JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

Komparasi Teknik Feature Selection Dalam Klasifikasi Serangan IoT Menggunakan Algoritma Decision Tree

Dicky Setiawan*, Adhitya Nugraha, Ardytha Luthfiarta

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202013089@mhs.dinus.ac.id, [2]adhitya@dsn.dinus.ac.id , [3]ardytha.luthfiarta@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202013089@mhs.dinus.ac.id

Abstrak−Kehadiran Internet of Things (IoT) telah merevolusi bagaimana cara kita berinteraksi dengan dunia dalam kehidupan sehari-hari dengan memungkinkan berbagai perangkat terhubung ke internet dan mentransmisikan data. Namun, semakin luasnya penggunaan dari teknologi IoT ini juga membawa ancaman serius terhadap keamanan siber dan meningkatkan adanya serangan IoT. Kebutuhan akan model klasifikasi yang kuat menjadi semakin jelas untuk mengantisipasi masalah tersebut. Penelitian ini berfokus pada pengembangan model klasifikasi serangan IoT dengan mengkomparasikan teknik feature selection yang memanfaatkan data dari CIC IoT Dataset 2023. Penelitian ini menghadapi tantangan seperti ketidakseimbangan data dan kompleksitas penanganan berbagai fitur. Untuk mengatasi tantangan tersebut, penelitian ini menggunakan teknik random undersampling untuk menyeimbangkan data dan memanfaatkan berbagai metode feature selection, antara lain filter based, wrapper based, dan embedded based. Selain itu, penelitian ini juga mencoba menggunakan algoritma decision tree. Temuan mengungkapkan bahwa penerapan teknik wrapper based sebagai feature selection bersama dengan algoritma decision tree menghasilkan akurasi tertinggi sebesar 87,32% dalam mengklasifikasikan jenis serangan IoT. Hal ini menekankan bahwa penggunaan teknik dan algoritma yang masih jarang digunakan sekalipun dapat memberikan hasil akurasi yang cukup baik.

Kata Kunci: Internet of Things (IoT); Serangan IoT; Imbalance Data; Seleksi Fitur; Klasifikasi; Decision Tree

Abstract−Presence of Internet of Things (IoT) has revolutionized how we interact with the world on our daily life by enabling various devices to connect the internet and transmit data. However, the increasingly widespread use of IoT technology also brings serious threats to cyber security and increases the number of IoT attacks. The need for robust classification models is becoming increasingly clear to anticipate these problems. This research focuses on developing an IoT attack classification model by comparing feature selection techniques that utilize data from the CIC IoT Dataset 2023. This research faces challenges such as data imbalance and the complexity of handling various features. To overcome these challenges, this research uses random undersampling techniques to balance the data and utilizes various feature selection methods, including filter based, wrapper based, and embedded based. Apart from that, this research also tries to use a decision tree algorithm. The findings reveal that the application of wrapper based techniques as feature selection together with a decision tree algorithm produces the highest accuracy of 87.32% in classifying IoT attack types. This emphasizes that the use of techniques and algorithms that are still rarely used can provide fairly good accuracy results.

Keywords: Internet of Things (IoT); IoT Attack; Imbalance Data; Feature Selection; Classification; Decision Tree

1. PENDAHULUAN

Sejatinya Internet Of Things (IoT) telah memainkan peran penting dalam kehidupan masyarakat di dunia pada saat ini. IoT merupakan konsep tentang revelousi cara interaksi manusia dengan dunia sekitar [1]. IoT memungkinkan berbagai perangkat untuk terhubung ke internet dan mengirim data tanpa interaksi manusia. IoT sendiri bertujuan untuk membantu kehidupan sehari-hari kita dengan memungkinkan perangkat cerdas di sekitar kita untuk bisa mengerjakan pekerjaan atau tugas sehari-hari [2]. Pemanfaatan IoT juga memungkinkan untuk dikembangkan lebih mendalam seiring berkembangnya era revolusi industri 4.0, dimana mesin dan manusia bekerja sama dengan bantuan Internet of Things (IoT) [3]. IoT memungkinkan terjalinnya komunikasi mesin antar mesin, manusia antar mesin, dan manusia dengan komputer, hal tersebut dapat dilakukan dengan mengkomunikasikan informasi atau data yang melalui internet [4]. Perkembangan dan implementasi IoT memberikan dampak seperti kemudahan dan kenyamanan dalam kehidupan manusia [5]. Namun seiring dengan perkembangan IoT, selain menjadi potensi untuk menjadi sumber inovasi, juga menjadi ancaman serius yang tak terhindarkan terhadap keamanan pada sistem yang rentan. Dengan masifnya IoT saat ini, tentu juga tidak terlepas dari konsekuensi berupa ancaman dan serangan bercakupan luas terhadap perangkat IoT [6]. Situasi ini merupakan potensi risiko besar dalam hal keamanan, karena banyak titik masuk yang mungkin memiliki kerentanan keamanan [7]. Kerentanan keamanan inilah yang memberikan peluang bagi penyerang untuk melancarkan aksinya.

Guna mengantisipasi beberapa serangan yang ada pada IoT, maka penelitian ini mencoba untuk merancang model klasifikasi menggunakan pendekatan machine learning yang sebelumnya dilakukan feature selection terlebih dahulu. Klasifikasi merupakan proses pemilahan dan membedakan kelas data melalui proses pencarian fungsi dan model sehingga sebuah model dapat digunakan untuk meramal atau memperkirakan kelas objek pengamatan yang belum diketahui [8]. Machine learning sendiri merupakan adopsi dari pembelajaran berdasarkan data dengan tujuan menghasilkan prediksi masa mendatang yang diaplikasikan melalui algoritma matematika, machine learning juga ditujukan untuk memperoleh kecerdasan melalui tahapan bernama pelatihan (training) dan pengujian (testing) [9]. Metode machine learning dapat digunakan untuk mendeteksi adanya serangan pada jaringan IoT dengan menganalisa perilaku dari perangkat IoT [10]. Berdasarkan hal-hal tersebut, machine learning dapat dimanfaatkan sebagai solusi dan pendekatan yang efektif dalam bidang klasifikasi.

