BAB I
PENDAHULUAN

1.1 Latar Belakang

Malware telah menjadi ancaman besar bagi pengguna teknologi saat ini. *Malware* atau *Malicious Software* adalah sebuah perangkat lunak bersifat intrusif yang dikembangkan oleh peretas dengan tujuan utama menginfeksi, menjelajah, mencuri, atau merusak suatu perangkat yang ditargetkan demi kepentingan peretas. Berbagai jenis perangkat dapat diinfeksi oleh malware, salah satunya adalah *smarphone*, di mana kasus *malware* terbanyak didominasi pada sistem operasi *Android*. Hal ini dikarenakan pengguna *smartphone* dengan sistem operasi *Android* lebih banyak dibandingkan sistem operasi lainnya yakni sebesar 72,11% berdasarkan data dari *StatCounter GlobalStats*.

Kasus *malware* berdasarkan laman AV Test, dilaporkan bahwa terdapat 1139,24 juta *malware* yang tersebar di berbagai perangkat pada tahun 2020 dibandingkan tahun 2019 sebesar 1001,52 juta . Kemudian, kasus *malware* berdasarkan McAfee tentang *mobile threat report*, dilaporkan bahwa terdapat peningkatan signifikan *malware* bagi perangkat *mobile* di tahun 2020 pada Q4 menjadi 43 juta malware dari yang sebelumnya sebanyak 40 juta *malware* pada Q3.

Metode pendeteksi *malware* telah dikembangkan sebagai antisipasi dalam menghadapi perkembangan *malware*. Penelitian ini berfokus pada klasifikasi malware berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* dengan *K-Means Clustering* sebagai metode pemilihan *centroid*-nya.

API *call* merupakan media interaksi bagi suatu program, di mana API (*Application Programming Interfaces*) adalah suatu cara dari sebuah program untuk berinteraksi dengan program lainnya. Sedangkan, Android *permissions* adalah izin yang dimiliki oleh aplikasi *smartphone* untuk memperoleh kontrol dan akses perangkat seperti akses kamera, mikrofon, pesan pribadi, percakapan, foto, dan sebagainya.

Radial Basis Function Network adalah bagian dari jaringan saraf tiruan dengan penggunaan fungsi berbasis radial sebagai fungsi aktivasinya. Struktur *Radial Basis Function Network* terdiri dari *input layer*, satu *hidden layer*, dan *output layer*. Secara konvensional, fungsi *Gaussian* digunakan sebagai implementasi fungsi aktivasinya dengan perhitungan *euclidean distance*. *Euclidean distance* di sini berfungsi sebagai pengukuran jarak antara titik data dengan *centroid*, di mana pada *Radial Basis Function Network centroid*-nya dapat dipilih secara acak atau dengan metode *clustering*. Salah satu metode *clustering* yang banyak digunakan adalah *K-Means Clustering*.

K-Means Clustering adalah algoritma pembelajaran *unsupervised* di mana cara kerja algoritma ini dengan pembagian set data menjadi K *cluster* berbeda. Tiap set data dengan karakteristik atau properti serupa dikelompokkan ke dalam satu *cluster*. *K-Means* menggunakan perhitungan *euclidean distance* sebagai pengukuran jarak antara titik data dengan *centroid*-nya. Data dengan nilai jarak terkecil dari *centroid cluster*-nya dijadikan sebagai anggota *cluster* tersebut. Penggunaan *clustering* sebagai metode pemilihan *centroid Radial Basis Function Network* dinilai baik bila dibandingkan dengan pemilihan secara acak.

Evaluasi model *Radial Basis Function Network* akan dinilai dengan *K-Fold Cross-Validation* dan *Confusion Matrix*. *K-Fold Cross-Validation* adalah metode pengambilan sampel ulang data di mana set pembelajaran dibagi menjadi K subset terpisah dengan *fold* sebagai representasi jumlah *subset* yang dihasilkan. *K-Fold Cross-Validation* biasa digunakan pada pembelajaran mesin terapan sebagai metode prediksi kemampuan model *machine learning*. Hasil yang dikeluarkan tidak banyak menghasilkan bias sehingga mudah dipahami.

