BAB II

TINJAUAN PUSTAKA

2.1 *Machine Learning*

Machine Learning adalah bidang ilmu komputer yang melibatkan studi dan konstruksi teknik yang memungkinkan komputer untuk belajar mandiri berdasarkan data input untuk memecahkan masalah spesifik.

Jenis-jenis permasalahan yang umumnya diselesaikan dengan pendekatan *machine learning* adalah klasterisasi dan klasifikasi. Klasterisasi adalah aktivitas yang bertujuan mengelompokan data berdasarkan kedekatan fitur yang dimilikinya, sedangkan klasifikasi bertujuan untuk memisahkan data menjadi kelas-kelas tertentu. Perbedaan yang mendasar antara dua buah permasalahan ini adalah, pada proses klasterisasi, data-data dikelompokan tanpa pelabelan, sedangkan klasifikasi mengelompokan data-data menjadi label tertentu (Laksana Utama, 2018).

2.2 *Distributed Denial of Service* (DDoS)

Serangan *Distributed Denial-of-Service* (DDoS) adalah serangan yang menyebabkan *crash* pada *server* dan sistem di jaringan dengan membanjiri paket atau permintaan di jaringan. Sebuah serangan DDoS terdiri dari empat elemen, yaitu penyerang, program kontrol utama, *daemon* serangan/*bots*, dan korban. Serangan DDoS dapat dibagi ke dalam jenis berikut (Geges & Wibisono, 2015) :

1. Serangan dengan basis *bandwith*, yaitu serangan dengan cara mengirimi pesan data sampah secara masal untuk menyebabkan *overload* yang mengakibatkan

berkurangnya *bandwith* jaringan yang tersedia atau berkurangnya sumber daya perangkat jaringan.

2. Serangan dengan basis lalu lintas jaringan, yaitu serangan yang dilakukan dengan cara mengirimkan sejumlah besar paket *TCP*, paket *UDP*, paket *ICPM* yang tampak sah kepada *host/server* target. Serangan dengan basis ini dapat menghindari sistem deteksi dengan teknologi kamuflase alamat asal yang menyebabkan permintaan sah tidak terlayani
3. Serangan dengan basis aplikasi, yaitu serangan jenis ini biasanya mengirim pesan data pada tingkat *layer* aplikasi sesuai dengan fitur bisnis yang spesifik (menggunakan fungsi tampaknya legal dan operasional, seperti akses *database*), sehingga semakin berkurangnya sumber daya tertentu pada lapisan aplikasi (seperti jumlah pengguna dan koneksi aktif yang diperbolehkan) dan layanan sistem tidak lagi tersedia.

2.3 *Intrusion Detection System* (IDS)

Intrusion Detection System (IDS) adalah proses pemantauan, mendeteksi, dan menganalisis peristiwa yang dianggap sebagai pelanggaran terhadap kebijakan keamanan lingkungan jaringan memperkenalkan konsep mendeteksi serangan berbasis *cyber* pada jaringan komputer dengan menyediakan kerangka kerja untuk *intrusion detection system* (IDS), yang didasarkan pada hipotesis bahwa pelanggaran keamanan dapat dideteksi dengan memantau catatan audit sistem untuk pola *abnormal system* (Ashfaq, Wang, Huang, Abbas, & He, 2017).

2.4 *Data Mining*

Data mining adalah proses yang menggunakan teknik statistik, perhitungan, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai basis data besar (Khatib et al., 2022)

Tahapan dalam *data mining* terbagi dalam beberapa langkah yang disebut *CRoss-Industry Standard Process for Data Mining* (CRISP-DM) yaitu :

1. *Business Understanding/Organizational Understanding* (Pemahaman bisnis/organisasi) : Tahap pemahaman sistem yang berjalan dan kebutuhan apa yang dibutuhkan dalam menyelesaikan masalah yang timbul didalamnya.
2. *Data Understanding* (Pemahaman data) : Tahap pemahaman dan pengumpulan data yang dibutuhkan untuk sebelum dilakukan persiapan untuk analisa. Pada tahap ini data yang dikumpulkan harus merupakan data yang tepat digunakan untuk proses penelitian dan mewakili masalah yang akan dipecahkan serta sesuai dengan kebutuhan dan kepentingan.
3. *Data Preparation* (Persiapan data) : Tahap persiapan dan seleksi data yang telah dikumpulkan dan diubah menjadi bentuk yang dapat diolah dalam model yang ditentukan selanjutnya.
4. *Modelling* (Pemodelan) : Proses Analisa dan pemodelan data yang telah disiapkan dimana dalam ini dilakukan penerapan atau penghitungan berdasarkan algoritma atau metode yang ditentukan untuk mendapatkan hasil yang diinginkan sesuai dengan kebutuhan pengguna dan melakukan representasi pemecahan masalah.

5. *Evaluation* (Evaluasi) : Melakukan analisa dan evaluasi dari hasil model yang telah dibuat apakah sudah sesuai standar dan telah memecahkan masalah atau memenuhi kebutuhan dari pengguna.
6. Deployment (Penerapan) : Tahap penerapan hasil dari model yang telah dievaluasi dan dianalisa untuk kemudian dijadikan bentuk yang dapat diolah Kembali.

2.5 *Naive Bayes*

Algoritma *Naïve Bayes* dapat dijadikan sebagai pengklasifikasi probabilistik sederhana yang akan memperkirakan himpunan peluang dengan memperhitungkan kemunculan serta kombinasi nilai dalam himpunan data tertentu. Algoritma *Naïve Bayes* menggunakan teorema *Bayes* yang menganggap bahwa semua atribut bersifat berdiri sendiri atau tidak berkaitan satu sama lain berdasarkan nilai variabel kelas. Secara matematis, algoritma ini dapat dituliskan sebagai berikut :

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)} \quad (2.1)$$

Dimana, $P(H|E)$ adalah kemungkinan atau peluang hipotesis berdasarkan kondisi (*posterior probability*), $P(E|H)$ adalah peluang parameter E berdasarkan kondisi pada hipotesis H, kemudian $P(H)$ adalah peluang hipotesis H (*prior probability*) dan $P(E)$ adalah peluang parameter E (*prior probability*).

2.6 *Random Forest*

Random Forest adalah pengklasifikasi yang bersifat *ensemble*, yaitu *Random Forest* akan menciptakan sebuah hutan (f*orest*) menggunakan sejumlah pohon keputusan (*decision tree*). Jumlah suara terbanyak (*voting*) dari seluruh pohon

keputusan akan digunakan untuk menentukan kelas dari sebuah *input data*. Hal ini secara langsung dapat mengatasi masalah ketika melakukan klasifikasi hanya menggunakan satu pohon keputusan saja sering kali tidak optimal, tetapi dengan memasukkan banyak pohon keputusan, maka akan diperoleh nilai akurasi yang optimal secara global (Kusumarini et al., 2021)

2.7 Algoritma *J48*

J48 merupakan implementasi dari algoritma *C4.5* yang memproduksi *decision tree*. Ini merupakan standar algoritma yang digunakan dalam *machine learning*. Dalam *k-fold cross validation*, data pengujian dipisah secara acak ke dalam k himpunan bagian yang *mutually exclusive* atau *folds* (lipatan), *D1*, *D2*,….., *Dk*, yang masing-masing kurang lebih berukuran sama. Pelatihan dan pengujian dilakukan sebanyak k kali. Pada *iterasi* ke-I, partisi Di, digunakan sebagai data ter dan partisi sisanya digunakan bersama untuk melatih model. Dalam iterasi pertama, yaitu himpunan bagian *D2*,…..,*Dk* secara bersama bertindak sebagai data pelatihan untuk memperoleh model pertama yang diuji pada *D1*, iterasi kedua dilatih pada himpunan bagian *D1*,*D2*, ….., *Dk* dan diuji pada *D2* dan seterusnya (Situmeang, 2019)

2.8 *Confusion Matrix*

Confusion matrix dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan analisis apakah *classifier* tersebut baik dalam mengenali *tuple* dari kelas yang berbeda. Nilai dari *TruePositive* dan *TrueNegative* memberikan informasi ketika classifier dalam melakukan klasifikasi data bernilai benar,

sedangkan *FalsePositive* dan *False-Negative* memberikan informasi ketika classifier salah dalam melakukan klasifikasi data.

TP (*True Positive*) merupakan Jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif, FP (*False Positive*) merupakan Jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif, FN (*False Negative*) merupakan Jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif, TN (*True Negative*) merupakan jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif (Fibrianda & Bhawiyuga, 2018).

