

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Sentimen

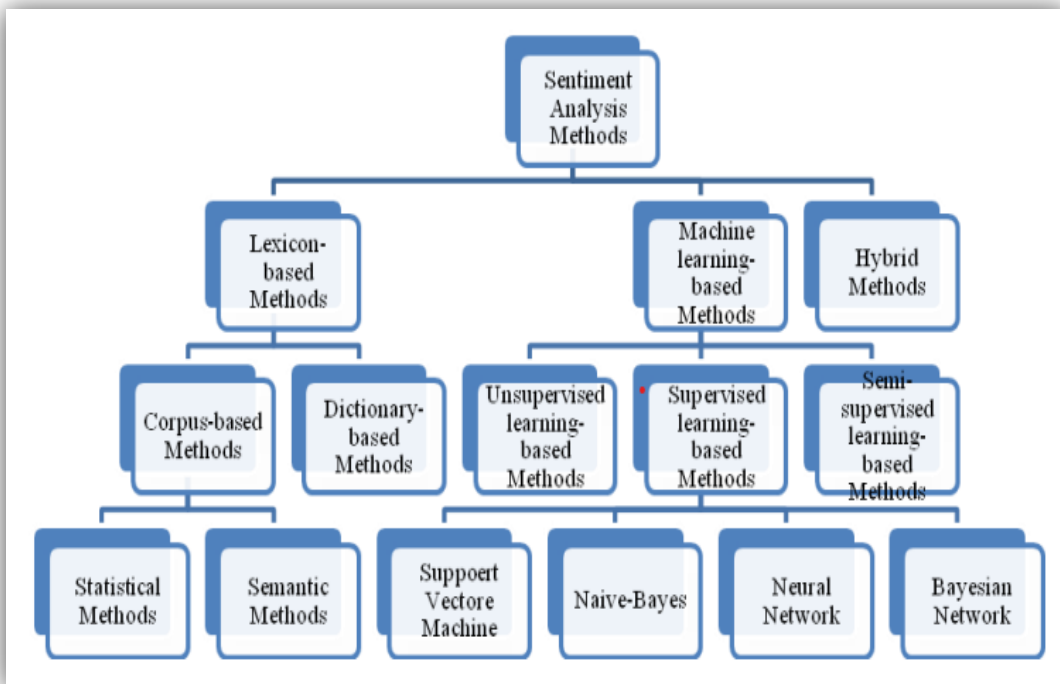
Analisis sentimen adalah bidang pemrosesan bahasa alami (*NLP*) yang membangun algoritma kognitif dan inferensial ke dalam format data. Informasi berbasis teks kini banyak tersedia secara *online* dalam bentuk forum diskusi dan situs ulasan. Dengan menggunakan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diubah menjadi data yang sangat terstruktur. Data ini dapat menggambarkan sikap masyarakat terhadap produk, merek, layanan, politik, dan topik lainnya. Perusahaan, pemerintah, dan lembaga lain menggunakan data ini untuk melakukan analisis penjualan, ulasan produk, evaluasi produk, dan layanan sosial [6].

Analisis sentimen juga dikenal opini mining adalah data yang mengevaluasi opini, perasaan, penilaian, sikap, dan perasaan seseorang untuk menentukan apakah pembicara atau penulis konsisten dengan pesan yang mereka maksudkan tentang topik, produk, layanan, organisasi, orang, atau aktivitas tertentu [7].

Analisis sentimen merupakan menganalisis data dalam format teks untuk memahami opini dan sentimen yang terkandung di dalamnya. Tujuan analisis sentimen adalah untuk menentukan apakah suatu teks (seperti ulasan, komentar, atau postingan) mengandung pesan positif, negatif tentang suatu topik, produk, layanan, atau hal lainnya.

2.1.1 Pendekatan Analisis Sentimen

Ada tiga pendekatan utama dalam analisis sentimen, yaitu pendekatan *Machine Learning*, *Lexicon Based* dan *Hybrid* [8].



Gambar 2.1 Pendekatan Analisis Sentimen

a. Pendekatan *Machine Learning*

Pendekatan ini menggunakan algoritma pembelajaran mesin untuk menganalisis dan mengklasifikasikan sentimen teks, seperti ulasan dan komentar pengguna. *Machine learning* dibagi menjadi tiga tipe, yaitu *supervised learning*, *unsupervised learning* dan *reinforcement learning*. *supervised learning* merupakan jenis pembelajaran yang melatih model pada kumpulan data berlabel. Adapun algoritma yang di gunakan dalam pembelajaran ini yaitu *Support Vector Machine* (SVM), *Naive Bayes* (NB), *decision Tree* (DT) dan lain-lain [9].

b. Pendekatan *Lexicon Based*

Pendekatan *Lexicon Based* merupakan suatu metode analisis sentimen yang menggunakan kamus yang berisi daftar kata-kata yang mengandung opini. Setiap kata dalam kamus mempunyai prioritas dan diberi nilai negatif dan positif.

c. Pendekatan *Hybrid*

Pendekatan *Hybrid* adalah pendekatan yang memperkuat kinerja analitis dengan memanfaatkan kekuatan berbagai metode[10].

2.2 Jaminan Kesehatan Nasional (JKN)

Jaminan kesehatan adalah bentuk perlindungan yang memastikan agar peserta memperoleh manfaat dalam pemeliharaan dan perawatan kesehatan untuk memenuhi kebutuhan dasar mereka. Jaminan ini diberikan kepada setiap individu yang telah membayar iuran, atau iuran tersebut dibayarkan oleh pemerintah[11]. Program JKN dirancang untuk memberikan perlindungan kepada semua peserta serta menjamin manfaat pelayanan kesehatan. Dengan demikian, diharapkan peserta JKN akan dapat meningkatkan produktivitas mereka, yang pada gilirannya dapat meningkatkan kesejahteraan hidup mereka.

JKN menawarkan berbagai keuntungan bagi anggotanya dengan biaya premi yang ekonomis. Program ini memastikan dukungan keuangan untuk kesehatan yang berkelanjutan serta menyediakan layanan berkualitas seperti imunisasi dasar, program keluarga berencana, dan pemeriksaan kesehatan. Salah satu keunggulan utama JKN adalah aksesibilitasnya yang meliputi seluruh wilayah Indonesia, sehingga peserta dapat dengan mudah memperoleh layanan kesehatan di mana pun mereka berada[12].

Di zaman digital saat ini, BPJS Kesehatan meluncurkan aplikasi JKN *mobile* guna mempermudah peserta Jaminan Kesehatan Nasional dalam mengakses layanan kesehatan. Aplikasi ini memberikan kemudahan bagi peserta untuk memeriksa status keanggotaan, menemukan rumah sakit terdekat, serta melakukan pendaftaran secara daring. Sejak diluncurkan, JKN *mobile* telah menerima sambutan yang baik dari masyarakat dan terus berkembang dengan penambahan fitur-fitur baru yang disesuaikan dengan kebutuhan para peserta. Hal ini menjadikannya sebagai sebuah inovasi yang signifikan dalam sejarah layanan kesehatan di Indonesia[13].



Gambar 2.2 Tampilan Menu *Mobile* JKN

2.3 Google Play Store

Google Play Store merupakan sebuah platform resmi yang disediakan oleh *Google*, yang sering digunakan untuk mengunduh dan membeli aplikasi, game, buku, serta musik pada perangkat *android*. *Google play store* menyediakan fitur ulasan, sistem penilaian di *Google Play Store* memberikan kesempatan kepada pengguna untuk memberikan umpan balik atau penilaian terhadap aplikasi yang telah mereka *install*. Setiap aplikasi umumnya memiliki bagian untuk memberikan ulasan, di mana pengguna dapat menuliskan pengalaman mereka, baik yang bersifat positif maupun negatif. Sistem ini memberikan dukungan kepada calon pengguna lainnya dalam mengambil keputusan apakah akan mengunduh aplikasi tersebut atau tidak.