Confusion Matrix adalah teknik ringkasan atau simpulan kinerja suatu model pembelajaran, terdiri dari *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). *Confusion matrix* digunakan dalam pengukuran seperti akurasi, *precision*, *recall*, dan F1 *score*. Akurasi adalah metrik dasar yang digunakan untuk evaluasi model atau penggambaran jumlah prediksi benar dari semua prediksi pada suatu model. *Precision* adalah pengukuran seberapa banyak prediksi positif yang dibuat benar. *Recall* adalah pengukuran

seberapa banyak kasus positif yang diprediksi benar dari semua kasus positif. Sedangkan, F1 *score* adalah gabungan dari *precision* dan *recall*.

Penelitian tentang metode *Radial Basis Function Network* dengan *centroid* yang dipilih secara acak pada klasifikasi *malware* berdasarkan fitur *permissions* menghasilkan akurasi sebesar 97,20%, (Abdulrahman, dkk, 2021). Adapun metode *Radial Basis Function Network* dengan *centroid* yang dipilih dengan *K-Means Clustering* menghasilkan akurasi sebesar 93,75% pada kasus prediksi penyakit ginjal kronik (Santosa, dkk, 2016)

1.2 Rumusan Masalah

Adapun rumusan masalah yang dapat diambil berdasarkan dari latar belakang penelitian adalah sebagai berikut :

1. Bagaimana klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* dengan *K-Means Clustering* sebagai metode pemilihan *centroid*-nya?
2. Bagaimana perbedaan performa antara klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* yang *centroid*-nya dipilih secara acak dengan yang dipilih menggunakan *K-Means Clustering*?

1.3 Ruang Lingkup Penelitian

Ruang lingkup penelitian ini terdiri dari beberapa cakupan sebagai berikut :

1. Penggunaan dataset *Malgenome-215-Dataset*, diunduh dari *figshare*
2. Penggunaan *Radial Basis Function Network* sebagai metode klasifikasi *malware*
3. Penggunaan *K-Means Clustering* sebagai pemilihan *centroid* pada *Radial Basis Function Network*
4. Penggunaan *K-Fold* dan *Confusion Matrix* sebagai penguji model

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui bagaimana klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* dengan *K-Means Clustering* sebagai metode pemilihan *centroid*-nya.
2. Mengetahui perbedaan performa antara klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* yang *centroid*-nya dipilih secara acak dengan yang dipilih menggunakan *K-Means Clustering*.

1.5 Manfaat Penelitian

Berikut adalah manfaat dari penelitian yang dilakukan :

1. Memberikan gambaran tentang bagaimana klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* dengan *K-Means Clustering* sebagai pemilihan *centroid*-nya.
2. Memberikan hasil perbedaan performa antara klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* menggunakan *Radial Basis Function Network* yang *centroid*-nya dipilih secara acak dengan yang dipilih menggunakan *K-Means Clustering*.

1.6 Luaran Yang Diharapkan

Pemberian gambaran terkait *Radial Basis Function Network* dengan *centroid* yang dipilih menggunakan *K-Means Clustering* sebagai metode klasifikasi *malware* berdasarkan fitur API *call* dan Android *permissions* yang diharapkan mampu dijadikan sebagai bahan pertimbangan dan evaluasi untuk penelitian di masa mendatang.

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini dijelaskan secara detail mengenai keseluruhan bab yang ada. Berikut adalah sistematikanya :

BAB I PENDAHULUAN

Penjelasan tentang Latar Belakang Permasalahan, Rumusan Masalah, Ruang Lingkup Penelitian, Tujuan Penelitian, Manfaat Penelitian, Luaran Yang Diharapkan, dan Sistematika Penulisan.

BAB II TINJAUAN PUSTAKA

Penjelasan tentang Landasan Teori yang relevan dengan topik penelitian serta Studi Literatur mengenai karya ilmiah, jurnal, dan penelitian terkait klasifikasi malware dan *Radial Basis Function Network*.

BAB III METODE PENELITIAN

Penjelasan tentang Kerangka Pikir atau urutan proses yang dilakukan dalam penelitian terkait metode, analisa, dan hasil akhir

BAB IV HASIL DAN PEMBAHASAN

Penjelasan tentang percobaan yang telah dilakukan terkait klasifikasi malware dengan *Radial Basis Function Network*, serta evaluasi hasilnya.

BAB V KESIMPULAN DAN SARAN

Penjelasan tentang kesimpulan dari hasil percobaan yang telah dilakukan terkait klasifikasi malware dengan *Radial Basis Function Network*, serta saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA