

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian berfokus pada Taman Safari Indonesia, yang merupakan salah satu destinasi wisata yang terkenal di Indonesia, terletak di Cisarua, Bogor. Taman Safari Indonesia menampilkan berbagai jenis hewan dari seluruh dunia, dengan konsep yang menekankan pada konservasi satwa dan memberikan pengalaman interaktif bagi pengunjung untuk melihat hewan-hewan tersebut dalam lingkungan yang menyerupai habitat aslinya. Selain itu, Taman Safari Indonesia juga menawarkan berbagai atraksi seperti pertunjukan hewan, zona interaktif dengan hewan, dan area bermain yang menjadikannya tempat yang sangat populer untuk keluarga.

Sebagai destinasi wisata yang besar, Taman Safari Indonesia memiliki ribuan pengunjung setiap tahunnya. Oleh karena itu, penting bagi pengelola untuk memahami umpan balik dan persepsi pengunjung mengenai pengalaman mereka di taman tersebut. Salah satu sumber utama untuk mendapatkan informasi mengenai kualitas layanan, fasilitas, dan pengalaman pengunjung adalah melalui ulasan yang ditinggalkan oleh pengunjung di *Google Maps*. Ulasan-ulasan ini dapat memberikan gambaran yang lebih mendalam tentang apa yang disukai atau tidak disukai oleh pengunjung, serta tentang kekuatan dan kelemahan yang ada pada layanan dan fasilitas yang disediakan oleh Taman Safari Indonesia.

Analisis sentimen terhadap ulasan pengunjung sangat penting untuk membantu pengelola taman safari memahami sentimen publik dan pengunjung terhadap taman tersebut. Melalui analisis sentimen, setiap ulasan dapat diklasifikasikan menjadi sentimen positif, negatif, atau netral yang menggambarkan kepuasan atau ketidakpuasan pengunjung. Selain itu, ulasan ini juga dapat digunakan untuk mengidentifikasi area-area yang perlu diperbaiki atau ditingkatkan oleh pengelola taman. Misalnya, jika banyak pengunjung memberikan ulasan positif terkait interaksi mereka dengan hewan atau atraksi tertentu, pengelola dapat mempertahankan dan mengembangkan lebih lanjut aspek tersebut. Sebaliknya, jika

banyak ulasan mengandung keluhan mengenai kebersihan atau pelayanan, hal ini dapat menjadi fokus perbaikan untuk meningkatkan kualitas pengalaman pengunjung.

Penelitian mengembangkan rekomendasi berbasis konten yang memberikan saran atau rekomendasi atraksi kepada pengunjung berdasarkan analisis sentimen ulasan yang ada. rekomendasi ini akan menggunakan penerapan *Term Frequency*, yang menganalisis ulasan pengunjung untuk menemukan kesamaan antara ulasan-ulasan yang diberikan dan kemudian merekomendasikan atraksi atau program yang sesuai dengan preferensi pengunjung. Dengan tujuan untuk memberikan dua manfaat utama: pertama, untuk meningkatkan pemahaman pengelola tentang persepsi pengunjung melalui analisis sentimen, dan kedua, untuk menyediakan rekomendasi yang dapat membantu pengunjung memilih atraksi yang sesuai dengan minat mereka berdasarkan ulasan yang ada.

3.2 Metode Penelitian

Berikut merupakan komparasi antara metode KDD dan CRISP-DM:

Tabel 3. 1 Tabel Metode Penelitian

	KDD	CRISP-DM
Definisi	KDD adalah proses yang melibatkan identifikasi pola-pola penting dalam data besar dan kompleks, dari pemilihan data hingga evaluasi hasil yang diperoleh.	CRISP-DM adalah metodologi <i>data mining</i> yang sistematis dan berbasis siklus yang banyak digunakan di berbagai industri untuk menemukan pengetahuan dalam data.