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

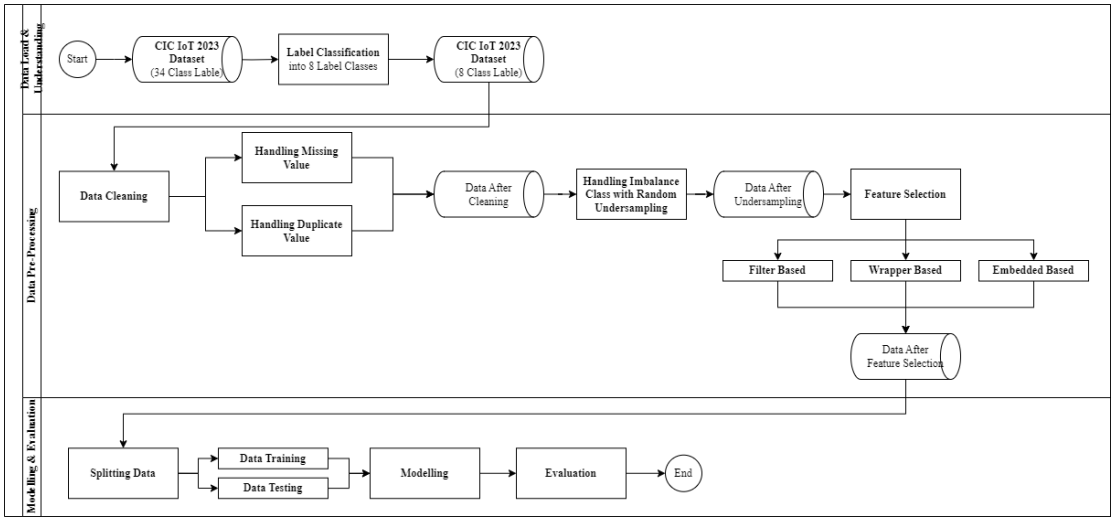
Dalam penerapan klasifikasi data, permasalahan yang kerap dijumpai adalah ketidakseimbangan jumlah data (imbalance data) antara kelas yang berbeda. Walaupun terdapat banyak metode klasifikasi yang efektif, sayangnya tidak semua metode sesuai ketika kondisi data tidak seimbang, hal ini dikarenakan metode-metode yang ada bergantung pada asumsi persebaran data pada kelas-kelas yang ada telah merata [8]. Kemudian juga data dengan dimensi dan fitur yang besar juga seringkali menjadi masalah dalam pengolahan data dimana komputasi menjadi lebih berat.

Sejumlah penelitian terkait klasifikasi serangan IoT dengan menggunakan algoritma machine leaning telah banyak diteliti dengan menggunakan beberapa algoritm. Adapun penelitian terdahulu dalam penelitian [11] dengan dataset yang sama (CIC IoT Dataset 2023) juga telah menggunakan beberapa algoritma dalam melakukan klasifikasi serangan IoT untuk delapan kelas tanpa penerapan teknik feature selection dengan hasil diantaranya logistic regression menghasilkan akurasi sekitar 83,31%, perceptron dengan akurasi sekitar 86,63%, adaboost dengan akurasi sekitar 35,13%, random forest dengan akurasi sekitar 99,43%, dan Deep Neural Network (DNN) menghasilkan akurasi sekitar 99,11%. Selanjutnya pada penelitian [3] juga melakukan klasifikasi serangan IoT dari dataset CICDDos 2017 dan CICDDoS 2019 untuk dua kelas yang telah dilakukan sampling dengan teknik Syntetic Minority Over-Sampling Technique (SMOTE) dimana dalam penelitian ini menggunakan perpaduan antara teknik Principal Component Analysis (PCA) sebagai feature selection (memilih 15 fitur saja) dengan algoritma random forest yang kemudian menghasilkan akurasi sebesar 99,94%. Kemudian pada penelitian [12], juga melakukan klasifikasi serangan IoT untuk empat kelas dan menghasilkan akurasi tertinggi yakni 98,3% dengan penggabungan teknik wrapper based (memilih enam fitur saja) dan algoritma K-Nearest Neighbor (KNN). Pada penelitian [13] dengan dataset CICIDS-2017 versi MachineLearningCSV menghasilkan akurasi sebesar 99,87% dari 15 fitur sebagai fitur-fitur yang optimal untuk deteksi anomali menggunakan algoritma random forest. Terdapat juga dalam penelitian [14] dengan menggunakan dataset BOT-IoT yang kemudian dipilih 10 fitur terbaik menghasilkan akurasi rata-rata sebesar 97% dengan menggunakan algoritma KNN.

Ditinjau dari beberapa permasalahan dan penelitian yang ada, maka pada penelitian ini akan melakukan eksperimen dengan melakukan penanganan imbalance data, dan mengkomparasikan beberapa teknik feature selection terhadap algoritma machine learning. Teknik feature selection yang digunakan seperti filter based, wrapper based, dan embedded based. Kemudian algoritma machine learning yang akan digunakan dalam penelitian ini adalah decision tree. Adapun data yang digunakan yaitu berkaitan dengan serangan IoT dengan dilakukan sampling menggunakan teknik random undersampling dan menggunakan delapan kelas. Kemudian, evaluasi model diuji berdasarkan classification report dan confusion matrix. Dari penelitian ini, bertujuan untuk mencari teknik feature selection yang tepat melalui penggabungan dengan algoritma decision tree untuk menghasilkan akurasi yang terbaik.

2. METODOLOGI PENELITIAN

Alur penelitian dari komparasi teknik feature selection dipadukan algoritma decision tree pada klasifikasi serangan IoT dapat dilihat pada gambar 1 terkait tahapan yang terlibat dalam alur penelitian ini



Gambar 1. Alur Penelitian

Fase pertama yaitu pengumpulan data penelitian yang relevan, dalam hal ini dataset CIC IoT Dataset 2023 dengan label kelas sebanyak 34 label, kemudian dilakukan klasifikasi lagi menjadi 8 label kelas. Dilanjut pada fase kedua yakni dengan melakukan data cleaning dengan deteksi nilai yang hilang (missing value) dan deteksi data duplikat (duplicate value). Setelah data dinyatakan bersih, maka fase ketiga dilakukan handling imbalance class dengan teknik random undersampling. Kemudian pada fase keempat dilakukan proses pemilihan fitur

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

(feature selection) menggunakan beberapa teknik diantaranya Filter Based, Wrapper Based, dan Embedded Based. Fase terakhir adalah modelling dan evaluation. Pada tahap modelling data akan terbagi menjadi training dan testing, bagian training digunakan dalam proses pelatihan model sedangkan testing digunakan dalam proses pengujian, setelah data terbagi maka dilakukan modelling dengan menggunakan beberapa algoritma machine learning, kemudian dilakukan evaluasi dan pengujian berdasarkan classification report dan confusion matrix secara berurutan.

2.1 Sumber Data

Pada penelitian ini dataset yang digunakan adalah dataset CIC IoT Dataset 2023 yang dihimpun oleh University of New Brunswick. CIC IoT Dataset 2023 berisikan beberapa jenis serangan. Dataset ini adalah bahan data utama dalam percobaan eksperimen terhadap teknik feature selection sebelum dilakukan permodelan dengan algoritma decision tree. Hal yang mendasari penggunaan dataset ini adalah karena dataset ini dinilai layak karena telah mencakup kriteria-kriteria seperti heterogenitas data, kompleksitas data, dan interaksi yang lengkap serta faktor-faktor pendukung lainnya.

2.2 Data Cleaning

Sebelum memasuki fase yang lebih mendalam, memastikan bahwa data telah bersih merupakan hal yang penting dalam pemrosesan data. Data cleaning merupakan tahapan pembersihan data dengan menghapus, data yang terdapat error, data tidak konsisten, dan data yang tidak lengkap dari koleksi data. Data yang sudah dibersihkan kemudian dapat digunakan untuk proses penggalian pengetahuan (discovery knowledge) [15]. Pada penelitian ini, untuk memastikan bahwa data yang ada telah bersih maka dilakukan penerapan handling missing value untuk deteksi nilai yang hilang dan handling duplicate value untuk mendeteksi dan menghapus data yang duplikat.

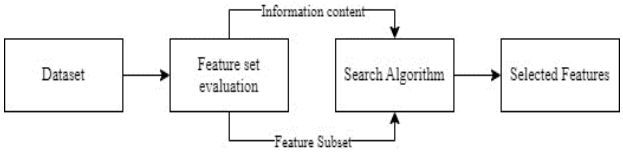
2.3 Handling Imbalance Class

Ketidakseimbangan data atau imbalance data adalah kondisi dimana distribusi data pada kelas data tidak seimbang sehingga mengakibatkan jumlah data pada suatu terlampau lebih banyak atau lebih sedikit daripada kelas data lainnya [16]. Penerapan algoritma klasifikasi secara langsung tanpa melakukan penanganan ketidakseimbangan kelas tersebut akan mengakibatkan penurunan performa algoritma serta menghasilkan hasil yang tidak maksimal [17]. Ketidakseimbangan data juga berpotensi menimbulkan bias atau kecenderungan nilai akurasi yang lebih tinggi pada kelas mayoritas [18]. Berkaitan dengan hal tersebut, maka perlu dilakukan teknik sampling agar data yang ada menjadi seimbang. Teknik yang digunakan dalam hal ini adalah random undersampling (RUS). RUS merupakan teknik sampling yang bertujuan menyamakan jumlah data, antara data mayoritas dan data minoritas dengan cara mengambil beberapa jumlah data dari data mayoritas [16]. Dengan demikian, seluruh label kelas akan disamakan jumlahnya dengan jumlah label kelas minoritas dengan memilih secara acak (random).

2.4 Feature Selection

Data dengan dimensi yang tinggi menyebabkan prevalensi data yang cenderung noise, tidak relevan, dan berlebihan [19]. Setelah dataset dinyatakan seimbang (balance), maka fase selanjutnya adalah dilakukan feature selection. Feature selection atau pemilihan fitur dilakukan untuk mengurangi dampak dimensi pada kumpulan data melalui pencarian subkumpulan fitur yang mendefinisikan data secara efisien [20]. Fase ini dilakukan untuk memilih fitur-fitur mana saja yang berkaitan erat dengan klasifikasi serangan IoT. Dalam penelitian ini, hanya dipilih 10 fitur saja untuk dapat dilakukan pemrosesan pada fase selanjutnya. Terdapat beberapa teknik feature selection yang nantinya akan digabungkan dengan algoritma decision tree guna membandingkan mana yang lebih efektif dengan akurasi terbaik.

Teknik yang akan dicoba pertama yaitu filter based, merupakan teknik feature selection yang selanjutnya juga digunakan dalam penelitian ini. Dengan teknik feature selection berbasis filter (filter based) berkontribusi terhadap pencapaian hasil fitur yang dihitung menggunakan metode statistik [12]. Teknik filter based akan memanfaatkan metode statistik dengan menilai atribut-atribut yang ada, filter based akan melakukan pengurutan atau rangking terhadap atribut berdasarkan cara kerja atribut tersebut. Filter based akan menghasilkan matriks yang berisi hasil identifikasi dari setiap atribut dimana atribut yang relevan akan disimpan dalam dataset dan atribut yang tidak relevan akan dibuang atau dihapus [12] (gambar 2).



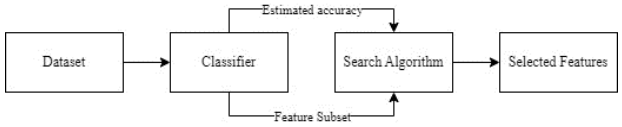
Gambar 2. Alur Filter Based

Pada penelitian ini, uji statistik yang diterapkan dalam teknik filter based ini adalah Analysis of Variance (ANOVA). ANOVA bekerja dengan menganalisa keterkaitan antara sampel data dengan kelas dan fitur, kemudian signifikansi suatu variabel dalam memprediksi nilai rata-rata fitur tersebut akan diukur atau dinilai [21].

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

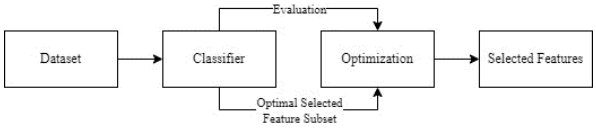
Teknik feature selection yang digunakan selanjutnya yaitu Wrapper Based. Dengan teknik ini, metode feature selection akan langsung digabungkan dengan menggunakan algoritma decision tree dalam pemrosesan. Teknik wrapper based bertujuan untuk mencari atribut yang paling ideal dengan mencoba seluruh fitur yang ada dalam klasifikasi algoritma [12]. Apabila kumpulan fitur ideal telah terpilih, maka proses selesai (gambar 3).



Gambar 3. Alur Wrapper Based

Metode ini mungkin lebih berhasil dibandingkan metode statistik, namun, tergantung pada algoritma klasifikasi yang digunakan [12]. Pada penelitian ini, teknik wrapper based yang ada menggunakan metode Recursive Feature Elimination (RFE) digabungkan dengan algoritma decision tree. Secara berulang (iteratif), RFE akan membangun model dan kemudian memilih yang terbaik berdasarkan koefisien [22].

Kemudian terdapat juga teknik embedded based dalam melakukan feature selection. Teknik ini bekerja sedemikian rupa sehingga fitur-fitur terbaik dipilih selama proses pembelajaran dengan menggunakan algoritma machine learning untuk menguji kualitas subset fitur yang dipilih [19]. Embedded based dikembangkan untuk menggabungkan keunggulan dari filter based dan wrapper based [12]. Teknik ini memilih fitur dengan mengidentifikasi fitur yang paling berkontribusi terhadap keakuratan model [12] (gambar 4).



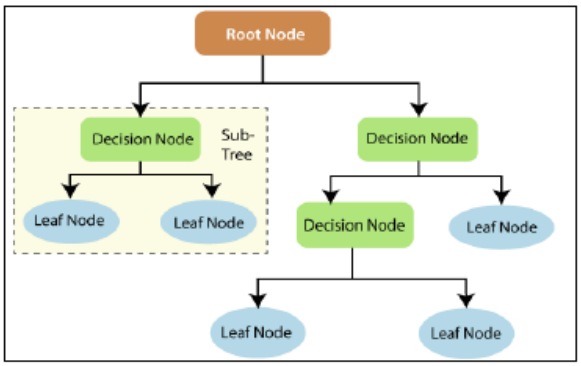
Gambar 4. Alur Embedded Based

Pada penelitian ini, teknik embedded based juga digabungkan algoritma decision tree untuk memilih fitur terbaiknya.

2.5 Modelling Dengan Algoritma Decision Tree

Sebelum melakukan modelling dari data yang ada, dilakukan splitting (dibagi) terlebih dahulu untuk data training (pelatihan) dengan porsi 80% dan testing (pengujian) dengan porsi 20%. Dari beberapa fase yang telah dilalui sebelumnya, fase selanjutnya adalah penerapan algoritma klasifikasi dalam melakukan modelling. Penggunaan algoritma klasifikasi didasarkan pada sejumlah penilaian seperti algoritma klasifikasi dinilai mampu bekerja dengan data informasi yang besar, kemudian digunakan dalam pembuatan asumsi kelas kategorikal yang selanjutnya diproses, untuk melakukan klasifikasi pengetahuan yang didasarkan pada pelatihan dan label kelas, dan melakukan klasifikasi terhadap data baru [23]. Pada penelitian ini berfokus pada penggunaan algoritma decision tree sebagai modelling dalam melakukan klasifikasi jenis serangan IoT.

Decision tree atau pohon keputusan merupakan salah satu metode dalam bidang pembelajaran mesin yang bekerja dengan melakukan pengujian perbandingan antara fitur numerik dengan nilai ambang batas, secara kohesif dan efisien serta dilakukan secara berturut-turut pada setiap pengujiannya [23]. Dari setiap pohon terdapat node dan cabang dimana setiap node merepresentasikan fitur dalam suatu kategori, node kemudian diklasifikasikan dan menghasilkan nilai yang dapat diambil oleh tiap subset [24]. Decision tree memiliki analisis yang sederhana dan memiliki ketepatan pada beragam jenis data sehingga decision tree menjadi salah satu metode yang umum diimplementasikan [25].



Gambar 5. Struktur Decision Tree [23]

Terdapat beberapa jenis algoritma decision tree seperti: Iterative Dichotomies 3 (ID3), penerus atau turunan dari ID3 yaitu algoritma C4.5, penerus atau turunan dari C4.5 yaitu algoritma Classification And Regression Tree

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

(CART), serta algoritma lain dalam kelompok yang sama seperti Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE) dan CHi-squared Automatic Interaction Detector (CHAID) [23]. Pada penelitian ini tipe decision tree yang akan digunakan adalah CART. Secara rekursif, CART mempartisi rekaman dalam dataset pelatihan menjadi beberapa subset rekaman dengan nilai yang sama untuk atribut target [26]. CART merupakan turunan dari algoritma lain bernama C4.5, sedangkan C4.5 sendiri merupakan turunan dari algoritma ID3 yang dibuat pada tahun 1980 [27]. CART lebih banyak digunakan dalam praktik karena fleksibilitasnya dalam menangani berbagai jenis masalah.

2.4 Evaluation

Fase terakhir dalam penelitian ini adalah evaluasi model. Evaluasi bertujuan untuk menilai dan memilih metode klasifikasi serta pembagian metode terbaik berdasarkan kinerja klasifikasi yang telah digunakan. Kinerja klasifikasi dapat diukur dengan menggunakan confussion matrix. Metode klasifikasi dapat dinilai berdasarkan confussion matrix dan menghasilkan nilai akurasi dalam mengenali objek pengamatan dari beberapa kelas berbeda [8].

Tabel 1. Confussion Matrix

Confussion Matrix		Predicted Class	
		P	**N**
Actual Class	P	TP (True Positive)	FN (False Negative)
	N	FP (False Positive)	TN (True Negative)

Akurasi (accuracy) digunakan untuk mengevaluasi model klasifikasi dengan menggambarkan proporsi prediksi yang benar dalam kumpulan data tertentu dan didasarkan pada persamaan (1) berikut [11].

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (1)$$

Kemudian terdapat juga nilai presisi atau precision. Precision merupakan rasio label yang teridentifikasi dengan benar terhadap jumlah total klasifikasi positif [11] dengan persamaan berikut.

$$Precision = \frac{TP}{(TP+FP)} \times 100\% \quad (2)$$

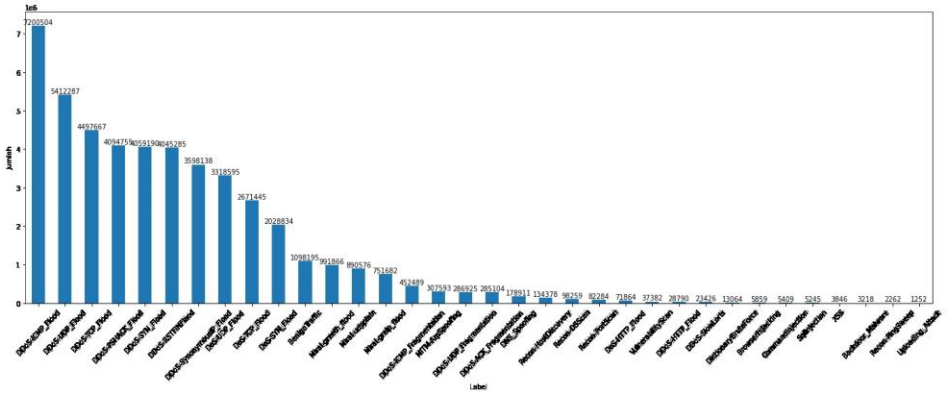
Recall merupakan rasio label yang diidentifikasi dengan benar terhadap jumlah total kemunculan label tersebut [11]. Recall juga merupakan salah satu bagian dari matriks evaluasi yang didasarkan pada persamaan (3) berikut.

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset

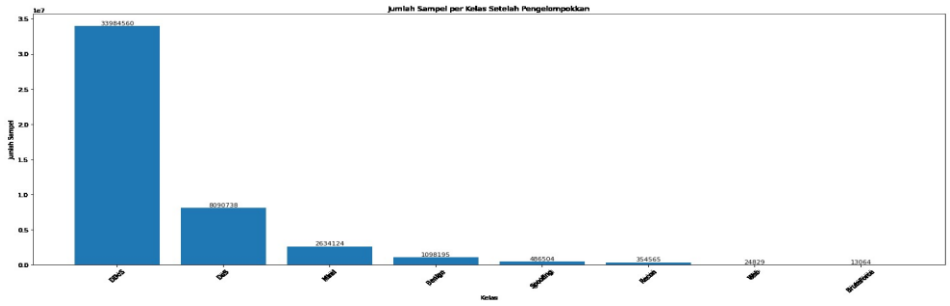
Penelitian ini menggunakan data dari dataset CIC IoT Dataset 2023 yang telah disediakan oleh University of New Brunswick, sebagai objek eksperimen untuk menganalisis serangan IoT. Dataset ini dipilih karena memiliki kelengkapan untuk menjawab kompleksitas data IoT, serta menyajikan interaksi data yang lengkap. Dengan 34 label kelas jenis serangan, seperti terlihat pada gambar 6, gambar 7, dan tabel 2, dataset ini memberikan fitur dan jenis yang bervariasi untuk digunakan dalam penelitian ini.



Gambar 6. Sebaran Dataset Berdasarkan 34 Label Kelas

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987



Gambar 7. Sebaran Dataset Berdasarkan 8 Label Kelas

Tabel 2. Jumlah Sebaran Dataset (8 Label Kelas)

Jenis	Jumlah
DDoS	33.984.560
DoS	8.090.738
Mirai	2.634.124
Benign	1.098.195
Spoofing	486.504
Recon	354.565
Web	24.829
BruteForce	13.064
Total	46.686.579

Dari data tersebut, terlihat bahwa sebaran antar label kelas sangat tidak seimbang. Sebagai contoh, kelas DDoS memiliki frekuensi yang jauh lebih besar dibandingkan dengan kelas lainnya. Ketidakseimbangan ini menimbulkan tantangan signifikan dalam permodelan klasifikasi, dan oleh karena itu, metode penanganan ketidakseimbangan kelas menjadi penting dan memerlukan pendekatan khusus dalam penanganannya. Maka dari itu, guna menangani permasalahan tersebut, penelitian ini menerapkan teknik random undersampling untuk menciptakan distribusi kelas yang lebih seimbang sebelum melanjutkan ke fase berikutnya.

3.2 Data Cleaning

Sebelum memasuki fase yang lebih mendalam, memastikan kualitas dataset merupakan hal yang sangat krusial untuk mendapatkan hasil yang akurat. Oleh karena itu, fase selanjutnya yang dilakukan adalah melakukan data cleaning, yang mencakup dua hal yaitu handling missing value dan handling duplicate value. Pertama, handling missing value menjadi langkah penting karena nilai yang hilang pada atribut dapat memengaruhi keakuratan analisis. Hasil dari penelitian ini menunjukkan bahwa dataset yang digunakan, tidak memiliki nilai kosong atau missing value, sehingga tidak perlu melakukan banyak hal atau pengisian nilai yang hilang atau kosong. Selain handling missing value, tahap handling duplicate value juga menjadi kunci dalam proses pembersihan data. Adanya duplikasi dapat memberikan hasil yang bias dan tidak akurat dalam analisis. Oleh karena itu, dilakukan identifikasi dan penghapusan nilai yang sama (duplikat). Pemilihan nilai yang akan dihapus didasarkan pada kesamaan di seluruh kolom atau hanya pada beberapa kolom tertentu. Penghapusan dilakukan dengan mempertimbangkan informasi yang relevan dan terbarukan, sehingga data yang tersisa mencerminkan kondisi sebenarnya dengan lebih baik. Hasil penelitian menunjukkan bahwa terdapat 47 data yang duplikat dan kemudian dihapus.

3.3 Handling Imbalance Class

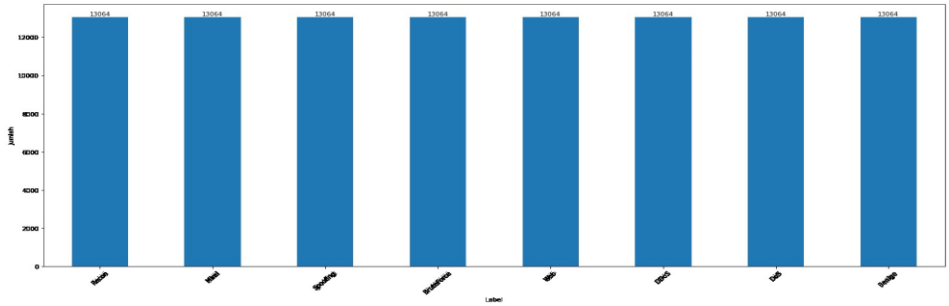
Pentingnya handling imbalance class atau penanganan ketidakseimbangan kelas dalam dataset juga menjadi aspek penting dalam mendapatkan hasil klasifikasi yang dapat diandalkan. Dalam penelitian ini, menghadapi ketidakseimbangan kelas pada dataset CIC IoT Dataset 2023, dimana beberapa kelas memiliki jumlah data yang jauh dibanding kelas lain. Tampak pada gambar 6 dan tabel 2 pada bagian sebelumnya bahwa terdapat ketidakseimbangan kelas (imbalance class) dimana DDoS merupakan kelas dengan jumlah data terbanyak sedangkan BruteForce sebaliknya. Untuk menangani permasalahan ini, maka pada penelitian ini menerapkan teknik random undersampling sebagai strategi utama dalam menyeimbangkan distribusi kelas.

Teknik random undersampling melibatkan pengurangan jumlah sampel dari kelas mayoritas (over-represented) untuk menyamakan dengan jumlah sampel pada kelas minoritas (under-represented). Dalam konteks klasifikasi serangan IoT, kelas serangan tertentu mungkin memiliki frekuensi yang signifikan, sementara kelas lain mungkin lebih langka. Hal tersebut dapat mengakibatkan model klasifikasi untuk memiliki kecendurungan memihak pada kelas mayoritas, sehingga identifikasi pada kelas minoritas yang sebenarnya penting menjadi kurang akurat, seperti serangan yang lebih jarang terjadi. Proses random undersampling ini dilakukan dengan cermat untuk mempertahankan proporsi antara kelas yang berbeda dan mengurangi potensi kehilangan informasi

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

penting. Oleh karena itu, setelah proses ini, dataset yang dihasilkan memiliki distribusi kelas yang lebih seimbang, menciptakan kondisi yang lebih baik untuk melatih model klasifikasi.



Gambar 8. Data Setelah Sampling (8 Label Kelas)

Penerapan teknik random undersampling pada penelitian ini menunjukkan hasil yang terlihat pada gambar 8 tersebut bahwa sebaran data telah berada dalam kondisi seimbang (balance) dimana setiap label kelas memiliki jumlah data yang sama 13.064. Sehingga dapat disimpulkan bahwa handling imbalance class dengan teknik random undersampling memberikan kontribusi positif dengan menciptakan data yang lebih seimbang dan dapat diandalkan.

3.4 Feature Selection

Tahapan selanjutnya dalam penelitian ini adalah feature selection yang juga merupakan tahapan yang tidak kalah dalam meningkatkan efisiensi dan kualitas model klasifikasi serangan IoT nantinya. Sebelumnya, CIC IoT Dataset 2023 memiliki fitur sebanyak 47 fitur (termasuk 1 fitur label) yang termuat dalam tabel 3 berikut ini dimana masing-masing fitur tersebut tentunya memberikan kontribusi terhadap proses klasifikasi. Namun, untuk mengurangi kompleksitas, maka perlu adanya penerapan teknik feature selection untuk memilih subset fitur yang dianggap paling informatif dan relevan.