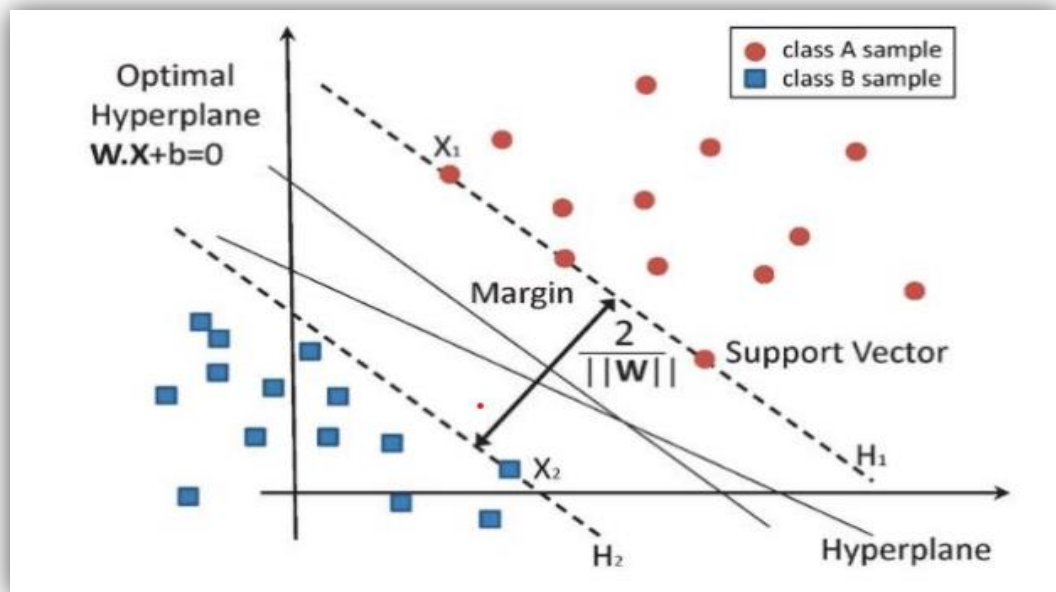
Terdapat dua kategori *review* pengguna menurut Kamal & Ratnasari (2021), yaitu *search product* dan *experience product*. *Search product* merupakan standar suatu produk dapat dievaluasi secara obyektif berdasarkan detail spesifikasinya, bahkan sebelum kita melakukan pembelian atau menggunakannya. Contohnya seperti pemilihan ponsel, Pengguna dapat mengamati spesifikasi kamera, daya tahan baterai, serta fitur-fitur tambahan lainnya sebelum akhirnya memutuskan untuk membeli. Sedangkan *Experince Product* merupakan produk yang tingkat kualitasnya lebih susah untuk dievaluasi sebelum digunakan. Dalam konteks ini, pengalaman pengguna merupakan elemen kunci yang mempengaruhi tingkat kepuasan mereka. Contohnya seperti aplikasi *mobile*, dimana pengguna dapat memberikan *review* setelah mereka menggunakan aplikasi tersebut[14].

Pada penelitian ini penulis mengambil data ulasan dari *google play store* yaitu data ulasan dari aplikasi *mobile JKN*. Dimana ulasan merupakan salah satu tempat pengguna aplikasi dapat memberikan penilaian serta masukan untuk aplikasi yang telah mereka gunakan. Dan tentunya ulasan yang ada sangat berharga untuk pengembang aplikasi agar mereka bisa meningkatkan kualitas aplikasi yang sudah diluncurkan.

2.4 Support Vector Machine

Dalam konteks konferensi Teori Pembelajaran (*COLT*), Boser Bernhard, Guyon dan Vapnik merupakan tokoh-tokoh terkemuka di bidang pembelajaran mesin dan teori statistik. Pada tahun 1992 konsep *Support Vector* diperkenalkan[15]. *Support Vector Machine (SVM)* adalah metode pembelajaran mesin yang mengklasifikasikan data ke dalam dua kelompok. Metode ini dilatih menggunakan algoritma yang berlandaskan teori, dan dalam konteks analisis sentimen, *SVM* digunakan untuk mengelompokkan sentimen berdasarkan frekuensi kata dalam dokumen. *SVM* termasuk dalam kategori *supervised learning*, di mana model dilatih menggunakan data berlabel.

SVM merupakan salah satu pendekatan yang telah banyak digunakan dalam berbagai studi di bidang data. *SVM* memiliki potensi untuk menghasilkan hasil yang lebih baik dalam proses penambangan data. Metode ini bertujuan untuk menentukan *hyperplane* optimal yang mempunyai *margin* maksimum. *Margin* dapat didefinisikan sebagai jarak antara titik terdekat dari setiap kelas atau *support vector* dengan *hyperplane*.



Gambar 2.3 Klasifikasi *Support Vector Machine*

Menurut Bhavsar & Panchal (2012), fungsi kernel yang biasa digunakan dalam *SVM* ada tiga yaitu, kernel *Linear*, Kernel *Radial Bias Function (RBF)* dan Kernel *Polynomial*[16]. *SVM Linear* merupakan alat pembelajaran mesin yang sangat baik, karena beroperasi secara signifikan lebih cepat daripada *SVM* kernel *nonlinier* sekaligus memberikan akurasi yang sebanding untuk tugas berdimensi tinggi seperti klasifikasi dokumen.

2.5 Pembagian Data Klasifikasi

Splitting adalah tahap pembagian dataset kedalam dua bagian yaitu *training* dan *testing*. *Training* data digunakan untuk mengembangkan dan melatih sistem klasifikasi, sementara *testing* data dimanfaatkan untuk mengevaluasi performa sistem klasifikasi yang telah dilatih dengan *training* data tersebut [17].

Pengujian dilakukan dengan membandingkan rasio data latih dan data uji dengan perbandingan 80:20. Persentase pertama pada setiap perbandingan merupakan data latih dan persentase selanjutnya merupakan data uji.

2.6 Confusion Matrix

Confusion matrix merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Tabel pada *confusion matrix* menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah diklasifikasikan.

ACTUAL CLASS	NEGATIF	TN = NETRAL TP = NEGATIF FN = POSITIF	FP = NETRAL FN = NEGATIF	FN = NEGATIF FP = POSITIF
	NETRAL	TN = NETRAL FP = NEGATIF	TP = NETRAL TN = NEGATIF FN = POSITIF	TN = NETRAL FP = POSITIF
	POSITIF	FP = NETRAL FN = POSITIF	FP = NETRAL FN = POSITIF	TN = NETRAL TN = NEGATIF TP = POSITIF
		NEGATIF	NETRAL	POSITIF
		PREDICTED CLASS		

Gambar 2.4 Confusion Matrix SVM

Dari gambar diatas didapatkan penjelasan yaitu sebagai berikut :

1. *True Positif (TP)* : Jumlah prediksi yang benar untuk kelas positif. Artinya model memprediksi positif dan nilai aktualnya juga positif.

2. *False Positif (FP)* : Jumlah kasus di mana model memprediksi kelas positif tetapi nilai aktualnya negatif (atau netral). Ini menunjukkan kesalahan dalam memprediksi.
3. *True Negatif (TN)* : Jumlah prediksi yang benar di mana model tidak memprediksi kelas positif, dan nilai aktualnya juga tidak positif (negatif atau netral). Ini menunjukkan bahwa model dengan benar mengidentifikasi kasus non-positif.
4. *False Negatif (FN)* : Jumlah kasus di mana model gagal memprediksi kelas positif, yaitu model memprediksi kelas lain (seperti negatif atau netral) padahal seharusnya positif.

Dengan menggunakan data yang dikumpulkan, kita dapat melakukan perhitungan untuk mengevaluasi kinerja model klasifikasi dengan cara berikut[18] :

1. *Accuracy* yaitu hasil dari jumlah data yang diprediksi mempunyai nilai positif dari total keseluruhan prediksi yang sudah dilakukan.
2. *Recall (true positif rate)* yaitu persentase data positif yang teridentifikasi dengan benar dibandingkan dengan seluruh data positif yang tersedia dalam proses klasifikasi. Matriks evaluasi ini digunakan untuk mengukur seberapa akurat model dalam melakukan klasifikasi, baik untuk kategori positif maupun negatif.
3. *Precision* yaitu perbandingan antara data yang teridentifikasi sebagai positif dengan keseluruhan data yang bernilai positif.

4. *F1-Score* yaitu Perhitungan nilai rata-rata yang mempertimbangkan *precision* dan *recall* secara bersamaan untuk mengevaluasi seberapa baik performa suatu metode dalam melakukan klasifikasi.

2.7 Python

Python adalah bahasa pemrograman tingkat tinggi yang diciptakan oleh Guido Van Rossum dan dirilis pada tahun 1991. Saat ini, *Python* sangat populer dan dikenal sebagai bahasa yang serbaguna. Misalnya, *Python* dapat digunakan dalam bidang *Machine Learning* dan *Deep Learning*. Bahasa ini dipilih untuk penelitian karena sintaksisnya yang sederhana, serta memiliki banyak *library* yang lengkap dan dukungan komunitas yang kuat, dari sifatnya yang *open source*. Menulis kode *Python* bisa menggunakan berbagai IDE, seperti *VS Code*, *Sublime Text* dan lainnya. Atau menggunakan IDE online seperti *Jupyter Notebook* dan *Google Collab*.

