

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terkait

Rujukan dari penelitian-penelitian sebelumnya sangat penting untuk memperkaya dan mendalami hasil penelitian ini. Berikut beberapa penelitian terdahulu yang relevan dengan penelitian ini.

Tabel 2.1 Penelitian Terkait

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
Aditya Hastami Ruger, M Suyanto, Mei P Kurniawan	Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma Naive Bayes	JIFOTECH (JOURNAL OF INFORMATIO N TECHNOLOGY ) Vol. 1, No. 2, September 2021, hlm. 26-29 [19]	Menganalisis jumlah ulasan positif dan negatif dalam grup Shopee di Twitter dengan algoritma <i>Naive Bayes</i> .	Penggunaan algoritma Naive Bayes dalam analisis sentimen pada data Shopee di Twitter menghasilkan akurasi sebesar 97%.
Dany Pratmanto, Rousyati Rousyati, Fanny Fatma Wati, Andrian Eko Widodo, Suleman Suleman, Ragil Wijianto	App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm	Journal of Physics: Conference Series Ser. 1641, 2020 [20]	Membantu pengelola Shopee terhadap pendapat positif atau negatif pengguna aplikasi dan dapat menyediakan tinjauan teorii yang berhubungan dengan algoritma <i>Naive Bayes</i> .	Berdasarkan pengujian menggunakan Algoritma <i>Naive Bayes</i> menggunakan data ulasan sentimen dari aplikasi Shopee di Playstore, dari 200 ulasan berisi 100 ulasan positif dan 100 ulasan negatif. Penerapan <i>data</i>

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
				<p><i>mining</i> dapat membantu dalam mengkategorikan review sentimen aplikasi Shopee. Algoritma <i>Naive Bayes</i> dengan teknik <i>partitioning</i> menghasilkan akurasi sebesar 96,667%, presisi 100%, recall 93,33%, dan AUC 1,00, sehingga termasuk dalam klasifikasi yang sangat baik. Analisis sentimen dengan <i>Naive Bayes</i> terbukti efektif dalam menghasilkan nilai akurasi yang tinggi</p>
Arif Rahman, Ema Utami, Sudarmawan	Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan	Jurnal Komtika (Komputasi dan Informatika), Vol. 5 No. 1  Mei 2021 [21]	Menganalisa sentimen review aplikasi Ruangguru, Shopee, Gojek,	Akurasi dari 4 aplikasi yang digunakan dalam penelitian ini jauh lebih

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
	Algoritma Naive Bayes dan Algoritma Genetika		dan Tokopedia dengan algoritma <i>Naive Bayes</i> dan algoritma Genetika.	tinggi daripada penelitian sebelumnya, dengan nilai akurasi untuk keempat aplikasi tersebut melebihi 90%. Hal tersebut dikarena klasifikasi <i>Naive Bayes</i> didukung oleh algoritma genetika.
Faiza R Irawan, Ahmad Jazuli, Tutik Khotimah	ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEAREST NEIGHBORS	JIKO (Jurnal Informatika dan Komputer) Vol. 5, No. 1, April 2022, hlm. 62-68 [18]	Menyederhanakan proses klasifikasi respon pengguna Gojek di Twitter dengan menggunakan metode <i>K-Nearest Neighbor</i> .	Akurasi yang dicapai adalah 79,43% dengan k=15. Sentimen positif sebanyak 47,59%, negatif 8,66%, dan netral 43,75%.
Fajar Sodik Pamungkasa, Iqbal Kharisudin	Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter	PRISMA 4(2021): 628-634 [22]	Mengetahui sentimen pendapat masyarakat mengenai pandemi Covid-19 yang disampaikan dalam sosial media dengan algoritma SVM,	Dari 10.000 data tanggapan yang diperoleh, 6.128 tanggapan memiliki sentimen positif dan 3.872 tanggapan memiliki sentimen negatif. Dengan

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
			<i>Naïve Bayes</i> , dan KNN	menggunakan algoritma <i>Naïve Bayes</i> , memperoleh akurasi 79,2%, sedangkan dengan algoritma <i>K-nearest neighbor</i> , akurasi yang dihasilkan adalah 62,1%.
Nico Nathanael Wilim, Raymond Sunardi Oetama	Sentiment Analysis about Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, and Decision Tree	IJNMT(International Journal of New Media Technology), Vol. 8, No. 1  June 2021 [23]	Menunjukkan bahwa opini publik di Twitter dapat diperiksa untuk mengetahui sentimen masyarakat pada acara bincang-bincang tv serta untuk mengkonfirmasi pemenang Penghargaan acara bincang-bincang TV dengan algoritma <i>K-Nearest Neighbor</i> , <i>Naïve Bayes Classifier</i> , dan <i>Decision Tree</i> .	Pemenang Panasonic Gobel Award diambil dari opini public di Twitter dimana ILC menang pada tahun 2018 dan kalah di tahun kemudian. Algoritma terbaik dari ketiga algoritma yang digunakan adalah algoritma KNN dengan akurasi 76,94%.

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
Tika Adilah M, Hendra Supendar, Rahayu Ningsih, Sri Muryani, Kusmayanti Solecha	Sentiment Analysis of Online Transportation Service using the Naïve Bayes Methods	Journal of Physics: Conference Series Ser. 1641, 2020 [24]	Mengelompokkan teks dalam kalimat guna mengidentifikasi opini yang terdapat pada kalimat atau dokumen tersebut, baik positif maupun negatif terhadap aplikasi Gojek dengan menggunakan metode <i>Naïve Bayes</i> .	Berdasarkan evaluasi sentimen terhadap layanan transportasi daring dengan metode <i>Naive Bayes</i> , peneliti menyimpulkan bahwa metode ini cukup efektif dalam melakukan klasifikasi <i>data mining</i> atau <i>text mining</i> . Hal ini terbukti dari akurasi yang cukup tinggi, yakni 81,00%, yang menunjukkan bahwa semua komentar di laman Instagram dapat terklasifikasi dengan baik, baik komentar negatif maupun positif.
Khofifah Diah Indarwati, Herny Februariyant	ANALISIS SENTIMEN TERHADAP KUALITAS PELAYANAN	JITASI (Jurnal Teknik Informatika dan Sistem Informasi) Vol	Melakukan penambangan teks untuk mengatasi masalah dengan	Nilai akurasi yang diperoleh dari hasil studi dengan klasifikasi <i>Naive</i>

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
	APLIKASI GO-JEK MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER	10 No 1, 2023 [17]	menganalisis sentimen opini publik terkait layanan dan tinjauan kepuasan pelanggan menggunakan algoritma <i>Naive Bayes</i> .	<i>Bayes</i> adalah 68%.
Nur Fitriyah, Budi Warsito, Di Asih I Maruddani	ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM)	JURNAL GAUSSIAN, Volume 9, Nomor 3, Tahun 2020, Halaman 376-390 [25]	Mengklasifikasi kan tanggapan pengguna Gojek ke dalam kategori sentimen positif dan negatif, dengan algoritma SVM.	Klasifikasi sentimen dengan metode SVM pada data Gojek menghasilkan akurasi keseluruhan tertinggi sebesar 79,19% dan nilai akurasi Kappa tertinggi bernilai 16,52%. Sementara itu, klasifikasi sentimen dengan pelabelan data berdasarkan skor sentimen menggunakan metode SVM pada Gojek menunjukkan akurasi keseluruhan tertinggi yang

Peneliti	Judul	Jurnal	Tujuan Penelitian	Hasil Penelitian
				sama, yaitu 79,19%, namun dengan nilai akurasi Kappa tertinggi sebesar 21%.
Rizki Wahyudi, Gilang Kusumawardhana	Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine	JURNAL INFORMATIK A, Vol. 8 No. 2 September 2021, Halaman 200~207 [6]	Melakukan analisis sentimen terhadap ulasan pengguna aplikasi Grab di Google Play Store menggunakan algoritma <i>Support Vector Machine</i> (SVM).	Analisis sentimen terhadap 900 data ulasan dengan metode SVM dengan pengujian 5-fold menghasilkan akurasi sebanyak 85,54%. Dari hasil pengujian tersebut, terdeteksi 59 ulasan positif dan 675 ulasan negatif.