Tabel 3. Fitur Pada Dataset

Nama Fitur		
flow_duration	fin_count	LLC
Duration	HTTP	Max
Protocol type	rst_count	Min
Header_Length	urg_count	Tot sum
Rate	HTTPS	AVG
Drate	Telnet	Tot size
Srate	DNS	Std
fin_flag_number	SMTP	IAT
syn_flag_number	SSH	Number
ack_flag_number	UDP	Covariance
rst_flag_number	IRC	Magnitue
psh_flag_number	TCP	Radius
ece_flag_number	DHCP	Variance
cwr_flag_number	ARP	Weight
syn_count	IPv	
ack_count	ICMP	

Pertimbangan utama dalam memilih fitur adalah relevansi dan signifikansi terhadap tujuan klasifikasi serangan IoT. Oleh karena itu, berbagai metode feature selection, seperti filter based, wrapper based, dan embedded based diterapkan dalam penelitian ini untuk mengevaluasi kinerja dan keefektifannya.

3.4.1 Filter Based

Teknik feature selection pertama yang dicoba dalam penelitian ini adalah filter based. Pada penelitian ini, teknik filter based memanfaatkan metode statistika, seperti Analysis of Variance (ANOVA), untuk menilai signifikansi setiap atribut terhadap label kelas. Melalui uji statistik ANOVA, penelitian ini berhasil mengidentifikasi 10 fitur terbaik yang memiliki pengaruh signifikan terhadap proses klasifikasi serangan IoT. Tabel 4 berikut ini merupakan hasil dari penerapan teknik filter based.

Tabel 4. Fitur Terpilih dari Teknik Filter Based

Fitur Terplih	
Protocol Type	Min

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

Fitur Terplih	
ack_flag_number	Std
rst_count	Magnitue
HTTPS	Radius
TCP	Variance

Keberhasilan teknik filter based dalam memilih fitur-fitur tersebut didasarkan pada signifikansinya dalam membedakan antar kelas serangan IoT. Seleksi fitur ini memiliki pengaruh yang positif yang tidak hanya memberikan efisiensi komputasional, tetapi juga menghasilkan fitur-fitur yang efektif dan akurat dalam mengklasifikasikan serangan IoT berdasarkan karakteristik yang relevan.

3.4.2 Wrapper Based

Dalam pengembangan model klasifikasi serangan IoT, teknik wrapper based juga menjadi fokus utama untuk memilih fitur-fitur yang memberikan kontribusi maksimal terhadap performa model. Wrapper based mencoba seluruh kombinasi fitur dalam suatu algoritma klasifikasi untuk menemukan subset fitur yang paling ideal. Dalam penelitian ini, metode Recursive Feature Elimination (RFE) digunakan dalam kombinasi dengan algoritma decision tree. Teknik wrapper based mengadopsi pendekatan iteratif, di mana setiap iterasi melibatkan penghapusan fitur terlemah dari model dan evaluasi ulang performa model. Pada akhirnya, hasil evaluasi ini membentuk subset fitur optimal yang memberikan kontribusi maksimal terhadap akurasi klasifikasi. Hasil dari penerapan teknik wrapper based pada penelitian ini ditampilkan dalam tabel 5 berikut yang memuat 10 fitur terpilih setelah proses seleksi.

Tabel 5. Fitur Terpilih dari Teknik Wrapper Based

Fitur Terpilih	
flow_duration	urg_count
Header_Length	rst_count
Protocol Type	Min
Rate	IAT
fin_count	Magnitue

Penerapan teknik wrapper based dalam penelitian ini tidak hanya memberikan informasi mengenai fitur mana saja yang relevan, tetapi juga meningkatkan interpretasi model. Dengan demikian, teknik wrapper based memainkan peran penting dalam mengoptimalkan konfigurasi fitur dan meningkatkan performa model klasifikasi serangan IoT secara keseluruhan.

3.4.3 Embedded Based

Teknik feature selection yang terakhir digunakan adalah embedded based. Teknik ini bekerja dengan mengidentifikasi fitur-fitur yang paling berkontribusi terhadap keakuratan model. Sama halnya dengan teknik wrapper based, teknik embedded based ini juga diimplementasikan dengan menggabungkannya bersama algoritma decision tree dalam upaya memilih fitur-fitur terbaik. Hasil dari penerapan teknik embedded based ini ditampilkan dalam tabel 6 berikut yang mencantumkan 10 fitur terpilih setelah proses seleksi.

Tabel 6. Fitur Terpilih dari Teknik Embedded Based

Fitur Terpilih	
IAT	flow_duration
Magnitue	fin_count
Protocol Type	rst_count
Header_Length	Srate
Min	syn_count

Teknik embedded based dalam penelitian ini memberikan keunggulan dalam mengidentifikasi fitur-fitur yang tidak hanya relevan tetapi juga memiliki dampak signifikan pada keakuratan klasifikasi serangan IoT. Dengan memanfaatkan kekuatan algoritma decision tree, teknik embedded based membantu meminimalkan overfitting dan memaksimalkan kemampuan model untuk mengenali pola serangan dengan tepat. Selain itu, penerapan teknik embedded based juga memudahkan pemahaman tentang kontribusi masing-masing fitur terhadap proses klasifikasi. Dengan demikian, teknik embedded based bersama algoritma decision tree juga berperan penting dalam mengoptimalkan fitur-fitur yang diperlukan.

3.5 Modelling Dengan Algoritma Decision Tree

Setelah melalui fase atau tahap sebelumnya sampai data yang ada dinyatakan bersih dengan pemilihan fitur-futur terbaik yang relevan, maka tahap selanjutnya adalah modelling. Pada penelitian ini menggunakan DecisionTreeClassifier sebagai algoritma untuk melatih model klasifikasi. Decision tree adalah model prediktif

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

dengan mengambil keputusan berdasarkan struktur pohon keputusan yang terbentuk dari fitur-fitur yang ada. Decision tree membagi dataset menjadi subgrup berdasarkan fitur-fitur yang paling signifikan untuk meningkatkan kemurnian. Proses ini terus berlanjut hingga mencapai titik di mana node atau daun yang dihasilkan memiliki kemurnian maksimal atau memenuhi kriteria penghentian yang ditentukan, seperti kedalaman maksimum atau jumlah sampel minimum di setiap node. Selain kemampuannya dalam memahami pola-pola data yang kompleks, decision tree juga memberikan informasi tentang kepentingan relatif setiap fitur dalam pengambilan keputusan. Ini membuatnya sangat sesuai untuk kasus seperti klasifikasi serangan IoT, di mana identifikasi fitur-fitur dapat mendukung upaya deteksi dan pencegahan yang efektif.

Pada tahap permodelan, dataset yang telah melalui proses data cleaning (handling missing value, handling duplicate value), dan balancing class dengan teknik random undersampling digunakan sebagai bahan latihan. Pada penelitian ini membagi data menjadi dua subset: data pelatihan (data train) sebesar 80% dan data pengujian (data test) sebesar 20%. Decision tree kemudian dilatih menggunakan data pelatihan untuk memahami pola-pola yang ada.

