

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Teori**

##### **2.1.1. Kehamilan**

Kehamilan adalah proses pembuahan sel telur oleh sperma yang kemudian berkembang menjadi janin di dalam rahim wanita. Proses kehamilan dimulai saat sel telur yang dilepaskan oleh indung telur wanita bertemu dengan sperma yang masuk melalui saluran reproduksi. Sel telur yang telah dibuahi akan menuju rahim dan menempel pada dinding rahim, kemudian akan berkembang menjadi embrio dan kemudian janin. Kehamilan memerlukan berbagai faktor yang mendukung seperti hormon, kondisi fisik dan psikologis wanita, serta asupan nutrisi yang cukup. Selain itu, proses kehamilan juga dipengaruhi oleh faktor lingkungan seperti polusi dan paparan bahan kimia yang dapat membahayakan kesehatan janin [16].

Selama kehamilan terjadi perubahan besar pada tubuh wanita termasuk perubahan hormonal yang mempengaruhi pertumbuhan janin dan kesehatan ibu. Perubahan hormon seperti estrogen dan progesteron sangat penting untuk menjaga kesehatan janin dan mempersiapkan tubuh ibu untuk melahirkan. Kehamilan juga memerlukan asupan nutrisi yang cukup seperti protein, zat besi, asam folat, dan kalsium yang diperlukan untuk pertumbuhan dan perkembangan janin yang sehat. Selain itu, kondisi fisik dan psikologis wanita juga berpengaruh pada proses kehamilan. Wanita yang sehat dan terlatih fisiknya cenderung memiliki kehamilan yang lebih baik dibandingkan dengan wanita yang kurang sehat [17].

Dalam hal ini, kehamilan merupakan suatu proses yang kompleks dan memerlukan perhatian khusus. Faktor-faktor seperti kondisi kesehatan ibu dan lingkungan harus diperhatikan untuk memastikan kesehatan janin dan

ibu selama kehamilan. Oleh karena itu, ibu hamil perlu memperhatikan pola makan, menjaga kesehatan fisik dan psikologis, serta berkonsultasi dengan dokter kandungan secara teratur untuk memastikan kehamilan berjalan dengan baik dan lancar.

### **2.1.2. Risiko Kehamilan**

Risiko kehamilan adalah kondisi dimana seorang ibu hamil mengalami komplikasi atau masalah kesehatan selama kehamilan, persalinan, atau pasca melahirkan. *World Health Organization (WHO)* mengemukakan ada beberapa faktor yang dapat meningkatkan risiko kehamilan, seperti usia ibu yang terlalu muda atau terlalu tua, riwayat penyakit atau masalah kesehatan sebelumnya, dan faktor gaya hidup yang tidak sehat. Selain itu, risiko kehamilan juga dapat dipengaruhi oleh faktor eksternal seperti lingkungan dan aksesibilitas pelayanan kesehatan [18].

Risiko kehamilan dapat berdampak negatif pada kesehatan ibu dan bayi. Beberapa komplikasi kehamilan yang umum meliputi preeklampsia, gestasional diabetes, infeksi, dan masalah pada plasenta. Jika tidak ditangani dengan baik, kondisi-kondisi ini dapat membahayakan nyawa ibu dan bayi, serta meningkatkan risiko kelahiran prematur atau bayi dengan berat lahir rendah [19].

Kesehatan ibu selama kehamilan memiliki peran penting dalam meminimalkan risiko kehamilan. Upaya pencegahan yang teratur dan pemantauan kesehatan selama kehamilan sangat penting untuk mengurangi risiko kehamilan. Konsultasikan dengan dokter dan perawat kesehatan setiap kali ada kekhawatiran atau gejala yang tidak biasa selama kehamilan [19].

### **2.1.3. Faktor Penyebab Risiko Kehamilan**

Risiko kehamilan dapat dipengaruhi oleh banyak faktor, termasuk usia ibu, riwayat kesehatan, gaya hidup, dan lingkungan sekitar. Menurut penelitian, faktor-faktor tersebut dapat meningkatkan kemungkinan

terjadinya komplikasi selama kehamilan, persalinan, atau pada bayi yang lahir.

Salah satu faktor yang dapat mempengaruhi risiko kehamilan adalah usia ibu. Ibu yang berusia di atas 35 tahun memiliki kemungkinan lebih tinggi untuk mengalami komplikasi selama kehamilan dan melahirkan dibandingkan dengan ibu yang berusia di bawah 35 tahun [20]. Selain usia, riwayat kesehatan ibu juga dapat mempengaruhi risiko kehamilan. Misalnya, ibu yang memiliki riwayat hipertensi atau diabetes sebelum hamil lebih rentan terhadap preeklampsia atau kelahiran prematur [21].

