BAB 1
PENDAHULUAN

1.1. Latar Belakang

Android adalah sistem operasi yang dirancang oleh Google yang saat ini mendominasi pasar global dengan pangsa sebesar 71.8% pada akhir tahun 2023. Keberhasilan ini sebagian besar disebabkan oleh sifat open-source Android yang memungkinkan berbagai produsen perangkat keras seperti Samsung, Xiaomi, Oppo, Vivo, dan lainnya untuk memodifikasi dan menggunakan sistem operasi ini secara bebas [1]. Di berbagai negara seperti India dan Indonesia, pangsa pasar Android bahkan mencapai lebih dari 85%. Keberhasilan Android juga didukung oleh beragam aplikasi yang tersedia di Google Play Store, yang mencapai lebih dari 2.67 juta aplikasi pada Maret 2023. Namun, popularitas Android juga menjadikannya target utama bagi serangan malware, dengan 97% dari semua serangan malware pada tahun 2022 ditujukan pada perangkat Android [2]. Menurut laporan dari Kaspersky, pada tahun 2023 terjadi peningkatan signifikan dalam jumlah serangan terhadap perangkat mobile, mencapai 33,8 juta serangan, yang merupakan peningkatan sebesar 50% dari tahun sebelumnya. Ancaman yang paling umum adalah adware, yang mencakup 40,8% dari semua ancaman yang terdeteksi [3][4].

Android sebagai sistem operasi yang paling banyak digunakan di dunia, sering menjadi target serangan malware. Malicious Software atau yang lebih dikenal sebagai malware adalah perangkat lunak yang dirancang untuk menyusup ke dalam sistem operasi, mengganggu fungsi normal, dan bahkan mencuri data penting dari perangkat korban. Malware memiliki beberapa jenis seperti ransomware, backdoor, adware, file infector, spyware, PUA, riskware, trojan, scareware, trojan-sms, trojan-banker, trojan- spy, dan trojan-dropper. Malware mempunyai cara kerja yang beragam, salah satu contoh serangan malware melalui aplikasi berbahaya yang melakukan akses permission secara ilegal tanpa izin dari pengguna dan sistem operasi. Serangan malware pada perangkat Android dapat menyebabkan kerugian finansial, pencurian data, dan kerusakan reputasi bagi individu maupun organisasi. Kerentanan terhadap serangan malware ini menimbulkan ancaman serius terhadap keamanan sistem operasi Android [5].

Untuk mengatasi permasalahan deteksi dan klasifikasi malware Android yang semakin kompleks, pendekatan menggunakan algoritma machine learning menjadi salah satu solusi yang menjanjikan. Salah satu algoritma yang telah terbukti efektif adalah XGBoost (Extreme Gradient Boosting). XGBoost merupakan algoritma machine learning berbasis pohon

keputusan yang dapat mendeteksi dan mengklasifikasi malware Android dengan kinerja tinggi. Algoritma XGBoost mampu menangani data yang kompleks dan nonlinear, serta dapat mengoptimalkan akurasi prediksi melalui proses boosting. Dengan menggunakan XGBoost, sistem dapat belajar dari data malware sebelumnya dan meningkatkan kemampuan deteksi terhadap varian malware yang baru muncul. Pendekatan machine learning berbasis XGBoost terbukti efektif dalam meningkatkan akurasi dan efisiensi dalam mengidentifikasi malware Android, memberikan lapisan perlindungan tambahan bagi pengguna perangkat Android [6].

Terdapat beberapa penelitian terdahulu menggunakan machine learning dalam pengklasifikasian dan pengidentifikasian malware Android khususnya melalui pendekatan metode algoritma XGBoost (Extreme Gradient Boosting). Penelitian yang dilakukan oleh Fauzi Mohd, dkk. Dengan judul "Android Malware Classification Using Xgboost On Data Image Pattern", dengan menggunakan XGBoost untuk mengklasifikasikan malware berdasarkan pola gambar data dex file, hasil pengujian menunjukkan bahwa model XGBoost yang dibangun mampu mencapai akurasi 98,2% dalam membedakan aplikasi malware dan benign pada platform Android [7]. Selain itu, penelitian lain oleh Jong wang, dkk. Dengan judul "XGBoost-Based Android Malware Detection" juga menunjukkan bahwa penggunaan model XGBoost dalam deteksi malware Android memberikan efisiensi komputasi yang baik dan akurasi klasifikasi yang tinggi. Dalam penelitian tersebut, eksperimen yang dilakukan menunjukkan bahwa XGBoost mampu mencapai akurasi deteksi sebesar 99% dengan waktu pelatihan yang lebih singkat dibandingkan dengan metode lain seperti Support Vector Machine [8]. Penelitian serupa juga dilakukan oleh Narayanan, M. E, dkk. Dalam judul "Malware Classification Using Xgboost With Vote Based Backward Feature Elimination Technique", pada penelitian tersebut bahwa hasil penelitian menunjukan Model XGB-VBFE yang diusulkan mampu mengklasifikasikan file malware dan benign dengan akurasi tertinggi sebesar 99.50%, presisi 0.99, dan recall 0.96. Model ini juga menunjukkan waktu pelatihan yang lebih singkat dibandingkan dengan algoritma lain seperti SVM dan Random Forest [9]. Selanjutnya penelitian yang dilakukan oleh Palša et al, dkk. Dengan judul "MLMD—A Malware-Detecting Antivirus Tool Based on the XGBoost Machine Learning Algorithm". Juga menunjukkan hasil penelitian menunjukkan bahwa model terbaik adalah yang menggunakan algoritma XGBoost, dengan akurasi deteksi mencapai 91,9% dan sensitivitas 98,2% pada dataset analisis statis, serta akurasi 96,4% dan sensitivitas 98,5% pada dataset analisis dinamis [10].

Dari uraian diatas, penulis ingin mengusulkan pada penelitian ini menggunakan metode XGBoost dalam mengklasifikasi malware Android dan juga dengan diterapkannya RFE dan Multikolinearitas Remove pada Feature Selection. Penelitian ini diharapkan dapat memberikan

kontribusi yang signifikan dalam bidang keamanan siber, khususnya dalam deteksi dan klasifikasi malware pada perangkat Android.

1.2. Rumusan Masalah

1. Bagaimana hasil klasifikasi yang diperoleh model deteksi malware android menggunakan algoritma XGBoost?
2. Bagaimana perbedaan akurasi hasil klasifikasi dengan menggunakan Feature Selection dan tanpa Feature Selection ?

1.3. Tujuan Penelitian

Tujuan penelitian ini untuk mengevaluasi dan analisis hasil klasifikasi model deteksi malware Android menggunakan algoritma XGBoost dengan implementasi Feature Selection.

1.4. Batasan Masalah

1. Model yang digunakan adalah XGBoost (Extreme Gradient Boosting).
2. Penelitian ini fokus pada klasifikasi malware android.
3. Hasil dari klasifikasi model yang diperoleh nantinya akan digunakan sebagai pembanding dengan hasil klasifikasi model pada penelitian terdahulu.
4. Dataset yang digunakan berasal dari Kaggel (Malware Detection Dataset).
5. dataset yang berjumlah 29.333 dengan 2 kategori benign & malware Android apps.