MIND (Multimedia Artificial Intelligent Networking Database) Journal

MIND Journal | ISSN (p): 2528-0015 | ISSN (e): 2528-0902 | Vol. 8 | No. 1 | Halaman 52 - 64
DOI : https://doi.org/10.26760/mindjournal.v8i1.52-64 Juni 2023

Teknik *SMOTE* Sebagai Solusi *Imbalance Class* dalam Model Deteksi Intrusi DDoS dengan Metode *PCA-Random Forest*

BACHTIAR RAMADHAN[1], DIASH FIRDAUS[2], ARGYA RIJAL RAFI[3]

[1,3] D4 Teknik Informatika, Universtias Logistik dan Bisnis Internasional
[2]Teknik Elektro, Telkom University
Email : 1204077@std.ulbi.ac.id

Received 30 November 201x **|** *Revised* 30 Desember 201x **|** *Accepted* 30 Januari 201x

ABSTRAK

Keamanan sistem informasi adalah faktor yang harus diperhatikan. Keamanan sistem informasi mampu mendeteksi serangan yang terjadi pada sistem informasi. Salah satunya adalah serangan DDoS. Hal ini disebabkan DDoS dapat menimbulkan ancaman dalam jumlah besar yang dapat menganggu sistem. Serangan DDoS di dunia meningkat 6% setiap tahunnya. Untuk mengatasi hal tersebut, dilakukan penelitian dengan pendekatan machine learning. Dataset yang digunakan adalah CICDDos 2017 dan CICDDoS 2019 dari University of New Brunswick. Untuk menghasilkan data yang baik, dilakukan SMOTE untuk mengatasi imbalance class, dan feature selection menggunakan PCA sehingga menghasilkan 15 fitur pilihan. Kemudian dilakukan pemodelan menggunakan Random Forest Classifier. Hasil penelitian ini adalah nilai akurasi sebesar 99.94%, presisi sebesar 99.90%, recall sebesar 99.97%, dan f1-score sebesar 99.94%. Dari hasil tersebut, dapat disimpulkan teknik PCA-Random Forest dapat mendeteksi serangan DDoS dengan baik.

***Kata kunci**: DDoS, SMOTE, PCA-Random Forest*

ABSTRACT

Information system security is a factor that must be considered. Information system security is able to detect attacks that occur on information systems. One of them is a DDoS attack. This is because DDoS can cause a large number of threats that can disrupt the system. DDoS attacks in the world are increasing 6% every year. To overcome this, we conducted research using a machine learning approach. The dataset used is CICDDoS 2017 and CICDDoS 2019 from the University of New Brunswick. To produce good data, SMOTE is performed to overcome class imbalance, and feature selection uses PCA to produce 15 selected features. Then modeling is done using the Random Forest Classifier. The results of this study are 99.94% accuracy, 99.90% precision, 99.97% recall, and 99.94% f1-score. From these results, it can be concluded that the PCA-Random Forest technique can detect DDoS attacks properly.

***Keywords**: DDoS, SMOTE, PCA-Random Forest*

1. PENDAHULUAN

Pada revolusi industri 4.0, kemajuan ilmu pengetahuan dan teknologi semakin cepat. Pada revolusi industri 4.0, komunikasi dan pergerakan antara mesin dan manusia didukung oleh kecerdasan dan *internet of things* **(Prasetyo & Trisyanti, 2018)**. Akibatnya, menimbulkan keadaan di mana manusia sangat bergantung pada teknologi informasi. Sistem keamanan siber menjadi krusial karena maraknya kebutuhan masyarakat akan penggunaan sistem informasi di berbagai bidang, antara lain perbankan, bisnis, pendidikan, dan lain-lain. Ini termasuk deteksi intrusi **(Zuech et al., 2015)**. Serangan *cyber* menjadi mimpi buruk bagi semua pihak yang menggunakan teknologi informasi **(Lin et al., 2022)**. *Traffic* lalu lintas yang padat serta keragaman dan pengenalan jenis intrusi baru memberikan tantangan kontemporer untuk deteksi intrusi. Serangan DDoS *(Distributed Denial of Service)* merupakan salah satu jenis serangan yang sering dialami. Pada tahun 2018, intrusi DDoS adalah yang paling umum dan terkenal **(Alison DeNisco Rayome, 2019) (Cisco, 2020)**. Serangan DDoS akan menyerang targetnya dengan cara membuat banyak permintaan dengan jumlah yang sangat besar ke target, sehingga target mengalami penolakan permintaan normal atau menurunkan kualitas layanan **(David & Thomas, 2015)**. Bagi sejumlah pihak, serangan ini mengakibatkan kerugian yang sangat besar, antara lain penurunan pendapatan, kegagalan manufaktur, rusaknya reputasi, kejahatan pencurian, dan lain-lain. Ini memotivasi banyak pihak untuk mengembangkan strategi deteksi dan pencegahan yang efektif **(Singh et al., 2016)**.