	KDD	CRISP-DM
Fokus Utama	Proses untuk mengekstrak pengetahuan dari data dengan penekanan pada tahap transformasi dan <i>data mining</i> .	Fokus pada penerapan proses yang terstruktur dalam seluruh proyek <i>data mining</i> , termasuk perencanaan dan pemantauan.
Tahapan Utama	<ol style="list-style-type: none"> 1. <i>Selection</i> (Pemilihan Data) 2. <i>Preprocessing</i> (Pemrosesan Data) 3. <i>Transformation</i> (Transformasi Data) 4. <i>Data Mining</i> 5. <i>Evaluation</i> (Evaluasi) 	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> (Pemahaman Bisnis) 2. <i>Data Understanding</i> (Pemahaman Data) 3. <i>Data Preparation</i> (Persiapan Data) 4. <i>Modeling</i> (Pemodelan) 5. <i>Evaluation</i> (Evaluasi) 6. <i>Deployment</i> (Penerapan)
Pendekatan	KDD berfokus pada proses penambahan data untuk mengekstrak pengetahuan baru dari data yang sudah ada.	CRISP-DM lebih fokus pada siklus hidup proyek <i>data mining</i> dengan pendekatan interaktif, melibatkan pemahaman bisnis dan implementasi di dunia nyata.
Iterasi	KDD cenderung lebih terstruktur dan tidak selalu memperhatikan interaksi antara tahapan yang berbeda. Prosesnya dapat lebih linier.	CRISP-DM bersifat lebih interaktif dan fleksibel, memungkinkan untuk kembali ke tahap sebelumnya jika diperlukan.

	KDD	CRISP-DM
Pemahaman Bisnis	Kurang menekankan pemahaman mendalam tentang bisnis pada tahap awal.	CRISP-DM dimulai dengan pemahaman bisnis yang mendalam sebagai langkah pertama, memastikan bahwa solusi <i>data mining</i> selaras dengan tujuan bisnis.
Fleksibilitas	KDD lebih berfokus pada teknik <i>data mining</i> dan evaluasi. Fleksibilitasnya terbatas pada pemilihan data dan pemrosesan.	CRISP-DM sangat fleksibel dan bisa diterapkan pada berbagai jenis proyek dan industri, dengan siklus yang dapat beradaptasi sesuai dengan kebutuhan.
Tujuan Utama	Menemukan pola dan pengetahuan tersembunyi dalam data yang dapat digunakan untuk analisis lebih lanjut.	Menyediakan solusi <i>data mining</i> yang praktis dan <i>actionable</i> yang dapat diterapkan untuk mendukung pengambilan keputusan di dunia nyata.
Tahap Evaluasi	Evaluasi dalam KDD cenderung berfokus pada kinerja model dan kualitas hasil akhir dari proses <i>data mining</i> .	Evaluasi dalam CRISP-DM lebih luas, mencakup tidak hanya hasil dari <i>data mining</i> , tetapi juga penilaian terhadap penerapan solusi dalam bisnis.
Fase Deployment	KDD tidak secara eksplisit menyebutkan penerapan hasil sebagai bagian dari metodologi.	CRISP-DM memiliki fase <i>Deployment</i> yang jelas, yang berfokus pada implementasi model dan hasil analisis untuk penggunaan nyata.

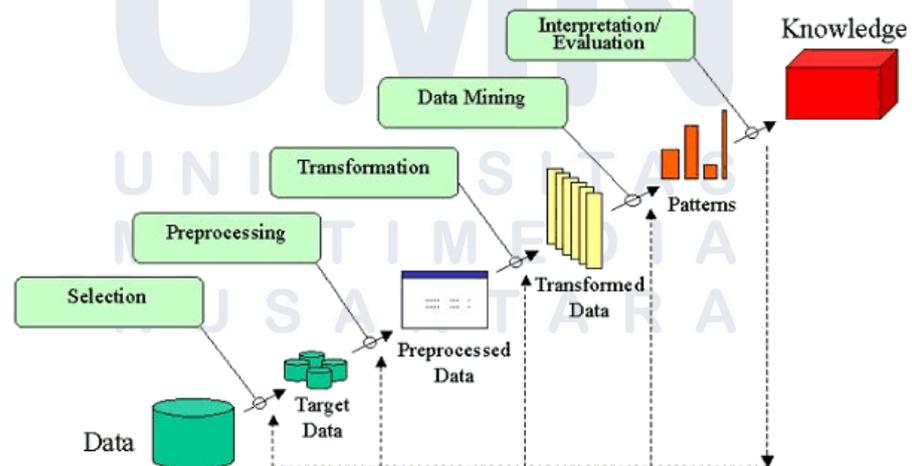
Tabel 3.1 merupakan perbandingan yang mendasari pemilihan KDD (*Knowledge Discovery in Databases*) dibandingkan CRISP-DM yang didasarkan pada fokus KDD yang lebih besar pada tahap *data mining* dan penemuan pola tersembunyi dalam data yang besar dan kompleks dan sangat berfokus pada analisis sentimen dan rekomendasi berbasis ulasan pengunjung.

