BAB 1
PENDAHULUAN

1.1 Latar Belakang

Malicious Software atau malware adalah perangkat lunak berbahaya yang dirancang untuk menyusup, merusak sistem, atau mencuri data dengan memanfaatkan celah keamanan (Yuniati dkk., 2022). Seiring dengan meningkatnya digitalisasi, ancaman malware terus berkembang, menyebabkan dampak signifikan seperti pencurian data pribadi, gangguan operasional, hingga sabotase sistem informasi (Hapsari & Pambayun, 2023). Badan Siber dan Sandi Negara (BSSN) mencatat bahwa pada tahun 2022 terdapat lebih dari 714 juta serangan siber di Indonesia, dengan 56,84% diantaranya berasal dari aktivitas malware (CNN, 2023). Perkembangan malware selama ini sering dikaitkan dengan perangkat komputer. Namun, dengan meningkatnya penggunaan perangkat *mobile*, khususnya Android, ancaman terhadap sistem operasi ini juga semakin besar. Android menjadi target utama malware karena sifatnya yang *open-source*, distribusi aplikasi yang luas, serta minimnya kesadaran pengguna terkait keamanan siber. Menurut laporan (Kaspersky, 2023), sebanyak 1.661.743 malware atau penginstal perangkat lunak yang tidak diinginkan terdeteksi pada tahun 2022. Jumlah serangan terhadap perangkat mobile pada tahun 2023 dilaporkan mencapai 33,8 juta kasus, meningkat hingga 50% dibandingkan tahun sebelumnya (Kaspersky, 2024). Fakta ini menunjukkan bahwa sistem keamanan Android masih memiliki celah yang dapat dimanfaatkan oleh peretas, sehingga diperlukan strategi deteksi yang lebih akurat dalam mengidentifikasi ancaman malware.

Salah satu tantangan dalam keamanan Android adalah fitur *permissions*. *Permissions* merupakan izin yang diberikan oleh pengguna kepada aplikasi untuk mengakses berbagai sumber daya perangkat, seperti kamera, lokasi, penyimpanan, dan pesan teks (Herwanto, 2022). Penyalahgunaan *permissions* telah menjadi strategi utama bagi malware untuk mencuri data atau melakukan tindakan berbahaya tanpa sepengetahuan pengguna (Wanli Sitorus dkk., 2021). Berdasarkan

penelitian (Hanifurohman & Hutagalung, 2020), analisis terhadap beberapa aplikasi *e-commerce* berbasis Android menunjukkan bahwa semua aplikasi yang diteliti menggunakan *dangerous permissions*, yaitu izin yang memungkinkan akses ke data sensitif pengguna. Penyalahgunaan izin ini dapat menyebabkan kebocoran informasi pribadi, terutama jika aplikasi mengandung kode berbahaya atau dikembangkan oleh pihak yang tidak bertanggung jawab. Penelitian (Jha & Lee, 2016) juga menyoroti bahwa mekanisme *permissions* di Android masih kurang ketat, sehingga dapat dimanfaatkan oleh malware untuk melakukan eskalasi hak istimewa, yang pada akhirnya memungkinkan eksploitasi sistem secara lebih luas.

Pendekatan deteksi malware yang memanfaatkan *permissions* telah dikembangkan dengan berbagai metode, namun masih memiliki tantangan tersendiri. Metode *signature-based detection*, yang mengandalkan basis data tanda tangan malware yang sudah ada, memiliki keterbatasan karena tidak mampu mendeteksi malware atau varian baru yang belum terdaftar (Matin dkk., 2023). Sementara itu, metode *heuristic-based detection* yang mencoba mengidentifikasi pola perilaku mencurigakan masih menghadapi tantangan berupa tingginya *False Positives*, yang menyebabkan banyak aplikasi non-malware diklasifikasikan sebagai malware (Berdaya, 2024). Karena keterbatasan tersebut, machine learning menjadi alternatif yang semakin banyak digunakan dalam deteksi malware di fitur *permissions*. Machine learning memungkinkan model untuk mengenali pola yang tidak dapat diidefntifikasi oleh metode konvensional (Judijanto dkk., 2024), serta mampu mempelajari hubungan kompleks antara fitur *permissions* dan klasifikasi aplikasi sebagai malware atau non-malware. Namun, tidak semua algoritma machine learning dapat diimplementasikan secara optimal dalam deteksi malware di fitur *permissions*. Data *permissions* bersifat string (hanya memiliki nilai "meminta izin" atau "tidak meminta izin") dan sering kali mengalami ketidakseimbangan kelas (*imbalanced class*), di mana jumlah aplikasi non-malware dalam dataset umumnya jauh lebih besar dibandingkan dengan aplikasi malware (Magnolia dkk., 2022). Kondisi ini dapat mempengaruhi performa model dalam mengklasifikasikan aplikasi dengan tingkat akurasi tinggi pada kelas mayoritas non-malware, tetapi lemah dalam mengenali kelas minoritas.

