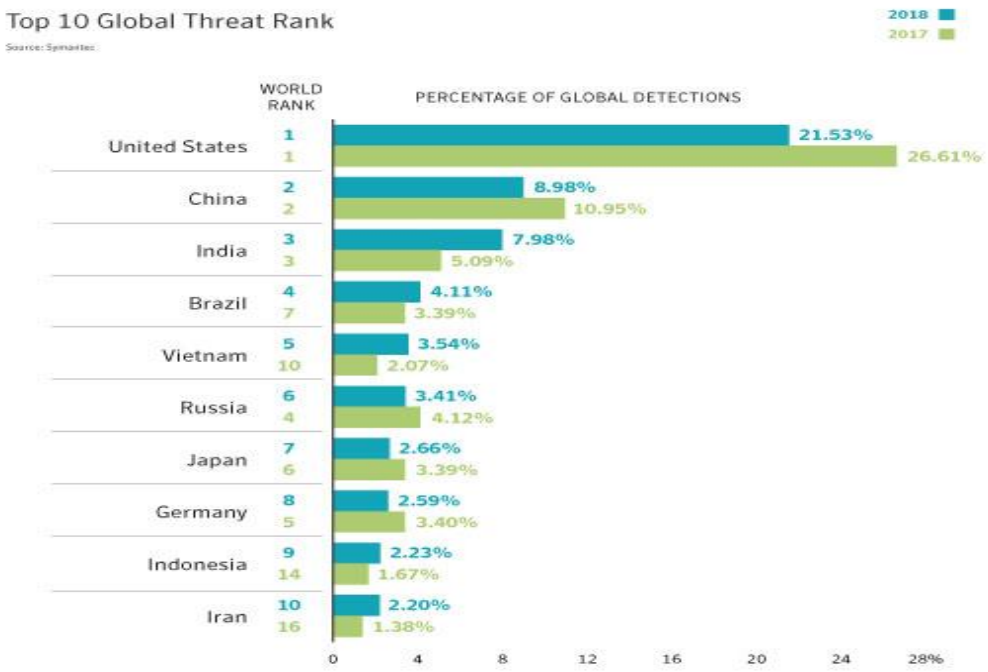
BAB I

PENDAHULUAN

1.1 Latar Belakang

Dengan semakin canggihnya infrastruktur, perangkat dan teknologi digital, kini semakin banyak masyarakat hidup berdampingan dengan dunia maya atau internet dengan beragam fungsi. Bak “pisau bermata dua”, internet punya fungsi dan peran yang beragam, pun demikian dengan ancaman di belakangnya yakni kejahatan siber salah satunya malware.



Gambar 1.1 Peringkat Ancaman Kejahatan Siber di Dunia (Top Ten)

Indonesia masuk peringkat ke-9 dari 157 negara yang terdeteksi mendapat ancaman kejahatan siber terbanyak selama tahun 2018. Ranking Indonesia ini naik dibandingkan tahun sebelumnya, yaitu urutan ke-14 dari 157 negara. Di atas peringkat Indonesia terdapat Jerman, Jepang, Rusia, Vietnam, Brasil, India, China, dan Amerika Serikat.

Dalam lingkup kawasan Asia Pasifik, Indonesia masuk peringkat ke-5 sebagai negara yang terdeteksi memperoleh ancaman kejahatan siber terbanyak pada 2018. Tahun sebelumnya, Indonesia menempati peringkat ke-6. Hal tersebut terangkum dalam laporan Symantec Internet Security Threat Report Volume 24, terbit Februari 2019. Laporan dirilis berdasarkan hasil riset Symantec terhadap 123 juta kasus serangan dan 142 juta ancaman kejahatan siber setiap hari di 157 negara.

Serangan atau ancaman kejahatan siber diteliti Symantec hanya fokus pada perangkat lunak perusak (malware), *spam, phising, hosts, bots,* serangan jaringan, serangan laman, aplikasi penyandera data (ransomware), dan cryptominers. Director of System Engineering Symantec ASEAN Halim Santoso mengatakan, kenaikan ranking Indonesia ada kaitannya dengan semakin banyaknya jumlah pengguna internet aktif. Ditambah lagi, mereka yang aktif tersebut juga semakin terbiasa bertransaksi apa pun melalui platform toko daring. Pelaku industri juga kian marak mengadopsi perangkat digital untuk meningkatkan produksi.

Di Indonesia, ancaman kejahatan siber berupa malware paling banyak menyasar sektor industri keuangan, asuransi, dan properti, diikuti manufaktur, konstruksi, serta jasa. Terkait pencurian identitas atau phising, sektor industri yang terkena ancaman terbesar adalah manufaktur, lalu konstruksi, jasa, serta keuangan, asuransi, dan properti. Tren global tahun 2018 menunjukkan, satu dari 10 kelompok penyerang menggunakan malware untuk menghancurkan atau sekadar mengganggu operasional bisnis. Soal spam, sektor industri yang mengalami ancaman terbanyak adalah jasa, kemudian manufaktur, keuangan, asuransi, dan properti, serta konstruksi.

Laporan Symantec menyebutkan, tren secara global pada tahun 2018 menunjukkan, satu dari sepuluh kelompok penyerang menggunakan malware untuk menghancurkan atau sekadar mengganggu operasional bisnis. Symantec telah menerka akan terjadi tren seperti itu sejak 2017. Mengutip *Cyber Crime* Kepolisian Negara RI pada 2018, Halim menyebutkan, mayoritas korban dari 4.000 laporan ancaman atau serangan kejahatan siber di Indonesia berasal dari perusahaan.

Beberapa alasan mendorong penjahat siber lebih memilih perusahaan dibandingkan konsumen individual sebagai target.

Alasan pertama, ada kecenderungan perusahaan malas membuat sistem cadangan, untuk dokumen strategis sekalipun. Alasan kedua, pemilik perusahaan lebih mudah memutuskan membayar nilai berapa pun yang diminta penjahat, terutama untuk kasus malware.

Top 10 Threats	Trend since 2016
Infiltration of Malware via Removable Media and External Hardware	
Malware Infection via Internet and Intranet	
Human Error and Sabotage	
Compromising of Extranet and Cloud Components	
Social Engineering and Phishing	
(D)Dos Attacks	
Control Components Connected to the Internet	
Intrusion via Remote Access	
Technical Malfunctions and Force Majeure	
Compromising of Smartphones in the Production Environment	

Gambar 1.2 Ancaman dan serangan pada ICS (publikasi BSI - Federal Office for Information Security)

Berdasarkan riset Federal Office for Information Security yang dipublikasikan pada BSI tanggal 6 Juni 2019, ancaman cyber khususnya tentang Malware mengalami tren peningkatan sejak tahun 2016 dimana ancaman malware berada pada peringkat ke-1 yaitu Infiltrasi Malware melalui media yang dapat dilepas dan perangkat keras eksternal sedangkan infeksi Malware melalui Internet dan Intranet menduduki peringkat ke-2.

Oleh karena itu malware membutuhkan perhatian khusus, karena menurutnya sebagian besar sumber ancaman di ruang siber dinilai berasal dari malware. Berdasarkan riset Security Endpoint Threat Report 2019 bahwa kasus Malware di

Indonesia tertinggi di Kawasan Asia Pasifik. Indonesia tercatat memiliki tingkat kasus Malware 10,68 persen pada 2019.

Malware sendiri merupakan perangkat lunak yang secara khusus dirancang untuk melakukan aktifitas berbahaya yang bisa merusak perangkat lunak lain. Contoh malware seperti Virus, Trojan, Spyware dan Exploit yang dibuat khusus agar tersembunyi sehingga mereka bisa tetap berada di dalam sistem komputer pada periode waktu tertentu tanpa sepengetahuan pemilik sistem (Cahyanto dkk., 2018). Beberapa Malware diciptakan dengan tujuan memata-matai seseorang, melakukan aktifitas merugikan seperti pencurian data dan informasi pribadi, membobol keamanan program dan sistem operasi serta banyak lagi. Untuk membobol suatu perangkat lunak atau sistem operasi dilakukan dengan menggunakan script yang diselipkan secara tersembunyi oleh penyerang (Sandag dkk., 2018).

Dengan banyaknya aktifitas penyebaran malware yang terjadi melalui jaringan internet membuat banyak pengguna menjadi resah. Pada umumnya malware dapat menyerang karena adanya transmisi secara public (C.Lupu et al.,2015). Serangan malware dapat menimbulkan kerugian pada perangkat (E.Mwangi et al.,2017). Untuk meminimalisir kerugian dan masalah pada perangkat telah banyak teknik yang dapat mendeteksi malware (A.Souri et al.,2018). Pendeteksian terhadap serangan malware ini diperlukan khususnya di jaringan agar pengguna bisa mengetahui apakah data yang berasal dari internet aman dari penyisipan malware atau tidak (Akbi & Rosyadi, 2018).

Teknik deteksi malware dibagi menjadi beberapa proses yaitu: yang pertama disassemble file, yang kedua feature extraction, dan yang ketiga yaitu feature selection (Yuxin et al.,2017). Berdasarkan penelitian yang dilakukan oleh Kateryna C. (2017) menggunakan pendekatan pembelajaran mesin/machine learning untuk pendeteksian malware menunjukkan akurasi pada angka 94,6% untuk decision tree dan 55 % untuk naïve bayes. Penelitian lain oleh Setiawan (2017), Pendeteksian Malware pada Lingkungan Aplikasi Web dengan Kategorisasi Dokumen, untuk algoritma Decision Tree menghasilkan tingkat akurasi 97 %, dengan pengujian yang sama tingkat akurasi algoritma Naive Bayes 83%.