Faktor gaya hidup juga berperan dalam menentukan risiko kehamilan. Merokok, minum alkohol, dan penggunaan obat-obatan terlarang dapat meningkatkan kemungkinan terjadinya kelainan bawaan pada bayi, persalinan prematur, atau kelahiran bayi dengan berat badan rendah [22].

Lingkungan sekitar juga dapat mempengaruhi risiko kehamilan. Paparan ibu pada polutan udara dapat meningkatkan kemungkinan terjadinya kelainan bawaan pada bayi atau kelahiran prematur. Dengan memperhatikan faktor-faktor tersebut, dapat membantu dalam menentukan risiko kehamilan dan mengambil tindakan pencegahan yang tepat untuk menjaga kesehatan ibu dan bayi [23].

#### **2.1.4. Algoritma Klustering**

Algoritma klustering adalah metode pengelompokan data tanpa supervisi di mana data yang serupa atau memiliki karakteristik yang sama dikelompokkan menjadi satu kelompok atau kluster. Algoritma klustering digunakan dalam berbagai bidang seperti pengolahan citra, data mining, dan analisis biologi.

Algoritma klustering dibedakan menjadi dua jenis, yaitu klustering hierarki dan klustering partisi. Pada klustering hierarki, data dikelompokkan dalam sebuah pohon kluster di mana setiap *node* pada pohon mewakili

sebuah kluster, sedangkan pada klustering partisi, data dikelompokkan dalam beberapa kluster yang saling eksklusif [24].

Salah satu algoritma klustering yang populer adalah K-means. Algoritma ini menggunakan jarak antara data untuk menentukan kluster. Data diatributkan ke kluster dengan mean yang terdekat. Menurut Dubey, K-means sering digunakan dalam pengolahan citra dan analisis biologi [25].

Selain K-means, terdapat pula algoritma klustering lain seperti *Hierarchical Clustering*, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), dan *Fuzzy C-Means*. *Hierarchical Clustering* membagi data menjadi beberapa kluster yang bertingkat berdasarkan jarak antara data, sementara DBSCAN menggunakan kerapatan data untuk membentuk kluster. *Fuzzy C-Means*, di sisi lain, memperhitungkan ketidakpastian dalam pembagian data ke kluster [26].

Secara umum, algoritma klustering sangat bermanfaat untuk mengelompokkan data berdasarkan karakteristik yang dimilikinya. Namun, pemilihan algoritma yang tepat tergantung pada jenis data dan tujuan klustering yang ingin dicapai [26].

#### **2.1.5. Algoritma Klasifikasi**

Algoritma klasifikasi merupakan metode pengolahan data yang digunakan untuk memisahkan data ke dalam kategori-kategori yang telah ditentukan sebelumnya. Algoritma ini bekerja dengan memanfaatkan informasi yang terdapat pada data untuk membangun model klasifikasi yang dapat memprediksi kategori dari data yang belum dikenal [27]. Klasifikasi juga merupakan suatu teknik analisis data yang digunakan untuk memilah atau mengelompokkan data ke dalam kategori-kategori tertentu berdasarkan fitur atau ciri yang ditemukan dalam data tersebut [28].

Algoritma klasifikasi memiliki dua tahap utama, yaitu tahap pembangunan model dan tahap penggunaan model. Pada tahap pembangunan model, algoritma memanfaatkan data latih yang telah

terkategori sebelumnya untuk membangun model klasifikasi. Model ini akan terus disempurnakan melalui proses *training* dengan data latih hingga mencapai tingkat akurasi yang diinginkan. Setelah model selesai dibangun, tahap selanjutnya adalah penggunaan model untuk memprediksi kategori dari data yang belum dikenal [29].

Beberapa algoritma klasifikasi yang sering digunakan antara lain k-NN, Naive Bayes, Decision Tree, Random Forest, SVM, dan Neural Network. Setiap algoritma klasifikasi memiliki keunggulan dan kelemahan masing-masing tergantung pada jenis data dan masalah klasifikasi yang dihadapi. Oleh karena itu, pemilihan algoritma klasifikasi yang tepat sangat penting dalam memastikan akurasi dan efisiensi dari model klasifikasi yang dibangun [30].

#### **2.1.6. Kombinasi Algoritma Klustering dan Klasifikasi**

Kombinasi algoritma klustering dan klasifikasi telah menjadi topik penelitian yang menarik dalam bidang *data mining* dan *machine learning*. Secara umum, kombinasi ini dapat membantu meningkatkan kualitas hasil prediksi dan pengelompokan data [31].

Integrasi algoritma klustering dan klasifikasi dapat memperbaiki kualitas klasifikasi dengan cara mengurangi heterogenitas antara data dalam kelompok yang berbeda, dan meningkatkan homogenitas antara data dalam kelompok yang sama. Algoritma klustering bertanggung jawab untuk mengelompokkan data yang serupa, sedangkan algoritma klasifikasi mengklasifikasikan data ke dalam kelas tertentu [31].

