BAB II

TINJAUAN PUSTAKA

2.1 Definisi Data Mining

Data mining adalah suatu istilah yang digunakan menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Maulida 2018).

2.1.1 Tahapan Data Mining

Berikut ini adalah beberapa tahapan atau langkah kerja yang harus dilakukan dalam data mining (Thi Bi Dan, Widya Sihwi, and Anggrainingsih 2016), yaitu :

1. *Data Selection*

   Pemilihan (seleksi) data dari sekumpulan data operasional. Data hasil seleksi akan digunakan untuk proses data mining, dan disimpan dalam suatu berkas dan terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

   Proses *cleaning* dilakukan dengan menghilangkan *noise*, membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. *Transformation Data*

Tahap ini merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

4. *Data mining*

   Data mining merupakan proses mencari pola atau informasi yang sangat menarik dalam data terpilih.

5. *Interpretation/Evalution*

   Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

1.1 Machine Learning

Machine Learning memberikan dasar teknis untuk penambangan data. Ini digunakan untuk mengekstrak informasi dari data mentah ke dalam database. Umumnya dalam bentuk yang dapat dipahami dan dapat digunakan untuk berbagai keperluan. Dalam banyak aplikasi, pembelajaran mesin menghasilkan gambar bentuk dari sebuah contoh. Jenis deskripsi yang ditemukan dapat digunakan untuk prediksi, menjelaskan apa dan mengapa, dan pemahaman. Beberapa aplikasi penambangan data berfokus pada prediksi: memprediksi apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu, dengan menebak dari contoh (Widodo et al. 2022).

Machine Learning telah digunakan untuk beberapa standar tugas yang telah dipelajari secara luas, diantaranya klasifikasi, regresi, clustering, pengelompokan, dan *dimensionality reduction* atau *manifold learning*. Klasifikasi merukapan permasalahan penentuan kategori untuk suatu data. Misalnya klasifikasi dokumen

yang terdiri dari data politik, bisnis, olahraga, dan lain-lain. Klasifikasi dalam jumlah sub yang besar tentu akan menyulitkan pembagian klasifikasi tersebut, maka dibutuhkan *Machine Learning* untuk mempermudah pengklasifikasian tersebut (Rabbani, Wahidah, and Santoso 2021).

2.2 Random Forest

Algoritma *Random Forest* merupakan metode ensemble learning yang diusulkan pertama kali oleh Breiman pada tahun 2001 yang terdiri dari kombinasi dari pohon klasifikasi sehingga pada setiap pohon bergantung pada nilai acak vektor sampel secara mandiri dengan distribusi yang sama untuk semua pohon (Syukron and Subekti 2018). Langkah-langkah pada *Random Forest* terdiri dari 3 tahap. Tahap pertama yaitu menghasilkan sampel acak, setiap sub pohon keputusan harus memasukkan set pelatihan acak yang akan diambil sampelnya dari kumpulan data asli. Tahap kedua yaitu membangun sub pohon keputusan, *Random Forest* terdiri dari beberapa pohon keputusan yang dapat tumbuh sempurna tanpa pemangkasan karena batas atas teoritis dalam kesalahan generalisasi. Tahap terakhir yaitu mensintesis hasil klasifikasi, setiap sub pohon akan menghasilkan hasil yang berbeda-beda. Oleh karena itu mekanisme *voting* diperlukan untuk membuat keputusan akhir (Rabbani, Wahidah, and Santoso 2021).

2.3 Algoritma K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* merupakan algoritma sederhana yang mengklasifikasikan data berdasarkan ukuran kesamaan. Klasifikasi dilakukan berdasarakan data pembelajaran dengan objek yang terdekat dari data pembelajaran

tersebut (Azis, Azhar, and Syaifuddin 2020). Prinsip dari *K-Nearest Neighbor* adalah jika memiliki sekumpulan sampel data sebagai data training, kita berikan label untuk seluruh data tersebut maka kita akan mengetahui data tersebut masuk kedalam kelas mana. Jika diberikan data baru tanpa label, maka akan dibandingkan data tersebut dengan data yang sudah ada lalu akan melihat kesamaan dan mencari labelnya (Muhammad, Ermatita, and Falih 2021).

