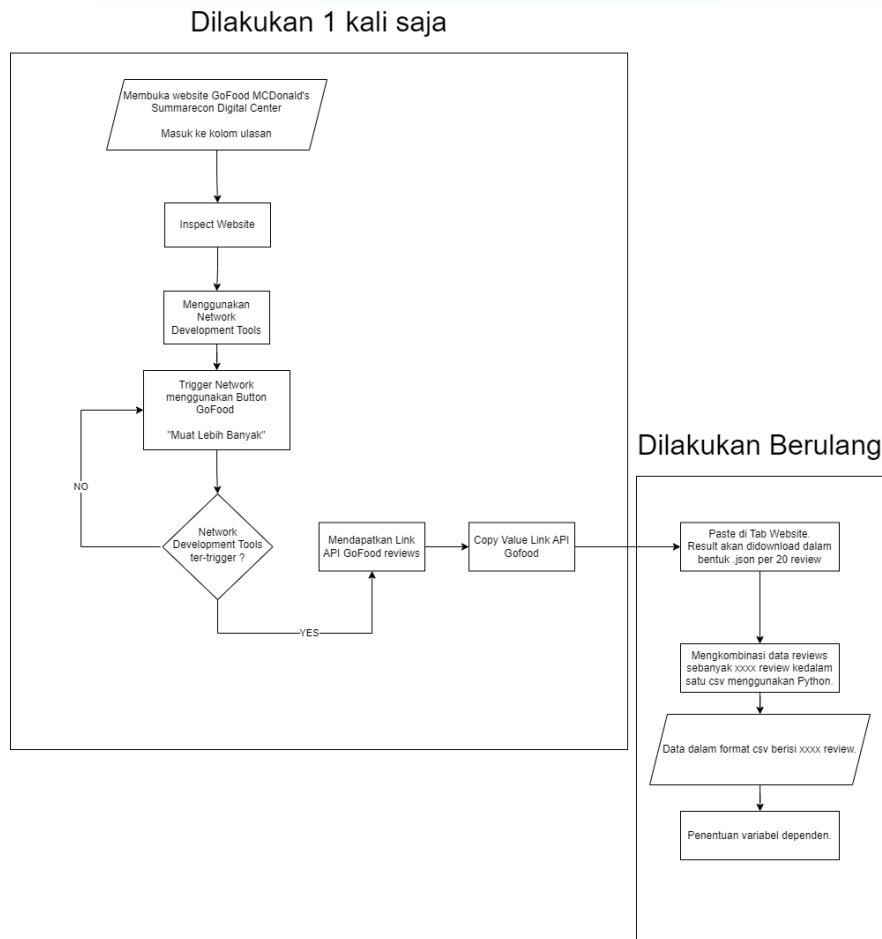
Berdasarkan Tabel 3.1, hanya SVM yang akan digunakan untuk pemodelan. SVM digunakan dengan alasan dapat digunakan untuk permasalahan klasifikasi. Kemampuan kerja pada ukuran data yang besar adalah kekurangan dari SVM tetapi kekurangan tersebut akan dibandingkan dengan bantuan *Latent Semantic Analysis* (LSA) dan *Latent Dirichlet Allocation* (LDA) sebagai pemecah dokumen dengan hasil yaitu kata dan kalimat yang memiliki hubungan kesamaan. Perbandingan dilakukan dengan alasan apakah kekurangan SVM dalam mendapatkan menggunakan data yang besar merupakan hal yang benar.

3.4 Teknik Pengumpulan Data

3.4.1 Data Collection

Teknik yang akan digunakan untuk mengumpulkan data pada penelitian ini adalah pengambilan data sekunder. Pengumpulan data sekunder yang digunakan yaitu teknik *crawling*. Tipe data yang dikumpulkan berupa teks berada pada website, sehingga *crawling* adalah teknik yang tepat. Pengumpulan data teks pada ulasan pelanggan di McDonald's Summarecon Digital Center dilakukan

