

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai referensi dalam mendukung penelitian yang dilakukan. Penelitian terkait deteksi kondisi kesehatan mental telah banyak dilakukan dalam beberapa tahun terakhir, seiring dengan meningkatnya perhatian terhadap isu kesehatan mental serta perkembangan teknologi. Berbagai penelitian telah memanfaatkan data, khususnya dari media sosial, sebagai sumber informasi untuk mendeteksi kondisi psikologis individu secara tidak langsung.

Dari sisi metode, penelitian terdahulu menunjukkan adanya variasi pendekatan yang digunakan, mulai dari metode Machine Learning tradisional hingga metode berbasis *deep learning*. Metode *Machine Learning* seperti *Logistic Regression* dan *Support Vector Machine (SVM)* merupakan algoritma yang paling banyak digunakan dalam klasifikasi karena kemampuannya dalam menangani data serta proses implementasi yang relatif sederhana. *Logistic Regression* dikenal memiliki keunggulan dalam hal interpretabilitas dan kestabilan model, sehingga sering digunakan sebagai baseline dalam berbagai penelitian. Sementara itu, *Support Vector Machine (SVM)* memiliki kemampuan dalam membentuk *hyperplane* optimal yang dapat memisahkan kelas dengan baik, sehingga sering menghasilkan performa yang tinggi dalam klasifikasi.

Lebih lanjut, beberapa penelitian juga mengembangkan pendekatan perbandingan antar algoritma untuk mengetahui metode yang paling optimal dalam melakukan klasifikasi dan juga menjadi aspek penting dalam penelitian. Sebagian besar penelitian menggunakan beberapa metode representasi numerik sebelum dilakukan proses klasifikasi

Meskipun demikian, masih terdapat beberapa keterbatasan dalam penelitian terdahulu. Sebagian besar penelitian masih berfokus pada penggunaan satu algoritma secara terpisah tanpa mengeksplorasi kombinasi metode secara optimal. Selain itu, beberapa penelitian yang menggunakan metode kompleks belum

mempertimbangkan aspek interpretabilitas model, yang menjadi faktor penting dalam domain kesehatan. Oleh karena itu, masih terdapat peluang untuk mengembangkan model klasifikasi yang tidak hanya memiliki performa yang baik, tetapi juga tetap sederhana dan mudah diinterpretasikan.

Berdasarkan uraian tersebut, penelitian ini berfokus pada pengembangan model klasifikasi berbasis *Machine Learning* dengan membandingkan algoritma *Logistic Regression* dan *Support Vector Machine (SVM)*, serta mengembangkan pendekatan *hybrid* yang menggabungkan kedua metode tersebut. Dengan pendekatan ini, diharapkan dapat diperoleh model yang lebih optimal dalam mendeteksi kondisi kesehatan mental. Tabel 2.1 menyajikan beberapa penelitian terdahulu yang relevan dengan topik penelitian ini.

Tabel 2.1 Penelitian Terdahulu

Referensi	Tools / Method	Hasil	Gap Analysis
Ahmed et al. (2025) [13]	<i>Explainable AI (XAI), Logistic Regression, Random Forest, XGBoost, SHAP Analysis</i>	framework Explainable AI untuk mendeteksi depresi dan mengklasifikasikan tingkat keparahan depresi berdasarkan data aktivitas pengguna. Hasil menunjukkan bahwa model machine learning mampu melakukan klasifikasi depresi dengan baik.	Penelitian berfokus pada aspek interpretabilitas dan klasifikasi tingkat keparahan depresi menggunakan Explainable AI, namun belum mengevaluasi performa Support Vector Machine (SVM) maupun pendekatan Hybrid Voting Classifier
Xing-Xuan Dong et al. (2025) [14]	<i>Logistic Regression, Random Forest, Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost</i>	Model machine learning menunjukkan kemampuan yang kompetitif. Beberapa metode machine learning menghasilkan performa yang lebih baik dibandingkan Logistic Regression pada dataset tertentu.	Penelitian membandingkan Logistic Regression dengan beberapa algoritma machine learning modern, namun tidak mengevaluasi penggunaan (SVM) maupun pendekatan hybrid yang menggabungkan beberapa model untuk meningkatkan performa klasifikasi.
S. Rakeen et al. (2024) [15]	<i>Logistic Regression, Support Vector Machine (SVM), Naive Bayes, Decision Tree, Gradient Boosting Machine (GBM)</i>	<i>SVM</i> memiliki performa terbaik diikuti oleh <i>Logistic Regression</i> Model lain seperti <i>GBM</i> , <i>Naive Bayes</i> , dan <i>Decision Tree</i> menunjukkan performa yang lebih rendah.	hanya menggunakan model tunggal dan tidak mengimplementasikan pendekatan <i>hybrid</i> juga belum mengeksplorasi kombinasi model seperti <i>Logistic Regression</i> dan <i>SVM</i> dalam bentuk voting classifier.

I. Mohammed & H. Hassani [16]	<i>Support Vector Machine (SVM) + Logistic Regression</i>	SVM & LR bervariasi tergantung dataset dan fitur yang digunakan	Penelitian belum membandingkan <i>hybrid</i> secara komprehensif.
W. Tahir et al. [17]	<i>Logistic Regression + SVM + Hybrid</i>	Berbagai metode ML dan DL dapat digunakan untuk deteksi depresi tergantung karakteristik data	Penelitian belum mengintegrasikan metode lain untuk meningkatkan akurasi klasifikasi.
Q. Nguyen et al. [18]	<i>Logistic Regression, SVM + NLP / TF-IDF</i>	SVM & LR mampu mengidentifikasi akurasi tertinggi mengidentifikasi tanda depresi melalui pola bahasa pada teks	Studi tidak mengembangkan model <i>hybrid</i>
M. Agirrezabal & J. Amann (2022) [19]	<i>Logistic Regression / SVM + NLP Pipeline</i>	<i>Logistic Regression & SVM</i> dapat menangkap pola linguistik yang berkaitan dengan depresi	Penelitian belum melakukan perbandingan dengan model <i>hybrid</i> .
Yan et al. (2024) [20]	<i>SMOTE Oversampling, Logistic Regression, Random Forest, XGBoost, Multi-Class Classification Framework</i>	Penelitian mengusulkan framework klasifikasi multikelas untuk prediksi status kesehatan pasien dengan mengatasi masalah ketidakseimbangan data menggunakan teknik oversampling.	Penelitian berfokus pada penanganan class imbalance menggunakan teknik oversampling pada klasifikasi multikelas, namun belum mengkaji penggunaan Support Vector Machine (SVM) maupun metode Hybrid Voting Classifier
M. De Choudhury et al. (2022) [21]	TF-IDF + SVM / <i>Logistic Regression</i>	SVM / LR dengan fitur TF-IDF menghasilkan performa klasifikasi yang tinggi pada dataset teks mental health.	Penelitian tidak membandingkan sebagai pembanding model.
H. Pu (2024) [22]	<i>Logistic Regression vs SVM</i>	SVM menunjukkan performa lebih baik pada data berdimensi tinggi dibanding <i>Logistic Regression</i> .	Penelitian belum mengembangkan perbandingan SVM, LR dan model <i>hybrid</i> untuk meningkatkan performa klasifikasi.
Jaufari D. & Wella (2025) [28]	<i>Hybrid Long Short-Term Memory + Support Vector Machine</i>	<i>hybrid</i> yang menggabungkan LSTM dan SVM mampu meningkatkan akurasi prediksi dibandingkan penggunaan model tunggal, karena mampu menangkap pola temporal dan non-linear secara lebih baik	Penelitian berfokus pada data time-series (harga saham) dan menggunakan kombinasi LSTM-SVM, sehingga belum relevan secara langsung untuk klasifikasi teks serta belum mengeksplorasi <i>hybrid</i> berbasis model klasik seperti <i>Logistic Regression</i> dan SVM

Berdasarkan Tabel 2.1, dapat dilihat bahwa seluruh penelitian terdahulu memiliki keterkaitan yang kuat dalam hal tujuan, yaitu memanfaatkan data untuk mendeteksi kondisi kesehatan mental [13], [14]. Sebagian besar penelitian menggunakan data yang berasal dari media sosial,

yang dianggap mampu merepresentasikan kondisi emosional pengguna secara spontan dan natural, sehingga dapat digunakan sebagai sumber data dalam mendeteksi indikasi gangguan mental [15].

Dari sisi permasalahan, penelitian-penelitian tersebut secara konsisten berfokus pada identifikasi gangguan mental seperti depresi, kecemasan, dan stres melalui pendekatan berbasis teks [16]. Hal ini menunjukkan adanya kesamaan pandangan bahwa bahasa yang digunakan individu dalam media sosial dapat menjadi indikator awal kondisi psikologis seseorang.

Dari sisi metode, pada Tabel 2.1 terlihat bahwa sebagian besar penelitian menggunakan metode *Machine Learning* tradisional, khususnya *Logistic Regression* dan *Support Vector Machine (SVM)*, yang dikenal efektif dalam klasifikasi teks berdimensi tinggi [17], [18]. Selain itu, beberapa penelitian mulai mengembangkan pendekatan *hybrid* dengan mengkombinasikan lebih dari satu metode klasifikasi guna meningkatkan performa model. Penggunaan teknik ekstraksi fitur juga menjadi pendekatan umum yang digunakan untuk merepresentasikan teks dalam bentuk numerik sebelum dilakukan proses klasifikasi [19].

Dari sisi hasil, sebagian besar penelitian menunjukkan bahwa metode *Support Vector Machine (SVM)* cenderung menghasilkan performa yang lebih tinggi dibandingkan metode klasifikasi lainnya pada data berdimensi tinggi [20]. Namun demikian, beberapa penelitian menunjukkan bahwa metode *Logistic Regression* memiliki keunggulan dalam hal stabilitas model dan kemudahan interpretasi hasil [21]. Selain itu, pendekatan *hybrid* yang menggabungkan *Logistic Regression* dan *SVM* dilaporkan mampu meningkatkan akurasi dan nilai *F1-score* dibandingkan penggunaan model Tunggal [22].

Berdasarkan keterhubungan antar penelitian tersebut, dapat disimpulkan bahwa masih terdapat peluang untuk mengembangkan model klasifikasi yang mengkombinasikan keunggulan dari *Logistic Regression*

dan *Support Vector Machine* dalam bentuk model *hybrid*. Oleh karena itu, penelitian ini berfokus pada perbandingan performa metode *Logistic Regression*, *Support Vector Machine (SVM)*, serta model *hybrid* dari kedua metode tersebut untuk mendeteksi indikasi kesehatan mental pada media sosial.

2.1.1 Penelitian yang Diadopsi

Berdasarkan Tabel 2.1, seluruh penelitian terdahulu memiliki keterkaitan dalam hal permasalahan, metode, serta hasil yang diperoleh [13]. Dari sisi permasalahan, sebagian besar penelitian berfokus pada pemanfaatan data dari media sosial untuk mendeteksi kondisi kesehatan mental seperti depresi, stres, dan kecemasan [15]. Penelitian-penelitian tersebut menunjukkan bahwa bahasa yang digunakan dalam media sosial dapat mencerminkan kondisi emosional pengguna, sehingga dapat dimanfaatkan sebagai indikator awal dalam mendeteksi gangguan kesehatan mental [16].

Dari sisi metode, penelitian-penelitian pada Tabel 2.1 umumnya menggunakan pendekatan *Machine Learning*, khususnya metode *Logistic Regression* dan *Support Vector Machine (SVM)*, yang banyak digunakan dalam klasifikasi teks karena kemampuannya dalam menangani data berdimensi tinggi [17], [18]. Beberapa penelitian juga mulai mengembangkan pendekatan *hybrid*, yaitu dengan menggabungkan lebih dari satu algoritma klasifikasi untuk meningkatkan performa model [19]. Selain itu, sebagian besar penelitian menggunakan teknik ekstraksi fitur seperti TF-IDF untuk mengubah data menjadi representasi numerik sebelum dilakukan proses klasifikasi.

Dari sisi hasil, penelitian-penelitian terdahulu menunjukkan bahwa metode *Support Vector Machine (SVM)* umumnya menghasilkan performa yang baik dalam klasifikasi dengan akurasi yang tinggi, sedangkan *Logistic Regression* menunjukkan performa yang stabil serta memiliki keunggulan dalam hal kesederhanaan model dan interpretasi hasil. Selain itu, penelitian

yang menggunakan pendekatan *hybrid Logistic Regression* dan *SVM* menunjukkan adanya peningkatan performa dibandingkan dengan penggunaan model tunggal, terutama pada nilai akurasi dan *F1-score* [28].

Berdasarkan analisis tersebut, penelitian yang akan diadopsi dalam skripsi ini adalah penelitian yang menggunakan pendekatan *Machine Learning* berbasis *Logistic Regression* dan *Support Vector Machine (SVM)*, khususnya penelitian yang mengembangkan pendekatan *hybrid* antara kedua metode tersebut. Pendekatan ini dipilih karena terbukti efektif dalam klasifikasi kesehatan mental serta memiliki keseimbangan antara performa model dan efisiensi komputasi.

2.1.2 Research Gap

Berdasarkan analisis terhadap penelitian-penelitian terdahulu pada Tabel 2.1, masih terdapat beberapa keterbatasan yang menunjukkan adanya peluang untuk dilakukan pengembangan penelitian lebih lanjut [13], [14]. Sebagian besar penelitian terdahulu masih menggunakan satu algoritma *Machine Learning* secara terpisah, seperti *Logistic Regression* atau *Support Vector Machine (SVM)* [17], sehingga potensi pemanfaatan keunggulan dari kombinasi beberapa metode belum dimaksimalkan secara optimal. Penggunaan satu metode saja berpotensi menghasilkan performa yang kurang optimal karena setiap algoritma memiliki karakteristik dan keterbatasan yang berbeda dalam menangani data teks.

Selain itu, beberapa penelitian yang melakukan perbandingan metode *Machine Learning* umumnya hanya membandingkan hasil performa model secara individual tanpa mengembangkan pendekatan *hybrid* yang menggabungkan beberapa algoritma dalam satu sistem klasifikasi [19]. Padahal, pendekatan *hybrid* memiliki potensi dalam meningkatkan performa klasifikasi dengan memanfaatkan kekuatan masing-masing metode, terutama dalam menghadapi kompleksitas data teks yang memiliki variasi bahasa yang tinggi.

Keterbatasan lainnya terlihat pada ruang lingkup penerapan metode yang masih bersifat umum dan belum secara khusus mengoptimalkan performa model pada domain analisis sentimen kesehatan mental [15], [16]. Hal ini menunjukkan bahwa masih terdapat peluang untuk mengembangkan model klasifikasi yang lebih optimal melalui pendekatan perbandingan serta penggabungan metode *Logistic Regression* dan *Support Vector Machine (SVM)* dalam bentuk model hybrid.

Berdasarkan kesenjangan tersebut, penelitian ini berfokus pada pengembangan model klasifikasi berbasis *Machine Learning* dengan melakukan perbandingan antara *Logistic Regression* dan *Support Vector Machine (SVM)*, serta mengembangkan pendekatan *hybrid Logistic Regression dan SVM* untuk meningkatkan performa analisis dalam mendeteksi indikasi kesehatan mental [22]. Pendekatan ini diharapkan dapat memberikan kontribusi dalam menghasilkan model klasifikasi yang memiliki performa lebih baik dibandingkan penggunaan satu algoritma secara tunggal.

2.2 Teori yang berkaitan

Pada bagian ini akan dibahas berbagai teori yang berkaitan dengan penelitian yang dilakukan. Teori-teori tersebut digunakan sebagai landasan dalam memahami konsep dasar serta mendukung proses analisis yang dilakukan dalam penelitian ini. Dengan adanya landasan teori yang kuat, diharapkan penelitian dapat dilakukan secara lebih terarah serta memiliki dasar ilmiah yang jelas.

Teori yang digunakan dalam penelitian ini mencakup konsep-konsep yang berkaitan dengan kesehatan mental, pengolahan data, serta metode machine learning yang digunakan dalam proses klasifikasi. Kesehatan mental menjadi fokus utama penelitian karena berbagai faktor psikologis, sosial, dan perilaku dapat memengaruhi kondisi mental seseorang. Selain itu, proses pengolahan data juga menjadi bagian penting dalam penelitian untuk mempersiapkan data sebelum digunakan dalam tahap pemodelan. Metode *machine learning* seperti *Logistic*

Regression, *Support Vector Machine (SVM)*, dan *Hybrid* digunakan untuk melakukan klasifikasi kondisi kesehatan mental berdasarkan atribut-atribut yang terdapat pada dataset. Teori-teori tersebut menjadi dasar dalam proses pengembangan, pelatihan, dan evaluasi model yang dilakukan pada penelitian ini.

2.2.1 Kesehatan Mental

Kesehatan mental merupakan kondisi kesejahteraan individu yang mencakup kemampuan untuk mengelola emosi, berpikir secara rasional, serta berinteraksi secara sosial dengan baik. Menurut *World Health Organization (WHO)*, kesehatan mental tidak hanya berarti tidak adanya gangguan mental, tetapi juga mencakup kemampuan individu dalam menghadapi tekanan hidup, bekerja secara produktif, dan berkontribusi dalam masyarakat.

Dalam konteks penelitian ini, kesehatan mental menjadi fokus utama karena meningkatnya kasus gangguan psikologis, khususnya depresi, di kalangan masyarakat modern. Depresi merupakan salah satu gangguan mental yang ditandai dengan perasaan sedih berkepanjangan, kehilangan minat terhadap aktivitas, serta penurunan fungsi kognitif dan sosial. Kondisi ini sering kali tidak terdeteksi secara dini karena individu cenderung tidak mencari bantuan profesional.

Seiring dengan perkembangan teknologi digital, media sosial menjadi salah satu sarana utama bagi individu untuk mengekspresikan perasaan dan kondisi emosional mereka. Ungkapan dalam bentuk teks, seperti status, komentar, dan postingan, dapat mencerminkan kondisi psikologis seseorang secara tidak langsung. Dalam penelitian ini, klasifikasi merupakan salah satu cara untuk mengidentifikasi kecenderungan emosi dalam teks yang dikategorikan menjadi positif dan negatif. Klasifikasi ini menjadi indikator awal dalam mendeteksi potensi gangguan mental, di mana dominasi negatif dapat mengindikasikan adanya tekanan psikologis atau gejala depresi.

2.2.2 Framework Data Mining

Dalam proses pengolahan data diperlukan suatu kerangka kerja (*framework*) yang sistematis agar penelitian dapat dilakukan secara terstruktur dan menghasilkan output yang valid. Beberapa *framework* yang umum digunakan dalam data mining antara lain **CRISP-DM**, **KDD**, dan **SEMMA**.

A. CRISP-DM

CRISP-DM merupakan salah satu *framework* yang paling banyak digunakan dalam penelitian data mining karena memiliki struktur yang sistematis dan fleksibel [23]. *Framework* ini terdiri dari enam tahapan utama, yaitu:

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modeling*
5. *Evaluation*
6. *Deployment*

Dalam penelitian ini, *framework* CRISP-DM digunakan sebagai acuan dalam proses analisis data, namun implementasinya hanya dilakukan hingga tahap *Evaluation*. Hal ini dikarenakan tujuan utama penelitian adalah untuk menganalisis dan membandingkan performa model machine learning, bukan untuk mengembangkan sistem yang siap digunakan secara langsung.

Tahap *evaluation* dianggap sudah cukup untuk menilai kinerja model melalui metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*, serta analisis *confusion matrix*. Oleh karena itu, penelitian ini bersifat eksperimental dan berfokus pada analisis hasil model. Meskipun demikian, hasil penelitian ini tetap dapat dikembangkan lebih lanjut ke tahap implementasi pada penelitian berikutnya.

B. KDD

KDD merupakan proses penemuan pengetahuan dari data yang terdiri dari beberapa tahapan, yaitu:

1. *Selection (pemilihan data)*
2. *Preprocessing (pembersihan data)*
3. *Transformation (transformasi data)*
4. *Data Mining (penerapan algoritma)*
5. *Interpretation/Evaluation (interpretasi hasil)*

KDD memiliki konsep yang mirip dengan CRISP-DM, namun lebih berfokus pada proses teknis dalam pengolahan data. Dalam penelitian ini, tahapan KDD terlihat terutama pada proses preprocessing, transformasi, serta penerapan algoritma *machine learning*.

C. SEMMA

SEMMA merupakan framework yang dikembangkan oleh SAS Institute yang terdiri dari lima tahapan:

1. *Sample* → *pemilihan subset data*
2. *Explore* → *eksplorasi data*
3. *Modify* → *transformasi data*
4. *Model* → *pembangunan model*
5. *Assess* → *evaluasi model*

SEMMA lebih berfokus pada aspek teknis analisis data dan kurang menekankan pada pemahaman masalah bisnis dibandingkan CRISP-DM.