Beberapa informasi yang terlihat di Tabel 2.1 merupakan penelitian terdahulu mengenai sentimen analisis dari berbagai aplikasi dengan memanfaatkan algoritma *Naive Bayes*, KNN, SVM, dan beberapa algoritma lainnya. Penelitian terdahulu tersebut menghasilkan beberapa informasi yakni pengguna aplikasi belum puas seratus persen akan aplikasi yang digunakannya sehingga aplikasi terkait dapat meningkatkan kualitasnya kembali sesuai dengan hasil dari analisis sentimen yang telah dilakukan. Tabel 2.1 memberikan gambaran rinci mengenai temuan-temuan tersebut, termasuk perbandingan kinerja masing-masing algoritma. Tinjauan kritis ini menekankan pentingnya analisis sentimen pada aplikasi transportasi online

untuk memprediksi keberlanjutannya. Oleh karena itu, mengkomparasikan ketiga algoritma, yaitu *Naïve Bayes*, KNN, dan SVM menjadi relevan dan penting untuk dilakukan guna menentukan algoritma yang paling efektif dalam analisis sentimen pada konteks ini. Hal ini didukung dengan nilai akurasi dari ketiga algoritma yang cukup tinggi.

Berbeda dengan penelitian sebelumnya, penelitian ini menitikberatkan pada keberlanjutan Gojek di pasar transportasi Indonesia berdasarkan kepuasan pelanggan Gojek terhadap layanannya. Penelitian ini memiliki peran bukan untuk memberikan *knowledge* kepada perusahaan tetapi untuk memberikan informasi kepada masyarakat agar mengetahui prediksi akan kondisi pasar Transportasi *Online* yang ada di Indonesia, khususnya Gojek. Penelitian ini juga merupakan perbandingan tiga algoritma untuk analisis sentimen yaitu *Naïve Bayes*, KNN, dan SVM. Sumber data yang digunakan pada penelitian ini adalah ulasan pelanggan Gojek pada *Google Play Store*.

## 2.2 Tinjauan Teori

### 2.2.1 Transportasi dan Transportasi Online

Transportasi merupakan sistem untuk perpindahan dari satu tempat ke tempat lain [1]. Transportasi dapat dilakukan dengan berbagai macam moda yaitu transportasi darat, air, dan udara. Transportasi menjadi kebutuhan penting bagi banyak orang karena mobilitas manusia yang terus berlangsung. Inilah yang membuat transportasi menjadi ladang bisnis yang menguntungkan.

Layanan transportasi *online* kini berkembang pesat sebagai salah satu bisnis yang cepat tumbuh. Gojek adalah salah satu penyedia jasa transportasi *online* terbesar di Indonesia [4]. Dengan layanan transportasi *online*, pengguna dapat memesan dan menggunakan transportasi melalui aplikasi di perangkat seluler seperti *smartphone* atau tablet, tanpa perlu berinteraksi langsung dengan sopir atau pergi ke stasiun atau terminal.

Layanan ini memungkinkan pengguna untuk memesan kendaraan dan melakukan pembayaran secara *online*.

### 2.2.2 Analisis Sentimen

Analisis sentimen merupakan metode mengumpulkan, mengolah, serta mengevaluasi opini atau emosi yang disampaikan melalui teks. Dalam analisis sentimen, teknik-teknik pemrosesan *Machine Learning* digunakan untuk mengidentifikasi sentimen yang terkandung dalam kalimat, baik itu positif, negatif, atau netral[21]. Tujuan dari analisis sentimen adalah untuk mendapatkan pemahaman yang lebih mendalam mengenai cara konsumen merespons produk, merek, atau peristiwa, dengan demikian, dapat mendukung perusahaan dalam membuat keputusan yang didasarkan pada data. Sebagai contoh, perusahaan dapat menggunakan analisis sentimen untuk mengevaluasi respons konsumen terhadap produk baru dan mengambil tindakan untuk meningkatkan kepuasan pelanggan.

Analisis Sentimen memiliki beberapa pendekatan, yaitu *Machine Learning, Lexicon, Embedded* [26]. Metode *machine learning* semakin banyak digunakan dalam analisis sentimen karena kemampuannya dalam mengidentifikasi sudut pandang atau sentimen dalam teks secara representatif. Di sisi lain, pendekatan berbasis *lexicon* menggunakan kumpulan kata yang telah diberi nilai polaritas guna mengevaluasi respon konsumen terhadap suatu objek. Selain itu, metode campuran yang mengombinasikan *machine learning* dan pendekatan *lexicon*, meskipun jarang diaplikasikan, sering kali memberikan hasil analisis yang lebih akurat dan efektif.

### 2.2.3 Text Mining

*Text Mining* merupakan cara analisis data yang dilakukan guna mengekstraksi informasi berharga dari teks atau dokumen [27]. *Text*

*mining* yang melibatkan algoritma komputasi diperlukan untuk mengekstrak informasi penting dari dokumen seperti topik, kata kunci, pola, dan hubungan antar kata. Selain itu, *text mining* juga berguna untuk menganalisis sentimen dan opini dalam teks [17]. Berbagai bidang seperti bisnis, kesehatan, media sosial, dan ilmu pengetahuan sosial telah memanfaatkan *text mining*.

