BAB I
PENDAHULUAN

1.1 Latar Belakang

Android, merupakan sebuah operasi sistem yang dirancang oleh Google. Saat ini di tahun 2023 sekitar 3,6 miliar orang di seluruh dunia menggunakan HP Android, dan diketahui bahwa dengan pangsa pasar global sebesar 71,75% pada akhir tahun 2022, Android dianggap sebagai sistem operasi yang paling banyak digunakan [1]. Android sendiri merupakan platform yang berbasis open-source yang mana gratis untuk di akses dan juga dikontrol oleh Original Equipment Manufacturer (OEM) seperti Samsung, Xiaomi, Oppo, Vivo, Huawei, Motorola, Google, dan masih banyak lagi. Pabrikan ini sering kali menjual ponselnya lebih murah, dengan harga jual rata-rata (ASP) $261 pada tahun 2021 dibandingkan Apple iOS, yang menjadikannya salah satu faktor mengapa Android sukses saat ini [1]. Namun dengan kesuksesan android saat ini menjadikannya target incaran bagi *cyber-crime* dalam melancarkan aksi malicious yang salah satunya melalui serangan *malware.*

Menurut penelitian dari university of cambridge menemukan bahwa 87 persen dari semua smartphone android terkena setidaknya *satu critical vulnerability* [2]. sementara Zimperium Labs menemukan bahwa 95 persen perangkat Android dapat diretas dengan *pesan teks sederhana* [3]. Apple juga tidak kebal. Pada tahun 2015, 40 aplikasi ditarik dari app store resmi karena terinfeksi *XcodeGhost*, suatu bentuk malware yang dirancang untuk mengubah perangkat Apple menjadi botnet berskala besar. Terlepas dari perlindungan Apple yang dibanggakan, malware tidak hanya menyelinap masuk tetapi juga berlapis di atas aplikasi yang terlihat sah atau normal, ini yang membuatnya sulit untuk terdeteksi [4]. Malware itu sendiri terdapat beberapa jenis seperti ransomware, backdoor, adware, file infector, spyware, PUA, riskware, trojan, scareware, trojan-sms, trojan-banker, trojan-spy, dan trojan-dropper [5]. Malware yang menyerang perangkat Android juga memiliki potensi dalam merugikan situasi finansial korban serta mendapatkan akses secara tidak sah untuk mencuri informasi pribadi korban. Sebagaimana jumlah serangan malware Android yang terus berlanjut meningkat, pentingnya memiliki metode

deteksi yang andal. Pada bidang IT saat ini *machine learning* telah menunjukkan hasil yang cukup efisien dalam mendeteksi malware Android. Karena machine learning dapat mengenali pola data yang kompleks dan belajar dari kumpulan data besar, algoritma machine learning termasuk ideal dalam mendeteksi malware Android [6]. Namun Karena kian waktu malware mengalami terus peningkatan yang menjadikannya semakin sulit untuk dideteksi. Seperti penjelasan dari Aslan, dkk [9] Saat ini klasifikasi malware semakin sulit karena beberapa instance malware dapat menampilkan karakteristik multiple classes pada waktu yang bersamaan dan selain itu malware juga dapat berjalan dalam mode kernel yang mana lebih destruktif dan lebih sulit dideteksi daripada malware tradisional.

Terdapat beberapa penelitian terdahulu menggunakan machine learning dalam pengklasifikasian dan pengidentifikasian malware Android khususnya melalui pendekatan metode algoritma *LightGBM*, ada sebuah penelitian yang dilakukan oleh Taha, dkk [24] penelitian ini memperkenalkan pendekatan Hybrid baru untuk klasifikasi malware Android dengan mengintegrasikan pengelompokan Fuzzy dan menggunakan LightGBM untuk klasifikasi yang tepat, metode ini mencapai akurasi yang bagus yaitu sebesar *94,63%*, Area di bawah kurva (AUC) sebesar *98,74%*, dan presisi sebesar *97,70%* dalam membedakan antara aplikasi android yang benign dan yang berbahaya (Malware). Kemudian penelitian yang dilakukan oleh Wang, dkk [25] Penelitian ini mengusulkan model deteksi malware android dengan menggunakan pendekatan LightGBM, yang menggabungkan teknik seleksi fitur inovatif (Chi2 dan Extra Trees). Dengan dataset yang terdiri dari 2000 sampel aplikasi malware dan sampel aplikasi benign, model ini menunjukkan akurasi yang mengesankan sekitar *96,4%* dan waktu train yang signifikan lebih singkat dibandingkan dengan model-model yang ada. Selanjutnya penelitian yang dilakukan oleh Sarah, dkk [26] pada penelitian ini dilakukan prediksi android malware menggunakan beberapa ensemble machine learning algorima dan diantaranya menggunakan algoritma LightGBM. Tujuan penelitian sarah, dkk sendiri melakukan analisa untuk mencari model prediksi yang terbaik dalam mendeteksi android malware dengan diterapkannya Recursive Feature Elimination (RFE) feature selections untuk mengurangi jumlah feature dalam men-optimisasi waktu dan resource. Dan hasil penelitiannya didapatkan bahwa LightGBM berhasil

memperoleh akurasi terbaik sebesar *99,4%* dengan optimal feature yang digunakan adalah berjumlah 100.

Dari penelitian-penelitian di atas dapat disimpulkan bahwa pendekatan metode LightGBM pada machine learning diharapkan dapat menjadi solusi yang andal dalam mendeteksi malware android dengan tingkat akurasi yang tinggi. Banyak boosting tools menggunakan algoritma berbasis pra-sortir[19, 20] (misalnya, algoritma bawaan pada XGBoost) untuk decisions tree learning, yang mana ini adalah solusi yang sederhana, namun tetapi tidak mudah untuk dioptimalkan. Sedangkan *LightGBM* menggunakan algoritma berbasis *Histogram*[21, 22, 23], yang mengelompokkan nilai fitur (atribut) kontinu ke dalam bin diskrit, yang mana hal ini dapat mempercepat training dan mengurangi penggunaan memori. Berbeda dengan algoritma Tradisional lainnya, LightGBM memberikan *Better Accuracy* dengan cara menghasilkan tree yang lebih kompleks melalui pendekatan *Leaf-wise (Best-first) Split* dibanding Level-wise Split seperti algoritma lainnya. Yang mana dengan pendekatan ini LightGBM memprioritaskan pada daun (leaf) yang memiliki *greatest error reduction* saat membuat keputusan, sehingga didapatkan model yang lebih *akurat* [29]. Selain itu perlunya penunjang lain yang dapat meningkatkan akurasi seperti diterapkannya feature selections, disini penulis mengusulkan pendekatan *Recursive Feature Elimination (RFE)* sebagai Feature Selection nantinya. Seperti yang dijelaskan pada penelitian sarah, dkk [26] sebelumnya, sebelum diterapkan RFE, jumlah feature pada dataset peneliti tersebut berjumlah 215 feature, setelah dilakukannya RFE maka didapatkan 100 optimal feature yang digunakan pada model LightGBM, dan dengan itu LightGBM berhasil memperoleh hasil akurasi tertinggi dari pada beberapa model lainnya yang penelitian Sarah gunakan, dengan akurasi LightGBM sebesar *99,4%*, Sarah, dkk juga menerangkan dengan menggunakan metode feature selection (RFE) untuk mengurangi jumlah feature, dapat men-optimisasi waktu dan resource [26].

Maka dari itu, penulis ingin mengusulkan pada penelitian kali ini menggunakan metode *LightGBM* sebagai pendekatan algoritma dalam mengklasifikasi malware android dan juga dengan diterapkannya RFE pada Feature Selection.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil *akurasi* yang diperoleh pada train model deteksi malware android menggunakan algoritma *LightGBM* ?
2. Apakah terdapat perbedaan signifikan dalam *akurasi* antara model sebelum dan setelah *Feature Selection* dilakukan ?

1.3 Tujuan Penelitian

Tujuan penelitian ini untuk mengetahui hasil nilai akurasi yang diperoleh oleh hasil train model pada deteksi malware android dengan menggunakan algoritma LightGBM, kemudian mengetahui perbedaan signifikan nilai akurasi yang diperoleh setelah diterapkan Feature Selection dan sebelum diterapkannya.

1.4 Batasan

Batasan pada penelitian kali ini hanya terbatas pada mengklasifikasikan malware android dengan menggunakan metode LightGBM. Kemudian pada dataset, menggunakan dataset yang berjumlah 29.332 permissions yang diambil dari Androzoo (untuk aplikasi benign) pada *androzoo.uni.lu/lists (diakses pada 29-Juli-2023)* dan Argus Lab's Android Malware Database (untuk aplikasi berbahaya) pada *impactcybertrust.org/dataset_view?idDataset=1275 (diakses pada 29-Juli-2023)*. Dan setelah dilakukannya train model nantinya, dari hasil akurasi model yang diperoleh akan dilakukan perbandingan dengan hasil akurasi model-model pada penelitian terdahulu.