2.7.1 Pustaka Python

Pustaka adalah sekumpulan kode yang sering digunakan dan dapat dimanfaatkan oleh pengembang dalam program *Python*, sehingga mereka tidak perlu menulis kode dari awal. Secara bawaan, *Python* dilengkapi dengan pustaka standar yang berisi banyak fungsi yang bisa digunakan kembali. Selain itu, terdapat lebih dari 137.000 pustaka *Python* yang tersedia untuk berbagai aplikasi, termasuk pengembangan *web*, ilmu data, dan *machine learning (ML)*[22]. Adapun beberapa pustaka dari python yaitu sebagai berikut :

1. *Pandas*

Pandas adalah *library Python* yang paling terkenal untuk ilmu data. *Library* ini memungkinkan pengguna untuk memuat data, memanipulasinya, menghitung berbagai properti statistik, dan bahkan menangani data duplikat. Selain itu, *Pandas* menyediakan struktur data yang cepat, fleksibel, dan mudah dimengerti, sehingga memudahkan dalam bekerja dengan data relasional dan terstruktur.

2. *Numpy*

NumPy (Numerical Python) adalah *library Python* yang digunakan untuk bekerja dengan *array* dan melakukan berbagai operasi matematis, seperti aljabar linier dan transformasi *Fourier*. Meskipun *Python* memiliki tipe data list untuk *array*, *NumPy* jauh lebih cepat, bisa hingga 50 kali lipat, sehingga sangat berguna untuk pengolahan data yang membutuhkan kecepatan.

3. *Matplotlib*

Matplotlib adalah salah satu *library Python* yang paling baik dan sering digunakan oleh ilmuwan data untuk visualisasi data, seperti membuat *plot*, grafik, *histogram* dan lainnya.

4. *Google Play Scraper*

Google Play Scraper adalah alat yang digunakan untuk mengambil data dari *Google Play Store*. Dengan ini kita bisa mendapatkan berbagai informasi tentang aplikasi, seperti nama, deskripsi, kategori, *rating* dan lainnya.

5. *NLTK*

NLTK, atau *Natural Language Toolkit*, adalah *library Python* yang digunakan untuk pemrosesan bahasa alami (*NLP*). *Library* ini menyediakan berbagai alat dan sumber daya untuk membantu dalam analisis teks, seperti *tokenisasi*, penghilangan *stopwords*, *stemming*, dan lainnya.

6. *Sastrawi*

Sastrawi adalah pustaka *Python* sederhana yang membantu menyederhanakan kata-kata yang mengalami *infleksi* dalam bahasa Indonesia ke bentuk dasarnya.

6. *Seaborn*

Seaborn adalah *library* digunakan oleh ilmuwan data untuk membuat visualisasi data yang lebih mudah dipahami. *Seaborn* membantu dalam membaca dan memahami data dengan menampilkan informasi dalam bentuk grafik statistik yang informatif, sehingga memudahkan analisis dalam proyek *machine learning*.

7. *Scikit Learn*

Scikit-learn adalah salah satu *library* terpenting dalam *Python*. Setelah membersihkan dan memanipulasi data dengan *Pandas* atau *NumPy*, *Scikit-learn* digunakan untuk membangun model *machine learning*. *Library* ini menawarkan banyak alat yang berguna untuk pemodelan prediktif dan analisis data.

8. *Library re(RegEx)*

Library ini digunakan untuk mencocokkan, mencari, mengganti, dan memanipulasi teks berdasarkan pola tertentu.

9. *Pillow*

Pillow merupakan *library* canggih untuk pemrosesan gambar dalam *Python*. *Library* ini memungkinkan untuk melakukan berbagai operasi pada gambar, seperti mengubah ukuran, memotong, dan memutar. *Pillow* merupakan pengembangan dari *Python Imaging Library (PIL)*, yang juga digunakan untuk mengolah gambar. Dengan *Pillow*, dapat dengan mudah mengedit dan memanipulasi gambar sesuai kebutuhan.

10. *Date time*

Date time merupakan sebuah *library* atau modul yang dipanggil jika anda membutuhkan segala operasi yang berhubungan demi waktu.

