

## BAB II

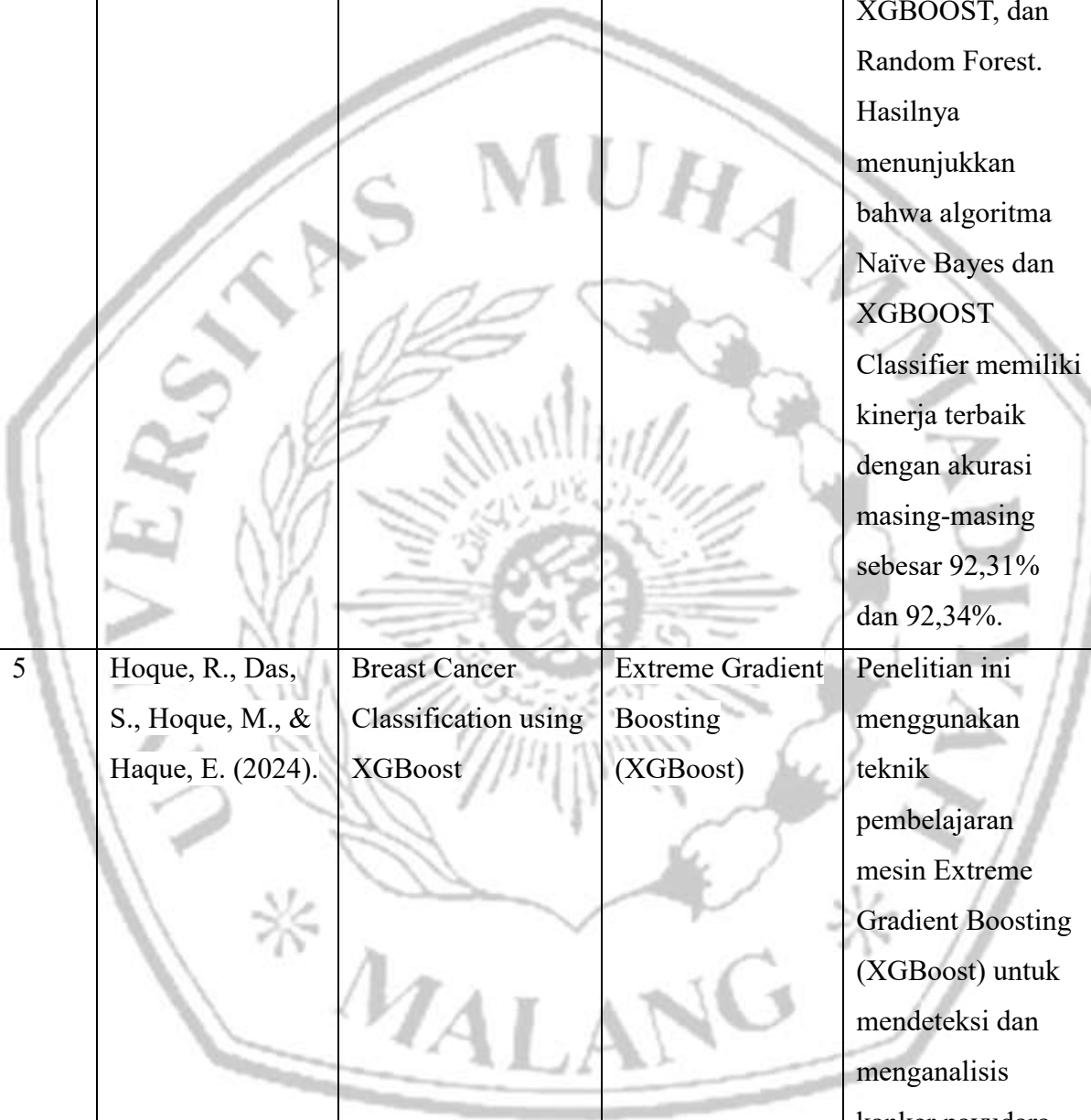
### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Tabel 2.1. Penelitian Terdahulu