Selanjutnya beberapa peneliti juga telah menggunakan teknik diskritisasi variabel pada algoritma naïve bayes dan decision tree untuk klasifikasi malware, dimana discretization (pendiskritan) atribut merupakan teknik untuk merubah sebuah fungsi atau nilai kontinu kedalam bentuk diskrit. Teknik ini dilakukan sebagai penyesuaian terhadap kemungkinan kemunculan nilai kontinu dalam fitur dataset yang sangat kecil sehingga akan mempengaruhi proses klasifikasi, meskipun pediskritan dari sebuah nilai kontinu tentu akan menimbulkan kesalahan berupa hilangnya beberapa informasi. Dalam penelitian Anggraini(2019) yang menerapkan diskritisasi variabel pada algoritma naïve bayes untuk deteksi malware menunjukkan akurasi 78,16 % sedangkan jika pengujian tanpa diskritisasi menghasilkan angka 69,72%. Ini memperlihatkan penggunaan teknik diskritisasi mampu meningkatkan pendeteksian secara signifikan jika dibandingkan dengan proses klasifikasi tanpa menggunakan teknik diskritisasi. Dengan Teknik pendiskritisasian menjadikan probabilitas dari Algoritme Naïve Bayes bisa lebih dihandalkan dalam penentuan kelas.

Kemudian dalam penelitian Revindar (2011) yang juga menerapkan teknik distritisasi tetapi tanpa melakukan normalisasi nilai pada atribut di datasetnya, peneliti hanya menghilangkan beberapa atribut/fitur fitur (attribute reduction) pada datasetnya, dimana proses pengujiannya dilakukan dengan teknik diskritisasi menggunakan metode *Entropy-Based Discretization menggunakan Tool Weka* pada 2 (dua) algoritma yaitu Decision Tree J48 dan Naïve Bayes. Selanjutnya diperoleh hasil decision tree J48 dan naïve bayes dengan akurasi decision tree J48 81,45% sedangkan naïve bayes 85,07%, tetapi jika tanpa menggunakan diskritisasi diperoleh hasil akurasi sebaliknya algoritma decision tree menghasilkan akurasi 81,45 % meskipun akurasinya tetap, tetapi lebih baik jika dibandingkan dengan naïve bayes dengan akurasi hanya 80,09%.

Oleh karena itu, pada penelitian ini akan dibahas analisis perbandingan algoritma decision tree dan naïve bayes dengan teknik diskritisasi dengan metode binning dengan Tool Rapidmner pada dataset malware (Saravana, 2018), tetapi pada dataset ini akan dilakukan normalisasi nilai pada atributnya dengan metode

z-score untuk menentukan algoritma mana yang memiliki kinerja paling baik pada pendeteksian malware.

1.2 Ruang Lingkup

Ruang lingkup pembahasan dalam penelitian ini dibatasi pada perbandingan metode klasifikasi data mining dengan algoritma decision tree dan Naïve Bayes pada pendeteksian malware dengan diskritisasi variabel, kemudian mengevaluasi hasil perbandingan untuk mengetahui metode klasifikasi data mining mana yang paling akurat.

1.3 Rumusan Masalah

Berdasarkan permasalahan yang sudah dianalisa, maka masalah-masalah yang ada di dalam penelitian ini: Bagaimana perbandingan hasil akurasi metode klasifikasi data mining dengan algoritma decision tree dan Naïve Bayes pada pendeteksian malware dengan diskritisasi variabel.

1.4 Tujuan dan Manfaat Penelitian

Berdasarkan rumusan masalah yang dikemukakan di atas tujuan dari penelitian ini adalah sebagai berikut :

a. Mengimplementasikan metode klasifikasi Decision Tree dan Naïve Bayes untuk pendeteksian Malware dengan Diskritisasi Variabel.
b. Melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan parameter akurasi (cross validation, confusion matrix, dan area under the curve/AUC) antara metode Decision Tree dan Naïve Bayes pada data training yang sudah ditentukan.
c. Menambah khazanah pengetahuan tentang menerapkan metode Decision Tree dan Naïve Bayes

1.5 Sistematika Penulisan

Agar memperoleh gambaran jelas mengenai penelitian ini, maka dibuatlah suatu sistematika penulisan yang berisi gambaran dalam tiap bab penelitian ini, yaitu:

a. **BAB I PENDAHULUAN**

   Bab ini menjelaskan tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat serta sistematika penulisan.

b. **BAB II TINJAUAN PUSTAKA**

   Bab ini menjelaskan mengenai literature review yang berhubungan dengan masalah dalam penelitian ini.

c. **BAB III METODOLOGI PENELITIAN**

   Bab ini menjelaskan tahapan atau metode secara rinci yang akan dilakukan dalam penelitian ini

d. **BAB IV ANALISA DAN PEMBAHASAN**

   Bab ini menjelaskan pembahasan dari pengujian serta analisa yang didapat dari data hasil pengukuran

e. **BAB V KESIMPULAN DAN SARAN**

   Bab ini menjelaskan kesimpulan dan saran dari hasil yang diperoleh, serta merupakan jawaban yang diperoleh dari tujuan pada bab 1.