

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini merujuk pada beberapa studi sebelumnya sebagai landasan atau acuan penting. Penelitian-penelitian tersebut memberikan kontribusi signifikan dalam pengembangan pemahaman tentang topik atau metode yang terkait dengan focus penelitian ini.

Peneliti [5] yang digunakan sebagai acuan dalam penelitian ini menerapkan beberapa algoritma *Machine Learning*, termasuk *Decision Trees*, *Random Forest*, *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Artificial Neural Network*, digunakan untuk memprediksi tingkat kemampuan beradaptasi pelajar dalam lingkungan pendidikan online. Dari berbagai algoritma yang diimplementasikan, pengklasifikasi *Random Forest* mencapai akurasi tertinggi sebesar 89,63%, mengungguli kinerja algoritma-algoritma lainnya. Peningkatan performa model *Machine Learning* terus dikembangkan dan ditingkatkan secara optimal.

Peneliti [7], menganalisis beberapa model klasifikasi seperti *Decision Tree*, *Extra Trees Classifier*, *Random Forest*, *Logistic Regression*, *Support Vector Machine*, *Naives Bayes*. Hasil prediksi dianalisis menggunakan parameter evaluasi seperti *Accuracy*, *Precision*, *Recall*, and *F1-score*. Dan hasilnya menunjukkan akurasi yang paling baik adalah *Extra Trees Classifier* dengan akurasi 90%, sedangkan yang lainnya menghasilkan 88%, 87%, 86%. *Extra Trees Classifier* bekerja jauh lebih cepat daripada *Random Tree Classifier*, bahkan bekerja 3 kali lebih cepat.

Peneliti [8], menganalisis sentimen tweet covid-19. Peneliti menggunakan teknik pra-pemrosesan data dan teknik ekstraksi fitur untuk melatih lima model *Machine Learning* yaitu *Random Forest*, *XGBoost*, *Support Vector Classifier*, *Extra Trees Classifier*, dan *Decision Trees*. Selain *Machine Learning*, penelitian ini juga mencoba pendekatan deep learning dengan menggunakan arsitektur *Long Short-Term Memory (LSTM)* untuk menganalisis dataset. Hasilnya menunjukkan

Extra Trees Classifier mencapai akurasi tertinggi sebesar 0.93, sedangkan LSTM memiliki akurasi lebih rendah dikarenakan ukuran dataset yang kecil.

Peneliti [9], mengevaluasi kinerja metode klasifikasi *Extra Trees Classifier* untuk mengklasifikasikan kanker payudara. Peneliti membandingkan dengan Metode *Decision Trees* dan lainnya, hasil perbandingan menampilkan bahwa metode *Extra Trees Classifiers* merupakan kinerja yang lebih unggul yaitu dengan akurasi 99.27%. Peneliti juga melakukan optimasi *hyperparameter tuning Cross Validation* untuk hasil yang lebih optimal. Metode yang digunakan berhasil mengurangi kesalahan dalam mengklasifikasikan pasien kanker payudara, hal ini menunjukkan metode tersebut lebih tepat dan efektif.

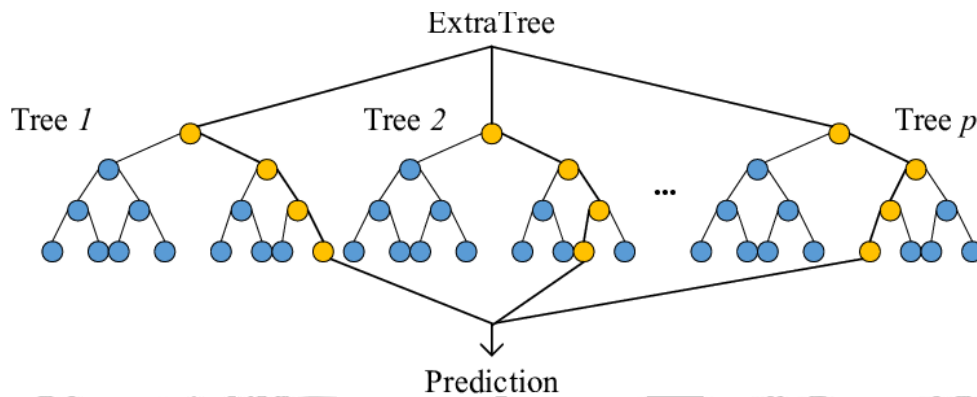
Peneliti [10], menggunakan 6 algoritma *ensemble learning* untuk mendeteksi penyakit pada hati. Algoritma yang digunakan melakukan berbagai tahapan dalam pemrosesan data dan hasil dianalisa menggunakan berbagai metrik evaluasi. Hasil menunjukkan bahwa *Extra Trees Classifier* memiliki akurasi pengujian tertinggi, yaitu 91.82%, diikuti oleh *Random Forest* dengan akurasi 86.06%. Ketika algoritma dikombinasikan dengan teknik validasi silang 10-fold, akurasi yang dicapai sebesar 93.15%.