Selain itu, kombinasi ini dapat meningkatkan efisiensi waktu komputasi. Hal ini terlihat pada penelitian yang dilakukan oleh Bashir, dimana kombinasi algoritma klustering dan klasifikasi berhasil meningkatkan efisiensi waktu komputasi hingga 75% dibandingkan dengan penggunaan algoritma klustering atau klasifikasi secara terpisah [32].

Dalam penelitian yang lebih baru, terdapat usulan sebuah *framework* yang memanfaatkan algoritma klustering dan klasifikasi untuk meningkatkan akurasi prediksi kelangsungan hidup pasien kanker payudara. Hasil eksperimen menunjukkan bahwa *framework* yang diusulkan mampu memberikan akurasi yang lebih tinggi dibandingkan dengan penggunaan algoritma klustering atau klasifikasi secara terpisah [33].

Secara keseluruhan, kombinasi algoritma klustering dan klasifikasi dapat meningkatkan kualitas hasil prediksi dan pengelompokan data serta efisiensi waktu komputasi. Oleh karena itu, kombinasi ini layak dipertimbangkan sebagai pilihan metode dalam penyelesaian masalah *data mining* dan *machine learning* [26].

## 2.2 Framework dan Algoritma

### 2.2.1 Algoritma K-Means

Algoritma k-means adalah salah satu metode dalam klustering atau pengelompokan data yang sering digunakan dalam analisis data. Metode ini bekerja dengan cara membagi data ke dalam beberapa kelompok berdasarkan kemiripan antar data di dalamnya. Algoritma ini bekerja dengan cara menentukan pusat *cluster* atau *centroid* dari kelompok data yang telah terbentuk, kemudian menghitung jarak antara setiap data dengan *centroid* dan memasukkan data tersebut ke dalam kelompok dengan *centroid* terdekat. Proses ini diulang hingga *centroid* dari setiap kelompok tidak berubah atau konvergen [34].

Algoritma k-means mempunyai kelebihan dalam mengelompokkan data yang berjumlah banyak dan ukuran variabelnya berbeda-beda. Selain itu, algoritma k-means juga mampu menyelesaikan permasalahan klustering dengan lebih cepat dibandingkan metode klustering lainnya [35]. Algoritma k-means masih menjadi salah satu algoritma klustering yang paling sering digunakan dalam praktik. Hal ini dikarenakan kecepatan dan kemampuan algoritma dalam menangani data yang kompleks. Namun, salah satu

kelemahan dari algoritma k-means adalah rentannya terhadap inisialisasi *centroid* yang buruk dan kemungkinan terjadinya lokal optima [36].

Rumus jarak yang umum digunakan dalam K-means adalah jarak Euclidean. Namun, rumus jarak dapat disesuaikan tergantung pada kebutuhan, seperti menggunakan jarak Manhattan atau jarak Minkowski. Dalam rumus jarak Euclidean, misalkan  $x$  merupakan titik data dan  $c$  merupakan pusat kluster, maka rumus jarak Euclidean adalah:

$$D(x, c) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2}$$

Rumus 2.1 Rumus Jarak Euclidean pada Algoritma K-Means

Di mana  $x_1, x_2, \dots, x_n$  adalah atribut-atribut dari titik data  $x$ , dan  $c_1, c_2, \dots, c_n$  adalah atribut-atribut dari pusat kluster  $c$  [37].

### 2.2.2 Algoritma Naïve Bayes

Algoritma Naive Bayes adalah sebuah metode klasifikasi yang berbasis probabilitas dengan mengasumsikan bahwa setiap fitur atau atribut yang ada pada sebuah data, saling bebas (*independen*) satu sama lain. Hal ini memungkinkan untuk menghitung probabilitas kelas (*class probability*) dari sebuah data dengan memanfaatkan hubungan antara setiap atribut dan kelas yang telah diketahui sebelumnya [38]. Algoritma Naive Bayes sangat efektif digunakan dalam bidang *text mining*, khususnya untuk klasifikasi dokumen, analisis sentimen, dan *spam filtering*. Di sisi lain, pada bidang keamanan siber, algoritma ini juga sering digunakan untuk deteksi intrusi (*intrusion detection*) dan identifikasi *malware* [39].

Dalam implementasinya, algoritma Naive Bayes terbagi menjadi tiga jenis, yaitu Naive Bayes Bernoulli, Naive Bayes Multinomial, dan Naive Bayes Gaussian. Setiap jenis memiliki keunggulan dan kelemahan masing-masing tergantung pada karakteristik dari *dataset* yang akan diolah [40]. Pada dasarnya, algoritma Naive Bayes dapat diterapkan pada berbagai jenis data yang memiliki banyak atribut atau fitur dengan cara menghitung nilai probabilitas kelas yang terbaik (*maximum a posteriori*) dari sebuah data

yang baru. Dalam hal ini, perhitungan probabilitas kelas dilakukan dengan memanfaatkan teorema Bayes (*Bayes' theorem*) [41].