3.6 Evaluation (Hasil Komparasi)

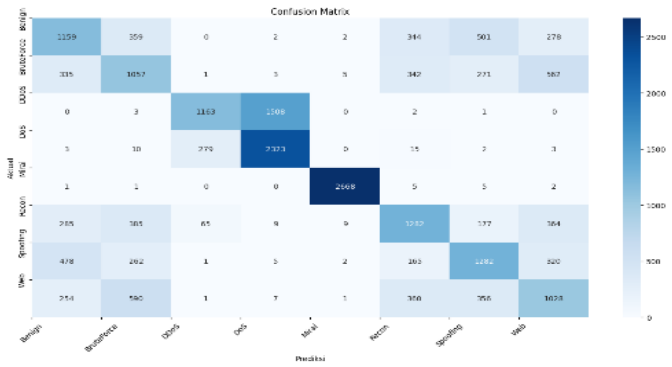
Setelah proses modelling selesai, model decision tree diuji menggunakan data pengujian yang belum pernah dilihat sebelumnya. Hasil evaluasi model mencakup metrik-metrik seperti accuracy, precision, dan recall. Pada penelitian ini juga menyajikan confusion matrix untuk memberikan pemahaman yang lebih mendalam tentang kinerja model.

Selama eksperimen, penelitian ini juga mengevaluasi berbagai skenario, termasuk penggunaan beberapa teknik feature selection dengan filter based, wrapper based, dan embedded based. Dari masing-masing penerapan teknik feature selection tersebut, kemudian menghasilkan 10 fitur yang dipilih berdasarkan penggunaan masing-masing tekniknya. Kemudian dari hasil tersebut dilakukan permodelan dengan algoritma decision tree yang kemudian menghasilkan matriks evaluasi sebagai berikut.

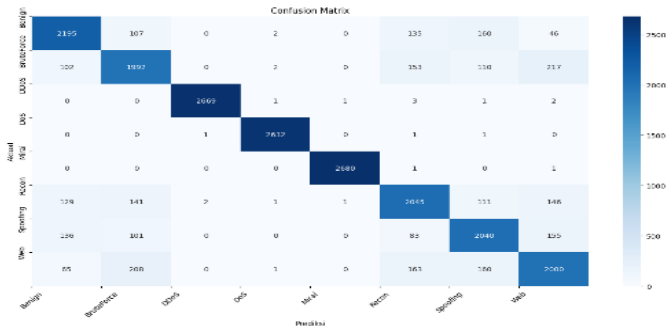
Tabel 7. Hasil Komparasi Teknik Feature Selection

Teknik	Accuracy	Precision	Recall
Filter Based	57,22%	58%	57%
Wrapper Based	87,32%	87%	87%
Embedded Based	87,10%	87%	87%

Hasil menunjukkan bahwa penggunaan teknik wrapper based dalam melakukan feature selection memberikan kontribusi terbesar terhadap peningkatan akurasi model. Fitur-fitur yang terpilih dari teknik ini membantu decision tree untuk lebih efektif dalam mengklasifikasikan serangan IoT. Kemudian untuk confussion matrix yang dihasilkan dari penerapan teknik feature selection dengan algoritma decision tree untuk masing-masing adalah sebagai berikut.



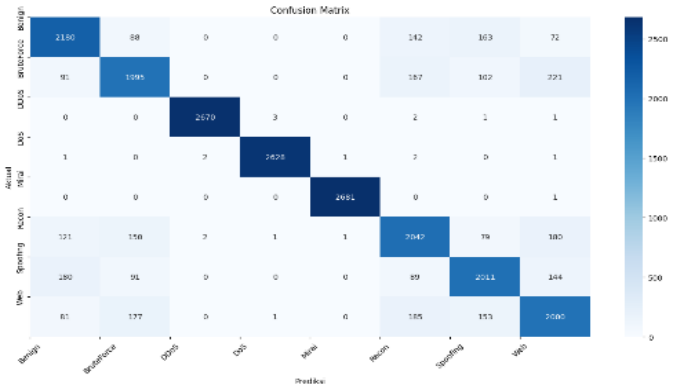
Gambar 9. Confussion Matrix Filter Based



Gambar 10. Confussion Matrix Wrapper Based

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987



Gambar 11. Confussion Matrix Embedded Based

Berdasarkan nilai akurasi yang ada pada tabel 7 dan confussion matrix di atas tampak bahwa penerapan teknik wrapper based dengan algoritma decision tree menghasilkan nilai akurasi yang mengungguli penggunaan teknik feature selection lainnya yakni sebesar 87,32%. Hal ini pun selaras dengan eksperimen pada penelitian [12] yang juga menunjukkan nilai akurasi terbaik dengan menggunakan teknik wrapper based. Maka secara keseluruhan, hasil dan pembahasan ini menunjukkan bahwa modelling dengan algoritma decision tree, terutama dengan penerapan teknik feature selection wrapper based, dapat meningkatkan kinerja model klasifikasi serangan IoT. Penggabungan dan penerapan teknik ini tentunya menawarkan kombinasi yang baik dalam hal kemampuan klasifikasi dan menjadikannya pilihan yang kuat dalam menghadapi kompleksitas dan variasi data serangan IoT.

4. KESIMPULAN

Dari hasil penelitian untuk menentukan teknik feature selection yang tepat dengan penggabungan algoritma decision tree pada dataset CIC IoT Dataset 2023, menghasilkan temuan yang dapat disimpulkan bahwa setelah melakukan handling imbalance class dengan teknik random undersampling pada total semula 46.686.579 data dengan 8 label kelas (awalnya 34 label kelas yang kemudian dikelompokkan menjadi 8 kelas), diperoleh dataset yang seimbang sebanyak 104.512 data dengan distribusi sekitar 13.064 data untuk setiap label kelas (8 label kelas). Dataset hasil sampling ini awalnya memiliki 47 fitur, termasuk 1 fitur label. Selanjutnya, diterapkan teknik feature selection untuk memilih 10 fitur terbaik yang memiliki pengaruh signifikan dalam permodelan klasifikasi dengan berbagai teknik yang digunakan seperti filter based, wrapper based, dan embedded based. Pengolahan data melibatkan pembagian (splitting data) menjadi data training sebesar 80% dan data testing sebesar 20%. Dari 10 fitur yang telah terpilih tersebut kemudian digunakan dalam permodelan klasifikasi dengan algoritma decision tree yang selanjutnya dievaluasi berdasarkan nilai classification report dan confussion matrix. Hasil penelitian menunjukkan bahwa penerapan teknik wrapper based dalam feature selection pada algoritma decision tree menghasilkan tingkat akurasi sebesar 87,32%. Nilai akurasi tersebut menjadi yang paling unggul dibandingkan penerapan teknik feature selection lainnya. Hasil ini menunjukkan bahwa teknik wrapper based efektif dalam meningkatkan performa model klasifikasi pada dataset IoT yang besar dan kompleks.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, tempat penulis menimba ilmu dan mendapatkan pengalaman melakukan penelitian bersama dosen-dosen di program studi Teknik Informatika. Kemudian penulis juga berterima kasih kepada IDSS Research Lab yang telah memfasilitasi dalam pengembangan penelitian ini. Yang terakhir, penulis juga berterima kasih kepada semua pihak yang telah memberikan kontribusi yang signifikan dalam penelitian ini.

REFERENCES

[1] E. Erwin dkk., PENGANTAR & PENERAPAN INTERNET OF THINGS: Konsep dasar & Penerapan IoT di berbagai Sektor. PT. Sonpedia Publishing Indonesia, 2023.
[2] H. Mantik, “Revolusi Industri 4.0: Internet of Things, Implementasi Pada Berbagai Sektor Berbasis Teknologi Informasi (Bagian 1),” JSI (Jurnal sistem Informasi) Universitas Suryadarma, hlm. 41–48, 2022, doi: https://doi.org/10.35968/jsi.v9i2.919.
[3] B. Ramadhan, D. Firdaus, dan A. R. Rafi, “Teknik SMOTE Sebagai Solusi Imbalance Class dalam Model Deteksi Intrusi DDoS dengan Metode PCA-Random Forest,” Journal MIND Journal, vol. 8, no. 1, hlm. 52–64, 2023, doi: 10.26760/mindjournal.v8i1.52-64.
[4] M. S. Gitakarma dan L. P. A. S. Tjahyanti, “Peranan Internet of Things dan Kecerdasan Buatan dalam Teknologi Saat Ini,” Jurnal Komputer dan Teknologi Sains (KOMTEKS), vol. 1, no. 1, hlm. 1–8, 2022.

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 1, Januari 2024, Page 83-93
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i1.6987

[5] W. Najib, T. Ancaman dan Solusi Keamanan, S. Sulistyo, dan K. Kunci, "Tinjauan Ancaman dan Solusi Keamanan pada Teknologi Internet of Things (Review on Security Threat and Solution of Internet of Things Technology)," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 9, no. 1, 2020.
[6] R. R. Krishna, A. Priyadarshini, A. V. Jha, B. Appasani, A. Srinivasulu, dan N. Bizon, "State-of-the-Art Review on IoT Threats and Attacks: Taxonomy, Challenges and Solutions," Sustainability (Switzerland), vol. 13, no. 16, Agu 2021, doi: 10.3390/su13169463.
[7] A. Sandriana dan F. Maulana, "Klasifikasi serangan Malware terhadap Lalu Lintas Jaringan Internet of Things menggunakan Algoritma K-Nearest Neighbour (K-NN)," E-JOINT (Electronica and Electrical Journal of Innovation Technology) , vol. 03, no. 1, 2022.
[8] P. R. Sihombing dan I. F. Yuliati, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 2, hlm. 417–426, Mei 2021, doi: 10.30812/matrik.v20i2.1174.
[9] A. Roihan, P. A. Sunarya, dan A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," IJCIT (Indonesian Journal on Computer and Information Technology), vol. 5, no. 1, hlm. 75–82, 2019.
[10] S. M. Tahsien, H. Karimipour, dan P. Spachos, "Machine learning based solutions for security of Internet of Things (IoT): A survey," Journal of Network and Computer Applications, vol. 161, hlm. 102630, 2020, doi: https://doi.org/10.1016/j.jnca.2020.102630.
[11] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, dan A. A. Ghorbani, "CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment," Sensors, vol. 23, no. 13, Jul 2023, doi: 10.3390/s23135941.
[12] H. Polat, O. Polat, dan A. Cetin, "Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models," Sustainability (Switzerland), vol. 12, no. 3, Feb 2020, doi: 10.3390/su12031035.
[13] K. Kurniabudi, A. Harris, dan E. Rosanda, "Optimalisasi Seleksi Fitur Untuk Deteksi Serangan Pada IoT Menggunakan Classifier Subset Evaluator," JURIKOM (Jurnal Riset Komputer), vol. 9, no. 4, hlm. 885, Agu 2022, doi: 10.30865/jurikom.v9i4.4618.
[14] J. Nicholas Sibarani, D. Ronaldo Sirait, dan dan Salma Safira Ramadhanti, "Intrusion Detection Systems pada Bot-IoT Dataset Menggunakan Algoritma Machine Learning," Jurnal Masyarakat Informatika, vol. 14, no. 1, hlm. 38-52, 2023.
[15] C. Zai, "IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA," JURNAL PORTAL DATA, vol. 2, no. 3, 2022.
[16] L. Qadrini, H. Hikmah, dan M. Megasari, "Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017," Journal of Computer System and Informatics (JoSYC), vol. 3, no. 4, hlm. 386–391, Sep 2022, doi: 10.47065/josyc.v3i4.2154.
[17] M. Sulistiyono, Y. Pristyanto, S. Adi, dan G. Gumelar, "Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi," SISTEMASI: Jurnal Sistem Informasi, vol. 10, no. 2, hlm. 445-459, 2021.
[18] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," Journal Information System Development (ISD), vol. 3, no. 1, hlm. 44-49, 2018.
[19] B. Venkatesh dan J. Anuradha, "A review of Feature Selection and its methods," Cybernetics and Information Technologies, vol. 19, no. 1, hlm. 3–26, 2019, doi: 10.2478/CAIT-2019-0001.
[20] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, dan J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," Journal of Applied Science and Technology Trends, vol. 1, no. 2, hlm. 56–70, Mei 2020, doi: 10.38094/jastt1224.
[21] A. Wenda, "Support Vector Machine untuk Pengenalan Bentuk Manusia Menggunakan Kumpulan Fitur yang Dioptimalkan," Jurnal Sains dan Teknologi, vol. 11, hlm. 77–84, 2022, doi: 10.23887/jst-undiksha.v11i1.
[22] W. Lian, G. Nie, B. Jia, D. Shi, Q. Fan, dan Y. Liang, "An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning," Mathematical Problems in Engineering, vol.2020, 2020, doi: 10.1155/2020/2835023.
[23] B. Charbuty dan A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, hlm. 20–28, Mar 2021, doi: 10.38094/jastt20165.
[24] H. Zhou, J. Zhang, Y. Zhou, X. Guo, dan Y. Ma, "A feature selection algorithm of decision tree based on feature weight," Expert Syst Appl, vol. 164, hlm. 113842, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113842.
[25] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," Jurnal SAINTEKOM, vol. 13, no. 1, hlm. 22–31, Mar 2023, doi: 10.33020/saintekom.v13i1.349.
[26] A. Saputra, "KAJIAN EMPIRIS METODE POHON KEPUTUSAN ROTATION FOREST," 2018.
[27] D. P. Sinambela, H. Naparin, M. Zulfadhilah, dan N. Hidayah, "Implementasi Algoritma Decision Tree dan Random Forest dalam Prediksi Perdarahan Pascasalin," Jurnal Informasi dan Teknologi, vol. 5, no. 3, hlm. 58–64, Sep 2023, doi: 10.60083/jidt.v5i3.393.

Submitted: **11/11/2023**; Accepted: **13/12/2023**; Published: **09/01/2024**