2.9 *Parameter Metric*

Accuracy merupakan tingkat keterhubungan antara suatu nilai yang diprediksi dengan nilai aktual yang ada (Devita et al., 2018). Persamaan atau formula dari *accuracy* dijelaskan pada persamaan 2.2.

$$Accuracy = \frac{True\ Positive+True\ Negative}{True\ Positive+False\ Negative+False\ Positive+True\ Negative} \quad (2.2)$$

Precision merupakan pengukuran tingkat ketepatan antara informasi yang diminta oleh pemohon dengan jawaban yang diberikan oleh sistem. Persamaan atau formula dari *precision* dijelaskan pada persamaan 2.3 (Yunus et al., 2019).

$$Precision = \frac{True\ Positive}{True\ Positive+False\ Positive} \quad (2.3)$$

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi dalam suatu pemrosesan data. Persamaan atau formula dari *recall* dijelaskan pada persamaan 2.4 (Yunus et al., 2019)

$$Recall = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad (2.4)$$

Sensitivity merupakan proses yang digunakan untuk mengukur pecahan pola positif yang diklasifikasian dengan benar. Persamaan atau formula dari *sensitivity* dijelaskan pada persamaan 2.5 (Pristyanto et al., 2018).

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2.5)$$

Specificity merupakan proses yang digunakan untuk mengukur pecahan pola negatif yang diklasifikasikan dengan benar. Persamaan atau formula dari *specificity* dijelaskan pada persamaan 2.6 (Pristyanto et al., 2018).

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (2.6)$$

Error Rate merupakan kesalahan klasifikasi mengukur rasio prediksi yang salah atas jumlah total contoh yang dievaluasi. Persamaan atau formula dari *Error Rate* dijelaskan pada persamaan 2.7 (Devita et al., 2018).

$$Error\ Rate = \frac{False\ Positive + False\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (2.7)$$

G-means merupakan indikator evaluasi performa dari algoritma. Persamaan atau formula dari *G-means* dijelaskan pada persamaan 2.8 (Pristyanto et al., 2018).

$$G\text{-}means = \frac{False\ Positive + False\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (2.8)$$

2.10 State Of The Art

Penelitian tentang klasifikasi *anomaly network traffic* dilakukan perbandingan metode yang dihasilkan dari proses klasifikasi berdasarkan nilai akurasi *confusion matrix, precision, recall*, dan f1 *score*. *Naive Bayes*, *Support*

Vector Machine (SVM) *Linear*, *SVM Polynomial* dan *SVM Sigmoid* menghasilkan persentase akurasi berturut-turut sebesar 85,055%, 99,995%, 99,999%, dan 99,995%. Persentase akurasi tertinggi diperoleh SVM *Polynomial*, sedangkan Naive Bayes menghasilkan persentase akurasi terendah (Fluorida Fibrianda & Bhawiyuga, 2018).

Algoritma *J48* memiliki akurasi yang cukup tinggi dibandingkan algoritma Naive Bayes dengan pengaturan *testing*, *atributte* dan *intances* yang sama, algoritma J48 mendapatkan nilai 81,85% sedangkan *Naive Bayes* 80,17% (Cendana & Permana, 2019).

Algoritma Random forest dalam melakukan deteksi serangan DDoS dengan cukup baik. Rata-rata akurasi yang dihasilkan adalah 92,8% dengan akurasi maksimum bisa mencapai 100%. Untuk 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 (Harto & Basuki, 2021).

2.11 Studi Literatur

Tabel 2.1 Studi Literatur

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
1	Muamar Zidane (2022)	Klasifikasi Serangan *Distributed Denial-of-Service* (DDoS) menggunakan Metode *Data Mining Naive Bayes*	*Distributed Deniel of Service* (DDoS) merupakan salah satu serangan yang bertujuan untuk menyebabkan *crash* pada sistem *server* dengan cara membanjiri paket ataupun permintaan pada jaringan. Karakteristik serangan DDoS sulit di bedakan dari arus lalu lintas jaringan normal, sehingga untuk mengidentifikasi serangan ini diperlukan sistem yang dapat mengklasifikasi serangan DDoS	Algoritma Naive Bayes	Tingkat akurasi dari hasil klasifikasi dihitung menggunakan *confusion matrix* dengan nilai akurasi sebesar 95% menggunakan metode naive bayes sehingga sistem dinilai cukup baik untuk melakukan klasifikasi serangan DDOS
2	Alvita I. Kusumarini, Pandu A. Hogantara, Muammar Fadhlurohman, Nurul Chamidah (2021)	Perbandingan Algoritma *Random Forest, Naive Bayes*, Dan *Decision Tree* Dengan *Oversampling* Untuk Klasifikasi Bakteri E. Coli	Suatu hal penting dalam mengidentifikasi bakteri adalah melalui karakteristik yang dapat diamati pada bakteri tersebut yang memanfaatkan ciri bentuk serta melalui perwarnaan sifat dari bakteri itu sendiri. Karakteristik yang dapat	Algoritma *Random Forest, Naive Bayes*, dan *Decision Tree*	Pengujian yang dilakukan menggunakan algoritma *Naive Bayes* memperoleh hasil akurasi yaitu sebesar 78%, *decision tree* dengan menggunakan parameter {*criterion: entropy, max_depth: 6, max_features: 5, min_samples_leaf: 3*} mendapatkan nilai akurasi sebesar 76%, serta *random forest*