2.4 Decision Tree

Decision Tree adalah metode pengklasifikasian yang menggunakan fungsi-fungsi pendekatan diskrit. Terdapat banyak algoritma *Decision Tree* seperti ID3, C4.5, C5.0, dan CART. Algoritma *Decision Tree* menggunakan perhitungan *enthropy* kemudian *information gain* untuk membentuk pohon keputusan. Pemilihan atribut diproses menggunakan *information gain*. Berikut adalah formula perhitungan nilai *entropy* dan *information gain* (Syahputra, Akbi, and Risqiwati 2020) :

$I(S1, S2, S3, \dots, Sm) = -\sum pi\ log2(pi)$.....(Persamaan 1)

S = Himpunan, m = Banyaknya kelas, p_i = Proporsi kelas

Berikut adalah formula untuk memperoleh informasi nilai subset dari A dan seterusnya.

$E(A) = \sum S1j + S2j + \dots S + mj\ S\ I(S1j, S2j, \dots, Smj)$.....(Persamaan 2)

S_1j = Sampel kelas i, j = Atribut A , E = Entropy

Setelah semua perhitungan *Entropy* total dan subset selesai selanjutnya akan dilakukan perhitungan nilai *Information gain* dengan formula :

$GAIN\ (A)\ =\ I(S1, S2, \dots\ , S1m)\ -\ E(A)$.....(Persamaan 3)

2.5 Malware Mirai

Malware Mirai adalah *malware linux* yang ditulis dalam bahasa C, tujuan utama tetapi tidak eksklusif dari *malware* ini adalah menginfeksi router dan kamera IP dengan terus memindai port 22, 23, 5747, dan lain-lain. Menggunakan teknik *brute force* untuk menebak kata sandi melalui serangan kamus. Setelah terhubung ke IoT, upaya untuk masuk, mendapatkan akses, dan menginfeksi perangkat. Perangkat yang terinfeksi kemudian memindai jaringan lain untuk mencari lebih banyak perangkat IoT dan meluncurkan serangan DDoS (Jaramillo 2018).

Mirai adalah *botnet* yang dirancang khusus mempenetrasi IoT yang memiliki sistem keamanan yang lemah. IoT yang dijadikan target biasanya adalah router dan kamera CCTV yang perangkatnya terkoneksi dengan jaringan internet. Cara *malware* ini mencari celah pada IoT adalah dengan mengeksploitasi data personal dari perangkat IoT seperti *username* atau *password* bawaan dari sparepart yang rentan sistem keamanannya. Setelah mendapatkan data perangkat IoT, Mirai *Malware* akan membuat perangkat IoT tersebut mengirimkan paket data ke server target serangan.

2.6 K-Fold Cross Validation

Cross validation adalah teknik yang digunakan untuk menganalisis apakah suatu model memiliki generalisasi yang baik (dapat bekerja dengan baik pada

contoh yang tidak terlihat). Dalam *cross validation*, sampel asli dibagi menjadi beberapa subsampel dengan pemisahan K-fold. Hasil dari setiap iterasi berikutnya dihitung dengan menggunakan rata-rata Persamaan (Rhohim, Suryani, and Nugroho 2021).

2.7 Confusion Matrix

Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah, urutan pengujian ditabulasikan dalam confusion matrix dimana kelas yang diprediksi ditampilkan dibagian atas matriks dan kelas yang diamati disisi kiri. Sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi.

2.8 Parameter Matrix

Accuracy merupakan tingkat keterhubungan antara suatu nilai yang diprediksi dengan nilai aktual yang ada (Devita, Herwanto, and Wibawa 2018). Berikut merupakan rumus atau formula dari *accuracy* dijelaskan pada rumus 4.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Negative + False\ Positive + True\ Negative} \text{ .(Persamaan 4)}$$

Precision merupakan pengukuran tingkat ketepatan antara informasi yang diminta oleh pemohon dengan jawaban yang diberikan oleh sistem. Rumus atau formula dari *precision* dijelaskan pada rumus 4 (Yunus et al. 2019).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \text{.....(Persamaan 5)}$$

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi dalam suatu pemrosesan data. Berikut merupakan rumus atau formula dari *recall* dijelaskan pada rumus 6 (Zidane 2021).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{(Persamaan 6)}$$

2.9 ROC Curve

Kurva ROC dibagi dalam dua dimensi, dimana tingkat TP diplot pada sumbu Y dan tingkat FP diplot pada sumbu X. Grafis yang menentukan klasifikasi mana yang lebih baik, digunakan metode yang menghitung luas daerah dibawah kurva ROC yang disebut AUC *(Area Under the ROC Curve)* yang diartikan sebagai probabilitas (Ardiyansyah, Rahayuningsih, and Maulana 2018).

AUC mengukur kinerja diskriminatif dengan memperkirakan probabilitas output dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar AUC, semakin kuat klasifikasi yang digunakan. Nilai AUC adalah bagian dari daerah unit persegi, nilainya akan selalu antara 0,0 dan 1,0.

Tabel 2.1 Nilai AUC

Nilai AUC	**Klasifikasi**
0.90 - 1.00	Paling Baik
0.80 - 0.90	Baik
0.70 - 0.80	Adil atau Sama
0.60 – 0.70	Rendah
0.50 – 0.60	Gagal

2.10 Studi Literatur

Tabel 2.2 Studi Literatur

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
1	Ilham Ramadhan, Parman Sukarno, Muhammad Arief Nugroho (2019)	Analisis Perbandingan Algoritma K-Nearest Neighbor dan Decision Tree Dalam Mendeteksi Distributed Denial of Service	Metode IDS non-machine learning saat ini tidak terlalu akurat, sehingga diperlukan metode IDS dengan machine learning (ML) yang lebih akurat dalam mendeteksi serangan.	Algoritma K-Nearest Neighbour dan Decision Tree	Hasil dari penelitian ini yaitu DT memiliki akurasi lebih tinggi dengan nilai akurasi sebesar 99,91% daripada KNN yang hanya mempunyai nilai akurasi sebesar 98,94% dalam mendeteksi serangan DDoS.
2	Alejandro Guerra-Manzanares, Jorge Medina-Galindo, Hayretdin Bahsi dan Sven Nomm (2020)	*MedBIoT: Generation of an IoT Botnet Dataset in a Medium-sized IoT Network*	Mengklasifikasikan botnet pada infrastruktut IoT dengan 80 perangkat yang terkoneksi dengan network traffic.	Algoritma K-NN, DT, RF	Pada proses pengklasifikasian Binary mendapatkan akurasi 90,25% untuk K-NN, 93,15% untuk DT, dan 95,32% untuk RF. Selain itu pada pengklasifikasian Multi-Class mendapatkan hasil akurasi sebesar 87,06% untuk KNN, 95,16% untuk DT, dan 97,66% untuk RF.
3	Aditya Dwi Afifaturahman,	Perbandingan Algoritma K-Nearest Neighbour (KNN)	Banyak tantangan muncul karena serangan jahat terus berubah dan terjadi dalam	Algoritma K-Nearest Neighbour	Pengujian dengan *percentage split* 60%, 70% dan 80% menunjukkan bahwa algoritma

