



**ANALISIS PERBANDINGAN ALGORITMA *HIDDEN*
MARKOV MODEL DAN ALGORITMA *NAÏVE BAYES* DALAM
MEMPREDIKSI SERANGAN PADA *WEB SERVER***

SKRIPSI

HILMY BASKORO

1610511048

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021



**ANALISIS PERBANDINGAN ALGORITMA *HIDDEN*
MARKOV MODEL DAN ALGORITMA *NAÏVE BAYES* DALAM
MEMPREDIKSI SERANGAN PADA *WEB SERVER***

SKRIPSI

HILMY BASKORO

1610511048

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021

**ANALISIS PERBANDINGAN ALGORITMA *HIDDEN*
MARKOV MODEL DAN ALGORITMA *NAÏVE BAYES* DALAM
MEMPREDIKSI SERANGAN PADA *WEB SERVER***

Hilmy Baskoro

1610511048

Abstrak

Website merupakan suatu aplikasi yang mudah diakses dimanapun dan kapanpun, dalam kemudahan tersebut terdapat suatu serangan yang dapat dilakukan pada *website* yang tertuju langsung pada *web server*. Seorang *user* yang mengakses ke sebuah *server* akan terekam oleh *system* dan akan tercatat menjadi sebuah *datalog*, begitu pula jika seorang penyerang yang ingin melakukan *hacking* dengan menembus celah keamanan yang terdapat pada sebuah *website* akan terekam sebagai *log* karena telah melakukan akses ke *website* tersebut. *Datalog* tersebut akan menjadi *dataset* yang dapat diolah dalam penelitian ini untuk klasifikasi serangan dan bukan serangan pada sebuah *web server*. Tetapi tidak semua algoritma *data mining* memiliki kinerja yang baik dalam mengklasifikasi jenis serangan. Maka dari itu penelitian ini mencoba untuk membandingkan dua algoritma klasifikasi yang terdapat pada *data mining* dan menggunakan *dataset log server website* SIAKAD UPNVJ yang sudah diperoleh. Data tersebut akan dilakukan *training* dan *testing* untuk klasifikasi serangan dan bukan serangan serta dibandingkan dengan menggunakan dua algoritma klasifikasi *data mining* yaitu algoritma *Hidden Markov Model* dan algoritma *Naïve Bayes* sehingga mendapat suatu hasil berupa *akurasi*, *precision* dan *recall* yang terbaik diantara kedua algoritma tersebut.

Kata kunci: *log server, Naïve Bayes, Hidden Markov Model, data mining*

**COMPARATIVE ANALYSIS OF HIDDEN MARKOV MODEL
AND NAÏVE BAYES ALGORITHM IN PREDICTING ATTACKS
ON WEB SERVER**

Hilmy Baskoro

1610511048

Abstract

The website is an application that is easily accessible anywhere and anytime, in that convenience there is an attack that can be carried out on a website that is aimed directly at the web server. A user who accesses a server will be recorded by the system and will be recorded as a datalog, as well as if an attacker who wants to do hacking by penetrating a security hole in a website will be recorded as a log because he has accessed the website. The datalog will be a dataset that can be processed in this research for classification of attacks and not attacks on a web server. But not all data mining algorithms have good performance in classifying types of attacks. Therefore, this study tries to compare the two classification algorithms contained in data mining and use the SIAKAD UPNVJ website server log dataset that has been obtained. The data will be trained and tested for attack and non-attack classification and compared using two data mining classification algorithms, namely the Hidden Markov Model algorithm and the Naïve Bayes algorithm so as to get a result in the form of accuracy, precision and recall which is the best between the two algorithms.

Keywords: *log server, Naïve Bayes, Hidden Markov Model, data mining*

PERNYATAAN ORISINALITAS

Skripsi ini adalah karya sendiri dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Hilmy Baskoro

NIM : 1610511048

Bilamana kemudian hari ditemukan tidak kesusaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 12 Januari 2022



(Hilmy Baskoro)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “ Veteran ” Jakarta,
saya yang bertanda tangan dibawah ini:

Nama : Hilmy Baskoro

Nim : 1610511048

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “ Veteran “ Jakarta Hak Bebas Royalti Non eksklusif karya ilmiah saya yang berjudul :

**ANALISIS PERBANDINGAN ALGORITMA *HIDDEN MARKOV MODEL*
DAN ALGORITMA *NAÏVE BAYES* DALAM MEMPREDIKSI SEBUAH
SERANGAN PADA *WEB SERVER***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “ Veteran “ Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pengkalan data (*database*), merawat dan mempublikasi Skripsi saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat : Jakarta

Pada tanggal : 12 Januari
2022

Yang Menyatakan



(Hilmy Baskoro)

PENGESAHAN

PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Hilmy Baskoro
NIM : 1610511048
Program Studi : S1 Informatika
Judul Skripsi : Analisis Perbandingan Algoritma *Hidden Markov Model* Dan Algoritma *Naive Bayes* Dalam Memprediksi Serangan Pada *Web Server*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Yuni Widiastwi, S.Kom, M.Si

Penguji I

Mavanda Mega Santoni, S.Komp., M.Kom.

Penguji II

Henki Bavu Seta, S.Kom., M.TI

Pembimbing I

Ika Nurlaili Isnainivah, S.Kom., M.Sc

Pembimbing II



Dr. Ernatha, M.Kom.

Dekan

Yuni Widiastwi, S.Kom, M.Si

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Persetujuan : 12 Januari 2022



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala nikmat dan karunia-Nya, sehingga Skripsi ini berhasil diselesaikan. Penulis ingin mengucapkan terima kasih kepada:

1. Orang Tua tercinta penulis yaitu Hendry Rumantyo dan Laela Hidayati yang terus selalu mendoakan, memberikan dorongan dan nasihat yang terbaik agar dapat menyelesaikan skripsi ini.
2. Bapak Henki Bayu Seta, M.Kom., MTI selaku dosen pembimbing I Skripsi yang membantu memberikan saran yang bermanfaat.
3. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc selaku dosen pembimbing II Skripsi yang membantu memberikan pembelajaran dan saran yang bermanfaat.
4. Rico Andreas, Ismoko Bayu Permana, Fajar Rizkiadi, Annisa Ramadhalia Kusumanti, Rialdi Varga dan Anggit Marthinory yang memberikan *support* dan membantu menyelesaikan Skripsi ini.
5. Ibu, Bapak Dosen Informatika UPN Veteran Jakarta atas segala pembelajaran dan ilmu-ilmu yang bermanfaat semasa perkuliahan.
6. Alumni teman-teman Informatika 2016 yang selalu mendukung dalam menyelesaikan Skripsi ini.
7. Alumni dan para senior Informatika angkatan 2012, 2013, 2014 dan 2015 yang memberikan dukungan moril semasa perkuliahan.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 12 Januari 2021

Penulis,

Hilmy Baskoro

Daftar Isi

Abstrak	iii
Abstract	iv
1. PERNYATAAN ORISINALITAS	v
2. PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI	vi
PENGESAHAN	vii
KATA PENGANTAR	viii
Daftar Isi.....	ix
Daftar Tabel.....	xii
Daftar Gambar	xiv
Daftar Simbol	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
1.5 Batasan Masalah	2
1.6 Luaran yang Diharapkan.....	3
1.7 Sistematika Penulisan	3
BAB 2 TINJAUAN PUSTAKA	5
2.1 Keamanan Jaringan.....	5
2.2 <i>Website</i>	5
2.3 <i>Web Server</i>	5
2.4 <i>Log</i>	6
2.4.1 Error log	6
2.4.2 <i>Access log</i>	6
2.5 Metode Klasifikasi <i>Data Mining</i>	7
2.6 <i>Naïve Bayes</i>	7
2.7 <i>Confusion Matrix</i>	8
2.8 <i>Hidden Markov Model</i>	10

2.9	<i>Imbalanced Data</i>	12
2.10	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	13
2.11	<i>Scikit - Learn</i>	14
2.11.1	<i>Library Pandas</i>	14
2.12	Penelitian Terkait.....	15
BAB III METODOLOGI PENELITIAN.....		18
3.1	Tahapan Penelitian.....	18
3.2	Metode Penelitian	18
3.2.1	Identifikasi Masalah.....	18
3.2.2	Perumusan Masalah	19
3.2.3	Studi Literatur.....	19
3.2.4	Pengumpulan Data.....	19
3.2.5	Praproses Data.....	19
3.2.6	<i>Modeling System</i>	20
3.2.7	Hasil.....	20
3.2.8	Evaluasi.....	21
3.3	Alat Bantu Penelitian	21
3.4	Jadwal Penelitian	22
BAB 4 HASIL DAN PEMBAHASAN		24
4.1	Akuisisi Data	24
4.2	Praproses Data	26
4.2.1	Transformasi Data	26
4.2.2	Balanced Data Training menggunakan Metode Synthetic Minority Oversampling Technique (SMOTE).....	28
4.3	<i>Modeling System</i>	29
4.3.1	<i>Modeling Naïve Bayes</i>	31
4.3.1.1	<i>Naïve Bayes</i> Skenario 1 (<i>data training 70% dan data testing</i> 30%) 31	
4.3.1.2	<i>Naïve Bayes</i> Skenario 2 (<i>data training 80% dan data testing</i> 20%) 36	
4.3.1.3	<i>Naïve Bayes</i> Skenario 3 (<i>data training 90% dan data testing</i> 10%) 41	
4.4	Hasil.....	45
4.4.1	Hasil Skenario 1 (<i>Data Training 70% dan Data Testing 30%</i>)	46