Sejumlah penelitian telah dilakukan untuk mengevaluasi berbagai algoritma ML dalam deteksi malware berbasis *permissions*. Algoritma berbasis pohon keputusan, seperti Random Forest dan Gradient Boosting, terbukti memiliki kinerja lebih baik dibandingkan metode lainnya (Meena dkk., 2023). Salah satu pengembangan gradient boosting yang telah menunjukkan hasil unggul adalah Extreme Gradient Boosting (XGBoost). XGBoost memiliki beberapa keunggulan, di antaranya kemampuannya menangani dataset dengan jumlah fitur yang besar (Sukmawati dkk., 2024), mengelola ketidakseimbangan kelas melalui *weighted loss function*, serta memiliki mekanisme regularisasi yang mampu mengurangi risiko *overfitting* (Aziz, 2024). Penelitian telah menunjukkan bahwa penerapan teknik *oversampling* , seperti *Random Over Sampling* (ROS), *Synthetic Minority Over-sampling Technique* (SMOTE), dan *Adaptive Synthetic Sampling* (ADASYN), dapat meningkatkan efektivitas XGBoost dalam mengatasi ketidakseimbangan kelas, terutama dalam klasifikasi penyakit jantung (M. Lutfi, 2024). Dalam konteks prediksi *churn*, metode *boosting* seperti XGBoost, LightGBM, dan CatBoost yang dikombinasikan dengan teknik *weighting* terbukti mampu meningkatkan performa model dalam mengenali kelas minoritas (Nugraha & Syarif, 2023).

Meskipun berbagai penelitian telah membuktikan efektivitas XGBoost dalam mengatasi ketidakseimbangan kelas, penelitian yang secara khusus berfokus pada deteksi malware yang memanfaatkan fitur *permissions* di Android masih terbatas. Penelitian ini menjawab tantangan-tantangan tersebut dengan melakukan pendekatan menggunakan algoritma XGBoost sebagai metode utama untuk menangani ketidakseimbangan kelas dalam deteksi malware Android yang memanfaatkan fitur *permissions*. Eksperimen dilakukan melalui penerapan beberapa teknik penguatan performa, seperti penyesuaian bobot kelas menggunakan parameter *scale_pos_weight*, pengaturan ambang keputusan (*threshold tuning*), serta metode *oversampling Synthetic Minority Over-sampling Technique* (SMOTE).

1.2 Perumusan Masalah

Ketidakseimbangan kelas dalam dataset untuk deteksi malware berbasis *permissions* pada sistem operasi Android dapat mempengaruhi akurasi model dalam membedakan aplikasi non-malware dan malware. Karakteristik data *permissions* yang bersifat string dan distribusi yang tidak merata semakin memperumit proses klasifikasi. Evaluasi diperlukan untuk memahami sejauh mana ketidakseimbangan data mempengaruhi performa deteksi malware.

1.3 Pertanyaan Penelitian

1. Bagaimana algoritma XGBoost dalam menangani ketidakseimbangan kelas pada deteksi malware Android berbasis fitur *permissions*?
2. Sejauh mana fitur *permissions* dalam dataset membantu XGBoost dalam membedakan aplikasi non-malware dan malware?
3. Bagaimana evaluasi performa XGBoost berdasarkan metrik akurasi, *Precision*, *Recall*, *F1-Score*, dan *Confusion matrix* dalam klasifikasi malware Android?

1.4 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan untuk mengatasi ketidakseimbangan kelas (*imbalanced class*) dalam pendeteksian malware berbasis fitur permissions pada sistem operasi Android dengan menerapkan algoritma Extreme Gradient Boosting (XGBoost). Untuk meningkatkan kemampuan model dalam mengenali kelas minoritas (malware), dilakukan beberapa pendekatan, yaitu penyesuaian bobot kelas menggunakan parameter *scale_pos_weight*, pengaturan ambang klasifikasi melalui teknik *threshold tuning*, serta penerapan metode *oversampling Synthetic Minority Over-sampling Technique* (SMOTE). Evaluasi performa model dilakukan menggunakan metrik *accuracy, precision, recall, F1-score,* dan *confusion matrix* guna mengukur efektivitas XGBoost dalam membedakan aplikasi non-malware dan malware pada kondisi distribusi data yang tidak seimbang.

1.5 Manfaat Hasil Penelitian

1. Menambah wawasan tentang penerapan machine learning, khususnya algoritma Extreme Gradient Boosting (XGBoost) dalam deteksi malware Android.
2. Memberikan referensi akademik mengenai efektivitas XGBoost dalam mendeteksi malware berbasis fitur *permissions*, yang dapat dijadikan dasar untuk penelitian lebih lanjut dalam bidang keamanan siber.
3. Menyediakan *insight* bagi praktisi keamanan siber dan pengembang aplikasi, khususnya dalam memilih model machine learning yang tepat untuk deteksi malware Android.

1.6 Batasan Penelitian

Klasifikasi dilakukan secara biner, yaitu hanya membedakan antara aplikasi malware dan non-malware, tanpa mempertimbangkan jenis atau varian malware. Fokus utama penelitian adalah pada penanganan ketidakseimbangan kelas (*imbalanced class*), sehingga eksperimen dilakukan menggunakan teknik seperti *scale_pos_weight*, *threshold tuning*, dan SMOTE.