Berdasarkan permasalahan yang dijelaskan, maka penelitian ini mencoba untuk merancang teknik deteksi intursi DDoS menggunakan metode pendekatan *machine learning. Machine learning* merupakan subbidang *artificial intelligence* yang berfokus pada pembelajaran data, sehingga tercipta sistem yang dapat belajar sendiri tanpa perlu terus menerus dilatih oleh manusia. Dengan demikian, dengan memanfaatkan *machine learning* sebagai solusi akan menciptakan sistem deteksi yang efisien **(Santoso et al., 2020) (Kurniawan & Yulianingsih, 2021) (Widodo, 2020)**. Metode *machine learning* yang akan digunakan adalah algoritma *Random Forest Classifier* yang digabung dengan algortima *Principal Component Analysis* (PCA). Algoritma PCA digunakan untuk mencari rekomendasi faktor-faktor yang mempengaruhi serangan DDoS. PCA adalah teknik yang handal untuk mengekstraksi struktur dari suatu dataset **(Ismawan, 2015) (Herianto, 2016)**. Sehingga pemilihan faktor-faktor spesifik menggunakan algoritma PCA dimaksudkan untuk menciptakan teknik deteksi yang memiliki akurasi yang baik. Gabungan antara *Random Forest Classifier* dengan algoritma PCA sebagai ekstraksor faktor akan menghasilkan deteksi yang ideal. Evaluasi model diuji berdasarkan *classification report* dan *confusion matrix* yang terdiri dari *precision, recall, f1-score,* dan *accuracy* **(Atimi & Enda Esyudha Pratama, 2022) (Fathan Hidayatullah & Sn, 2014)**.

Pada penelitian **(Chris et al., 2019)** digunakan algoritma SVM yang dapat memisahkan *traffic* normal dan *traffic* serangan. Pada penelitian tersebut, SVM melakukan deteksi serangan DDoS pada arsitektur *Software Defined Network* (SDN) dan menghasilkan akurasi sebesar 98,87%. Kemudian pada penelitian **(Wahyuni & Adytia, 2018)** untuk mendeteksi DDoS digunakan metode *XGBoost* yang merupakan salah satu metode yang *powerfull* dan *Decision tree* yang merupakan metode yang cukup mudah digunakan dan keakuratannya yang baik. Pada penelitian tersebut, *XGBoost* menghasilkan akurasi sebesar 99,92% dan *Decision tree* sebesar 99,87%. Kemudian pada penelitian yang dilakukan **(Firmansyah et al., 2022)**, digunakan metode *Naïve Bayes* untuk mendeteksi DDoS. *Naïve Bayes* adalah salah satu algoritma kategorisasi. Prosedur algoritma ini mampu memprediksi probabilitas informasi anggota kelas. Keuntungan menggunakan *Naïve Bayes* adalah bahwa prosedur ini hanya membutuhkan data

training yang sedikit untuk mengidentifikasi tujuan standar yang diperlukan. Dari penelitian ini dihasilkan akurasi model sebesar 82,45%.

Kemudian pada penelitian **(Harto & Basuki, 2021)** digunakan model *Random Forest. Random Forest* merupakan metode *ensamble* dari *decision tree*. *Decision tree* berfungsi sebagai diagram alir yang berbentuk seperti pohon yang memiliki akat yang digunakan untuk mengumpulkan data yang digunakan untuk mengambil keputusan sementara. Pada tahap akhir, *random forest* adalah suara terbanyak dari setiap keputusan tersebut. Dari penelitian tersebut, dihasilkan akurasi model sebesar 92,80%.

Pada penelitian ini, dataset yang digunakan adalah kombinasi dari dataset CICDDoS 2017 dan CICDDoS 2019 **(University of New Brunswick, 2019)**. Namun, ditemukan bahwa kumpulan data yang dimaksud memiliki kelas yang tidak seimbang *(imbalance class)*. Tentu saja, kumpulan data yang seimbang diperlukan agar model pembelajaran mesin dapat memberikan hasil yang efektif. Oleh karena itu, diperlukan suatu metode untuk mengimbangi perbedaan kelas tersebut. Teknik *Over-Sampling* Minoritas Sintetis adalah salah satu metode *over-sampling* (SMOTE). Kapasitas algoritma klasifikasi untuk mengatasi ketidakseimbangan kelas dapat ditingkatkan dengan penggunaan SMOTE **(Suryana et al., 2020) (Sutoyo & Asri Fadlurrahman, 2020)**. Dengan teknik ini, SMOTE akan melakukan duplikasi data secara sintetis sehingga permasalahan *over sampling* dapat teratasi dengan baik **(Rais & Subekti, 2019)**.

Berdasarkan hasil penelitian di atas, maka pada penelitian ini merancang strategi pemodelan yang baik dalam mendeteksi DDoS dengan menggunakan algoritma *Random Forest Classifier* sebagai model prediksi, *Principal Component Analysis* sebagai penyeleksi faktor-faktor yang mempengaruhi terhadap serangan DDoS, dan *Syntetic Minority Over-Sampling Technique* sebagai solusi dalam mengatasi *over-sampling* data. Kemudian, model yang dibuat akan dievaluasi berdasarkan standar evaluasi yaitu *Classification Report,* dan *Confusion Matrix.*