4.4.2	Hasil Skenario 2 (<i>Data Training</i> 80% dan <i>Data Testing</i> 20%)	46
4.4.3	Hasil Skenario 3 (<i>Data Training</i> 90% - <i>Data Testing</i> 10%).....	47
4.5	Evaluasi.....	47
4.5.1	Perhitungan Akurasi, Presisi dan Recall.....	48
4.5.1.1	Perhitungan Akurasi, Presisi dan <i>Recall</i> Skenario 1	48
4.5.1.2	Perhitungan Akurasi, Presisi dan <i>Recall</i> Skenario 2	49
4.5.1.3	Perhitungan Akurasi, Presisi dan <i>Recall</i> Skenario 3	49
BAB 5 PENUTUP		53
5.1	Kesimpulan	53
5.2	Saran	54
DAFTAR PUSTAKA		55
RIWAYAT HIDUP		57
3.	LAMPIRAN	58

Daftar Tabel


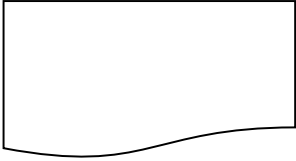
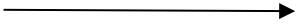

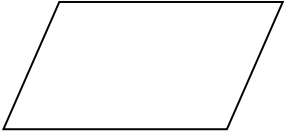
Tabel 2.1 Predicted Class Naïve Bayes	9
Table 3.1 Jadwal Penelitian.....	23
Tabel 4.1 Pembagian data dengan <i>hold out estimation</i>	24
Tabel 4.2 Fulldata webservlog_train_90%.csv	24
Tabel 4.3 Transofmasi Method	26
Tabel 4.4 Transformasi HTTP Status Code	27
Tabel 4.5 Transformasi Error.....	27
Tabel 4.6 Transformasi Class	28
Tabel 4.9 Model yang digunakan.....	29
Tabel 4.10 Train-Test Split	30
Tabel 4.11 Data training 70% setelah dilakukan transformasi dan oversampling	32
Tabel 4.12 Method Naïve Bayes Skenario 1	32
Tabel 4.13 <i>HTTP Status Code Naïve Bayes</i> Skenario 1	33
Tabel 4.14 <i>Error Status Code Naïve Bayes</i> Skenario 1	33
Tabel 4.15 <i>Prior Probability Class</i> Skenario 1	34
Tabel 4.16 Tabel Normalisasi Skenario 1	36
Tabel 4.17 Data training 80% setelah dilakukan transformasi dan <i>oversampling</i>	36
Tabel 4.18 Method Naïve Bayes Skenario 2.....	37
Tabel 4.19 <i>HTTP Status Code Naïve Bayes</i> Skenario 2	37
Tabel 4.20 <i>Error Status Code Naïve Bayes</i> Skenario 1	38
Tabel 4.21 <i>Prior Probability Class</i> Skenario 2.....	38
Tabel 4.22 Tabel Normalisasi Skenario 2	41
Tabel 4.23 <i>Data training</i> 90% setelah dilakukan transformasi dan <i>oversampling</i>	41
Tabel 4.24 <i>Method Naïve Bayes</i> Skenario 3	42

Tabel 4.25 <i>HTTP Status Code Naïve Bayes</i> Skenario 3	42
Tabel 4.26 <i>Error Status Code Naïve Bayes</i> Skenario 3	43
Tabel 4.27 <i>Prior Probability Class</i> Skenario 3	43
Tabel 4.28 Tabel Normalisasi Skenario 3	45
Tabel 4.29 <i>Confusion Matrix</i> Untuk <i>Train</i> 70% dan <i>Test</i> 30%	47
Tabel 4.30 <i>Confusion Matrix</i> Untuk <i>Train</i> 80% dan <i>Test</i> 20%	47
Tabel 4.31 <i>Confusion Matrix</i> Untuk <i>Train</i> 90% dan <i>Test</i> 10%	48
Tabel 4.32 Nilai akurasi, presisi dan <i>recall</i> tiap skenario	50
Tabel 4.33 Perbandingan Akurasi, Presisi, Recall tiap algoritma	51

Daftar Gambar

Gambar 2.1 Hidden Markov model.....	10
Gambar 2.2	12
Gambar 3.1 Tahapan penelitian.....	18
Gambar 4.1 hasil dari jupyter notebook untuk skenario 1	46
Gambar 4.2 hasil dari jupyter notebook untuk skenario 2	46
Gambar 4.3 hasil dari jupyter notebook untuk skenario 3	47
Gambar 4.4 Hasil evaluasi	50

Daftar Simbol

Simbol	Nama Simbol	Keterangan
	Simbol Proses	Menggambarkan Proses
	Simbol Dokumen	Dokumen yang dibutuhkan dalam proses sistem
	Simbol arah data atau arus data	Sebagai petunjuk arah data dan arus data pada proses
	Simbol Terminator	Simbol untuk permulaan atau akhir dari suatu kegiatan
	Simbol Data	Simbol sebagai masukan atau keluaran data untuk suatu proses