KDD memberikan pendekatan yang lebih mendalam dalam tahap *preprocessing* dan transformasi data, yang penting untuk menangani data teks yang bervariasi, seperti ulasan di *Google Maps*. Metodologi ini memungkinkan eksplorasi lebih dalam terhadap data guna menemukan pola yang tidak terduga. Sementara itu, CRISP-DM lebih berorientasi pada penerapan solusi bisnis secara keseluruhan dan penerapan model di dunia nyata.

Selain itu, KDD lebih terstruktur dalam tahap evaluasi hasil *data mining*, yang krusial untuk memastikan akurasi dan relevansi analisis yang dilakukan. Oleh karena itu, KDD lebih cocok karena berfokus pada eksplorasi dan penambangan data untuk menemukan pengetahuan baru dari data besar.

3.2.1 Knowledge Discovery in Databases (KDD)

Proses KDD dapat digambarkan dalam lima tahap utama yang saling terkait, yaitu *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation*.



Gambar 3. 1 Metode *Knowledge Discovery in Databases* (KDD)

a) Selection

Pemilihan data adalah tahap pertama dalam proses KDD yang berfokus pada pengumpulan data yang relevan dengan tujuan penelitian. Pada tahap ini, data yang digunakan dalam diperoleh dari *platform Google Maps*, yaitu ulasan pengunjung Taman Safari Indonesia. Data yang relevan dipilih berdasarkan kriteria tertentu, seperti kategori sentimen, *rating*, serta aspek tertentu yang ingin dianalisis, seperti kualitas layanan, kebersihan, dan interaksi pengunjung dengan hewan. Proses seleksi data ini sangat penting untuk memastikan bahwa data yang digunakan dalam analisis adalah data yang paling relevan dengan tujuan penelitian dan tidak membingungkan analisis yang akan dilakukan pada tahap berikutnya.

b) Preprocessing

Preprocessing adalah tahap yang bertujuan untuk membersihkan dan mempersiapkan data agar siap untuk digunakan dalam analisis lebih lanjut. Data yang dikumpulkan dari *Google Maps* sering kali memiliki masalah seperti data yang tidak lengkap, duplikat, atau dalam format yang tidak konsisten. Oleh karena itu, proses ini bertujuan untuk meningkatkan kualitas data agar lebih siap untuk diproses dengan algoritma *machine learning*.

Beberapa langkah dalam *preprocessing* data meliputi:

- a. Penghilangan Data yang Tidak Relevan: Menghapus ulasan atau data yang tidak relevan dengan analisis sentimen, seperti ulasan yang tidak mengandung informasi penting atau terlalu singkat.
- b. Pemeriksaan Data Duplikat: Memeriksa data yang terduplikat dan menghapus data tersebut.
- c. Tokenisasi: Proses memecah teks ulasan menjadi kata atau token agar dapat dianalisis lebih lanjut.

- d. Penghilangan *Stopwords*: Menghapus kata-kata yang tidak memiliki makna penting dalam analisis, seperti "dan," "atau," dan "yang".
- e. *Stemming*: Mengubah kata-kata ke bentuk dasar untuk mengurangi variasi kata yang ada dalam teks.

Setelah proses *preprocessing*, data ulasan menjadi lebih terstruktur dan siap untuk digunakan dalam tahap analisis sentimen.

c) Transformation

Setelah data diproses, langkah selanjutnya adalah transformasi data. Pada tahap ini, data yang telah dibersihkan diubah menjadi format yang lebih berguna untuk analisis selanjutnya. Data teks dapat diubah menjadi representasi numerik yang dapat digunakan oleh algoritma *machine learning*. Salah satu teknik yang digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*), yang digunakan untuk mengekstrak fitur dari teks ulasan. Teknik ini membantu mengukur pentingnya kata-kata tertentu di seluruh *dataset*. Transformasi juga melibatkan konversi teks ke dalam vektor atau matriks untuk memudahkan pemrosesan lebih lanjut oleh algoritma *machine learning* yang digunakan dalam analisis sentimen, seperti SVM, *Logistic Regression*, Naïve Bayes, KNN, dan *Random Forest*.

d) Data Mining

Tahap Data Mining merupakan inti dari proses KDD, di mana teknik analisis digunakan untuk menemukan pola yang berguna dalam data yang telah diproses. Pada tahap ini, data ulasan yang telah dianalisis akan dikelompokkan ke dalam kategori sentimen positif, negatif, atau netral menggunakan lima algoritma *machine learning* yang berbeda: *Logistic Regression*, Naïve Bayes, *Support Vector Machine* (SVM), *Random Forest* dan *K-Nearest Neighbors* (KNN).