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
	Firmansyah Maulana (2021)	dan Naive Bayes pada IDS	volume yang sangat besar yang membutuhkan solusi yang dapat diskalakan.	(KNN) dan Naive Bayes	KNN mendapatkan nilai yang lebih tinggi dari Naive Bayes kecuali *error rate*
4	Riyan Eko Putri, Suparti, Rita Rahmawati (2014)	Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012	Status kerja penduduknya apakah menganggur atau tidak menganggur (bekerja) dimana ketika tidak diimbangi dengan lapangan kerja yang tersedia dapat menyebabkan tingkat pengangguran yang tinggi.	Algoritma K-Nearest Neighbour (KNN) dan Naive Bayes	Perhitungan Press's Q dapat dikatakan bahwa pengklasifikasian status kerja di Kabupaten Demak tahun 2012 dengan metode Naïve Bayes dan metode K-Nearest Neighbor sudah baik atau sudah akurat.
5	Nur Widiyasono, Ida Ayu Dwi Giriantari, Made Sudarma, L Linawati (2021)	*Detection of Mirai Malware Attacks in IoT Environment Using Random Forest Algorithms*	Investigasi Forensik Jaringan membutuhkan Algoritma Random Forest yang digunakan untuk melakukan teknik klasifikasi dan deteksi serangan Malware Mirai.	Algoritma RF	Hasil percobaan menunjukkan bahwa algoritma RF mencapai performa optimal dengan rata-rata nilai akurasi 95,01%, recall 90,82%, F1 Score 93,85% dan nilai presisi terbaik 99,23%.
6	Ardiyansyah, Panny Agustia Rahayuningsih, Reza Maulana (2018)	Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner	Dibutuhkan teknik penyimpanan dan pengumpulan data agar menghasilkan informasi dari data yang telah ada	Algoritma KNN, Naive Bayes	Pengujian menggunakan validasi 10-fold cross validation dan uji t-test menghasilkan nilai akurasi tertinggi sebesar 85.00% yaitu algoritma KNN.

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
7	Riyani Wulan Sari, Anjar Wanto, Agus Perdana Windarto (2018)	Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)	Penyakit campak masih merupakan masalah kesehatan di Indonesia dalam upaya menurunkan angka kesakitan dan angka kematian	Algoritma K-Means	persentase balita yang pernah mendapat imunisasi campak berdasarkan provinsi dengan cluster tinggi (c1) sebanyak 21 provinsi untuk cluster sedang (c2) sebanyak 12 provinsi dan untuk cluster rendah (c3) sebanyak 1 provinsi.
8	Ahmad Riyadh Al Faathin, Eko Sakti Pramukantoro, Fariz Andri Bakhtiar (2019)	Pengembangan Personal Data Analitik Menggunakan PHP-ML dan Apache Spark pada IoT Cloud Apps	membangun sebuah framework yang mampu menjawab permasalahan tersebut, sehingga dibangunlah sebuah data storage, data analytic dan juga visualization webservice yang di dalamnya akan dikembangkan menggunakan pendekatan personal.		Dengan kemampuan personalisasi sistem maka sistem ini dapat menerima tidak hanya data yang heterogen, akan tetapi juga dapat menangani pengguna, topik, perangkat dan juga data yang heterogen. Sistem dapat meningkatkan utitisasi sumberdaya pada lingkungan cloud, ditandai dengan penambahan kemampuan data analytic yang mengimplementasikan machine learning.
9	Fery Antony1, Rendra Gustriansyah	Deteksi Serangan *Denial of Service* pada *Internet of*	Penelitian ini bertujuan untuk mendeteksi dan mencegah serangan *denial of service*	bash-iptables	Implementasi bash-iptables berhasil mengurangi serangan synchronize flooding dengan

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
		Things Menggunakan Finite-State Automata	berupa *synchronize flooding* dan *ping flooding* pada jaringan *internet of things* dengan pendekatan *finite-state* automata		efisiensi waktu pencegahan sebesar 55,37% dan mengurangi serangan ping flooding sebesar 60% tetapi dengan waktu yang tidak signifikan.
10	Muhammad Khairullah Harto, Achmad Basuki (2021)	Deteksi Serangan DDoS Pada Jaringan Berbasis SDN Dengan Klasifikasi Random Forest	Penelitian ini bertujuan melakukan klasifikasi data dalam jaringan yang merupakan DDoS masih perlu dilakukan dalam peningkatan jumlah kasus serangan DDoS yang signifikan.	Random Forest	kinerja Random Forest dalam melakukan deteksi serangan DDoS dengan cukup baik. Rata-rata akurasi yang dihasilkan adalah 92,8% dengan akurasi maksimum bisa mencapai 100%. presisi dan recall memiliki rata-rata hasil 0.90 untuk presisi dan 0.93 untuk recall. Untuk rata-rata f1_score nya didapatkan nilai 0.9 dengan false rate 7.2%. Waktu pengambilan keputusannya juga terbilang singkat dengan rata-rata 282.4 ms atau sekitar 0.3 detik.
11	Thanh Thi Bi Dan, Sari Widya Sihwi, Rini	Implementasi iterative dichotomiser 3 pada data kelulusan	Penelitian ini Berfokus pada data wisuda yang menumpuk kemudian dilakukan pengolahan data wisuda agar	Iterative Dichotomiser 3 (ID3)	Hasil pengujian data testing yang diuji menggunakan algoritma ID3 memiliki nilai rata-rata maksimal precision

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
	Anggrainingsih (2015)	mahasiswa s1 di universitas sebelas maret	memberi nilai tambah, salah satunya dengan memanfaatkan datamining agar megatasi masalah tersebut.		63.96% dan ratarata maksimal recall 62.47% dari sepuluh kali pengujian dengan data yang diperoleh secara random.
12	Vincentius Riandaru Prasetyo, Hamzah Lazuardi, Aldo Adhi Mulyono, Christian Lauw (2021)	Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier	Penelitian ini mencoba memprediksi nilai tukar rupiah terhadap US Dollar dengan memanfaatkan aplikasi RapidMiner	linear regression	Dari hasil ujicoba pada 100 data testing, didapatkan nilai akurasi sebesar 95% dengan nilai threshold sebesar 30. Untuk performa model linear regression yang dihitung menggunakan root mean square error dihasilkan nilai sebesar 14,951. Hal tersebut menandakan performa model yang dihasilkan kurang baik.
13	Dwiki Bayu Satmoko, Parman Sukarno, Erwid Musthofa Jadied (2018)	Peningkatan Akurasi Pendeteksian Serangan DDoS Menggunakan Multiclassifier Ensemble Learning dan Chi-Square	DDoS merupakan serangan yang dapat mengakibatkan habisnya sumber daya yang dimiliki komputer sehingga komputer tidak dapat diakses bahkan hingga dapat menghilangkan data yang terdapat didalamnya	knn, naive bayes, random forest	Hasil akurasi dapat meningkatkan akurasi pendeteksian serangan DDoS menjadi sebesar 99,68%.
14	Rohan Doshi, Noah Apthorpe,	*Machine Learning DDoS Detection for*	Meningkatnya jumlah perangkat *Internet of Things* (IOT) terhubung ke Internet,	Neural Network, KNN,	Pengujian lima pengklasifikasi ML yang berbeda pada kumpulan data lalu lintas

NO	Peneliti/ Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
	Nick Feamster (2018)	*Consumer Internet of Things Devices*	namun banyak dari perangkat ini pada dasarnya tidak aman	LSVM, Decision tree, Random Forest	serangan normal dan DoS dikumpulkan dari eksperimen jaringan perangkat IoT konsumen. Kelima algoritma memiliki akurasi set tes lebih tinggi dari 0,99. Hasil awal ini memotivasi penelitian tambahan tentang anomali pembelajaran mesin deteksi untuk melindungi jaringan dari perangkat IoT yang tidak aman.
15	Mohd Faizal Ab Razak, Nor Badrul Anuar, Fazidah Othman, Ahmad Firdaus, F. Afifi1, Rosli Salleh [2017]	*Bio-inspired for Features Optimization and Malware Detection*	Data sensitif pada perangkat seluler Android menimbulkan ancaman serius bagi pengguna, dan serangan yang berbaahaya	Algoritma: Random forest, J48, K-nearest neighbors, multilayer perceptron, AdaBoost	Hasil percobaan menunjukkan tingkat deteksi 95,6% untuk TPR menggunakan classifier AdaBoost pada sampel malware Drebin yang dianalisis menggunakan optimasi fitur PSO.

Tabel 2.3 Matriks Penelitian Terkait

No	Peneliti (Tahun)	Judul	Ruang lingkup						
			Parameter Metric				Algoritma		
			Accuracy	*Precision*	*Recall*	AUC	KNN	DT	RF
1	Nur Widiyasono, Ida Ayu Dwi Giriantari, Made Sudarma, L Linawati (2021)	*Detection of Mirai Malware Attacks in IoT Environment Using Random Forest Algorithms*	✓	✓	✓	-	-	-	✓
2	Ridwan Hidayat (2022)	Perbandingan Algoritma *Random Forest, K-Nearest Neighbor* (Knn) Dan *Decision Tree* Pada Dataset N-BaIoT	✓	✓	✓	✓	✓	✓	✓

Tabel 2.3 merupakan penelitian terkait yang telah dilakukan sebelumnya mengenai deteksi *botnet*. Penelitian tersebut berjudul "*Detection of Mirai Malware Attacks in IoT Environment Using Random Forest Algorithms*" (Widiyasono et al. 2021) yaitu penelitian dengan mendeteksi serangan Malware Mirai dengan menggunakan Algoritma *Random Forest* pada dataset N-BaIoT. Penelitian dilakukan karena besarnya potensi serangan cyber terhadap infrastruktur *Internet of Thing* (IoT) dalam perangkat yang berjalan pada infrastruktur jaringan yang sudah ada sebelumnya, contohnya serangan *Malware Mirai*. Forensik jaringan investigasi membutuhkan algoritma *Random Forest* yang digunakan untuk melakukan klasifikasi dan mendeteksi teknik serangan malware mirai. Percobaan telah dilakukan dengan menggunakan 5 skenario serangan dan jenis perangkat. Hasil percobaan menunjukkan bahwa Algoritma *Random Forest* mencapai kinerja optimal dengan rata-rata nilai akurasi 95,01%, recall 90,82%, F1 Score 93,85% dan nilai presisi terbaik 99,23%. Selain itu, algoritma *Random Forest* cocok untuk pengolahan data yang sanbat besar.

Penelitian ini memiliki kesamaan dengan penelitian tersebut yaitu dalam mendeteksi serangan *malware mirai* menggunakan parameter *accuracy, presicion,* dan *recall.* Selain itu, pada penelitian ini juga menggunakan algoritma yang sama yaitu algoritma *Random Forest*.

Kelebihan penelitian ini dibandingkan dengan penelitian tersebut adalah adanya pengujian pada parameter AUC (*Area Under the ROC Curve*) dan menggunakan tiga algoritma yaitu *Random Forest, K-Nearest Neighbour,* dan D*ecision Tree*. Sedangkan kekurangan penelitian ini dibangdingkan dengan

penelitian yang dilakukan oleh Nur Widiyasono, Ida Ayu Dwi Giriantari, Made Sudarma, L Linawati adalah tidak adanya parameter F1 Score.

Penelitian dengan judul "*Detection of Mirai Malware Attacks in IoT Environment Using Random Forest Algorithms*" cukup sebagai acuan dan sudah memenuhi aspek yang dibutuhkan untuk penelitian yang akan dilakukan, terutama penelitiannya cukup mendekati dalam hal teknik dasar yang akan digunakan dengan judul penelitian "Perbandingan Algoritma *Random Forest, K-Nearest Neighbor*, Dan *Decision Tree* Pada Dataset *N-Baiot*".

Perbedaan dengan penelitian yang akan dilakukan terdapat pada dataset publik yang berbeda, selain itu penelitian ini juga melakukan pengujian pada parameter AUC (*Area Under the ROC Curve*) serta menggunakan tiga algoritma yaitu *Random Forest, K-Nearest Neighbour,* dan D*ecision Tree.*