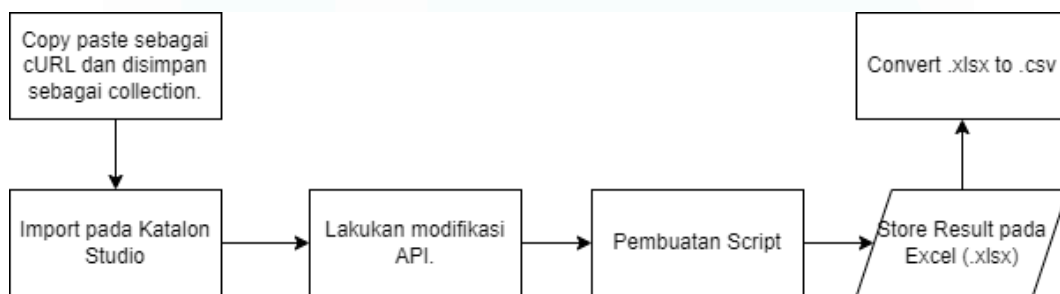
dengan teknik *web crawling* dan *Automated scraping*, dimana hasil ulasan diambil dari ulasan pengguna pada website GoFood.



Gambar 3. 4 Pembagian proses pekerjaan berdasarkan *Flowchart Data Collection*.

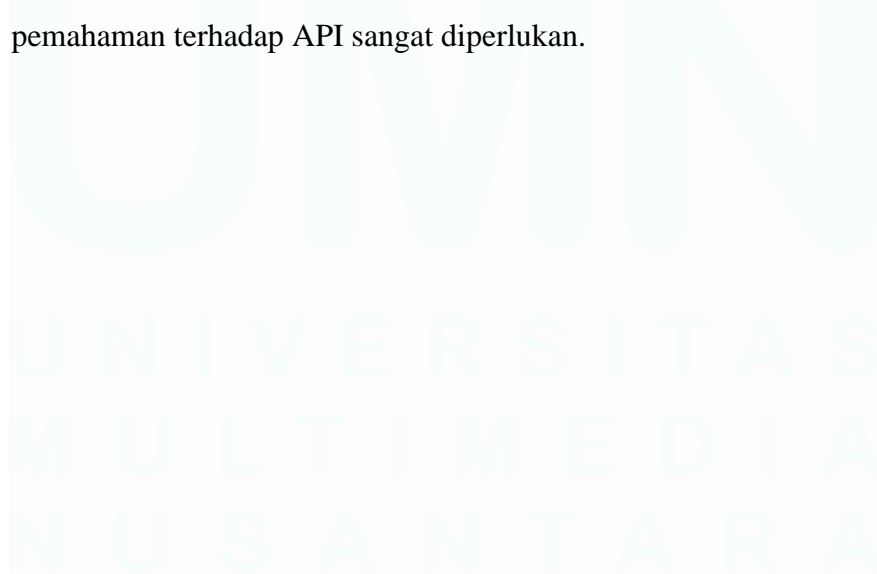
Terlihat pada Gambar 3. 4 bahwa *Flowchart Data Collection* telah terpampang dan telah dipisahkan antara pengerjaan satu kali dan repetitif. Proses dilakukan satu kali saja dengan mendapatkan hasil yaitu sebuah API GoFood yang dapat digunakan seterusnya. API tersebut akan dilakukan pengambilan keseluruhan data. Proses pengambilan data dilakukan secara otomatis menggunakan *Robotic Process Automation* atau dikenal sebagai RPA. Ada 2 dari banyak *tools* yang dapat digunakan pada RPA, yaitu Katalon Studio dan UiPath. 2 *tools* ini terkenal pada sisi *Robotic Process*, dimana penggunaannya digunakan berbeda. Katalon Studio berfokus kepada pengetesan, sedangkan UiPath berfokus kepada otomatisasi proses bisnis.

Katalon Studio adalah sebuah *tools* untuk menjalankan proses oleh robot sehingga akan berjalan secara otomatis. Katalon Studio menggunakan bahasa pemrograman Groovy, yaitu salah satu bahasa pemrograman *Object Oriented Language* dengan menggunakan *Java Platform*[26]. Katalon Studio adalah *tools* yang tepat untuk dilakukannya pengumpulan data walaupun Katalon Studio dikhususkan untuk pengetesan.



Gambar 3. 5 Flowchart proses detail pengolahan API menjadi data yang siap digunakan.