- a. *Logistic Regression* digunakan untuk mengklasifikasikan ulasan menjadi kategori positif atau negatif berdasarkan probabilitas yang dihitung dari fitur yang ada dalam ulasan.
- b. *Naïve Bayes* adalah algoritma probabilistik yang mengklasifikasikan ulasan berdasarkan probabilitas kata-kata dalam ulasan tersebut.
- c. SVM (*Support Vector Machine*) digunakan untuk membedakan ulasan positif dan negatif dengan mencari hyperplane optimal yang memisahkan kelas-kelas tersebut.
- d. *Random Forest* digunakan untuk meningkatkan akurasi klasifikasi dan regresi dengan menggabungkan hasil dari banyak pohon keputusan, sehingga mengurangi risiko overfitting. Metode ini juga memberikan wawasan tentang pentingnya fitur yang berkontribusi dalam prediksi.
- e. KNN (*K-Nearest Neighbors*) digunakan untuk klasifikasi dan regresi dengan mengukur jarak antara titik data, sehingga dapat menentukan kelas atau nilai berdasarkan tetangga terdekat. Algoritma ini efektif dalam mengidentifikasi pola dalam data yang terdistribusi secara alami.

e) Evaluation

Pada tahap evaluasi, analisis dilakukan untuk menilai efektivitas algoritma yang diterapkan dalam pengolahan data. Metode evaluasi yang digunakan mencakup pengukuran akurasi dari berbagai algoritma, sehingga memungkinkan untuk melakukan perbandingan yang adil dan objektif. Setelah menerapkan teknik *data mining*, berbagai algoritma seperti *Support Vector Machine* (SVM), *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *K-Nearest Neighbors* (KNN) diuji untuk mengidentifikasi model dengan performa terbaik. Evaluasi dilakukan dengan cara membandingkan akurasi model sebelum dan sesudah penerapan teknik optimasi seperti SMOTE dan *GridsearchCV*. Hasil pengukuran akurasi ini dicatat dalam tabel perbandingan yang

memberikan gambaran jelas tentang efektivitas masing-masing algoritma. Dengan cara ini, kesimpulan yang meyakinkan mengenai algoritma mana yang akan memberikan hasil yang paling akurat dan konsisten, serta memberikan dasar yang kuat untuk pemilihan model yang akan digunakan dalam aplikasi lebih lanjut. Melalui pendekatan evaluasi ini, dapat diketahui juga faktor-faktor yang mempengaruhi performa model dan mengidentifikasi area yang memerlukan perbaikan untuk pengembangan analisis yang lebih mendalam di masa depan.

3.2.2 Term Frequency

Term Frequency adalah metode yang digunakan untuk memberikan rekomendasi berdasarkan penghitungan seberapa sering suatu kata muncul dalam teks ulasan pengunjung mengenai hewan di Taman Safari. Proses ini dimulai dengan pengumpulan data, di mana data yang digunakan berasal dari ulasan pengunjung yang diperoleh dari *Google Maps Reviews*. Data ini mencakup berbagai karakteristik dan atribut dari hewan-hewan yang ada di taman safari.

Setelah mengumpulkan data, langkah selanjutnya adalah melakukan pembersihan dan pra-pemrosesan pada teks ulasan. Pada tahapan ini, teknik *Natural Language Processing* (NLP) diterapkan untuk membersihkan data dari kata-kata umum atau *stopwords* yang tidak memiliki makna penting. Proses ini juga mencakup menormalkan teks dengan mengubah semua huruf menjadi kecil dan melakukan tokenisasi, yang memecah ulasan menjadi kata atau frasa yang lebih kecil untuk mempermudah analisis lebih lanjut.

Setelah data bersih, sistem akan menganalisis kata-kata yang sering muncul untuk menentukan frekuensi kemunculan kata tersebut. Dalam hal ini, *Term Frequency* digunakan untuk menghitung seberapa penting suatu kata dalam konteks ulasan yang diberikan. Kata-kata yang sering muncul dalam ulasan tentang hewan tertentu akan mendapatkan nilai *Term Frequency* yang lebih tinggi, mencerminkan tingkat ketertarikan pengunjung terhadap hewan tersebut. Dengan cara ini, sistem dapat

memberikan skor bagi setiap hewan berdasarkan seberapa sering dan seberapa positif hewan tersebut disebutkan dalam ulasan. Skor ini akan digunakan untuk menyusun daftar rekomendasi yang menyoroti hewan-hewan yang paling menarik bagi pengunjung.

3.3 Teknik Pengumpulan Data

Berikut adalah komparasi dari teknik pengumpulan data melalui Apify dan Selenium:

Tabel 3. 2 Tabel Teknik Pengumpulan Data

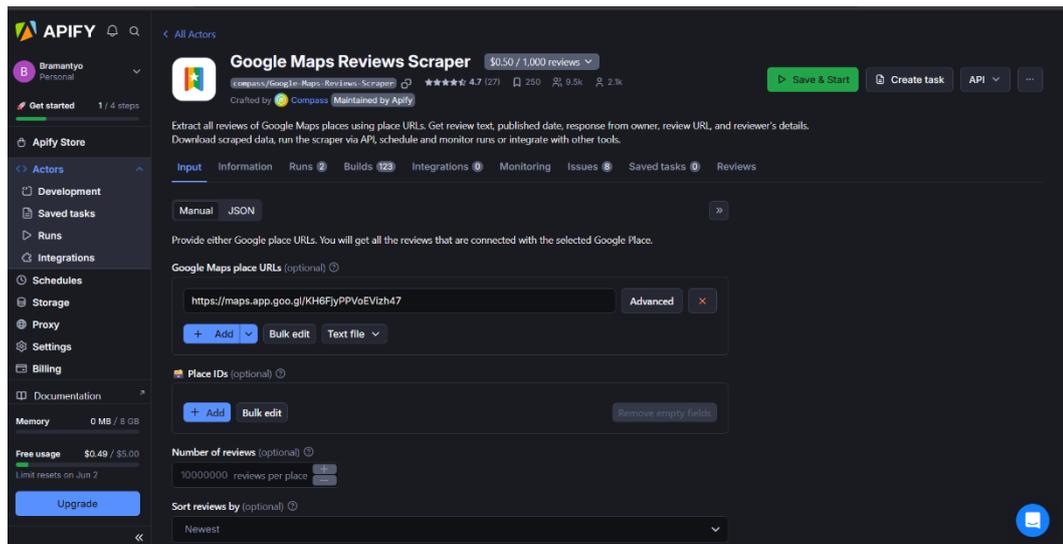
	Apify	Selenium
Kemudahan Penggunaan	Mudah digunakan dengan antarmuka grafis.	Memerlukan keterampilan pemrograman lebih tinggi dan pengetahuan tentang pengujian perangkat lunak.
Fleksibilitas	Menyediakan berbagai alat <i>built-in</i> untuk <i>scraping</i> dan otomatisasi.	Sangat fleksibel untuk pengujian dan bisa digunakan untuk web <i>scraping</i> , tetapi memerlukan banyak konfigurasi.
Kecepatan Implementasi	Cepat dalam hal setup dan implementasi karena tidak memerlukan pengaturan lingkungan.	Lebih lambat untuk disiapkan karena memerlukan instalasi dan konfigurasi browser.
Dukungan untuk Javascript	Mampu menangani banyak konten dinamis dengan baik	Dapat menangani aplikasi web yang sangat dinamis, tetapi dengan lebih banyak konfigurasi
Ketersediaan	Berbasis <i>cloud</i> sehingga dapat diakses dari mana saja; ada biaya yang tergantung pada penggunaan	Gratis, tetapi harus di <i>install</i> pada mesin lokal
Ketersediaan Sumber Daya	Dokumentasi yang baik dan tutorial yang mudah diikuti	Komunitas besar dan banyak sumber daya, tetapi lebih teknis

	Apify	Selenium
Penggunaan Daya	Menggunakan lebih sedikit sumber daya karena berjalan di <i>cloud</i>	Membutuhkan lebih banyak sumber daya lokal, terutama saat menjalankan browser
Skalabilitas	Sangat mudah untuk menangani jumlah data yang besar	<i>Scalability</i> terbatas tergantung pada spesifikasi lokal