2.8 *Machine Learning*

Machine Learning atau pembelajaran mesin adalah cabang dari kecerdasan buatan yang digunakan untuk meniru cara manusia menyelesaikan masalah dan otomatisasi. *Machine Learning* berusaha meniru proses belajar manusia dengan menggeneralisasi informasi. Terdapat dua aplikasi utama dalam *Machine Learning*, yaitu klasifikasi dan prediksi. Proses pelatihan atau *training* sangat penting dalam *Machine Learning*, karena mesin membutuhkan data untuk belajar, yang disebut data *training*. Beberapa metode *Machine Learning* yang populer termasuk sistem pengambil keputusan, *Support Vector Machine (SVM)*, dan *Neural Network*[23]. *Machine learning* terbagi menjadi tiga kategori yaitu sebagai berikut[24]:

1. *Supervised Learning*, menggunakan metode klasifikasi di mana semua data sudah diberi label untuk membantu mengidentifikasi kelas yang belum diketahui.
2. *Unsupervised Learning*, juga dikenal sebagai *clustering*, tidak memerlukan label pada data. Hasilnya tidak mengidentifikasi contoh dalam kelas yang sudah ditentukan.
3. *Reinforcement Learning*, berada di antara *Supervised* dan *Unsupervised Learning*. Teknik ini beroperasi dalam lingkungan yang dinamis dan berusaha mencapai tujuan tanpa pemberitahuan eksplisit dari komputer tentang pencapaian tujuan tersebut.

2.9 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Peneliti	Judul	Hasil Pembahasan
Tinaliah & Triana	Analisis Sentimen	Berdasarkan hasil pengujian dapat disimpulkan bahwa <i>SVM</i> dapat melakukan analisis sentimen
Elizabeth (2022)	Ulasan Aplikasi Primaku Menggunakan Metode Support Vector Machine.	ulasan pengguna aplikasi PrimaKu dengan baik menggunakan <i>linear</i> kernel dengan nilai akurasi, yaitu 97.5% dibandingkan dengan menggunakan <i>polynomial</i> kernel atau <i>RBF</i> kernel.[19][1]
Hendriyanto et al (2022)	Analisis Sentimen Ulasan Aplikasi	Hasil penelitian mengenai analisis sentimen terhadap aplikasi MOLA menunjukkan bahwa dari 520 ulasan yang dianalisis, terdapat 312

	<p>Mola Pada <i>Google Play Store</i> Menggunakan Algoritma <i>Support Vector Machine</i></p>	<p>ulasan positif dan 208 ulasan negatif. Penelitian ini menggunakan metode <i>SVM</i> dengan pembagian data dalam tiga skenario, yaitu 90:10, 80:20, dan 70:30. Hasil terbaik diperoleh pada skenario 90:10 dengan kernel <i>RBF</i>, yang mencapai akurasi sebesar 92,31% dan <i>f1 score</i> 92,86%. Selain itu, analisis kata-kata yang sering muncul dalam ulasan positif mencakup kata-kata seperti "bagus", "mantap", dan "gratis", sedangkan ulasan negatif sering mengandung kata-kata seperti "aplikasi", "error", dan "macet". [20]</p>
<p>Fitriyana et al. (2023)</p>	<p>Analisis Sentimen Ulasan Aplikasi Jamsostek <i>Mobile</i> Menggunakan Metode <i>Support Vector Machine</i></p>	<p>Penelitian ini bertujuan untuk menganalisis kualitas aplikasi JMO dan mengklasifikasikan ulasan atau opini kedalam kategori positif, negatif dan netral melalui analisis sentimen. Metode <i>Support Vector Machine</i> digunakan pada proses analisis ini dengan pendekatan kernel <i>linear</i> untuk mengetahui tingkat akurasi dari pengklasifikasian ulasan aplikasi JMO tersebut. Penelitian menunjukkan bahwa pengklasifikasian metode <i>SVM</i> terhadap analisis sentimen ulasan atau <i>review</i> aplikasi JMO</p>

		menghasilkan nilai akurasi terbaik, didapatkan hasil dengan <i>accuracy</i> 96%, <i>precision</i> 92%, <i>recall</i> 96%, dan <i>f1-score</i> 94%, sedangkan untuk hasil ulasan terbanyak adalah ulasan berkategori positif dengan jumlah 17.571.[21]
--	--	---

Berdasarkan hasil *studi literature* dari tiga penelitian terdahulu, terdapat kesamaan pada metode yang digunakan yaitu *Support Vector Machine*. Metode *SVM* dipilih karna memiliki beberapa kelebihan seperti akurasi tinggi dalam klasifikasi termasuk untuk analisis sentimen serta efektifitasnya dalam menangani data berdimensi tinggi. Dan pada penelitian sebelumnya menunjukkan bahwa *SVM* memberikan hasil yang konsisten dan akurat dalam analisis sentimen. Kebaruan pada penelitian ini yaitu terdapat pada data yang digunakan, yaitu ulasan pengguna aplikasi *mobile JKN* pada *google play store*.