#### 2.2.4 *Confusion Matrix*

*Confusion matrix* ialah alat untuk mengevaluasi performa model klasifikasi dengan menghitung akurasi setiap model yang diaplikasikan [28]. Matriks ini mencakup empat jenis nilai, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Nilai-nilai tersebut merepresentasikan jumlah data yang benar-benar positif, benar-benar negatif, salah diprediksi sebagai positif, dan salah diprediksi sebagai negatif oleh model. Metrik evaluasi kinerja dapat dihitung berdasarkan nilai-nilai tersebut [28]. *Confusion matrix* menjadi penting dalam evaluasi model klasifikasi dan membantu pengembang model dalam meningkatkan performanya. Metrik evaluasi kinerja pada *Confusion Matrix* memiliki rumus yang berbeda-beda dalam perhitungannya. Rumus dari metrik evaluasi kinerja pada *Confusion Matrix* antara lain sebagai berikut:

Keterangan :

TP : True Positive

FP : False Positive

FN : False Negative

TN : True Negative

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN}$$

Rumus 2.1 Rumus Akurasi Pada *Confusion Matrix* [29]

$$Precision = \frac{TP}{TP + FP}$$

Rumus 2.2 Rumus *Precision* Pada *Confusion Matrix* [29]

$$Recall = \frac{TP}{TP + FN}$$

Rumus 2.3 Rumus *Recall* Pada *Confusion Matrix* [29]

$$F1-Score = \left( \frac{Precision \cdot recall}{precision + recall} \right)$$

Rumus 2.4 Rumus *F1-Score* Pada *Confusion Matrix*

### 2.2.5 Web Scraping

*Web Scraping* merupakan proses yang dilakukan oleh sebuah program untuk mengekstrak data dari halaman web secara sistematis [6]. Proses *web scraping* dilakukan secara otomatis oleh sebuah program yang disebut *web scraper*. *Web scraping* sering digunakan untuk mengumpulkan data dari berbagai halaman web, seperti data produk dari situs e-commerce, data harga saham dari situs keuangan, atau data berita dari situs media. Data yang berhasil dikumpulkan kemudian dapat diolah dan dianalisis untuk berbagai tujuan, seperti riset pasar, pengembangan produk, atau analisis tren. Namun, beberapa halaman web membatasi atau melarang penggunaan *web scraping*, sehingga etika dan kebijakan penggunaan data harus selalu diperhatikan.

## 2.3 Framework dan Algoritma yang digunakan

### 2.3.1 CRISP-DM (Cross Industry Standard Process – Data Mining)

CRISP-DM (*Cross Industry Standard Process – Data Mining*) merupakan model standar pembangunan proyek *data mining* dan *knowledge discovery* yang memiliki keunggulan dalam memecahkan

masalah yang ada pada proyek penambangan data. CRISP-DM memiliki 6 tahapan [30] yaitu:

1. *Business Understanding*

Proses ini merupakan proses memahami bisnis secara mendalam mulai dari apa yang ingin dicapai, bagaimana cara melakukannya, hingga model apa yang akan diterapkan dalam pemrosesan data.

2. *Data Understanding*

Proses ini adalah proses pemahaman data yang ada mulai dari kelebihan dan kekurangannya sebelum data yang ada diproses.

3. *Data Preparation*

Proses ini merupakan proses mempersiapkan data. Data yang kurang bagus akan diperbaiki. Misalnya data yang dimiliki masih memiliki missing value atau nilai *null*, maka nilai *null* yang ada pada data harus dihilangkan terlebih dahulu agar pemrosesan data menjadi lebih mudah. Selain itu, data akan melalui *proses sampling* yang kemudian akan menjadi dua yaitu *data training* dan *data testing*.

4. *Modeling*

Proses ini merupakan proses mengolah data yang telah diperbaiki pada data preparation dengan menggunakan metode dan model yang telah ditentukan.

5. *Evaluation*

Proses ini merupakan proses mengevaluasi data yang telah diproses pada tahap modeling. Jika hasil evaluasi kurang baik, maka data akan dikembalikan dan akan diproses ulang mulai dari business understanding sampai evaluation kembali. Jika hasil evaluasi baik, maka data akan melalui proses deployment.

6. *Deployment*

Proses ini merupakan proses mengevaluasi data yang telah diproses pada tahap modeling. Jika hasil evaluasi kurang baik,

maka data akan dikembalikan dan akan diproses ulang mulai dari business understanding sampai evaluation kembali. Jika hasil evaluasi baik, maka data akan melalui proses deployment.

Proses ini merupakan proses penggunaan model dari data yang telah diproses dan dievaluasi sebelumnya.

### 2.3.2 *Naïve Bayes*

*Naïve Bayes* adalah salah satu teknik yang dilakukan untuk klasifikasi data [19]. Klasifikasi *Bayesian* adalah pendekatan statistik yang mampu memperkirakan probabilitas suatu data termasuk dalam kelas tertentu [7]. Teori ini disampaikan oleh Thomas Bayes dimana *Naïve Bayes* yang digunakan untuk memperkirakan probabilitas di masa yang akan datang sesuai dengan pengalaman atau data yang ada di masa lalu [16]. Dengan supervised learning, *Naïve Bayes* dapat dilatih secara efisien. Terdapat keuntungan dari teknik klasifikasi ini yaitu kebutuhan akan data latih yang cukup minim untuk menghitung parameter yang diperlukan pada proses klasifikasi [31].

Cara untuk memahami algoritma *Naive Bayes* adalah dengan menggunakan Teorema Bayes. Teorema Bayes merupakan sebuah rumus matematika yang digunakan untuk menentukan probabilitas suatu peristiwa dengan memanfaatkan informasi atau bukti yang tersedia. Rumus dari Teorema Bayes adalah:

$$P(B) = \frac{P(A) \cdot P(A)}{P(B)}$$

Rumus 2. 5 Rumus Teorema Bayes [32]

Keterangan:

$P(A|B)$  : probabilitas kejadian A terjadi, jika diketahui kejadian B telah terjadi

$P(B|A)$  : probabilitas kejadian B terjadi, jika diketahui kejadian A telah terjadi

$P(A)$  : probabilitas kejadian A terjadi secara umum

$P(B)$  : probabilitas kejadian B terjadi secara umum

Teorema Bayes banyak diterapkan di berbagai bidang, termasuk statistik, ilmu komputer, dan kecerdasan buatan. Salah satu contohnya adalah dalam pembelajaran mesin, di mana teorema Bayes dapat digunakan untuk memperkirakan probabilitas suatu data termasuk ke dalam suatu kelas tertentu berdasarkan fitur-fiturnya.

### 2.3.3 *K-Nearest Neighbors* (KNN)

Algoritma *K-Nearest Neighbors* adalah teknik klasifikasi yang bergantung pada label kategori dari dokumen pelatihan, dengan menggunakan contoh dasar tanpa adanya representasi deklaratif eksplisit dari kategori tersebut [28]. Metode ini bergantung pada kesamaan antara dokumen pelatihan dan dokumen tes dalam hal label kategori yang melekat pada dokumen tersebut [28]. Algoritma ini bekerja dengan mencari sejumlah  $k$  pola yang paling dekat dengan pola masukan, lalu menentukan kelas keputusan berdasarkan kelas yang paling dominan di antara  $k$  pola tersebut [22].

Algoritma *K-Nearest Neighbors* menggunakan *Euclidean Distance* untuk menghitung jarak antara dua titik. Metode tersebut dapat diterapkan pada ruang satu dimensi, dua dimensi, maupun ruang berdimensi lebih tinggi. Formula *Euclidean distance* adalah sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Rumus 2.6 Rumus Euclidean Distance[33]

Keterangan:

dis(x,y) : Jarak Sampel

x<sub>i</sub> : Data Sampel Pengetahuan

y<sub>i</sub> : Data input var

n : Jumlah Sampel

#### 2.3.4 Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah cara klasifikasi dalam penambangan data. Digunakan pemetaan non-linear pada Teknik ini guna mengganti data latih ke dimensi yang lebih tinggi. Pada dimensi tersebut, SVM berusaha menemukan *hyperplane* yang dapat memisahkan data secara linier [8]. SVM beroperasi dengan mencari *hyperplane* yang dapat memperbesar margin, yaitu jarak antara titik data yang paling dekat. Titik-titik data yang berada paling dekat dengan *hyperplane* disebut *support vectors*, dan mereka memiliki peran krusial dalam menentukan posisi serta orientasi *hyperplane* tersebut. Dengan memaksimalkan margin ini, SVM berusaha untuk meningkatkan generalisasi model pada data baru.