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
			diamati dari suatu bakteri dapat diklasifikasi dengan memanfaatkan algoritma-algoritma klasifikasi		dengan menggunakan parameter {*criterion: gini, n_estimators: 250, max_depth: 6, random_state: 42*} mendapatkan nilai akurasi tertinggi, yaitu mendapatkan nilai 84%.
3	Ayu Pangestu, Taufik Ridwan (2021)	Penerapan *Data Mining* Menggunakan Algoritma *K-Means* Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka	Permasalahan yang muncul pada penggunaan air yaitu adanya ketidaklancaran air yang mengalir pada rumah warga yang berada di dataran tinggi. Beberapa penelitian yang berkaitan dengan pengolahan data penggunaan air pada perusahaan pengairan menggunakan data mining metode clustering	Algoritma *K-Means*	didapatkan nilai *centroid* 0 (46,6), *centroid* 1 (13,6), dan *centroid* 2 (25,4). Kelompok cluster 0 merupakan *cluster* boros yaitu sebanyak 9 orang, *cluster* 1 meupakan *cluster* sedang, dan *cluste*r 2 merupakan pelanggan yang hemat.
4	Aditya Dwi Afifaturahman, Firmansyah Maulana (2021)	Perbandingan Algoritma *K-Nearest Neighbour* (KNN) dan *Naive Bayes* pada *Intrusion Detection System* (IDS)	Banyak tantangan muncul karena serangan jahat terus berubah dan terjadi dalam volume yang sangat besar yang membutuhkan solusi yang dapat diskalakan.	Algoritma *K-Nearest Neighbour* (KNN) dan *Naive Bayes*	Pengujian dengan *percentage split* 60%, 70% dan 80% menunjukkan bahwa algoritma *K-Nearest Neighbour* (KNN) mendapatkan nilai yang lebih tinggi dari *Naive Bayes* kecuali *error rate*
5	Fery Antony, Rendra Gustriansyah (2021)	Deteksi Serangan *Denial of Service* pada *Internet of Things*	Penelitian ini bertujuan untuk mendeteksi dan mencegah serangan *denial of service* berupa *synchronize flooding*		Implementasi *bash-iptables* berhasil mengurangi serangan *synchronize flooding* dengan efisiensi waktu pencegahan sebesar 55,37% dan

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
		Menggunakan *Finite-State Automata*	dan *ping flooding* pada jaringan *internet of things* dengan pendekatan *finite-state automata*		mengurangi serangan *ping flooding* sebesar 60% tetapi dengan waktu yang tidak signifikan.
6	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*	Berdasarkan pengujian yang telah dilakukan, 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%. Untuk *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.
7	Suci Dilasari Kamil, Didit Widiyanto, Nurul Chamidah (2020)	Perbandingan Metode *Decision Tree* Dengan *Naive Bayes* Dalam Klasifikasi Tumor Otak Citra MRI	Langkah dalam identifikasi citra MRI otak masih kurang baik dalam hal mendiganosis, oleh karena adanya kualitas hasil pencitraan *Magnetic resonance imaging* (MRI) yang susah untuk dibaca oleh para ahli dokter dan radiolog, misalnya karena terdapat noise	Algoritma Decision Tree dan Naive Bayes	nilai *accuracy*, *specificity* dan *sensitivity decision tree* lebih tinggi dibanding metode naive bayes yaitu 96% *accuracy*, 96% *specificity* dan 96% *sensitivity* dan metode naive bayes 91% *accuracy*, 90% *specificity* dan 93% *sensitivity*.

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
			yang mengganggu penglihatan ataupun keterbatasan mesin *Magnetic resonance imaging* (MRI), maka memungkinkan dapat mempengaruhi keakuratan diagnosis. Dari data tersebut membutuhkan teknologi mesin yang dapat meminimalisir kesalahan dalam diagnosis tumor otak dan dapat dihindari dan juga dalam ketidakakuratan dari mendeteksi penyakit tumor otak pasien dengan lebih baik.		
8	Maya Cendana, Silvester Dian Handy Permana (2019)	Analisis Perbandingan Algoritma *Naive Bayes, J48,* dan *Random Forest Tree* Dalam Peningkatan Loyalitas Pelanggan UMKM Dengan Voucher Belanja	Voucher belanja diberikan kepada pelanggan lama yang mempunyai potensial untuk berbelanja kembali ditoko online yang ditentukan dengan algoritma penambangan data untuk mencari informasi yang tepat. Namun kesalahan memilih algoritma dapat mengakibatkan tidak optimalnya pendapatan yang diproyeksikan.	Algoritma *Naive Bayes, J48,* dan *Random Forest Tree*	Dari hasil penelitian ini didapatkan *random forest tree* adalah algoritma terbaik untuk menentukan potensial dari pelanggan toko online, yaitu dengan tingkat akurasi sebesar 99,38%, diikuti oleh algoritma J48 dengan tingkat akurasi 81,85%, dan *Naive Bayes* sebesar 80,17%

No	Peneliti/Tahun	Judul	Masalah Penelitian	Metode /Algoritma	*State of The Art*
9	Rini Indrayani (2018)	Analisa Perbandingan Algoritme *Naive Bayes* Dan *Decision Tree* Pada Klasifikasi Data Transfusi Darah	Banyak aspek yang menjadi pertimbangan saat uji kelayakan pendonor darah sehingga dilakukan klasifikasi *data mining* dengan berbagai metode.	Algoritma *Naïve Bayes* dan *Decision Tree*	Hasil analisa menunjukkan bahwa metode *Decision Tree* menunjukkan akurasi yang lebih tinggi sebesar 77.8075 % dibandingkan metode klasifikasi *Naïve Bayes* yang menunjukkan nilai akurasi sebesar 75.4011 %.
10	Rohan Doshi, Noah Apthorpe, Nick Feamster (2018)	*Machine Learning DDoS Detection for Consumer Internet of Things Devices*	Meningkatnya jumlah perangkat *Internet of Things* (IOT) terhubung ke Internet, namun banyak dari perangkat ini pada dasarnya tidak aman	Metode : *Anomali detection* Algoritma : *Neural Network, KNN, LSVM, Decision tree, Random Forest*	Pengujian lima pengklasifikasi ML yang berbeda pada kumpulan data lalu lintas 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.
11	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	Metode : *Bio-inspired* 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 2.2 Matriks Penelitian

No	Peneliti (Tahun)	Judul	Ruang Lingkup											
			Parameter Metric							Algoritma				
			Accuracy	Precision	Recall	Sensitivity	Specificity	F -measure	Error rate	NB	KNN	RF	DT	J48
1	Muammar Zidane (2022)	Klasifikasi *Serangan Distributed Denial-of-Service* (DDoS) menggunakan Metode *Data Mining Naïve Bayes*	✓	✓	✓	-	-	-	-	✓	-	-	-	-
2	Aditya Dwi Afifaturahman, Firmansyah Maulana (2021)	Perbandingan *Algoritma K-Nearest Neighbour* (KNN) dan *Naive Bayes* pada *Intrusion*	✓	✓	✓	✓	✓	-	-	✓	✓	-	-	-

No	Peneliti (Tahun)	Judul	Ruang Lingkup											
			Parameter Metric							Algoritma				
			Accuracy	Precision	Recall	Sensitivity	Specificity	F -measure	Error rate	NB	KNN	RF	DT	J48
		Detection System (IDS)												
3	Bustami Yusuf, Muthmainna Qalbi, Basrul, Ima Dwitawati, Malahayati, Mega Ellyadi (2020)	Implementasi Algoritma *Naive Bayes* Dan *Random Forest* Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh	-	✓	✓	-	-	✓	-	✓	-	✓	-	-
4	Rini Indrayani (2018)	Analisa Perbandingan Algoritme *Naïve Bayes*	✓	-	-	-	-	-	-	✓	-	-	-	-

No	Peneliti (Tahun)	Judul	Ruang Lingkup											
			Parameter Metric							Algoritma				
			Accuracy	Precision	Recall	Sensitivity	Specificity	F -measure	Error rate	NB	KNN	RF	DT	J48
		Dan *Decision Tree* Pada Klasifikasi Data Transfusi Darah												
5	Irfan Pratama (2022)	Perbandingan Algoritma *Random Forest, Naive Bayes* dan *J48* pada *Dataset Anomaly Network Traffic (Alldays Ddos)*	✓	✓	✓	✓	✓	-	✓	✓	-	✓		✓