Tabel 3.2 merupakan perbandingan yang digunakan untuk menemukan aplikasi yang layak untuk penelitian, selain menggunakan Apify untuk *scraping* data, aplikasi *Google Maps Reviews Scraper* yang tersedia di GitHub juga dimanfaatkan. Aplikasi ini dikembangkan secara *open-source* dan dirancang khusus untuk mengambil ulasan dari *Google Maps* secara otomatis. Salah satu keunggulan utama dari *Google Maps Reviews Scraper* adalah kemampuannya dalam mengumpulkan data ulasan dengan cepat dan efisien, memungkinkan pengambilan informasi berharga tentang pengalaman pengunjung Taman Safari.

Dengan menggunakan aplikasi ini, tidak hanya dapat mengakses ulasan terbaru, tetapi juga berpotensi melakukan analisis mendalam terhadap sentimen dan opini pengunjung. Aplikasi ini ditulis dalam bahasa pemrograman *Python*, sehingga memberikan fleksibilitas untuk memodifikasi dan mengonfigurasi skrip sesuai kebutuhan penelitian. Selain itu, komunitas pengembang aktif di GitHub memfasilitasi dukungan dan berbagi pengalaman dengan pengguna lain, sehingga mempermudah dalam menyempurnakan teknik pengumpulan data.

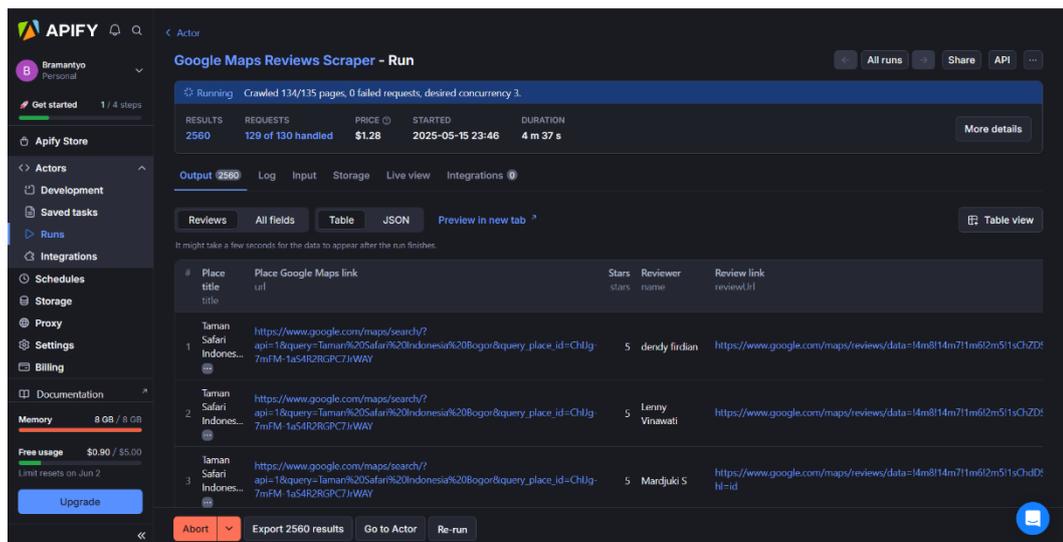
Dengan mengintegrasikan *Google Maps Reviews Scraper* dalam metode pengumpulan data, dapat dipastikan bahwa data yang diperoleh lebih komprehensif dan representatif.