Rumus dasar algoritma Naïve Bayes adalah sebagai berikut:

$$P(C|X) = (P(X|C) * P(C)) / P(X)$$

Rumus 2.2 Rumus Algoritma Naïve Bayes

dimana:

$P(C|X)$  adalah probabilitas kelas target C, diberikan data fitur X.

$P(X|C)$  adalah probabilitas data fitur X diberikan kelas target C.

$P(C)$  adalah probabilitas kelas target C.

$P(X)$  adalah probabilitas data fitur X.

Pada tahap pelatihan, algoritma Naive Bayes menghitung probabilitas  $P(C)$  untuk setiap kelas target dan probabilitas  $P(X|C)$  untuk setiap kombinasi fitur dan kelas target dari data pelatihan yang tersedia. Pada tahap pengujian atau prediksi, algoritma Naive Bayes menggunakan rumus di atas untuk menghitung probabilitas  $P(C|X)$  untuk setiap kelas target C dan memilih kelas target dengan probabilitas tertinggi sebagai hasil prediksi [42].

### 2.2.3 Algoritma Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah salah satu algoritma pembelajaran mesin yang populer untuk klasifikasi dan regresi. SVM bekerja dengan mencari *hyperplane* optimal untuk memisahkan dua kelas atau menghubungkan data dalam kasus regresi. *Hyperplane* ini dipilih berdasarkan margin terbesar di antara dua kelas, yang disebut sebagai *Maximum Margin Classifier*. SVM juga dapat melakukan klasifikasi non-linear dengan menggunakan kernel, yang memetakan data ke dalam dimensi yang lebih tinggi sehingga dapat dipisahkan oleh *hyperplane* linear [43].

SVM terkenal karena kemampuannya dalam menangani masalah klasifikasi data yang kompleks dan besar. Selain itu, SVM juga memiliki keuntungan dalam meminimalkan *overfitting*, yaitu kemampuannya untuk menemukan *hyperplane* terbaik dengan margin terbesar. Hal ini

memungkinkan SVM untuk menghasilkan model yang lebih umum dan dapat digeneralisasi dengan baik ke data baru [44].

SVM juga telah dikembangkan untuk kasus klasifikasi multi-kelas dengan menggunakan metode One-vs-All atau One-vs-One. SVM dengan metode One-vs-All lebih efektif untuk klasifikasi multi-kelas dengan kinerja yang lebih baik daripada metode One-vs-One. Namun, kinerja SVM sangat dipengaruhi oleh pemilihan parameter kernel dan C. Kernel yang berbeda dapat digunakan untuk menangani data yang berbeda, dan C merupakan parameter yang mengontrol pengaruh margin dan kesalahan klasifikasi pada hasil akhir SVM [45].

Secara keseluruhan, SVM adalah algoritma yang sangat berguna dan efektif untuk klasifikasi dan regresi dalam pembelajaran mesin. Namun, perlu diingat bahwa pemilihan parameter kernel dan C yang tepat sangat penting untuk mencapai kinerja yang optimal [44].

Rumus matematis untuk SVM bergantung pada formulasi klasifikasi atau regresi yang digunakan. Terdapat beberapa formulasi SVM yang umum digunakan, seperti SVM linear, SVM non-linear menggunakan kernel, SVM multi-kelas, dan sebagainya. Berikut adalah rumus umum untuk SVM linear:

Untuk klasifikasi biner:

$$\arg \min ||w||^2 + C * \Sigma(\max(0, 1 - y_i * (w * x_i + b)))$$

Rumus 2.3 Rumus Algoritma SVM untuk Klasifikasi Biner

Untuk regresi:

$$\arg \min ||w||^2 + C * \Sigma(\max(0, |y_i - (w * x_i + b)| - \epsilon))$$

Rumus 2.4 Rumus Algoritma SVM untuk Regresi

Di dalam rumus ini:

w adalah vektor bobot yang harus diestimasi.

b adalah bias.

$x_i$  adalah vektor fitur dari sampel ke-i.

$y_i$  adalah label kelas dari sampel ke-i.

C adalah konstanta yang mengontrol *trade-off* antara margin dan jumlah kesalahan klasifikasi/regresi.

$\epsilon$  adalah batas toleransi pada regresi *epsilon-insensitive*.

Pada formulasi ini,  $w$  dan  $b$  akan diestimasi dengan meminimalkan fungsi objektif yang terdiri dari dua bagian: bagian pertama ( $\|w\|^2$ ) bertanggung jawab untuk memaksimalkan margin, sedangkan bagian kedua bertanggung jawab untuk mengendalikan kesalahan klasifikasi/regresi [46].

#### 2.2.4 Algoritma K-Nearest Neighbour (KNN)

Algoritma K-Nearest Neighbors (K-NN) adalah salah satu algoritma *Machine Learning* yang paling sederhana dan populer. Metode ini digunakan dalam klasifikasi data dan regresi dengan cara mencari  $k$  tetangga terdekat dari suatu data *point* dan kemudian mengambil mayoritas label kelas dari tetangga-tetangganya sebagai prediksi kelas data *point* tersebut. Algoritma K-NN memanfaatkan konsep jarak antar data *point* dalam pengelompokan data, sehingga perlu dilakukan normalisasi data agar setiap *feature* memiliki bobot yang sama dalam menentukan jarak antar data *point* [47].

Kelebihan algoritma K-NN adalah mudah diimplementasikan dan dapat digunakan untuk *dataset* yang kompleks. Namun, algoritma K-NN memiliki kekurangan yaitu sensitif terhadap nilai  $k$  (jumlah tetangga terdekat) dan rentan terhadap data yang *noise*. Selain itu, kecepatan komputasi algoritma K-NN juga tergantung pada jumlah data dan jumlah *feature* yang dimiliki [47].

Untuk mengatasi kekurangan algoritma K-NN, beberapa penelitian telah dilakukan dengan mengembangkan variasi algoritma K-NN seperti Weighted K-NN (W-KNN) dan *Adaptive* K-NN (A-KNN). W-KNN memberikan bobot yang berbeda untuk setiap tetangga terdekat sehingga tetangga dengan jarak yang lebih dekat memiliki pengaruh yang lebih besar pada prediksi kelas. Pada A-KNN mengubah nilai  $k$  sesuai dengan kepadatan data pada sekitar data *point* yang akan diprediksi [48].

Berdasarkan penelitian terbaru, algoritma K-NN masih menjadi pilihan yang baik dalam memecahkan masalah klasifikasi pada beberapa *dataset* seperti *iris dataset*, *wine dataset*, dan *breast cancer dataset* [49]. Selain itu, algoritma K-NN juga memiliki performa yang baik dalam memprediksi kecelakaan jalan raya. Namun, penggunaan algoritma K-NN perlu diperhatikan dengan cermat pada *dataset* yang memiliki fitur yang tidak relevan atau memiliki *noise* yang tinggi [50].

Rumus dasar algoritma KNN yang melibatkan jarak Euclidean adalah sebagai berikut:

1. Hitung jarak Euclidean antara data baru ( $x$ ) dan setiap titik data pelatihan ( $x_i$ ):

$$D(x_i, x) = \sqrt{\sum (x_i - x)^2}$$

Rumus 2.5 Rumus Jarak Euclidean pada Algoritma KNN

2. Pilih  $k$  tetangga terdekat dengan jarak terkecil.
3. Untuk  $k$  tetangga terdekat yang dipilih, lakukan voting mayoritas untuk menentukan label atau nilai regresi data baru.
4. Berikan label atau nilai regresi yang diprediksi sebagai hasil dari algoritma KNN.

Jarak Euclidean digunakan dalam algoritma KNN karena metrik ini memberikan pengukuran jarak yang sederhana dan intuitif antara dua titik dalam ruang fitur. Jarak Euclidean menghitung jarak linear langsung antara dua titik dalam ruang fitur, dihitung sebagai akar kuadrat dari jumlah kuadrat perbedaan antara koordinat titik-titik tersebut. Penggunaan jarak Euclidean dalam KNN memungkinkan kita untuk mengukur kedekatan antara data baru dan data pelatihan dengan mudah, dan dengan menggunakan  $k$ -nearest neighbors terdekat, kita dapat membuat prediksi berdasarkan mayoritas label atau nilai regresi dari tetangga terdekat tersebut [51].

## 2.3 Tools Penelitian

### 2.3.1 Python

Python adalah salah satu bahasa pemrograman yang populer digunakan di seluruh dunia. Bahasa pemrograman ini diciptakan pada tahun 1989 oleh Guido van Rossum dan kini menjadi bahasa pemrograman *open-source* yang paling banyak digunakan. Python dapat digunakan untuk berbagai keperluan, seperti pemrograman web, analisis data, pengembangan *game*, dan masih banyak lagi [52].

Python adalah bahasa pemrograman yang mudah dipelajari dan sangat ramah bagi pemula. Python memiliki sintaks yang mudah dipahami, sehingga para pemula dapat dengan mudah mempelajari bahasa ini. Python juga memungkinkan pengembang untuk menghasilkan kode yang bersih dan mudah dibaca, sehingga mempermudah dalam pengembangan perangkat lunak yang kompleks [53].

Selain itu, Python juga memiliki banyak *library* dan *framework* yang dapat digunakan untuk mempercepat pengembangan perangkat lunak. Beberapa contoh *library* dan *framework* yang populer di antaranya adalah Pandas untuk analisis data, Django untuk pengembangan web, dan Pygame untuk pengembangan *game* [54].

Dalam beberapa tahun terakhir, Python semakin populer di kalangan pengembang perangkat lunak. Python merupakan bahasa pemrograman yang paling banyak digunakan di dunia, diikuti oleh JavaScript dan SQL. Hal ini menunjukkan bahwa Python semakin menjadi pilihan utama bagi pengembang perangkat lunak [55].

Secara keseluruhan, Python adalah bahasa pemrograman yang populer dan mudah dipelajari bagi pemula. Python juga memiliki banyak *library* dan *framework* yang dapat digunakan untuk mempercepat pengembangan perangkat lunak. Hal ini membuat Python semakin populer di kalangan pengembang perangkat lunak dalam beberapa tahun terakhir [56].

### 2.3.2 Google Colab

Google Colab adalah platform pengembangan dan penelitian yang disediakan oleh Google secara gratis. Platform ini menggunakan infrastruktur cloud dan memungkinkan pengguna untuk menulis dan mengeksekusi kode Python melalui browser web. Dengan menggunakan Google Colab, pengguna dapat mengakses dan menjalankan kode mereka secara *online*, tanpa perlu menginstal perangkat lunak atau mengatur lingkungan pengembangan mereka sendiri. Google Colab juga menyediakan akses ke sumber daya komputasi yang kuat, termasuk unit pemrosesan grafik (GPU) dan unit pemrosesan tensor (TPU), yang memungkinkan pengguna untuk melatih dan menjalankan model pembelajaran mesin dengan cepat [57].

Google Colab menyediakan banyak fitur yang berguna bagi para peneliti dan pengembang. Salah satu fitur penting adalah kemampuan untuk menyimpan, berbagi, dan mengelola notebook Jupyter secara *online*. Pengguna dapat menyimpan *notebook* di Google Drive mereka dan berbagi dengan orang lain untuk kolaborasi. Selain itu, Google Colab juga mendukung penggunaan berbagai *library* dan *framework* populer seperti TensorFlow, PyTorch, dan scikit-learn, yang memudahkan pengembangan dan implementasi model pembelajaran mesin. Dengan Google Colab, pengguna juga dapat mengakses dataset publik, melakukan visualisasi data, dan menulis laporan penelitian dalam bentuk *notebook* yang dapat dibagikan [57].

Google Colab juga mengintegrasikan alat pembelajaran mesin seperti TensorFlow dan Keras, yang memungkinkan pengguna untuk melakukan pelatihan model secara efisien. Dengan dukungan untuk GPU dan TPU, Google Colab memberikan akses ke daya komputasi yang tinggi untuk melatih model secara cepat. Selain itu, platform ini juga menyediakan akses ke Google Cloud BigQuery, yang memungkinkan pengguna untuk mengakses dan menganalisis dataset besar dalam proyek mereka. Dalam beberapa tahun terakhir, Google Colab telah menjadi pilihan populer bagi

para peneliti dan pengembang dalam komunitas pembelajaran mesin karena keterjangkauan, keandalan, dan kemudahan penggunaannya [58].

Google Colab juga telah terus mengalami perkembangan dan peningkatan fitur sejak tahun 2018. Pada tahun 2021, Google mengumumkan pengenalan Colab Pro, yang merupakan langganan berbayar yang memberikan akses lebih cepat, lebih banyak sumber daya, dan fitur tambahan kepada pengguna. Fitur-fitur ini meliputi akses prioritas ke GPU dan TPU, penggunaan lebih lama untuk sesi komputasi, dan dukungan teknis yang lebih baik. Selain itu, Google Colab juga terus memperbarui versi Python dan pustaka yang tersedia, sehingga pengguna dapat memanfaatkan fitur dan perbaikan terbaru. Semua perkembangan ini menjadikan Google Colab sebagai salah satu platform terkemuka untuk pengembangan dan penelitian dalam bidang pembelajaran mesin dan ilmu data [59].

### 2.3.3 Flask

Flask adalah salah satu kerangka kerja (*framework*) yang populer untuk pengembangan aplikasi web dengan menggunakan bahasa pemrograman Python. Dikembangkan oleh Armin Ronacher pada tahun 2010, Flask dirancang dengan prinsip-prinsip kecil dan sederhana, yang memberikan fleksibilitas kepada pengembang dalam membangun aplikasi web yang ringan dan skalabel. Flask menggunakan basis Werkzeug, yang menyediakan implementasi yang efisien untuk protokol HTTP dan *routing* URL. Flask juga memanfaatkan Jinja2 sebagai mesin *template* untuk memisahkan logika aplikasi dan tampilan [60].

Salah satu keunggulan utama Flask adalah kemudahan penggunaannya. Dibandingkan dengan beberapa kerangka kerja web Python lainnya, Flask memiliki kurangnya ketergantungan dan struktur yang ringan, sehingga lebih mudah dipelajari dan digunakan oleh pengembang pemula. Flask juga mendukung pengembangan berbasis modul, di mana pengembang dapat

menambahkan fungsionalitas tambahan melalui ekstensi yang tersedia secara luas.

Selain itu, Flask memiliki komunitas pengembang yang aktif, yang memungkinkan pengguna untuk mendapatkan dukungan dan sumber daya yang melimpah. Dalam beberapa tahun terakhir, Flask terus berkembang dan memperkenalkan fitur-fitur baru, seperti dukungan untuk asinkronisme dengan menggunakan Python 3.7 dan lebih baiknya pengelolaan sumber daya statis. Dengan demikian, Flask tetap menjadi salah satu pilihan yang populer untuk pengembangan aplikasi web menggunakan Python [61].

## 2.4 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Artikel Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
Clustering based semi-supervised machine learning for DDoS attack classification [9]	Journal of King Saud University – Computer and Information Sciences	Aamir M., Zaidi S. M. A. / 2019	Penelitian ini berhasil mengembangkan model klasifikasi serangan DDoS dengan tingkat akurasi sebesar 98%. Selain itu, teknik <i>semi-supervised clustering</i> yang digunakan dalam penelitian ini juga mampu meningkatkan akurasi klasifikasi dibandingkan dengan metode <i>supervised learning</i> tradisional.	Teknik pembelajaran mesin <i>semi-supervised</i> berbasis klustering dapat menjadi pilihan yang efektif untuk mengklasifikasikan serangan DDoS dan dapat membantu meningkatkan keamanan jaringan.
An accurate diabetes prediction system based on K-means clustering and proposed classification approach [11]	Journal of Ambient Intelligence and Humanized Computing	Maqbool S., Raza S., Shafique M., Javed M. Y. / 2019	Penulis berhasil mengembangkan sistem prediksi diabetes yang akurat dengan menggunakan metode K-means <i>clustering</i> dan pendekatan klasifikasi. Dalam penelitian ini, fitur-fitur yang paling signifikan dalam memprediksi diabetes adalah usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah puasa, dan	Hasil eksperimen menunjukkan bahwa sistem yang diusulkan mampu memberikan hasil prediksi yang lebih akurat dibandingkan dengan metode yang sudah ada sebelumnya. Oleh karena itu, penggunaan sistem prediksi yang diusulkan ini dapat membantu dalam mendiagnosis diabetes secara dini dan memungkinkan

			riwayat keluarga. Sistem yang dikembangkan memiliki akurasi sebesar 91.15%, sensitivitas sebesar 91.25%, dan spesifisitas sebesar 91.05%.	penanganan yang lebih efektif dan tepat waktu.
Automated Complaints Classification Using Modified Nazief-Adriani Stemming Algorithm and Naïve Bayes Classifier [62]	Journal of Theoretical and Applied Information Technology	Marcel B. K., Seng H., Vannia F. / 2019	Hasil penelitian menunjukkan bahwa algoritma Naive Bayes Classifier (NBC) yang dikombinasikan dengan algoritma stemming Nazief-Adriani yang dimodifikasi mampu melakukan klasifikasi dengan baik. Hal ini terlihat dari nilai presisi sebesar 91,86%, nilai <i>recall</i> sebesar 84,48%, dan nilai f-1 score sebesar 86,29% untuk rasio data pelatihan dan data pengujian 90:10, serta akurasi rata-rata sebesar 86%. Hasil ini menunjukkan bahwa sistem klasifikasi otomatis dapat membantu dalam mengatasi masalah pengelompokan keluhan secara manual yang memakan banyak waktu dan sumber daya manusia. Dengan adanya sistem ini, proses pengelompokan keluhan dapat dilakukan secara efisien dan akurat.	Penggunaan algoritma Naive Bayes Classifier (NBC) dengan algoritma stemming Nazief-Adriani yang dimodifikasi dapat meningkatkan efisiensi dalam pengelompokan keluhan mahasiswa. Hal ini dibuktikan dengan tingginya nilai presisi, <i>recall</i> , dan f-1 score, serta tingkat akurasi yang mencapai 86%. Dengan adanya sistem klasifikasi otomatis, waktu dan sumber daya manusia yang diperlukan oleh DKBM UMN dalam pengelompokan keluhan dapat dikurangi, sehingga proses penanganan keluhan dapat dilakukan dengan lebih cepat dan efektif.
MRI brain tumour segmentation using hybrid clustering and classification by back propagation algorithm [12]	Asian Pacific Journal of Cancer Prevention	M Malathi, P Sinthia / 2018	Hasil penelitian menunjukkan bahwa metode <i>hybrid clustering</i> dan klasifikasi dengan algoritma <i>backpropagation</i> dapat digunakan untuk segmentasi tumor otak pada citra MRI dengan	Metode yang diusulkan memiliki tingkat keberhasilan yang lebih baik daripada teknik segmentasi tumor otak lainnya. Oleh karena itu, metode ini dapat menjadi alternatif yang potensial dalam diagnosis dan

			tingkat akurasi yang tinggi.	perawatan tumor otak di masa depan.
Prediction of Heart Disease by Clustering and Classification Techniques [13]	International Journal of Computer Sciences and Engineering	R Singh, E Rajesh / 2019	Penelitian ini menunjukkan bahwa teknik klustering dan klasifikasi dengan menggunakan <i>machine learning</i> dapat digunakan untuk memprediksi risiko penyakit jantung pada pasien dengan akurasi yang cukup tinggi. Hasil dari algoritma <i>Decision Tree</i> menunjukkan akurasi tertinggi sebesar 85,47% dalam memprediksi risiko penyakit jantung pada pasien.	Teknik klustering dan klasifikasi dapat digunakan untuk memprediksi penyakit jantung dengan akurasi yang cukup baik. Kedua teknik ini dapat membantu dalam mendiagnosis dan mengobati penyakit jantung secara lebih efektif dan efisien. Namun, penelitian lebih lanjut diperlukan untuk meningkatkan kinerja dan validitas model yang digunakan.
A New Intelligent Medical Decision Support System Based on Enhanced Hierarchical Clustering and Random Decision Forest for the Classification of Alcoholic Liver Damage, Primary Hepatoma, Liver Cirrhosis, and Cholelithiasis [14]	Journal of Healthcare Engineering	Singh A., Pandey B. / 2021	Penelitian ini berhasil membangun sistem <i>pendukung</i> keputusan medis cerdas yang dapat <i>mengklasifikasikan</i> kerusakan hati akibat alkohol, hepatoma primer, sirosis hati, dan kolesistiasis dengan akurasi tinggi. Sistem ini dapat membantu dokter dalam melakukan diagnosis yang lebih akurat dan efisien.	Artikel jurnal <i>ini</i> menunjukkan potensi teknik yang lebih baik dalam bidang medis untuk membantu dokter dalam diagnosis dan pengobatan penyakit hati.

Berdasarkan Tabel 2.1 Penelitian Terdahulu, dapat disimpulkan bahwa teknik klustering dan klasifikasi dalam *machine learning* memiliki potensi yang cukup baik dalam berbagai bidang, seperti keamanan jaringan, prediksi penyakit, dan dukungan keputusan medis. Teknik-teknik ini telah berhasil digunakan dalam mengklasifikasikan serangan DDoS, memprediksi diabetes, segmentasi tumor otak

pada citra MRI, prediksi penyakit jantung, serta mengklasifikasikan berbagai jenis penyakit hati.

Penggunaan teknik klustering dan klasifikasi dalam *machine learning* dapat meningkatkan akurasi, efisiensi, dan efektivitas dalam mendiagnosis dan mengobati berbagai kondisi kesehatan. Selain itu, sistem yang dikembangkan berdasarkan teknik ini dapat membantu para profesional dalam mengambil keputusan yang lebih tepat dan tepat waktu. Namun, penelitian lebih lanjut diperlukan untuk terus mengoptimalkan dan mengembangkan model yang digunakan, sehingga bisa mencapai hasil yang lebih baik dan lebih valid.

Keterbaruan dari penelitian “*Model for Predicting Risk Levels in Maternal Healthcare*” [4] pada dataset yang sama adalah penelitian ini menggunakan metode klustering terlebih dahulu yaitu penggunaan algoritma K-Means sebelum dilakukan klasifikasi menggunakan algoritma Naive Bayes, SVM (Support Vector Machine), dan KNN (K-Nearest Neighbors) dalam penentuan level risiko kehamilan pada seorang ibu. Pada penelitian terdahulu tersebut, hanya menggunakan metode klasifikasi saja dengan algoritma KNN, Neural Network, Random Forest, serta AdaBoost. Berdasarkan kesimpulan yang didapatkan dari penelitian menggunakan teknik klustering dan klasifikasi dalam *machine learning*, diharapkan penerapan teknik klustering dan klasifikasi dapat meningkatkan akurasi yang dihasilkan dalam memprediksi level risiko kehamilan pada seorang ibu.

U M N  
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A