API hasil *crawling* akan dilakukan proses pengolahan oleh Katalon Studio. Pada Gambar 3. 5 terlihat proses pengolahan API. Hasil yang dilakukan satu kali saja pada Gambar 3. 4 yaitu sebuah API akan diolah untuk mendapatkan sumber data. API tersebut akan diambil dengan *Copy* dan *Paste* cURL dan akan disimpan sebagai *collection*. Setelah menjadi *collection*, Katalon Studio dapat menerima API tersebut sebagai *import*. Katalon Studio dapat melakukan modifikasi API secara dinamis, sehingga pemahaman terhadap API sangat diperlukan.



3.5 Teknik Analisis Data

Untuk penelitian ini, analisis data akan mengikuti *framework* sehingga sesuai dengan analisa. Ada banyak *framework* yang dapat melakukan khususnya pada pemodelan data serta evaluasi contohnya CRISP-DM dan SEMMA. Berikut tabel perbandingan kedua *framework* tersebut:

Table 3. 2 Perbandingan antara CRISP-DM dan SEMMA

Kategori	CRISP-DM	SEMMA
Tahapan	<ol style="list-style-type: none">1. Business Understanding2. Data Understanding3. Data Preparation4. Modelling5. Evaluation6. Deployment	<ol style="list-style-type: none">1. Sample2. Explore3. Modify4. Model5. Assess
Keterikatan	Tidak terikat dengan apapun	Terikat dengan perusahaan SAS.

Penelitian ini akan menggunakan *framework* CRISP-DM dengan alasan penelitian ini tidak menggunakan produk dari perusahaan SAS dan penelitian ini tidak mengikat akan suatu instansi atau organisasi. Mengenai alur penelitian, kurang lebih memiliki kemiripan antara CRISP-DM dan SEMMA, yaitu *modify* dan *model*.

Untuk analisis data, penelitian ini menggunakan platform Google Colab dengan bahasa pemrograman yaitu Python. Ada platform lain yang dapat menggunakan *Machine Learning* serta menggunakan bahasa pemrograman yang sama, yaitu Python contohnya Jupyter Notebook.

Berikut tabel perbandingan kedua platform tersebut :

Table 3. 3 Perbandingan antara Google Colab dan Jupyter Notebook

Kategori	Google Colab	Jupyter Notebook
Penyimpanan data	Penyimpanan data dilakukan melalui Cloud	Penyimpanan data dilakukan di internal storage.
Fleksibilitas dalam pengerjaan program	Fleksibel karena dapat bekerja dan dapat berbagi program yang telah dibuat.	Tidak fleksibel karena program yang dibuat hanya ada di internal storage, sehingga tidak dapat berbagi program yang telah dibuat.
Instalasi library	Tidak melakukan instalasi library.	Melakukan instalasi library.

Berdasarkan Tabel 3. 3, Google Colab merupakan pilihan terbaik untuk penelitian ini dengan alasan mempermudah melakukan pekerjaan dan penyimpanan data tanpa memikirkan internal storage. Selain itu, instalasi library dari Python juga tidak perlu dilakukan.

Beberapa *library* yang akan digunakan yaitu :

1. NumPy, adalah sebuah *library* fundamental dimana digunakan untuk komputasi ilmiah dalam Python. Beberapa fungsi yang dapat dijalankan yaitu objek multidimensi array, operasi dan array, perhitungan statistika dasar, dan masih banyak lagi[27].
2. Scikit-learn, adalah sebuah *library* yang dibuat diatas NumPy, SciPy, dan matplotlib. Scikit-learn berfokus pada pekerjaan pada *machine learning*, seperti klasifikasi dengan algoritma seperti SVM, *Gaussian Naïve Bayes*, *Decision Tree*, dan sebagainya[28].
3. Sastrawi, adalah sebuah *library* dimana memberikan fungsi untuk menjadikan bahasa Indonesia sesuai dengan KBBI dimana mengubah teks menjadi bentuk dasar[29].
4. Matplotlib, adalah sebuah *library* dimana berfungsi untuk visualisasi pada Python.

5. Re, adalah sebuah library berfungsi sebagai manipulasi, mencari, mengganti teks yang akan berguna untuk *regular expression*[30].
6. NLTK, adalah library digunakan untuk menangani data khusus yaitu bahasa manusia[31].