Gambar 3.2 *Tools Apify*

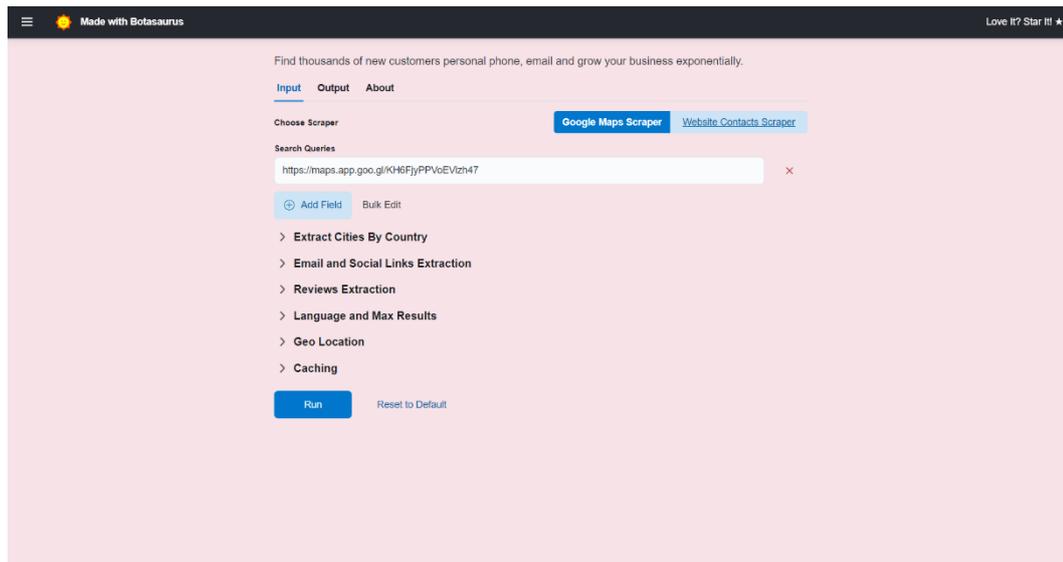
Gambar 3.1 menjelaskan bahwa Apify yang digunakan untuk mengumpulkan data ulasan dari *Google Maps*. Pengguna dapat melakukan *input link* tempat yang relevan dari *Google Maps* ke dalam fitur *Google Maps Reviews Scraper*. Proses ini memungkinkan pengambilan informasi mengenai ulasan, termasuk teks ulasan, tanggal publikasi, dan respons dari pemilik.

Pengaturan yang ada juga memungkinkan pengguna untuk menentukan jumlah maksimum ulasan yang akan diambil, hingga 1.000.000 ulasan per tempat. Dengan menggunakan alat ini, data yang diperoleh akan digunakan untuk analisis sentimen dan untuk mengidentifikasi tema-tema penting dalam ulasan, yang selanjutnya dapat membantu dalam memberikan rekomendasi berbasis data kepada pengunjung Taman Safari. Ini sejalan dengan tujuan penelitian untuk meningkatkan pengalaman pengunjung melalui pemahaman yang lebih baik terhadap umpan balik yang diberikan.



Gambar 3.3 Insert Data di Apify

Gambar 3.2 merupakan *output* dari *Google Maps Reviews Scraper* di Apify, yang saat ini sedang berjalan untuk mengumpulkan data ulasan dari berbagai lokasi, termasuk Taman Safari Indonesia. Proses ini tidak hanya menghasilkan total 2.560 ulasan, tetapi masih berjalan untuk mengambil lebih banyak data hingga total 66.383 ulasan. *Output* ini mencakup informasi penting seperti nama tempat, tautan ke lokasi di *Google Maps*, *rating* yang diberikan oleh pengunjung, nama pengulas, dan tautan ke ulasan lengkap. Efisiensi pengambilan data ini, yang berlangsung selama sekitar 4 menit dan 37 detik, memberikan gambaran yang baik tentang pengalaman pengunjung. Data yang terkumpul akan digunakan untuk analisis sentimen, yang bertujuan untuk mengidentifikasi pola dan tren dalam ulasan, serta memberikan rekomendasi berbasis data untuk meningkatkan pengalaman bagi pengunjung di Taman Safari.



Gambar 3. 4 *Tools Google Maps Extractor*

Gambar 3.3 menjelaskan bahwa pengguna dapat memasukkan *search query* dalam bentuk *link* lokasi spesifik, dan alat ini menawarkan berbagai fitur, seperti ekstraksi data kota berdasarkan negara, pengumpulan alamat email dan tautan media sosial, serta pengambilan ulasan dari lokasi yang di-*input*. Selain itu, pengguna dapat mengatur bahasa dan jumlah hasil yang diinginkan, serta menyertakan informasi geolokasi. Dengan menekan tombol *Run*, proses ekstraksi akan dimulai, memungkinkan pengumpulan data yang diperlukan untuk analisis lebih lanjut. Data ini akan membantu dalam memahami pengalaman pengunjung dan umpan balik mereka terhadap layanan yang ditawarkan, sejalan dengan tujuan penelitian untuk meningkatkan kualitas layanan di Taman Safari.

Place Id	Place Name	Review Id	Review Link	Name	Reviewer Id	Reviewer	Rating	Review Text	Published At	Published	Responses
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Moon Moon	111673525...	https://www...	5	저문기보고...	5 jam lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Adzey M	108264934...	https://www...	1	Nggk habis...	7 jam lalu	2025-05-02...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Alifa Vitalo...	111664825...	https://www...	5		8 jam lalu	2025-05-02...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Purwanto R...	115493234...	https://www...	5		10 jam lalu	2025-05-02...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	محمد اوفارس	110156245...	https://www...	5	جاء زور ووجه	15 jam lalu	2025-05-02...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	rini angraini	100540299...	https://www...	4	Kesini pas l...	1 hari lalu	2025-04-09...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Muhamad ...	108720020...	https://www...	5		1 hari lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	ka na	116168179...	https://www...	2		1 hari lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Egga Putri	104193374...	https://www...	5		1 hari lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Fahad	107138641...	https://www...	4	يحيى اول مرة	1 hari lalu	2019-08-25...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Boby Brahim	108452118...	https://www...	5	Always feel...	1 hari lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	abadi kaaby	112044237...	https://www...	5		1 hari lalu	2025-05-01...	
ChIJg-7mF...	Taman Saf...	ChZDSUh...	https://www...	Jasmine putri	100735987...	https://www...	5		1 hari lalu	2025-05-01...	

Gambar 3. 5 Input Data Google Maps Extractor

Gambar 3.4 adalah *output* dari aplikasi *Google Maps Extractor* yang menampilkan *Detailed Reviews* dari berbagai tempat, termasuk Taman Safari. *Output* ini memberikan informasi terperinci mengenai ulasan, mencakup *Place Id* dan *Place Name* untuk kejelasan lokasi, serta *Review Id* untuk melacak setiap ulasan. Terdapat pula *Review Link* yang memungkinkan akses langsung ke ulasan, serta nama dan ID pengulas yang memberikan pemberi ulasan. *Rating* yang diberikan mencerminkan tingkat kepuasan pengunjung, sedangkan *Review Text* menyajikan *insights* tentang pengalaman mereka. Tanggal publikasi ulasan tersaji untuk memahami waktu *feedback*, serta bagian *Responses* menunjukkan adanya tanggapan dari pemilik tempat. Informasi ini sangat berharga untuk analisis sentimen dan pengembangan rekomendasi berbasis data, membantu meningkatkan pengalaman pengunjung di Taman Safari dan memperdalam pemahaman tentang kepuasan pelanggan.

3.4 Teknik Analisis Data

Analisis data dilakukan melalui beberapa tahapan, dimulai dengan prapemrosesan data untuk memastikan kualitas data yang baik. Tahap ini mencakup pembersihan teks, tokenisasi, serta *stemming* untuk menghilangkan elemen yang tidak relevan dan memastikan teks siap dianalisis. Selanjutnya, data teks akan direpresentasikan dalam bentuk

vektor menggunakan metode seperti *TF-IDF* atau *Word2Vec* agar dapat digunakan dalam model analisis sentimen dan rekomendasi.

Untuk analisis sentimen, beberapa algoritma *machine learning*, yaitu *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)*, *Logistic Regression*, *Naive Bayes*, dan *Random Forest* digunakan untuk mengklasifikasikan sentimen ulasan pengunjung (positif, negatif, atau netral). *KNN* akan menghitung jarak antara data *input* dan data latih menggunakan jarak *Euclidean*, kemudian mengklasifikasikan ulasan berdasarkan mayoritas kelas dari tetangga terdekat. Algoritma lainnya, seperti *SVM* dan *Logistic Regression*, akan menganalisis pola kata atau frasa yang relevan dalam teks untuk meningkatkan akurasi klasifikasi sentimen.

Pada bagian rekomendasi, *Term Frequency* diterapkan untuk memberikan rekomendasi ulasan yang relevan kepada pengguna berdasarkan ulasan yang sebelumnya dipilih atau diberi peringkat. Dengan menghitung kesamaan antara ulasan baru dan ulasan yang ada menggunakan metrik seperti *Cosine Similarity*, sistem dapat menyarankan ulasan yang paling sesuai dengan preferensi pengguna. Pengguna juga akan dibangun profilnya berdasarkan ulasan yang telah diberikan peringkat, dan rekomendasi akan disesuaikan dengan profil tersebut.