2.2.3 Logistic Regression

Logistic Regression merupakan salah satu algoritma klasifikasi yang digunakan untuk memprediksi kelas berdasarkan probabilitas. Algoritma ini bekerja dengan membentuk hubungan antara fitur input dan output melalui fungsi sigmoid, sehingga menghasilkan nilai probabilitas yang berada pada rentang antara 0 hingga 1.

Dalam proses klasifikasi, nilai probabilitas tersebut akan dibandingkan dengan suatu *threshold*, yang umumnya bernilai 0.5, untuk menentukan apakah suatu data termasuk ke dalam kelas positif atau negatif. Jika nilai probabilitas lebih besar dari *threshold*, maka data akan diklasifikasikan ke dalam kelas positif, sedangkan jika lebih kecil maka diklasifikasikan sebagai kelas negatif. Dengan pendekatan ini, *Logistic Regression* tidak hanya memberikan hasil klasifikasi, tetapi juga tingkat keyakinan terhadap prediksi yang dihasilkan.

$$P(y = 1 | x) = \frac{1}{2a1 + e^{-(w \cdot x + b)}}$$

Rumus 2.1 Sigmoid function (Logistic Regression)

Pada penelitian ini, *Logistic Regression* digunakan untuk mengklasifikasikan data yang akan direpresentasikan. Representasi ini menghasilkan data dengan dimensi yang tinggi, sehingga diperlukan algoritma yang mampu menangani kondisi tersebut secara efisien. *Logistic Regression* dipilih karena memiliki kemampuan yang baik dalam menangani data, serta memiliki proses komputasi yang relatif cepat dibandingkan metode yang lebih kompleks.

Selain itu, *Logistic Regression* juga dikenal sebagai model yang stabil dan mudah diimplementasikan, sehingga sering digunakan sebagai baseline dalam berbagai penelitian klasifikasi. Model ini mampu memberikan hasil yang cukup baik dalam khususnya ketika hubungan antar fitur bersifat linear. Namun demikian, *Logistic Regression* memiliki keterbatasan dalam menangkap pola data yang bersifat kompleks dan tidak linear.

Hal ini dapat mempengaruhi performa model ketika data memiliki distribusi yang lebih variatif. Oleh karena itu, dalam penelitian ini *Logistic Regression* tidak hanya digunakan sebagai model tunggal, tetapi juga dikombinasikan dengan *Support Vector Machine* dalam bentuk

model hybrid untuk meningkatkan performa klasifikasi serta menghasilkan prediksi yang lebih optimal.

2.2.4 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma *Machine Learning* yang banyak digunakan dalam permasalahan klasifikasi. Algoritma ini bekerja dengan cara mencari batas pemisah terbaik yang dapat memisahkan data ke dalam kelas yang berbeda secara maksimal. Tujuan utama dari *SVM* adalah untuk menemukan *hyperplane* yang memiliki margin terbesar, yaitu jarak terjauh antara batas pemisah dengan titik data terdekat dari masing-masing kelas yang disebut sebagai *support vectors*.

SVM juga memiliki kemampuan untuk menangani data yang tidak dapat dipisahkan secara linear dengan memanfaatkan konsep *kernel trick*. Kernel memungkinkan data diproyeksikan sehingga data yang awalnya tidak dapat dipisahkan secara linear menjadi dapat dipisahkan dengan menggunakan *hyperplane*. Beberapa jenis kernel yang umum digunakan antara lain linear kernel, polynomial kernel, dan *radial basis function (RBF)*. Namun demikian, *SVM* juga memiliki beberapa keterbatasan, di antaranya adalah proses komputasi yang relatif lebih kompleks dibandingkan algoritma sederhana seperti *Logistic Regression*, terutama pada dataset dengan jumlah data yang sangat besar. Selain itu, pemilihan parameter seperti jenis kernel dan nilai parameter lainnya juga dapat mempengaruhi performa model, sehingga memerlukan proses tuning yang tepat.

$$f(x) = \text{sign}(w \cdot x + b)$$

Rumus 2.2 *Decision function (SVM hyperplane)*

Dalam rumus tersebut algoritma ini memiliki keunggulan dalam menghasilkan model dengan tingkat akurasi yang tinggi, terutama pada dataset dengan jumlah fitur yang besar. Selain itu, penggunaan kernel

pada *SVM* memungkinkan model untuk menangani hubungan non-linear pada data. Meskipun memiliki performa yang baik, *SVM* memiliki kelemahan dalam hal kompleksitas komputasi serta memerlukan pemilihan parameter yang tepat agar dapat menghasilkan performa optimal.

Dalam konteks penelitian ini, *Support Vector Machine* digunakan sebagai salah satu algoritma utama dalam melakukan klasifikasi kondisi kesehatan mental [25]. Pemilihan *SVM* didasarkan pada kemampuannya dalam menangani data serta performanya yang baik dalam berbagai penelitian sebelumnya. Selain itu, *SVM* juga digunakan sebagai bagian dari model hybrid yang dikombinasikan dengan *Logistic Regression* untuk meningkatkan performa klasifikasi.

2.2.5 Hybrid

Hybrid merupakan pendekatan dalam *Machine Learning* yang menggabungkan dua atau lebih algoritma untuk meningkatkan performa klasifikasi. Pendekatan ini dilakukan dengan tujuan memanfaatkan keunggulan masing-masing algoritma serta mengurangi keterbatasan yang dimiliki oleh model tunggal.

Dalam penelitian ini, *hybrid* model yang digunakan adalah kombinasi antara *Logistic Regression* dan *Support Vector Machine (SVM)* dan model ini menggunakan mekanisme untuk menentukan hasil klasifikasi akhir. Pendekatan ini bertujuan memanfaatkan keunggulan masing-masing algoritma sehingga dapat meningkatkan stabilitas dan konsistensi performa klasifikasi kesehatan mental.

Logistic Regression merupakan algoritma klasifikasi yang memiliki keunggulan dalam hal kesederhanaan model serta kemudahan dalam interpretasi hasil. Algoritma ini bekerja dengan menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan fungsi logistik. sementara itu, *Support Vector Machine (SVM)* merupakan algoritma

yang bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam kelas yang berbeda. SVM dikenal memiliki kemampuan yang baik dalam menangani data, serta mampu menghasilkan performa klasifikasi yang tinggi. dengan menggabungkan kedua metode tersebut, diharapkan model yang dihasilkan memiliki performa yang lebih baik dibandingkan penggunaan satu algoritma saja.

$$y = \arg \max_c \sum_{m=1}^M 1[y_m = c]$$

Rumus 2.3 Majority voting (Hybrid model)

Pendekatan *hybrid* juga memungkinkan peningkatan nilai *accuracy*, *precision*, *recall*, dan *F1-score* dalam proses klasifikasi. Oleh karena itu, penggunaan model hybrid menjadi salah satu strategi yang relevan dalam meningkatkan kinerja sistem untuk deteksi kesehatan mental berbasis teks.

Penggunaan kedua dan penggabungan metode ini didasarkan pada asumsi bahwa setiap algoritma memiliki kelebihan dan kekurangan masing-masing. *Logistic Regression* unggul dalam hal interpretabilitas dan kestabilan, sedangkan *SVM* unggul dalam kemampuan generalisasi dan pemisahan kelas. Dengan menggabungkan kedua metode tersebut, diharapkan dapat diperoleh model yang memiliki performa yang lebih baik dalam melakukan klasifikasi kondisi kesehatan mental.

2.2.6 Python

Python merupakan bahasa pemrograman yang digunakan dalam penelitian ini untuk melakukan proses pengolahan data, pemodelan *Machine Learning*, serta evaluasi hasil. *Python* dipilih karena memiliki sintaks yang sederhana, mudah dipahami, serta didukung oleh berbagai library yang lengkap dalam bidang *data science* dan *Machine Learning*.

Dalam penelitian ini, *Python* digunakan melalui lingkungan pengembangan *Google Collab*, yang memungkinkan peneliti untuk

menuliskan kode program, menampilkan hasil, serta mendokumentasikan proses analisis secara terstruktur dalam satu platform. beberapa *library* yang digunakan dalam penelitian ini antara lain *Pandas* dan *NumPy* untuk pengolahan data, serta *Scikit-learn* untuk implementasi algoritma *Machine Learning* seperti *Logistic Regression* dan *Support Vector Machine (SVM)*. Selain itu, digunakan juga *library* untuk evaluasi model seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

Dengan menggunakan *Python* dan berbagai *library* pendukungnya, proses analisis data dalam penelitian ini dapat dilakukan secara lebih efisien, sistematis, dan mudah untuk direplikasi.

