1. Pendahuluan

1.1 Latar Belakang

Pada kuartal ketiga tahun 2020, terdapat total 2,86 juta aplikasi Android yang tersedia, dengan rata-rata 482.579 contoh *malware* diidentifikasi setiap bulan, atau sekitar 16.000 contoh per hari [1][2]. Serangan *malware* ini bertujuan untuk mengendalikan atau mengganggu jaringan seluler serta membahayakan data pribadi pengguna [3]. *Malware* sering kali menyamar sebagai file atau tautan yang tampak sah untuk menipu pengguna agar mengunduhnya, sehingga memberikan penyerang akses ke perangkat atau jaringan korban [4][5][6]. Tingginya insiden *malware* ini menyoroti kebutuhan mendesak akan metode deteksi yang lebih canggih untuk melawan ancaman, baik di masa kini maupun di masa depan. Salah satu pendekatan yang menjanjikan untuk deteksi *malware* adalah penggunaan teknik pembelajaran *ensemble*.

Pembelajaran *ensemble* adalah metode yang menggabungkan beberapa model pembelajaran mesin untuk meningkatkan kinerja prediktif dan ketahanan sistem. Teknik ini dapat digunakan untuk meningkatkan deteksi *malware* pada perangkat Android dengan mengatasi kelemahan masing-masing model, sehingga menghasilkan hasil yang lebih akurat dan lebih tahan terhadap variasi data [7]. Dalam penelitian ini, beberapa metode *ensemble* seperti *Random Forest*, *XGBoost*, *Extremely Randomized Trees*, dan *Histogram-based Gradient Boosting* diterapkan untuk mengidentifikasi algoritma yang paling efektif dalam mendeteksi *malware* Android. Setiap metode memiliki keunggulan unik, dan kombinasi dari metode-metode ini bertujuan untuk menyediakan solusi deteksi *malware* yang lebih komprehensif.

Pendekatan ini juga ditingkatkan melalui teknik pemilihan fitur menggunakan *Random Forest Feature Importance*. Proses ini bertujuan untuk mengidentifikasi dan memanfaatkan fitur yang paling relevan yang memiliki kontribusi signifikan terhadap deteksi *malware*, sehingga meningkatkan efisiensi dan akurasi proses dengan memusatkan perhatian algoritma pada data yang paling penting [8]. Selain itu, transparansi model ditingkatkan dengan menggunakan SHAP (*SHapley Additive Explanations*), sebuah teknik *Explainable AI* (XAI) yang memberikan pemahaman mendalam tentang keputusan model. Hal ini memungkinkan pengguna untuk memahami alasan di balik setiap deteksi, sehingga meningkatkan kepercayaan pengguna terhadap sistem dan membantu mereka mengambil tindakan pencegahan yang tepat terhadap serangan *malware* [9].

Penelitian ini bertujuan untuk mengembangkan alat deteksi *malware* berbasis sistem operasi Android dengan menggunakan metode *Machine Learning* (ML) berbasis pembelajaran *ensemble*, yang dikombinasikan dengan teknik pemilihan fitur dan didukung oleh wawasan dari *Explainable AI*. Pendekatan ini tidak hanya menghasilkan sistem deteksi malware yang lebih akurat tetapi juga memberikan transparansi yang lebih tinggi dalam proses pengambilan keputusan model. Transparansi ini memungkinkan identifikasi dan mitigasi ancaman *malware* secara lebih proaktif. Dataset CICMalDroid2020 digunakan dalam penelitian ini untuk pelatihan dan validasi model. Dataset ini menyediakan sampel malware Android yang beragam dan komprehensif, sehingga mendukung pengembangan alat deteksi yang andal [10].

1.2 Rumusan Masalah

Berikut adalah rumusan masalah yang disimpulkan berdasarkan latar belakang:

1. Bagaimana penerapan pembelajaran *ensemble* dan pemilihan fitur *Random Forest Feature Importance* untuk mendeteksi *malware* pada Android?
2. Bagaimana XAI menjelaskan pembelajaran *ensemble* dan pemilihan fitur *Random Forest Feature Importance* untuk deteksi *malware* pada Android yang lebih informatif dan transparan?

1.3 Tujuan

Berikut adalah tujuan yang akan dicapai dari penelitian ini berdasarkan perumusan masalah:

1. Merancang dan mengintegrasikan metode pembelajaran *ensemble* dengan teknik pemilihan fitur *Random Forest Feature Importance*, sehingga sistem deteksi *malware* dapat lebih akurat dalam mengidentifikasi serangan *malware* pada perangkat Android.
2. Memahami dan menjelaskan keputusan yang diambil oleh XAI untuk memperkuat keandalan dan transparansi sistem deteksi *malware* pada perangkat Android. Tujuan ini dicapai dengan memberikan wawasan yang jelas dan mudah dimengerti mengenai bagaimana dan mengapa keputusan deteksi diambil, sehingga meningkatkan kepercayaan pengguna dan kemampuan untuk mengambil tindakan yang tepat terhadap serangan *malware* pada perangkat Android.