| No. | Penulis   | Judul  | Metode                 | Hasil  |
|-----|---|--|------------------------|--|
| 1   | Erlin, Yulvia Nora<br>Marlim, Junadhi,<br>Laili Suryati, &<br>Nova Agustina.<br>(2022). | Early Detection of<br>Diabetes Using<br>Machine Learning<br>with Logistic<br>Regression<br>Algorithm | Logistic<br>Regression | Hasil penelitian menggunakan Logistic Regression awalnya menunjukkan akurasi sebesar 77%. Namun, setelah menerapkan teknik SMOTE dan melakukan optimasi hyperparameter dengan GridSearchCV, kinerja model meningkat dengan akurasi mencapai 82%. |
| 2   | Aamir, K. M.,<br>Sarfriz, L.,<br>Ramzan, M.,<br>Bilal, M., Shafi,                       | A Fuzzy Rule-<br>Based System for<br>Classification of<br>Diabetes                                   | Logika Fuzzy           | Penelitian ini mengembangkan sistem berbasis aturan fuzzy untuk prediksi dini  |

|   |  |   |  |   |
|---|--|---|--|---|
|   | J., & Attique, M. (2021)   |   |  | diabetes, menggunakan dua pengklasifikasi fuzzy dengan jarak Euclidean dan amplitudo kosinus. Hasilnya menunjukkan akurasi sebesar 96,47% dan 95,38%.                             |
| 3 | Hairani Hairani, H., & Dadang Priyanto, D. P. (2023).                    | A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data | Support Vector Machine (SVM) dan Random Forest   | Hasil Penelitian menunjukkan metode Random Forest dengan SMOTE-ENN mencapai akurasi 95,8%, sensitivitas 98,3%, dan spesifisitas 92,5%, mengungguli SVM dan penelitian sebelumnya. |
| 4 | Kanwal, F., Abid, M. K., Maqbool, M. S., Aslam, N., & Fuzail, M. (2023). | Optimized Classification of Cardiovascular Disease Using Machine Learning Paradigms Fouzia                                      | Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), XGBOOST dan Random Forest | Penelitian ini menggunakan lima algoritma berbasis ML untuk mendiagnosis penyakit kardiovaskular (CVD), yaitu   |



|   |  |  |                                     |   |
|---|--|--|-------------------------------------|---|
|   |  |  |                                     | <p>Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), XGBOOST, dan Random Forest.</p> <p>Hasilnya menunjukkan bahwa algoritma Naïve Bayes dan XGBOOST Classifier memiliki kinerja terbaik dengan akurasi masing-masing sebesar 92,31% dan 92,34%.</p> |
| 5 | Hoque, R., Das, S., Hoque, M., & Haque, E. (2024). | Breast Cancer Classification using XGBoost | Extreme Gradient Boosting (XGBoost) | <p>Penelitian ini menggunakan teknik pembelajaran mesin Extreme Gradient Boosting (XGBoost) untuk mendeteksi dan menganalisis kanker payudara.</p> <p>Hasilnya menunjukkan bahwa XGBoost mencapai akurasi</p>   |

|  |  |  |  |   |
|--|--|--|--|---|
|  |  |  |  | sebesar 94,74%<br>pada dataset<br>diagnostik kanker<br>payudara<br>Wisconsin. |
|--|--|--|--|---|

## 2.2 Diabetes

Diabetes mellitus merupakan penyakit jangka panjang yang ditandai oleh tingginya kadar gula darah (hiperglikemia) akibat masalah dalam produksi atau fungsi insulin pada tubuh [14]. Insulin adalah hormon yang diproduksi oleh pankreas dan berperan penting dalam mengatur metabolisme glukosa dalam tubuh [15]. Kekurangan insulin menyebabkan peningkatan gula darah yang dapat menyebabkan komplikasi serius seperti kerusakan pada mata, ginjal, saraf, jantung, dan pembuluh darah.

Diabetes terbagi menjadi dua jenis utama yaitu, tipe 1 dan tipe 2. Diabetes tipe 1 terjadi karena kerusakan pada sel-sel beta pankreas yang menghasilkan insulin, biasanya akibat reaksi autoimun. Ini mengakibatkan ketidakmampuan tubuh untuk memproduksi insulin, yang mengharuskan pasien mendapatkan suntikan insulin eksternal. Biasanya, diabetes tipe 1 muncul pada usia muda dan hanya mencakup sekitar 5-10% dari semua kasus diabetes [16]. Sementara itu, diabetes tipe 2 adalah jenis diabetes yang paling umum, melibatkan sekitar 90-95% dari semua kasus diabetes. Pada diabetes tipe ini, tubuh tidak dapat menggunakan insulin dengan efektif atau pankreas tidak menghasilkan insulin yang cukup untuk memenuhi kebutuhan tubuh. Faktor risiko utamanya termasuk obesitas, kurangnya aktivitas fisik, pola makan yang tidak sehat, dan riwayat keluarga dengan diabetes [17]. Oleh karena itu, deteksi dini dan pengelolaan yang tepat sangat penting untuk mencegah komplikasi dan meningkatkan kualitas hidup pasien diabetes.

## 2.3 Hyperparameter Tuning

Hyperparameter tuning adalah proses mencari kombinasi hyperparameter terbaik untuk meningkatkan kinerja model machine learning. Proses ini penting karena hyperparameter yang tepat dapat meningkatkan kinerja model, sementara yang kurang optimal dapat menyebabkan overfitting, underfitting, atau konvergensi lambat [18]. Hyperparameter Tuning bisa dilakukan manual melalui trial-and-error, tetapi lebih efisien menggunakan metode otomatis seperti GridSearchCV dan RandomSearchCV.

### **2.3.1 GridSearchCV**

GridSearchCV adalah salah satu algoritma yang dapat digunakan untuk menemukan parameter optimal bagi suatu model. Metode ini bekerja dengan mencoba semua kombinasi yang mungkin berdasarkan nilai parameter yang telah ditentukan oleh pengguna. Metode ini menyimpan hasil dari berbagai kombinasi parameter dalam sebuah grid, kemudian memilih parameter terbaik dengan mencari kombinasi yang menghasilkan nilai error terkecil [19].

### **2.3.2 RandomSearchCV**

RandomSearchCV adalah metode tuning hyperparameter yang menggunakan pencarian acak untuk menemukan kombinasi hyperparameter terbaik. Berbeda dengan GridSearchCV yang mencoba semua kemungkinan kombinasi dalam grid yang telah ditentukan, RandomSearchCV secara acak memilih kombinasi hyperparameter dari distribusi yang telah diberikan [20].

## **2.4 Extreme Gradient Boosting (XGBoost)**

XGBoost adalah metode yang dikembangkan oleh Chen dan Guestrin (2016) yang menerapkan konsep Gradient Boosting (GB) yang efisien, cepat, dan terukur [21]. Kelebihan algoritma ini terletak pada kecepatannya dalam proses optimasi, mengungguli metode Gradient Boosting lainnya, baik dalam masalah klasifikasi maupun regresi [22]. XGBoost menggunakan pendekatan pertumbuhan berbasis level (level-wise growth), di mana serangkaian pohon keputusan dibangun dengan model yang bergantung pada model sebelumnya. Model pertama dalam XGBoost memulai dengan prediksi yang lemah, kemudian bobot pada setiap model diperbarui untuk menghasilkan prediksi yang lebih kuat dan akurat [23]. Oleh karena itu, penelitian ini akan menggunakan metode XGBoost untuk klasifikasi deteksi penyakit diabetes dengan melakukan hyperparameter tuning menggunakan GridSearchCV dan RandomSearchCV.