

## BAB II

### TINJAUN LITERATUR

#### 2.1. Penelitian Terdahulu

Tabel 1.2 *literature review* (SLR)

| No | Peneliti & Tahun                          | Objek Penelitian           | Metode / Algoritma                             | Dataset                                        | Tujuan Penelitian                                                                                              | Hasil Penelitian                                                                                                |
|----|-------------------------------------------|----------------------------|------------------------------------------------|------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| 1  | (Herliawan et al., 2020)                  | Penyakit Liver             | Random Forest + Backward Elimination           | Indian Liver Patient Dataset (ILDLP), 583 data | Mengklasifikasikan penyakit liver dengan optimasi seleksi fitur                                                | Akurasi 76.00% dan AUC 0.758 (fair classification)                                                              |
| 2  | (Kesuma et al., 2023)                     | Penyakit Jantung           | Random Forest Classifier                       | Dataset penyakit jantung (UCI)                 | Memprediksi penyakit jantung berdasarkan faktor klinis pasien                                                  | Akurasi tinggi (>80%), Random Forest efektif dalam klasifikasi penyakit jantung                                 |
| 3  | (Rozy, 2024)                              | Penyakit Virus Hepatitis C | Random Forest                                  | UCI Machine Learning Repository                | Memprediksi infeksi virus Hepatitis C (HCV) berdasarkan pemeriksaan darah menggunakan algoritma Random Forest. | Akurasi tinggi dalam mendeteksi HCV, dengan akurasi 91.87%, presisi 91.20%, recall 93.50%, dan F1-score 92.34%, |
| 4  | (Anggi Mei Sarah1 & Endang Warsini, 2023) | Anemia                     | Berbagai Metode Regresi (Linear, DT, RF, dll.) | Dataset Anemia (Kaggle), 1.441 data            | Memprediksi anemia secara dini dan menentukan metode terbaik                                                   | Decision Tree Regression terbaik dengan akurasi 99%                                                             |
| 5  | (Celomitha et al., 2025)                  | Stunting Balita            | Random Forest                                  | Data Balita Desa Kesenana                      | Mengidentifikasi faktor risiko utama stunting secara lokal                                                     | Random Forest efektif mengidentifikasi faktor dominan stunting                                                  |

## 2.2. Penyakit Liver

Penyakit liver merupakan gangguan pada organ hati yang menyebabkan penurunan fungsi metabolisme, detoksifikasi, dan sintesis protein esensial dalam tubuh. Organ hati berperan penting dalam berbagai fungsi fisiologis termasuk metabolisme nutrien, eliminasi toksin, serta produksi protein penting untuk homeostasis tubuh, sehingga kerusakan fungsinya dapat berdampak luas pada kesehatan pasien (Rosida, 2016). Penyakit liver mencakup berbagai kondisi, seperti hepatitis, sirosis, dan penyakit hati berlemak non-alkoholik, yang dapat berkembang menjadi kanker hepatoseluler jika tidak ditangani dengan tepat (Wazir et al., 2023). Selain itu, penyakit hati seringkali menunjukkan gejala yang tidak spesifik pada tahap awal, sehingga deteksi dini menjadi krusial untuk mencegah perkembangan penyakit ke tahap yang lebih berat dan meningkatkan peluang keberhasilan terapi (Pusporani & Qomariyah, 2019). Dengan meningkatnya digitalisasi data medis dan kemajuan komputasi, analisis penyakit liver kini dapat dibantu oleh metode *machine learning* yang mampu memproses data numerik medis secara cepat dan objektif untuk mendukung proses diagnosis.

## 2.3. Mesin Learning

Machine Learning (ML) merupakan cabang dari *Artificial Intelligence* (AI) yang memungkinkan komputer untuk belajar dari data dan menghasilkan prediksi tanpa diprogram secara eksplisit untuk setiap aturan prediksi. Dalam konteks diagnosis medis, pendekatan *supervised learning* terutama digunakan karena model dilatih menggunakan data yang sudah berlabel untuk mempelajari hubungan antara fitur klinis dan kondisi penyakit tertentu, sehingga mampu melakukan klasifikasi kondisi kesehatan secara otomatis (Dritsas & Trigka, 2023). Berbagai algoritma machine learning telah diaplikasikan dalam prediksi risiko dan klasifikasi penyakit liver dengan tujuan meningkatkan akurasi diagnosis berdasarkan parameter medis, seperti kadar enzim hati, bilirubin, dan fitur klinis lain yang direkam dalam basis data pasien (Mane, 2023). Penelitian dalam domain hepatologi menunjukkan bahwa penggunaan ML termasuk supervised learning dapat membantu mengidentifikasi pola kompleks dalam data medis yang tidak

mudah ditangkap oleh metode statistik konvensional, sehingga mempercepat dan memperbaiki proses diagnosis serta stratifikasi risiko pasien (Ali & Maseeh, 2025).

#### **2.4. Random Forest**

Random Forest adalah algoritma *ensemble learning* yang bekerja dengan membangun banyak *decision tree* independen dan menggabungkan hasil prediksi setiap pohon melalui mekanisme *majority voting*, sehingga dapat meningkatkan stabilitas prediksi sekaligus mengurangi risiko *overfitting* bila dibandingkan dengan satu pohon keputusan tunggal (Miftakhuddin et al., 2025). Keunggulan ini membuat Random Forest cocok untuk digunakan pada dataset medis yang kompleks dan bersifat non-linear, di mana banyak atribut klinis saling berinteraksi secara dinamis. Beberapa penelitian dalam bidang kesehatan telah melaporkan performa yang baik dari algoritma Random Forest dalam klasifikasi data medis. Misalnya, penelitian yang membandingkan algoritma Random Forest dan Decision Tree dalam diagnosis penyakit hepatitis menunjukkan bahwa Random Forest mampu mencapai tingkat akurasi yang lebih tinggi, yaitu 90,32%, dibandingkan Decision Tree (Riansyah et al., 2025). Selain itu, studi lain dalam prediksi penyakit jantung melalui Random Forest juga menemukan akurasi sampai sekitar 91% dalam mengklasifikasikan data pasien berdasarkan parameter klinis mereka (Firmansyah & Yulianto, 2023). Temuan-temuan ini memperkuat pemahaman bahwa Random Forest dapat menjadi alat yang efektif dan andal untuk mendukung diagnosis penyakit berbasis data medis.

#### **2.5. Google Colab**

Google Colab merupakan platform *cloud-based* yang memungkinkan eksekusi kode Python dan pembangunan model machine learning secara interaktif tanpa instalasi lokal. Colab menyediakan akses GPU dan TPU sehingga dapat mempercepat proses pelatihan model (Carneiro et al., 2018). Selain itu, platform ini telah digunakan secara luas dalam riset dan pendidikan data science karena integrasinya dengan pustaka ML seperti Scikit-learn, Pandas, dan TensorFlow.