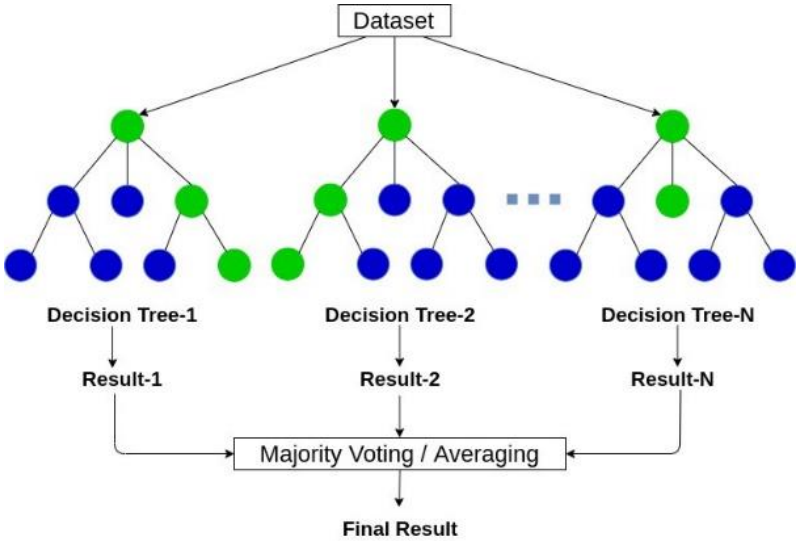
2. METODE PENELITIAN

2.1. DDoS *(Denial Distributed of Services)*

Serangan DDoS adalah tindakan mengirimkan banyak paket melalui jaringan dalam upaya membanjirinya dengan data dan membuat situs tidak dapat dijangkau oleh pengguna yang berwenang **(Yasin & Mohidin, 2018)**. Serangan ini merupakan ancaman yang berat bagi keamanan sistem dan melelahkan sistem target dengan lalu lintas yang berbahaya. Hampir setiap lapisan dalam tumpukan TCP/IP rentan terhadap beragam serangan DDoS yang mengancam keamanan dan ketersediaan jaringan. Serangan DDoS dikategorikan sebagai serangan konektivitas, serangan *bandwidth,* serangan *limitation exploitation,* serangan *resource exhaustion,* serangan *data corrupt,* proses disrupsi, dan lain-lain. Tingkat ancaman dari suatu serangan tergantung pada faktor-faktor seperti jenis serangan, jumlah *host* yang disusupi dan menyerang, sumber daya penyerang, protokol serangan, sumber daya *destination,* dan keamanan situs *destination* **(Lin et al., 2022)**.

2.2. *Random Forest Classifier*

Random Forest merupakan metode *machine learning* yang berbasis klasifikasi dan regresi dimana terdapat proses agregasi *decision tree* **(Dhawangkhara & Riksakomara, 2017) (Meng et al., 2021)**.



Gambar 1. Arsitektur *Random Forest*
(Rahul, 2020)

Gambar 1 merupakan arsitektur dari RF. RF menggunakan teknik *bootstrap* untuk mengekstrak sampel acak dari sampel asli dan membangun pohon keputusan tunggal **(Fouedjio, 2020)**. Pada setiap *node decision tree,* pemilihan titik pengurutan menggunakan *subspace* fitur acak **(Nehra & Nagaraju, 2022)**. Terakhir, gabungan *decision tree* ini mendapatkan hasil prediksi akhir dengan suara terbanyak. RF ditentukan dengan mencari nilai entropi terlebih dahulu, kemudian mencari nilai informasi gain. Oleh karena itu, *random forest* menghasilkan sedikit kesalahan, memiliki akurasi klasifikasi yang baik, dapat menangani volume data pelatihan yang sangat besar, dan efisien dalam menangani data yang tidak lengkap.

2.3. PCA *(Principal Component Analysis)*

Dalam PCA, variabel awal digabungkan secara linear untuk membuat sistem koordinat baru yang berasal dari sistem rotasi asli. Jika data yang ada memiliki banyak variabel dan terdapat keterkaitan antar variabel, pendekatan *principal component analysis* sangat membantu. Nilai *costs* yang terkait dengan komputasi nilai eigen dan vektor eigen, yang mewakili distribusi data dari kumpulan data, untuk PCA **(Fauzi et al., 2020)**. Penggunaan PCA sebagai *feature selection* adalah untuk memilih variabel sesuai dengan besarnya koefisien **(Astuti & Adiwijaya, 2019)**. PCA akan menggantikan variabel yang lebih atau kurang berkorelasi dengan kombinasi linier dari variabel asli **(Marestiani & Surono, 2022)**.

2.5 SMOTE *(Syntetic Minority Over-sampling Technique)*

Data dari kelas minoritas akan diduplikasi menggunakan data sintetik yang dibuat dengan mereplikasi data dari kelas minoritas melalui pendekatan *over-sampling* yang dikenal dengan SMOTE. Dalam SMOTE, contoh kelas dari kelas minoritas diambil, dan setelah menemukan *k-nearest neighbors* setiap *instance*, *over-sampling* membuat *instance* sintetik yang diturunkan dengan menduplikasi *instance* kelas minoritas. Akibatnya, ini dapat mencegah masalah *overfitting* yang ekstrim **(Erlin et al., 2022)**. Langkah pertama untuk SMOTE adalah mengalikan nilai varians antara nilai tetangga terdekat dari kelas minoritas dan nilai vektor fitur di kelas minoritas dengan angka acak antara 0 dan 1. Vektor fitur digabungkan dengan komputasi hasil untuk membuat nilai vektor baru **(Jishan et al., 2015)**. Perhitungan matematisnya bisa dilihat pada Persamaan (1).

$$X_{new} = X_i + (X'_i - X_i) \times \delta \quad (1)$$

Keterangan :
X_i = vektor dari fitur pada kelas minoritas
X'_i = *k-nearest neighbors* untuk X_i
δ = angka acak antara 0 sampai 1

2.6 Evaluasi Model

Evaluasi model yang dilakukan untuk melihat seberapa baik model dalam melakukan deteksi. Model tersebut dievaluasi dengan melihat nilai *error* pada deteksi. Terdapat beberapa perhitungan untuk melihat nilai *error* tersebut, yaitu dengan *Classification Report* dan *Confusion Matrix.* Untuk menguji klasifikasi pada algoritma yang dibangun, perhitungan tersebut merupakan perhitungan yang tepat untuk melihat deteksi intrusi DDoS **(Fathan Hidayatullah & Sn, 2014) (Hastuti, 2012)**. Menghitung nilai presisi, recall, dan ukuran f1 merupakan tahapan yang diperlukan untuk menilai kinerja algoritma guna menentukan nilai akurasi. Kemampuan sistem untuk membedakan antara data benar dan salah dikenal sebagai akurasi, dan tujuan pengukuran f1 adalah untuk mengevaluasi kinerja sistem secara keseluruhan dengan menghitung akurasi dan nilai perolehan. Perhitungan ditampilkan dalam Gambar 2 padat berikut.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Gambar 2. *Confusion Matrix*
(Mohd, 2019)

TP (*True Positive*) adalah prediksi positif dan nilai sebenarnya positif, TN (*True Negative*) adalah prediksi negatif dan nilai sebenarnya negatif, FP (*False Positive*) merupakan prediksi positif dan nilai sebenarnya negatif, dan FN (*False Negatif*) merupakan prediksi negative dan nilai sebenarnya positif **(Devella & Adi Putra, 2021) (Khasanah et al., 2021)**. Rumus dari evaluasi klasifikasi dapat dilihat pada Persamaan (2), (3), dan (4) berikut.

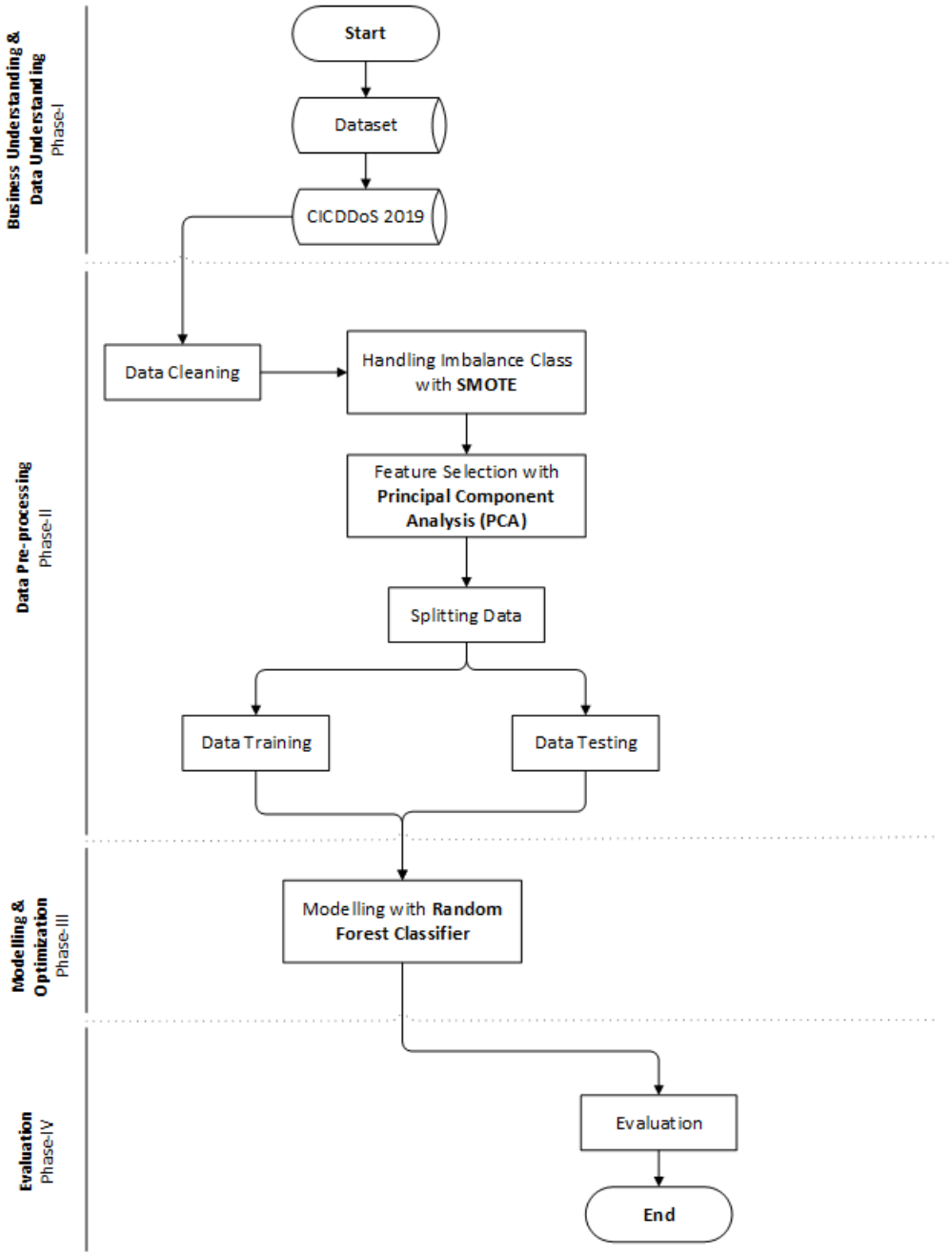
$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

2.7 Metodologi Penelitian

Adapun metode penelitian yang digunakan pada penelitian ini dapat dilihat pada Gambar 3 sebagai berikut.



Gambar 3. Metodologi Penelitian

Untuk melakukan pemodelan *machine learning* untuk melakukan deteksi intrusi DDoS terdapat beberapa tahapan sebagai berikut.

1. *Phase-I*, mengumpulkan data yang relevan dengan objek penelitian. himpunan data berdasarkan data yang sudah ada yang diambil dari CICDDoS 2019.
2. *Phase-II*, melakukan analisa dan *cleaning* data. Setelah itu, dilakukan proses pemilihan fitur menggunakan algoritma *principal component analysis.* Diketahui bahwa data yang digunakan mengalami *imbalanced class,* dilakukan teknik *over-sampling* menggunakan *Syntetic Minority Over-Sampling Technique* (SMOTE). Setelah data *balanced,* data

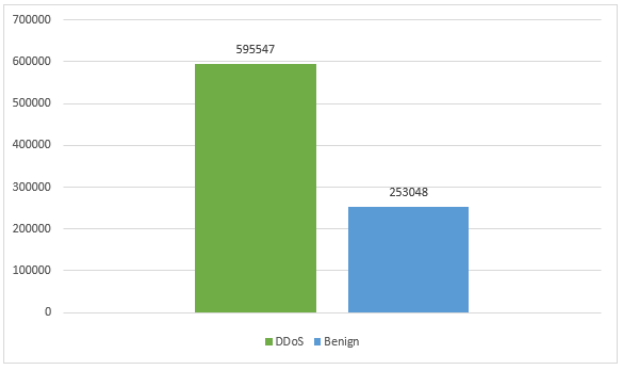
dibagi menjadi dua, yaitu data *training* untuk melakukan *training model,* dan data *testing* untuk melakukan *testing model.*

3. *Phase-III*, data yang telah disiapkan, digunakan untuk pemodelan *machine learning* menggunakan *random forest classifier.* Sehingga terbentuk model deteksi intrusi.
4. *Phase-IV*, mengevaluasi model dari hasil prediksi yang menggunakan data training menggunakan *classification report, r-squared, mean absolute error* (MAE), *mean squared error* (MSE) untuk melihat apakah model *machine learning* menghasilkan hasil yang baik atau tidak.

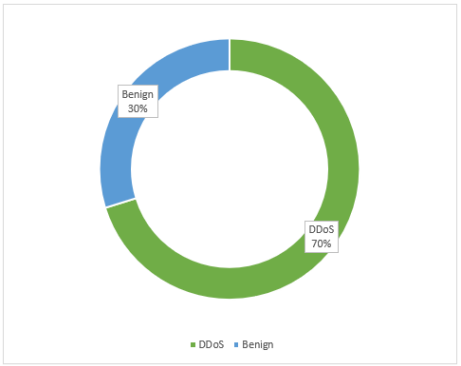
3. HASIL DAN PEMBAHASAN

3.1 Dataset

Pada penelitian ini, dataset yang digunakan adalah kombinasi dari dataset CICDDoS 2017 dan CICDDoS tahun 2019 yang dihimpun oleh *University of New Brunswick.* CICDDoS 2019 berisikan trafik normal dan serangan. Dataset ini digunakan sebagai bahan untuk melakukan *training machine learning.* Hal tersebut disebabkan dataset ini mewakili beberapa hal seperti kompleksitas, heterogenitas, interaksi yang lengkap dan lain sebagainya. Dalam himpunan data ini terdapat serangan DDoS termasuk *LDAP*, *MSSQL*, *NetBIOS*, *UDP*, *UDP-Lag*, dan *Syn*. Data yang diambil dari semua serangan sebanyak 848595 data dan 84 fitur yang digunakan untuk melakukan *training model* dan *testing model.* Berikut perbandingan data serangan dan data normal pada Gambar 4 dan Gambar 5.



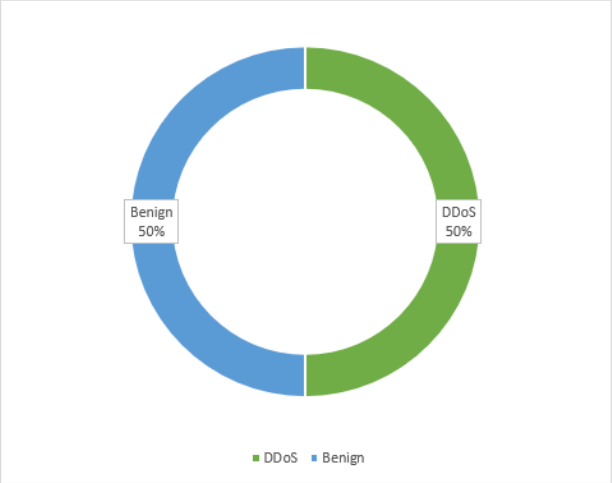
Gambar 4. Jumlah Data Serangan DDoS dan Normal



Gambar 5. Perbandingan Data Serangan DDoS dan Normal

3.2 *Over-sampling* menggunakan SMOTE

Pada Gambar 4 dan Gambar 5 dapat dilihat bahwa data serangan lebih banyak dari pada data normal. Sebaran distribusi data antar kelas yang sangat besar dapat menyebabkan *training model* yang kurang maksimal. Kelas minoritas, yang memiliki sampel lebih sedikit, akan dikalikan dengan data buatan yang dihasilkan oleh SMOTE, menghasilkan distribusi data yang lebih merata saat menggunakan SMOTE.

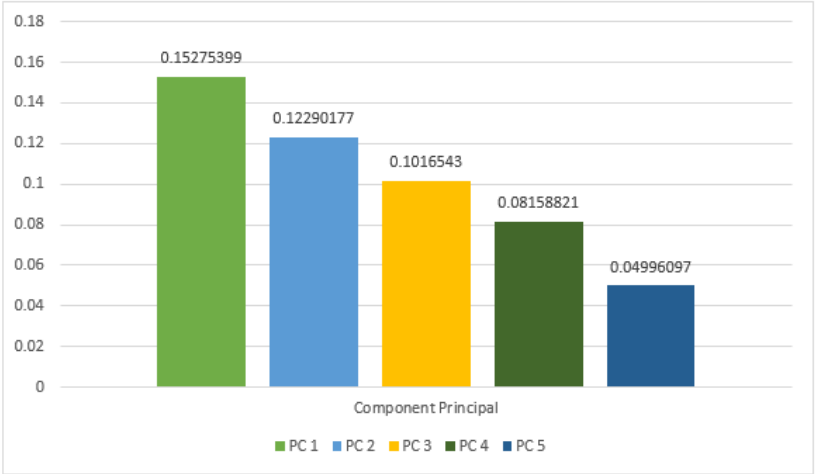


Gambar 6. Perbandingan Data Setelah *Over-sampling*

Sebelum melakukan *over-sampling,* dataset dibagi menjadi 2 yaitu data *training* yang berjumlah 80% dari jumlah data dan data *testing* yang berjumlah 20% dari dataset. Untuk menghasilkan akurasi yang baik, data *training* akan dilakuka proses *over-sampling.* Setelah melewati proses *over-sampling* menggunakan SMOTE, data serangan DDoS dengan data normal memiliki jumlah yg sama, terlihat pada Gambar 6 masing-masing kelas berjumlah 50% data atau berjumlah 454848 data. Dengan kata lain, *over-sampling* SMOTE berhasil digunakan untuk meningkatkan jumlah dataset untuk mencapai dataset yang seimbang.

3.3 *Feature Selection* menggunakan PCA

Setelah data *balance,* proses selanjutnya adalah melakukan *feature selection. Feature selection* digunakan untuk memilih faktor yang paling berpengaruh terhadap deteksi DDoS. Untuk memilih *feature selection* digunakan algoritma *Principal Component Analysis* sebagai algoritma penyeleksi faktor-faktor tersebut. Proses perhitungannya diawali dengan normalisasi data agar data berada dalam *range* tertentu. Kemudian dilakukan perhitungan nilai *variance* antar variabel, dan nilai vector *eigen* dihitug dan menentukan *component principal.*



Gambar 7. Nilai *Eigen*

Hasil perhitungan PCA menghasilkan nilai eigen 0.15275399, 0.12290177, 0.1016543, 0.08158821, 0.04996097. Kemudian untuk menentukan variabel apa saja yang benar benar mempengaruhi serangan DDoS dengan menggunakan metodologi rotasi faktor *varimax* yang menjelaskan hubungan antara variabel asli (korelasi) antara variabel baru (*principal*

component) yang dibentuk dengan PCA yang disebut sebagai nilai *loading.* Nilai *loading* yang dipilih adalah nilai *loading* yang di atas 0.2 (rendah - tinggi) yang dianggap mempengaruhi deteksi DDoS. Variabel lain yang di bawah 0.2 dianggap tidak atau kurang berpengaruh.

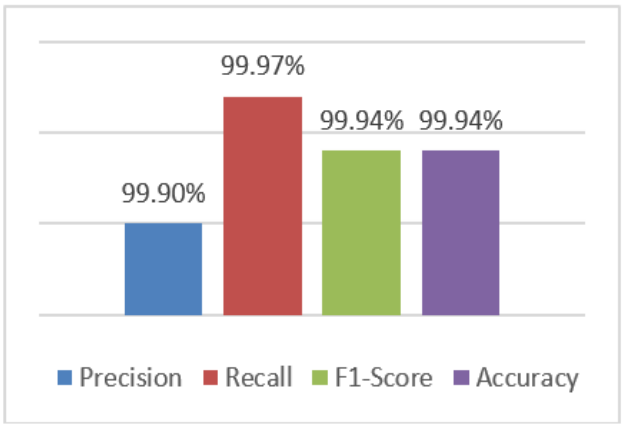
Tabel 2. Fitur Pilihan Hasil *Feature Selection* Dengan PCA

No	Nama Variabel	No	Nama Variabel
1	Fwd Pkt Len Max	9	Pkt Len Std
2	Fwd Pkt Len Min	10	Syn Flag Cnt
3	Fwd Pkt Len Mean	11	Fwd Seg Size Avg
4	Bwd Pkt Len Mean	12	Bwd Seg Size Avg
5	Bwd Pkt Len Std	13	Idle Mean
6	Flow Byts/s	14	Idle Max
7	Flow IAT Std	15	Idle Min
8	Pkt Len Min		

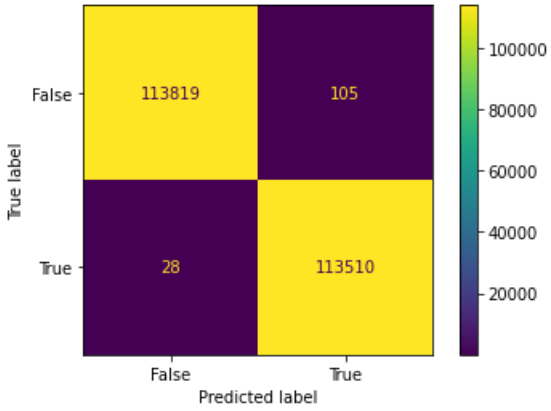
Dari analisis tersebut terlihat pada tabel 2 dihasilkan 15 variabel yang dianggap mempengaruhi dari 77 variabel. Kemudian *feature selection* ini akan digunakan pada proses *training model machine learning.*

3.4 *Modelling* menggunakan *Random Forest Classifier*

Setelah melewati proses *feature selection,* data yang tersebut dibagi menjadi 2, yaitu data *traning* sebanyak 80% dan data *testing* sebesar 20%. Setelah itu, dilakukan pemodelan dengan algoritma *Random Forest Classifier* untuk deteksi DDoS. Kemudian hasil pemodelan dievaluasi menggunakan *confusion matrix.*

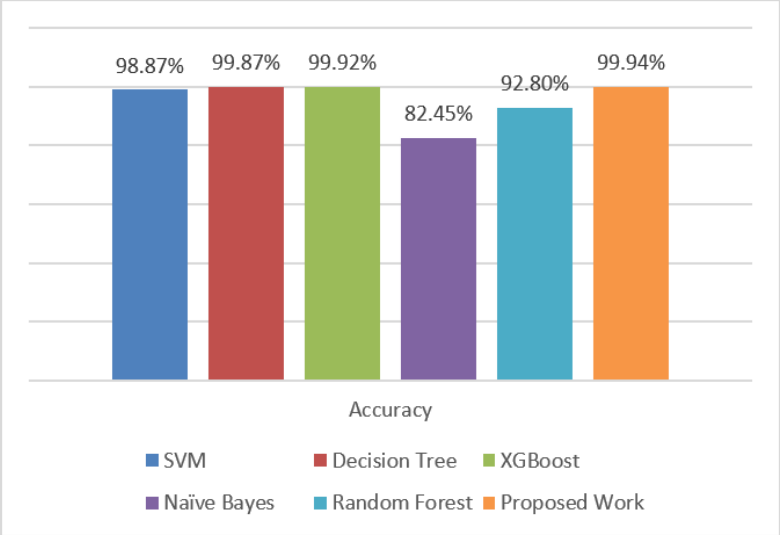


Gambar 8. *Classification Report*



Gambar 9. *Confusion Matrix*

Berdasarkan hasil dari pemodelan *PCA-Random Forest* menggunakan dataset hasil *SMOTE,* hasil yang didapat adalah nilai akurasi sebesar 99,94%, presisi sebesar 99,90%, *recall* sebesar 99,97%, dan *f1-score* sebesar 99,94%. Kemudian hasil tersebut dibandingkan dengan penelitian yang sudah ada. Kemudian pada Gambar 9, diketahui data yang *true positive* sebanyak 113510 data, *true negative* sebanyak 113819 data, *false positive* sebanyak 105 data, dan *false negative* sebanyak 28 data. Kemudian hasil pemodelan *PCA-Random Forest* dibandingkan dengan algoritma berdasarkan *State-of-the-art* pada Gambar 10 yang terdiri dari SVM, *Decision Tree,* XGBoost, *Naïve Bayes,* dan *Random Forest.*



Gambar 10. Perbandingan Akurasi Model

4. KESIMPULAN

Dalam penelitian ini, dikumpulkan dataset terkait serangan DDoS dan normal. Dataset yang berjumlah 84 fitur di seimbangkan jumlah data serangan dan normal dengan metode *syntetic minority over-sampling technique* (SMOTE), kemudian fitur-fitur tersebut diseleksi oleh algoritma *principal component analysis* (PCA) sehingga dihasilkan 15 fitur pilihan. Dari dataset tersebut dilakukan *training* dan *testing* dengan menggunakan salah satu model *machine learning* yaitu *Random Forest Classifier*. Dari penelitian ini dapat disimpulkan bahwa nilai akurasi sebesar 99.94%, presisi sebesar 99.90%, *recall* sebesar 99.97%, dan *f1-score* sebesar 99.94%. Kemudian hasil tersebut dibandingkan dengan penelitian yang sudah ada. Kemudian pada Gambar 8, diketahui data yang *true positive* sebanyak 113510 data, *true negative* sebanyak 113819 data, *false positive* sebanyak 105 data, dan *false negative* sebanyak 28 data. Artinya, teknik *PCA-Random Forest* dengan metode SMOTE sebagai *over-sampling* data mampu mendeteksi serangan DDoS secara akurat.

DAFTAR RUJUKAN

Alison DeNisco Rayome. (2019, March 18). *DDoS attacks on the rise: Largest attack ever hit 1.7 Tb/second*.

Astuti, W., & Adiwijaya, A. (2019). Principal Component Analysis Sebagai Ekstraksi Fitur Data Microarray Untuk Deteksi Kanker Berbasis Linear Discriminant Analysis. *Jurnal Media Informatika Budidarma*, *3*(2), 72–77. https://doi.org/10.30865/mib.v3i2.1161

Atimi, R. L., & Enda Esyudha Pratama. (2022). Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia. *Jurnal Sains Dan Informatika*, *8*(1), 88–96. https://doi.org/10.34128/jsi.v8i1.419

Chawla, N. v., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chris, J., Sihombing, J., Kartikasari, D. P., & Bhawiyuga, A. (2019). Implementasi Sistem Deteksi dan Mitigasi Serangan Distributed Denial of Service (DDoS) menggunakan SVM Classifier pada Arsitektur Software-Defined Network (SDN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(10), 9608–9613. http://j-ptiik.ub.ac.id

Cisco. (2020, March 9). *Cisco Annual Internet Report (2018–2023) White Paper*.

David, J., & Thomas, C. (2015). DDoS attack detection using fast entropy approach on flow-based network traffic. *Procedia Computer Science*, *50*, 30–36. https://doi.org/10.1016/j.procs.2015.04.007

Devella, S., & Adi Putra, C. (2021). Penggunaan Fitur Saliency-SURF Untuk Klasifikasi Citra Sel Darah Putih Dengan Metode SVM. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(4), 1998–2009. http://jurnal.mdp.ac.id

Dhawangkhara, M., & Riksakomara, E. (2017). Prediksi Intensitas Hujan Kota Surabaya dengan Matlab Menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya). *Jurnal Teknik ITS*, *6*(1), A94–A99.

Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(3), 677–690. https://doi.org/10.30812/matrik.v21i3.1726

Fathan Hidayatullah, A., & Sn, A. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. *Seminar Nasional Informatika*. http://www.situs.com

Fauzi, A., Supriyadi, R., & Maulidah, N. (2020). Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest. *Jurnal Infortech*, *2*(1). http://ejournal.bsi.ac.id/ejurnal/index.php/infortech96

Firmansyah, R., Utami, E., & Pramono, E. (2022). Evaluation of Naive Bayes, Random Forest and Stochastic Gradient Boosting Algorithm on DDoS Attack Detection. *1st International Conference on Science and Technology Innovation (ICoSTEC)*.

Fouedjio, F. (2020). Exact Conditioning of Regression Random Forest for Spatial Prediction. *Artificial Intelligence in Geosciences*, *1*, 11–23. https://doi.org/10.1016/j.aiig.2021.01.001

Harto, M. K., & Basuki, A. (2021). Deteksi Serangan DDoS Pada Jaringan Berbasis SDN Dengan Klasifikasi Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(4), 1329–1333. http://j-ptiik.ub.ac.id

Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*.

Herianto, dan. (2016). Pemilihan Fitur untuk Monitoring dan Klasifikasi Kondisi Pahat. In *Forum Teknik* (Vol. 37, Issue 1).

Ismawan, F. (2015). Hasil Ekstraksi Algoritma Principal Component Analysisi (PCA) untuk Pengenalan Wajah dengan Bahasa Pemrograman Java Eclips IDE. *Jurnal Sisfotek Global*, *5*(1), 26–30.

Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, *2*(1). https://doi.org/10.1186/s40165-014-0010-2

Khasanah, N., Komarudin, R., Afni, N., Maulana, Y. I., & Salim, A. (2021). Skin Cancer Classification Using Random Forest Algorithm. *Jurnal SISFOTENIKA*, *11*(2), 137–147. https://doi.org/10.30700/jst.v11i2.1122

Kurniawan, A., & Yulianingsih, Y. (2021). Pendugaan Fraud Detection pada kartu kredit dengan Machine Learning. *KILAT*, *10*(2), 320–325. https://doi.org/10.33322/kilat.v10i2.1482

Lin, H., Wu, C., & Masdari, M. (2022). A comprehensive survey of network traffic anomalies and DDoS attacks detection schemes using fuzzy techniques. *Computers and Electrical Engineering*, *104*, 108466. https://doi.org/10.1016/j.compeleceng.2022.108466

Marestiani, F., & Surono, S. (2022). Forecasting Using K-means Clustering and RNN Methods with PCA Feature Selection. *Journal of Data Science*, *4*. https://ipublishing.intimal.edu.my/jods.html

Meng, F., Tan, Y., & Bu, Y. (2021). Target Aggregation Regression based on Random Forests. *Procedia Computer Science*, *199*, 517–523. https://doi.org/10.1016/j.procs.2022.01.063

Mohd, Z. (2019, December 10). *Demystifying the Confusion Matrix Using a Business Example*. Toward Data Science.

Nehra, P., & Nagaraju, A. (2022). Host utilization prediction using hybrid kernel based support vector regression in cloud data centers. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 6481–6490. https://doi.org/10.1016/j.jksuci.2021.04.011

Prasetyo, B., & Trisyanti, U. (2018). Revolusi Industri 4.0 dan Tantangan Perubahan Sosial. *Prosiding SEMATEKSOS 3*, 22–27.

Rahul, R. (2020, June 16). *Random Forest Classification and it's Mathematical Implementation*. Medium.

Rais, A. N., & Subekti, A. (2019). Integrasi SMOTE dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing. *JURNAL INFORMATIKA*, *6*(2), 278–285. http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Santoso, B., Azminuddin, & Zohrahayaty. (2020). *Machine Learning & Reasoning Fuzzy Logic Algoritma, Manual, Matlab, & Rapid Minier*. Penerbit Deepublish.

Singh, N. A., Singh, J., & De, T. (2016). Distributed denial of service attack detection using naive bayes classifier through info gain feature selection. *ACM International Conference Proceeding Series*, *25-26-August-2016*. https://doi.org/10.1145/2980258.2980379

Suryana, N., Pratiwi, & Tri Prasetio, R. (2020). Penanganan Ketidakseimbangan Data pada Prediksi Customer Churn Menggunakan Kombinasi SMOTE dan Boosting. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(1), 31–37. https://creativecommons.org/licenses/by-sa/4.0/

Sutoyo, E., & Asri Fadlurrahman, M. (2020). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi Dan Penelitian Informatika*, *6*(3), 379–385.

University of New Brunswick. (2019). *DDoS Evaluation Dataset (CIC-DDoS2019)*.

Wahyuni, & Adytia, P. (2018). Perbandingan Algoritma Machine Learning Dalam Mendeteksi Serangan DDOS. *Jurnal Teknologi Informasi Komunikasi (e-Journal)*, *10*(1), 161–166. https://doi.org/10.38204/tematik.v9i2.1070

Widodo, E. (2020). Prediksi Penjurusan IPA, IPS dan BAHASA dengan Menggunakan Machine Learning. *Jurnal Pelita Teknologi*, *15*(1), 37–48.

Yasin, A., & Mohidin, I. (2018). Dampak Serangan DDoS pada Software Based Openfow Switch di Perangkat HG553. *Jurnal Technopreneur (JTech)*, *6*(2), 72. https://doi.org/10.30869/jtech.v6i2.206

Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion detection and Big Heterogeneous Data: a Survey. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-015-0013-4