2.2 *Extra Trees Classifier*

Metode *Extra Trees Classifier* menampilkan struktur dengan *Decision Tree* (DT) yang saling bergabung, mencakup elemen seperti root node, child/split nodes, dan leaf nodes. *Decision Trees* membentuk struktur hierarkis untuk membuat keputusan prediktif. Dalam proses pengambilan keputusan, *Extra Trees Classifier* menggunakan aturan split yang ditentukan oleh subset acak dari karakteristik (fitur) dan titik potong acak parsial pada set data X. Proses ini dimulai di root node dan berlanjut di setiap child node, di mana keputusan dibuat berdasarkan aturan split yang dihasilkan, hingga mencapai leaf node.

Tiga parameter utama *Extra Trees Classifier* mempengaruhi kinerjanya. Jumlah pohon dalam ensemble (k) menentukan seberapa banyak *Decision Trees* digunakan. Jumlah fitur yang dipilih secara acak (f) pada setiap split node membawa variasi dan kompleksitas pada model. Jumlah minimum sampel atau instance (n_{min}) yang diperlukan untuk membagi sebuah node mempengaruhi

kedalaman dan kerumitan pohon keputusan. Dengan memanfaatkan acak parsial, *Extra Trees Classifier* menawarkan fleksibilitas dalam pembuatan keputusan, menghasilkan model yang lebih baik dan mampu menangani variasi data dengan baik [11]. Ilustrasi dari algoritma *Extra Trees Classifier* akan ditunjukkan dengan Gambar 1.



Gambar 1. Algoritma *Extra Trees Classifier*

2.3 *Hyperparameter tuning*

Ada berbagai metode hyperparameter tuning yang tersedia, seperti *Grid Search*, *Random Search*, *Particle Swarm Optimization*, *Bayesian Optimization*, *Genetic Algorithm*, dan lain-lain. Dalam penelitian ini, teknik optimasi model *Machine Learning* yang diterapkan adalah *Grid Search Cross Validation (Grid Search CV)*. Metode ini merupakan suatu pendekatan untuk mengoptimalkan parameter pada model *Machine Learning* dengan cara membangun dan mengevaluasi setiap kombinasi parameter algoritma yang telah ditentukan dalam suatu grid. Pendekatan ini mencakup pencarian menyeluruh terhadap semua kombinasi parameter yang telah ditetapkan, dengan tujuan mendapatkan pengaturan parameter yang memberikan performa model optimal.

Dengan penerapan *Grid Search Cross Validation (Grid Search CV)*, peneliti memiliki kemampuan untuk menjalankan pemindaian pada sejumlah hyperparameter yang ditetapkan. Dalam metode ini, berbagai kombinasi hyperparameter diterapkan pada model, dan setiap kombinasi tersebut dievaluasi untuk performa menggunakan teknik *Cross Validation*. Seluruh rangkaian hyperparameter diuji melalui cross-validation untuk memastikan hasil yang dapat

diandalkan. Kombinasi hyperparameter yang menghasilkan kinerja terbaik, sesuai dengan metrik evaluasi yang dipilih, akan dipilih sebagai konfigurasi hyperparameter optimal untuk model yang dikembangkan. Adanya fleksibilitas pada konfigurasi parameter yang dapat disesuaikan dan divariasikan juga menjadi salah satu keunggulan dari penerapan *Grid Search CV* [12].

Penting untuk dicatat bahwa metode statistik yang digunakan untuk mengevaluasi akurasi model *Machine Learning* adalah *cross-validation (CV)* [6], [13]. Melalui *cross-validation*, model dinilai secara menyeluruh dengan membagi dataset menjadi subset training dan testing secara bergantian, sehingga memastikan bahwa model dapat memberikan performa yang konsisten dan handal pada berbagai subset data.