Ketika data tidak dapat dipisahkan secara linear, SVM menerapkan teknik pemetaan non-linear untuk memproyeksikan data ke dimensi yang lebih tinggi sehingga memungkinkan pemisahan secara linear. Teknik ini dikenal dengan istilah "*Kernel Trick*" [34], yang memungkinkan SVM untuk bekerja dengan berbagai jenis data dan pola. Kernel yang sering dipakai antara lain adalah linear, polynomial,

*Radial Basis Function* (RBF), dan sigmoid. Dengan menggunakan kernel ini, SVM mampu menangani berbagai masalah klasifikasi dengan batas pemisah yang kompleks dan non-linear, menjadikannya alat yang kuat dan fleksibel dalam bidang pembelajaran mesin.

Penelitian ini dalam pemrosesan SVM menggunakan kernel linear. Kernel linear adalah jenis kernel yang diterapkan dalam SVM pada data yang dapat dipisahkan secara linear atau ketika hubungan antara fitur bersifat linier [34]. Adapun rumus kernel linear SVM sebagai berikut.

$$K(x_i, x_j) = x_i^T \cdot x_j$$

Rumus 2.7 Rumus Kernel Linear SVM [25]

Keterangan:

$K(x, y)$  : Nilai yang dikalikan antara kedua vektor

“.” : Operasi dot product

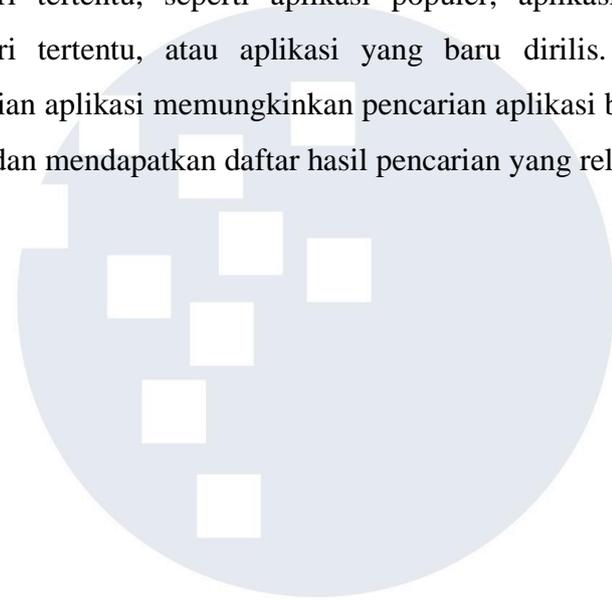
## 2.4 Tools yang digunakan

### 2.4.1 Python

*Python* adalah bahasa pemrograman yang lahir pada tahun 1990. Bahasa ini populer karena memiliki sintaks yang sederhana, mudah dipahami, dan didukung oleh komunitas yang besar [35]. *Python* digunakan secara luas di berbagai bidang, seperti pengembangan web, *data science*, kecerdasan buatan, dan web scraping. Khusus untuk *web scraping*, *Python* menjadi pilihan utama karena sintaksnya yang mudah, pustaka yang kaya fitur, serta dukungan komunitas yang solid.

Pada penggunaan *Python* pada *web scraping*, digunakan pustaka yang dibutuhkan untuk mempermudah aktivitas *scraping* yaitu *google-play-scraper*. *Library google-play-scraper* memiliki beberapa fitur utama yang sangat berguna untuk mengambil data dari *Google Play Store*. Pertama, *library* ini dapat mengambil informasi lengkap

mengenai sebuah aplikasi, termasuk nama aplikasi, pengembang, deskripsi, rating, jumlah unduhan, dan lain-lain. Kedua, fitur ulasan pengguna memungkinkan pengambilan ulasan pengguna untuk aplikasi tertentu, termasuk teks ulasan, rating yang diberikan oleh pengguna, tanggal ulasan, dan informasi pengguna jika tersedia. Selain itu, *google-play-scraper* juga dapat mengambil daftar aplikasi berdasarkan kategori tertentu, seperti aplikasi populer, aplikasi teratas dalam kategori tertentu, atau aplikasi yang baru dirilis. Terakhir, fitur pencarian aplikasi memungkinkan pencarian aplikasi berdasarkan kata kunci dan mendapatkan daftar hasil pencarian yang relevan.



UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA