SISTEM PENDETEKSI SERANGAN BRUTEFORCE DENGAN INTRUSION DETECTION SYSTEM (IDS) MENGGUNAKAN METODE LSTM (LONG SHORT TERM MEMORY)

SKRIPSI

Oleh :

AGUNG AL HAFIZIN

09011281823040

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2022

HALAMAN PENGESAHAN

Sistem Pendeteksi Serangan Bruteforce Dengan *Intrusion Detection System* (IDS) Menggunakan Metode LSTM (*Long Short Term Memory*)

TUGAS AKHIR

Program Studi Sistem Komputer
Jenjang S1

Oleh

Agung Al Hafizin
09011281823040

Indralaya, Juni 2022

Mengetahui,

Ketua Jurusan Sistem Komputer	Pembimbing Tugas Akhir
Dr. Ir. H. Sukemi, M.T.	**Ahmad Heryanto, S.Kom., M.T.**
NIP. 196612032006041000	**NIP. 198701222015041002**

AUTHENTICATION PAGE

BRUTEFORCE ATTACK DETECTION SYSTEM WITH INTRUSION DETECTION SYSTEM (IDS) USING LSTM (LONG SHORT TERM MEMORY) METHOD

FINAL TASK

Submitted to Complete One of the Conditions
Obtaining Strata 1 Degree

By

Agung Al Hafizin
09011281823040

Indralaya, June 2022

Acknowledge,

Head of Computer System Department	Supervisor
Dr. Ir. H. Sukemi, M.T.	**Ahmad Heryanto, S.Kom., M.T.**
NIP. 196612032006041000	**NIP. 198701222015041002**

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada

Hari : Kamis

Tanggal : 16 Juni 2022

Tim Penguji :

1. **Ketua : Ahmad Zarkasi, M.T.**
2. **Seketaris : Adi Hermansyah, M.T**
3. **Penguji : Huda Ubaya, M.T**
4. **Pembimbing : Ahmad Heryanto, M.T**

Mengetahui,
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi M.T
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Agung Al Hafizin

Nim : 09011281823040

Judul : Sistem Pendeteksi Serangan Bruteforce dengan *Instrusion Detection System* (IDS) Menggunakan Metode LSTM (*Long Short Term Memory*)

Hasil Pengecekan Software Ithenticate/Turnitin : 17%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas sriwijaya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Indralaya, Juni 2022

Agung Al Hafizin
09011281823040

KATA PENGANTAR

Alhamdulillahirabbil'alamin. Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-nya sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir ini dengan judul **"Sistem Pendeteksi Serangan Bruteforce Dengan Intrusion Detection System (IDS) Menggunakan Metode LSTM (Long Short Term Memory)"**

Pada kesempatan ini penulis ingin mengucapkan terimkasih kepada beberapa pihak atas ide dan saran serta bantuanyya dalam menyelesaikan penulisan skripsi tugas akhir ini. Oleh kerena itu penulis ingin mengucapkan rasa syukur kepada Allah SWT dan terima kasih kepada yang terhormat :

1. Tuhan yang Maha Esa Allah SWT Sebagai wujud dari rasa syukur atas segala nikmat, karunianya serta hidayah-Nya. Sehingga saya dapat penyelesaikan skripsi ini.
2. Rasullah SAW junjungan seluruh umat yang membawa dunia dari kegelapan menuju dunia yang terang benderang.
3. Skripsi ini adalah hadiah terindah dari saya untuk kedua orang tua saya. Ibu dan ayah yang telah mengorban segalanya, berkerja keras dan selalu mendoakan saya sehingga dipermudahkan Allah SWT dalam menyelesaikan pendidikan Strata-1 saya di Program Studi Sistem Komputer Universitas Sriwijaya. Karena kalian berdua, hidup terasa begitu mudah dan penuh kebahagiaan. Terimakasih karena selalu menjaga saya dalam doa-doa ayah dan ibu serta selalu membiarkan saya mengejar impian saya apapun itu. Saya berjanji tidak akan membiarkan semua itu sia-sia, saya ingin melakukan yang terbaik untuk setiap kepercayaan yang diberikan. Saya akan tumbuh, untuk menjadi yang terbaik yang saya bisa. pencapaian ini adalah persembahan istimewa saya untuk ibu dan ayah.
4. Sahabat Lillah, sahabat surga Muhammad Al Varez yang selalu ada menemani saya pada saat saya down prustasi di kosan. Ryan, Eza dan Jumhadi atas bantuan kalian dalam perjalanan PP dari layo bukit bukit layo atas boncengannya Saya bahkan tidak dapat menjelaskan betapa bersyukurnya saya memiliki kalian dalam hidup saya. Suka dan duka dilewati bersama, yang

saling mengingatkan dalam kebaikan, yang selalu mendoakan saya, yang selalu membantu saya serta selalu memotivasi saya ketika saya dalam keadaan sulit. Kebaikan kalian tiada bandingnya. Kalian menjadi salah satu orang yang layak kupersembahkan bentuk perjuanganku ini.

5. Bapak Deris Stiawan M.T P.hd selaku Dosen pembimbing Akademik saya yang telah memberikan informasi dan motivasi terbaiknya untuk kebaikan serta kemajuan dalam menjalani masa-masa kuliah di Fakultas Ilmu Komputer.
6. Bapak Ahmad Heryano M.T selaku Dosen pembimbing Tugas Akhir saya yang telah memberikan kritik, saran, dan motivasi terbaiknya untuk kebaikan serta kemajuan dalam penyusunan skripsi ini.
7. Terimakasih kepada teman-teman di lab jarkom dan comnets selaku aslab tersebut kiki, alfat, rifky, josman, ajie, robi, mais, garinang dan rafi yang telah menemani saya dan memberikan support dan dorongan terhadap keberhasilan saya dalam menalani penulisan tugas akhir ini.
8. Teman-teman seperjuangan Sistem Komputer (SK) angkatan 2018 yang telah menjadi bagian dari kisah hidup saya dan banyak sekali pelajaran hidup yang telah saya dapat dari teman-teman SK 2018.
9. Teman-teman bedeng kades (Okta, Fathur, Fahmi, sepa, Kak Ahmad, Kak Defrian, Kak Tommy, Kak Adi, Kak Ogi, Kak Tri, Kak Taufik dan lain-lain) yang sangat baik dalam pertemanan, mengarahkan dalam istiqomah dalam beribadah, dan telah memberikan rasa nyaman mengekos di bedeng kades.
10. Semua pihak yang tidak dapat disebut satu persatu

SISTEM PENDETEKSI SERANGAN BRUTEFORCE DENGAN INTRUSION DETECTION SYSTEM (IDS) MENGGUNAKAN METODE LSTM (LONG SHORT TERM MEMORY)

AGUNG AL HAFIZIN (09011281823040)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : agungalhafizin1@gmail.com

ABSTRAK

Bruteforce adalah serangan peretas kata sandi terhadap sebuah sistem keamanan komputer Dengan mengirimkan kombinasi berulang kali berupa angka, huruf, dan symbol-symbol yang berbeda. *Bruteforce attack* sebenarnya adalah metode serangan lama dan juga cukup sederhana, tetapi jenis serangan ini mempunyai *Succes Rate* yang cukup tinggi dan dinilai sangat efektif. itulah mengapa serangan ini masih populer sampai saat ini dan banyak digunakan oleh para hacker untuk melakukan tindakan kriminalnya. metode yang digunakan pada penelitian ini adalah *Long Short Term Memory* (LSTM). Keuunggulan LSTM adalah dapat mempertahankan error yang terjadi ketika melakukan *backpropagation* sehingga tidak memungkinkan loss pada traning meningkat. Pada penelitian ini dilakukan pendeteksian untuk 2 kelas serangan *FTP* dan *SSH Bruteforce* dengan validasi hasil dari 10% sampai 90% data latih, terhadap parameter Jumlah *layer* dan *node*, aktifasi *tanh* dan *softmax*, *learning rate*, *batch size*, fungsi *optimizer*, dan *loss*. berdasarkan pengujian yang telah dilakukan didapatkanlah hasil terbaik dengan tingkat akurasi sebesar 99.9995, presisi 100%, spesifitas 99.9984%, presisi 99.9992%, dan F1-Score 99.9996%.

Kata Kunci : Deteksi, *Bruteforce, Intrusion Detection System, Long Short Term Memory.*

Mengetahui,

Ketua Jurusan Sistem Komputer	**Pembimbing Tugas Akhir**
Dr. Ir. H. Sukemi, M.T.	**Ahmad Heryanto, S. Kom, M.T**
NIP. 196612032006041000	**NIP. 198701222015041002**

BRUTEFORCE ATTACK DETECTION SYSTEM WITH INTRUSION DETECTION SYSTEM (IDS) USING LSTM (LONG SHORT TERM MEMORY) METHOD

AGUNG AL HAFIZIN (09011281823040)
Computer Engineering Department, Computer Science Faculty, Sriwijaya University
Email : agungalhafizin1@gmail.com

ABSTRACT

Bruteforce is a password cracking attack against a computer security system by sending repeated combinations of numbers, letters, and different symbols. Bruteforce attack is actually an old attack method and also quite simple, but this type of attack has a fairly high success rate and is considered very effective. that's why this attack is still popular today and is widely used by hackers to carry out their criminal acts. The method used in this research is Long Short Term Memory (LSTM). The advantage of LSTM is that it can maintain errors that occur when doing backpropagation so that it does not allow the loss of training to increase. In this study, detection for 2 classes of FTP and SSH Bruteforce attacks with validation results from 10% to 90% of training data, on the parameters of the number of layers and nodes, activation of tanh and softmax, learning rate, batch size, optimizer function, and loss. based on the tests that have been carried out, the best results were obtained with an accuracy rate of 99,995, 100% precision, 99,9984% specificity, 99,9992% precision, and 99,9996% F1-Score.

Keywords: Detection, Bruteforce, Intrusion Detection System, Long Short Term Memory.

Acknowledge,

Head of Computer System Department	**Supervisor**
Dr. Ir. H. Sukemi, M.T.	**Ahmad Heryanto, S.Kom., M.T.**
NIP. 196612032006041000	**NIP. 198701222015041002**

DAFTAR ISI

HALAMAN SAMPUL i

HALAMAN PENGESAHAN ii

HALAMAN PENGESAHAN iii

HALAMAN PERSETUJUAN iv

HALAMAN PERNYATAAN v

KATA PENGANTAR vi

ABSTRACK viii

ABSTRAK ix

DAFTAR ISI x

DAFTAR GAMBAR xiv

DAFTAR TABEL xviii

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan, masalah, ... 3

1.3 Batasan Masalah 4

1.4 .Tujuan .Penelitian 4

1.5 .Manfaat. penelitian 4

1.6 Metodologi Penelitian 4

1.7 Sistematika penulisan 5

BAB II TINJAUAN PUSTAKA 7

2.1. Pendahuluan 7

2.2. Bruteforce Attack 10

2.2.1 FTP SSH Bruteforce Attack 14

2. SSH Bruteforce Attack 15

2.2.1 Sistem Kerja Serangan brute force 16

2.2.3 Metode Serangan Brute Force 17
2. Simple brute force attacks 17
2.2.4 BruteForce Attack Tools 18
2.2.3 Proteksi serangan brute force 21
2.3 Dataset CSE-CIC-IDS 2018 24
2.4 Ekstraksi Dataset 27
2.4.1 CICFLowmater 27
2.5 *Feature Selection* 28
2.5.1 *Correlation-based Feature Selection* 28
2.6 Artificial Intelegence 29
2.6.1 Aplikasi *Artificial Intelegence* (AI) 30
2.6.2 AI dan Cybersecurity 31
2.7 Machine Learning (ML) 35
2.7.1 NN (*Neural Network*) 37
2.7.2 Deep Learning (DL) 38
2.7.3 Reccurent Neural Network (RNN) 40
2.7.4 Long Short Term Memory (LSTM) 41
2.8 Confusion Matrix 45
2.8.1. Akurasi 46
2.8.2 Sensitivitas 46
2.8.3 Spesifitas 47
2.8.4. Presisi 47
2.8.5. F1-Score 47
2.9. Evaluasi BACC dan MCC 48
2.10. python 48
BAB III METODOLOGI PENELITIAN 51
3.1 Pendahuluan 51
3.2 Desain Penelitian 51
3.3 Tahap Ideation 53
3.4 Kerangka Kerja metodologi Penelitian 54

3.5 Kebutuhan Perangkat Keras Dan dan Perangkat Lunak 55
3.6 Persiapan Dataset 55
3.7 Ekstraksi Data 56
3.8 Seleksi Fitur 58
3.9 Arsitektur LSTM 59
3.10 Validasi Hasil 60
3.11 Skenario pengujian terhadap metode LSTM 61
BAB VI HASIL DAN ANALISA 69
4.1 Pendahuluan 69
4.2 Hasil Ekstraksi Dataset 69
4.3 Proses pendeteksian serangan pada Jupyter 71
4.3.1 Input dataset CIC-IDS-2018 71
4.3.2 Pembagian dataset menjadi data *training* dan *testing* 72
4.3.3 Encoder data 74
4.3.3 Seleksi Fitur 74
4.3.4 Membalancekan data (SMOTE) 77
4.3.5 Normalisasi Data 79
4.3.8 Hyperparamater pada model LSTM 80
4.3.9 Testing Dataset 82
4.4. Validasi Hasil 83
4.4.1 Validasi hasil data *training* 10% dan data *testing* 90% 83
4.4.2 Validasi hasil data *training* 20% dan data *testing* 80% 86
4.4.3 Validasi hasil data *training* 30% dan data *testing* 70% 89
4.4.4 Validasi hasil data *training* 40% dan data *testing* 60% 92
4.4.5 Validasi hasil data *training* 50% dan data *testing* 50% 95
4.4.6 Validasi hasil data *training* 60% dan data *testing* 40% 98
4.4.7 Validasi hasil data *training* 70% dan data *testing* 30% 101
4.4.8 Validasi hasil data *training* 80% dan data *testing* 20% 104
4.4.9 Validasi hasil data *training* 90% dan data *testing* 10% 107
4.5 Korelasi Hasil Deteksi Terhadap Kelas Label 110

4.5.1 Hasil Deteksi Data *Training* 10% dan Data *Testing* 90%..... 110
4.5.2 Hasil Deteksi Data *Training* 20% dan Data *Testing* 80%..... 111
4.5.3 Hasil Deteksi Data *Training* 30% dan Data *Testing* 70%..... 112
4.5.4 Hasil Deteksi Data *Training* 40% dan Data *Testing* 60%..... 113
4.5.5 Hasil Deteksi Data *Training* 50% dan Data *Testing* 50%..... 114
4.5.6 Hasil Deteksi Data *Training* 60% dan Data *Testing* 40%..... 115
4.5.7 Hasil Deteksi Data *Training* 70% dan Data *Testing* 30%..... 116
4.5.8 Hasil Deteksi Data *Training* 80% dan Data *Testing* 20%..... 117
4.5.9 Hasil Deteksi Data *Training* 70% dan Data *Testing* 30%..... 118
4.6. Analisa Hasil Validasi Keseluruhan..... 119
4.7 Perbandingan Berdasarkan Penelitian Terkait 121
BAB V KESIMPULAN DAN SARAN..... 123
5.1. Kesimpulan 123
5.2. Saran..... 123
DAFTAR PUSTAKA 124

DAFTAR GAMBAR

Gambar 2.1 Contoh serangan FTP Bruteforce 12

Gambar 2.2 Arsitektur jaringan pada dataset CSE-CIC-IDS 2018 24

Table 2.2 Features Used in CIC-AWS Dataset [17] 25

Gambar 2.3 Komponen Kerja pada AI meliputi ML dan DL 29

Gambar 2.4 Underfitting and Overfitting[29] 36

Gambar 2.5 Arsitektur Neural Network 37

Gambar 2.6 arsitektur deep lerningm 39

Gambar 2.7 Arsitektur Reccurent Neural Network 40

Gambar 2.8 Arsitektur Unit LSTM [41] 41

Gambar 2.10 Confusion Matrix[28]. 45

Gambar 3.1 Kerangka Kerja Penelitian 52

Gambar 3.2 Tahap Ideation 53

Gambar 3.3 Kerangka Kerja Metodologi Penelitian 54

Gambar 3.4 Flowchart seleksi Fitur 59

Gambar 3.5 Arsitektur LSTM 60

Gambar 3.6 Flowchart deteksi LSTM 61

Gambar 4.1 Data PCAP pada Komputer(172.31.64.17) 70

Gambar 4.2 Hasil ekstraksi data 70

Gambar 4.3 Proses ekstaksi data 70

Gambar 4.4 Grafik dataset berdasarkan Label 71

Gambar 4.5 Input Dataset 72

Gambar 4.6 Command split data *training* dan *test* 73

Gambar 4.7 Mengencoder Dataset 74

Gambar 4.8 *Command* visualisasi data 75

Gambar 4.9 Grafik korelasi Heatmap..... 75

Gambar 4.10 Proses Seleksi Fitur 76

Gambar 4.11 Proses membalance dataset 79

Gambar 4.12 Command split data *training* dan *test*..... 80

Gambar 4.13 Command model Hyperparameter 81

Gambar 4.14 *training* Dataset 81

Gambar 4.16 Command *testing* Dataset..... 82

Gambar 4.17 Grafik Akurasi dan *Loss* Data *training* 10% dan data *testing* 90%83

Gambar 4.18 Nilai Confusion Matrik Data *training* 10% dan data *testing* 90% 84

Gambar 4.19 Grafik Kurva ROC data t*raining* 10% dan data *testing* 90% 85

Gambar 4.20 Grafik Kurva *Precision-Recall* data *training* 20% dan data *testing* 80%..... 85

Gambar 4.21 Grafik Akurasi dan *Loss* validasi *training* 20% dan *testing* 80% . 86

Gambar 4.22 Nilai Confusion Matrik Data *Training* 20% dan data *testing* 80% 87

Gambar 4.23 Grafik Kurva ROC data t*raining* 20% dan data *testing* 80% 88

Gambar 4.24 Grafik Kurva *Precision-Recall* data *training* 20% dan data *testing* 80%..... 88

Gambar 4.25 Grafik Akurasi dan *Loss* validasi data *training* 30% dan data *testing* 70%..... 89

Gambar 4.26 Nilai Confusion Matrik Data *Training* 30% dan data *testing* 70% 90

Tabel 4.6 Hasil Validasi Data *training* 30% dan data *testing* 70%..... 90

Gambar 4.27 Grafik Kurva ROC data t*raining* 30% dan data *testing* 70% 91

Gambar 4.28 Grafik Kurva *Precision-Recall* data *training* 30% dan data *testing* 70%..... 91

Tabel 4.7. Hasil Validasi BACC dan MCC Data *training* 30% dan *testing* 70% 91

Gambar 4.29 Grafik Akurasi dan *Loss* validasi data training 40% dan *testing* 60% 92

Gambar 4.30 Nilai Confusion Matrik Data *Training* 40% dan data *testing* 60% 93

Tabel 4.8 Hasil Validasi Data *training* 40% dan data *testing* 60%..... 93

Gambar 4.31 Grafik Kurva ROC data t*raining* 40% dan data *testing* 60%..... 94

Gambar 4.32 Grafik Kurva *Precision-Recall* data *training* 40% dan data *testing* 60%..... 94

Gambar 4.33 Grafik Akurasi dan *Loss* validasi data training 50% dan data *testing* 50%..... 95

Gambar 4.34 Nilai Confusion Matrik Data *Training* 50% dan data *testing* 50% 96

Tabel 4.10 Hasil Validasi Data *training* 50% dan data *testing* 50%..... 96

Gambar 4.35 Grafik Kurva ROC data t*raining* 50% dan data *testing* 50%..... 97

Gambar 4.36 Grafik Kurva *Precision-Recall* data *training* 50% dan data *testing* 50%..... 97

Gambar 4.37 Grafik Akurasi dan *Loss* validasi *training* 60% dan *testing* 40% . 98

Gambar 4.38 Nilai Confusion Matrik Data *training* 60% dan data *testing* 40% 99

Tabel 4.12 Hasil Validasi Data *training* 60% dan data *testing* 40%..... 99

Gambar 4.39 Grafik Kurva ROC data t*raining* 60% dan data *testing* 40%..... 100

Gambar 4.40 Grafik Kurva *Precision-Recall* data *training* 60% dan data *testing* 40%..... 100

Tabel 4.13. Hasil Validasi BACC dan MCC Data *Training* 60% dan *testing* 40%..... 100

Gambar 4.41 Grafik Akurasi dan *Loss* validasi data *training* 70% dan data *testing* 30%..... 101

Gambar 4.42 Nilai Confusion Matrik Data *Training* 70% dan data *testing* 30%..... 102

Tabel 4.14 Hasil Validasi Data *training* 70% dan data *testing* 30%..... 102

Gambar 4.43 Grafik Kurva ROC data t*raining* 70% dan data *testing* 30%..... 103

Gambar 4.44 Grafik Kurva *Precision-Recall* data *training* 70% dan data *testing* 30%..... 103

Gambar 4.45 Grafik Akurasi dan *Loss* validasi *training* 80% dan *testing* 20% 104

Gambar 4.46 Nilai Confusion Matrik Data *Training* 80% dan data *testing* 20%..... 105

Gambar 4.47 Grafik Kurva ROC data t*raining* 80% dan data *testing* 20% 106

Gambar 4.48 Grafik Kurva *Precision-Recall* data *training* 80% dan data *testing* 20% 106

Gambar 4.49 Grafik Akurasi dan *Loss* validasi data *training* 90% dan data *testing* 10% 107

Gambar 4.50 Nilai Confusion Matrik Data *Training* 90% dan data *testing* 10% 108

Gambar 4.51 Grafik Kurva ROC data t*raining* 90% dan data *testing* 10% 109

Gambar 4.52 Grafik Kurva *Precision-Recall* data *training* 20% dan data *testing* 80% 109

Gambar 4.53 Korelasi keseluruhan Data *Training* 10% dan data *testing* 90%. 110

Gambar 4.54 Korelasi *False Negatif* Data *Training* 10% dan data *testing* 90%110

Gambar 4.55 Korelasi *False Positif* Data *Training* 10% dan data *testing* 90% 111

Gambar 4.56 Korelasi keseluruhan Data *Training* 20% dan data *testing* 80%. 111

Gambar 4.57 Korelasi *False Positif* Data *Training* 60% dan data *testing* 40% 112

Gambar 4.58 Korelasi *False Negatif* Data *Training* 20% dan data *testing* 80%112

Gambar 4.73 Grafik Hasil Validasi 121

DAFTAR TABEL

Tabel 2.1 Penelitian terkait yang di jadikan rujukan 7

Table 2.2 Features Used in CIC-AWS Dataset 25

Table 2.3 Label dari dataset pada penelitian ini 27

Tabel 3.1 Spesifikasi Perangkat Keras 55

Tabel 3.2 Spesifikasi Perangkat Lunak 55

Tabel 3.3 Atribut Feature Extraction [17] 56

Tabel 3.5 hasil pengujian untuk nilai pada hiden layer 62

Tabel 3.6 hasil pengujian untuk nilai pada batchsize 62

Tabel 3.7 hasil pengujian menggunakan optimizer adadelta 63

Tabel 3.8 hasil pengujian menggunakan optimizer adam 63

Tabel 3.9 hasil pengujian menggunakan optimizer adagrad 64

Tabel 3.10 hasil pengujian menggunakan optimizer adamax 64

Tabel 3.11 hasil pengujian menggunakan optimizer ftrl 65

Tabel 3.12 hasil pengujian menggunakan optimizer nadam 65

Tabel 3.13 hasil pengujian menggunakan optimizer RMSprop 66

Tabel 3.14 hasil pengujian menggunakan optimizer SGD 66

Tabel 3.15 *HyperParameter* pada LSTM 67

Tabel 3.16 Pembagian data untuk proses deteksi 68

Tabel 4.1 Hasil Seleksi Fitur 76

Tabel 4.2 Hasil Validasi Data *training* 10% dan data *testing* 90% 84

Tabel 4.3 Hasil Validasi BACC dan MCC Data training 10% dan data *testing* 90% 85

Tabel 4.4 Hasil validasi data *training* 20% dan *testing* 80% 87

Tabel 4.5. Hasil Validasi BACC dan MCC Data *Training* 20% dan data *testing* 80% 88

Tabel 4.9 Hasil validasi BACC dan MCC Data *Training* 40% dan data *testing* 60% 94

Tabel 4.11. Hasil Validasi BACC dan MCC Data training 50% dan *testing* 50% 97

Tabel 4.15 Hasil Validasi BACC dan MCC Data *Training* 70% dan *testing* 30% 103

Tabel 4.16 Hasil Validasi Data *training* 80% dan data *testing* 20% 105

Tabel 4.17 Hasil Validasi BACC dan MCC Data *Training* 80% dan *testing* 20% 106

Tabel 4.18 Hasil Validasi Data *training* 90% dan data *testing* 10% 108

Tabel 4.19 Hasil Validasi BACC dan MCC Data *Training* 90% dan *testing* 10% 109

Tabel 4.20 Hasil Performa Validasi Keseluruhan 120

Tabel 4.21 Perbandingan dengan penelitian terkait 122

BAB I
PENDAHULUAN

1.1 Latar Belakang

"Brute force attack" merupakan metode yang digunakan untuk mendapatkan informasi pribadi seperti nama pengguna, kata sandi, frasa sandi, dan informasi lainya. Dengan berulang kali mengirimkan kombinasi berupa angka, huruf, dan symbol-symbol yang berbeda, penyerang pada akhirnya dapat menebaknya dengan benar dan mendapatkan akses ke data yang dilindungi tersebut. Target umum untuk serangan brute force adalah kata sandi, kunci enkripsi serta kunci API dan login SSH.

Serangan brute force sering disebut "brute force cracking", karena dalam hal ini serangan brute force mencoba memecahkan sistem kredensial yang menjaga data sensitif atau data apa pun yang berharga bagi penyerang, Pendekatan ini awalnya mengarah pada pemrograman komputer yang menggunakan kekuatan dari komputasi komputer itu sendiri. Misalnya, untuk menyelesaikan sebuah persamaan kuadrat seperti “$x^2 + 4x - 50 = 0$”, di mana nilai x adalah bilangan berupa integer, dengan menggunakan algoritma dari teknik Bruteforce, pengguna hanya perlu menulis program untuk memeriksa semua kemungkinan nilai bilangan bulat untuk persamaan nilai x agar mendapatkan jawaban yang benar. Istilah dari “bruteforce” itu sendiri diperkenalkan oleh Kenneth Thompson dengan mottonya: “When in doubt, use brute-force” (bila ragu, gunakan brute force)[1].

Serangan brute force menggunakan sekumpulan teks, karakter atau angka yang akan digunakan untuk menunjukkan kata sandi yang ingin Anda crack/hack. Himpunan dari simbol yang digunakan akan menjadi indikator efektifitas dari algoritma itu sendiri. Semakin Banyak kombinasi dari kumpulan karakter ini, semakin tinggi persentase peretasan kata sandi untuk kata sandi yang dapat dibobol. Tetapi jika semakin banyak koombinasi karakter yang dibuat maka semakin lama juga waktu pemrosesan yang dibebankan.

Serangan bruteforce biasanya digunakan untuk mengganggu akses kepada host (jaringan / server / workstation), membobol komputer, jaringan, atau sumber daya IT yang dilindungi kata sandi atau data terenkripsi, seperti DES, Triple DES, AES, Blowfish dll., Cracker menggunakan metode ini untuk mendapatkan akun korban secara illegal [2]

Seranngan bruteforce memiliki beberapa metode dalam penyerangannya salah satu metode yang sering digunakan adalah serangan kamus atau *Dictinory attack* , *Dictinory attack* menggunakan kumpulan kata dan juga frasa yang telah dipilih sebelumnya untuk menebak kata sandi yang kemungkinan dipakai oleh pengguna. contohnya bahwa pengguna cenderung memakai dari daftar dasar kata sandi, seperti "kata sandi", "qwerty" "123abc" dan "123456." [3].

Pada penelitian [4] melakukan pendeteksian Serangan *Bruterforce Attack* (pada protokol SSH) menggunakan pendekatan *Mechine Learning* dengan metode *discussion tree*, Dari penelitian ini dijelaskan bahwa penggunaan port pada set fitur saat melatih model menunjukkan performa akurasi yang sangat baik dalam mendeteksi serangan brute force SSH dengan peningkatan max dari 0,9933 menjadi 0,9999. Namun penggunaan fitur tersebut dapat mempengaruhi proses training data yang mengabaikan beberapa fitur bagus lainnya diabaikan saat membangun model tersebut. pengklasifikasian kemungkinan akan kesulitan memberi label pada instance tertentu.

Pada penelitian [5] mendeteksi Serangan SSH Brute Force attack dengan Menggunakan Data Netflow Agregat. pada penelitian ini menggunakan dataset realtime, Hasil percobaan dari penelitian tersebut didapatkan bahwa Fitur yang diekstraksi dari Netflows agregat dapat membedakan antara trafik SSH serangan brute force dan trafik SSH normal, terutama trafik login dari pengguna yang sah dapat berupa trafik serangan apabila melebihi batasan dalam melakukan login.

Pada penelitian [6] menyajikan sistem deteksi SSH and FTP brute-force Attacks dengan menggunakan pendekatan *Mechine Learning* dengan metode *Long Short Term Memory.* Penelitian tersebut menggunakan dataset CIC-IDS-2017. Berdasarkan analisis dan hasil penelitian yang telah dilakukan mengungkapkan

bahwa pendeteksian serangan bruteforce memberikan akurasi yang lebih tinggi dengan menggunakan fungsi aktivasi *tanh* dan menngunakan fungsi pengoptimal c*atagorical crossentropi* mendapatkan akurasi 99,88%.

Pada penelitian [7] Deteksi Serangan SSH-Brute Force Menggunakan Deep learning. Penelitian ini menggunakan dataset CIC-IDS 2018, Hasil percobaan dari penelitian tersebut menunjukkan bahwa CNN lebih unggul dari pada metode Deep Learning yang lain dibandingkan dengan hasil eksperimen dari 5 algoritma *mechine learning* yaitu Naive Bayes, Logistic Regression, Decision Tree, k-Nearest Neighbor, dan Support Vector Machine. dengan akurasi 94,3%, tingkat presisi 92,5%, tingkat recall 97,8% dan F1-score 91,8% dalam segi kemampuan mendeteksi serangan SSH-Brute force.

Berdasarkan berbagai penelitian terkait diatas maka pada penelitian ini akan menggunakan metode LSTM (*Long Short-Term Memory*) untuk mendeteksi serangan Bruteforce attack dangan menggunakan dataset CIC-IDS-2018, tujuan utama dalam menggunakan Motode LSTM yaitu untuk mendapatkan hasil prediksi yang terbaik dalam mendeteksi suatu serangan, hasil predeksi yang baik tergantung pada tingkat kesalahannya semakin kecil error dalam prediksi maka semakin tepat metode tersebut dalam prediksi.

1.2 Rumusan masalah

Di bawah ini adalah rumusan masalah yang akan dilakukan dalam mengimplementasikan skripsi ini, yaitu:

1. Bagaimana penerapan seleksi fitur untuk memperoleh fitur penting dalam deteksi serangan Bruteforce?
2. Bagaimana cara mendeteksi serangan Bruteforce dengan menerapkan metode LSTM ?
3. Bagaimana hasil kinerja deteksi LSTM mempengaruhi nilai akurasi, sensitivi tas, spesifisitas, presisi, F1-Score, BACC, dan MCC ?

1.3 Batasan Masalah

Di bawah ini adalah batasan masalah dalam melakukan skripsi ini yaitu :

1. Serangan yang digunakan pada penelitian ini adalah FTP dan SSH bruteforce
2. Metode yang dipakai untuk penelitian ini adalah LSTM (*Long Short Term Memory)*
3. Dataset yang digunakan CIC IDS 2018

1.4 Tujuan Penelitian

Di bawah ini adalah Tujuan dalam melakukan penelitian skripsi ini, yaitu :

1. Penerapan Pemilihan Fitur Berbasis Korelasi (CFS) untuk mendapatkan fungsionalitas penting dalam proses deteksi serangan Bruteforce.
2. Penerapan metode Long Short-Term Memory untuk mendeteksi serangan Bruteforce.
3. Mengukur hasil kinerja dalam hal presisi, spesifisitas, sensitivitas, presisi, F1-Score, BACC dan MCC.

1.5 Manfaat penelitian

Di bawah ini adalah manfaat dari penelitian skripsi ini sbagai berikut :

1. Optimalisasi waktu proses komputasi.
2. Dapat menerapkan metode long short-term memory untuk mendeteksi serangan Bruteforce.
3. Mendapatkan performa terbaik dari proses LSTM Pada skripsi ini.

1.6 Metodologi Penelitian

Di bawah ini merupakan tahapan metodeloogi dalam penulisan skripsi ini sebagai berikut:

1) Metode studi pustaka dan studi literatur

   Pada metode ini peneliti mencari informasi mengenai klasifikasi dan pendetesian serangan dengan menggunakan metode LSTM (*Long*

Short Term Memory) melalui beberapa materi pembelajran dari buku, internet, jurnal, dan artikel yang berhubungan dengan penusan skripsi ini.

2) Metode konsultasi

Dalam metode ini penulis melakukan konsultasi dengan pihak yang mempunyai pengetahuan dan pemahaman terhadap penulisan skripsi ini untuk mengatasi masalah-masalah yang sedang dihadapi.

3) Metode pengumpulan data

Pada metode ini peneliti mengumpulkan data terkait dengan serangan Bruteforce, IDS (Instrusion Detection System), dan pendeteksian serta pengklasifikasian serangan.

4) Metode pengujian

Dalam tahap ini akan dilakukan pembuatan rancangan sistem yang digunakan untuk mendapatkan hasil dari pendeteksian serangan Bruteforce dengan melatih model pembelajaran.

5) Metode analisis dan penarikan kesimpulan.

Setelah mendapatkan hasil dari tahap pengujian pada skripsi ini, selanjutnya hasil dari proses pendeteksian serangan tersebut akan dianalisa kemudian dibuat kesimpulan pada penelitian ini.

1.7 Sistematika penulisan

Berikut ini adalah sistematika penulisan pada penelitian skripsi ini meliputi:

BAB I PENDAHULUAN

Untuk Bab I pada penelitian ini terdiri dari latar belakang sejarah,rumusan masalah,tujuan penelitian, hasil penelitian,batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Untuk Bab II terdiri dari penjelasan teori-teori utama berupa penjelasan tentang bruteforce, LSTM *(Long short-term memory)*, dan teori-teori lain yang terkait dengan penelitian Skripsi.

BAB III. METODOLOGI PENELITIAN

Untuk Bab III terdiri dari tahapan-tahapan penelitian yang dilakukan berupa proses perancangan sistem pendeteksian serangan dan penerapan metode penelitian yang diganakan pada skripsi ini.

BAB IV. HASIL DAN ANALISIS PENELITIAN

Untuk Bab IV terdiri dari proses tahapan-tahapan penelitian serta analisis hasil pendeteksian serangan pada dataset dengan menggunakan metode LSTM *(Long short-term memory).*

BAB V. KESIMPULAN DAN SARAN

Pada tahap bab V ini peneneliti akan menarik beberapa kesimpulan yang didapatkan dari hasil penjelasan pada bab-bab sebelumnya dan memberikan saran yang nantinya akan digunakan untuk penelitian selanjutnya.

.

DAFTAR PUSTAKA

[1] K. E. Pramudita and S. Teknik, “Brute Force Attack dan Penerapannya pada Password Cracking,” 2011.

[2] - Syaifuddin, D. Risqiwati, and E. A. Irawan, “Realtime Pencegahan Serangan Brute Force dan DDOS Pada Ubuntu Server,” *Techno.Com*, vol. 17, no. 4, pp. 347–354, 2018, doi: 10.33633/tc.v17i4.1766.

[3] B. Danczul, S. Gradinger, B. Greslehner-nimmervoll, W. Kastl, and F. Wex, “Cuteforce Analyzer: A Distributed Bruteforce Attack on PDF Encryption with GPUs and FPGAs,” pp. 720–725, 2013, doi: 10.1109/ARES.2013.94.

[4] M. M. Najafabadi, T. M. Khoshgoftaar, C. Kemp, N. Seliya, and R. Zuech, “Machine learning for detecting brute force attacks at the network level,” *Proc. - IEEE 14th Int. Conf. Bioinforma. Bioeng. BIBE 2014*, pp. 379–385, 2014, doi: 10.1109/BIBE.2014.73.

[5] M. M. Najafabadi, T. M. Khoshgoftaar, C. Calvert, and C. Kemp, “Detection of SSH brute force attacks using aggregated netflow data,” *Proc. - 2015 IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015*, pp. 283–288, 2016, doi: 10.1109/ICMLA.2015.20.

[6] M. D. Hossain, H. Ochiai, F. Doudou, and Y. Kadobayashi, “SSH and FTP brute-force attacks detection in computer networks: Lstm and machine learning approaches,” *2020 5th Int. Conf. Comput. Commun. Syst. ICCCS 2020*, pp. 491–497, 2020, doi: 10.1109/ICCCS49078.2020.9118459.

[7] S. K. Wanjau and G. M. Wambugu, “SSH-Brute Force Attack Detection Model based on Deep Learning,” *Int. J. Comput. Appl. Technol. Res.*, vol. 10, no. 01, pp. 42–50, 2021, doi: 10.7753/IJCATR1001.1008.

[8] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, “Performance Comparison of Support Vector Machine , Random Forest , and Extreme Learning Machine for Intrusion Detection,” *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.

[9] X. Li, W. Chen, Q. Zhang, and L. Wu, “Computers & Security Building Auto-Encoder Intrusion Detection System based on random forest feature selection,” *Comput. Secur.*, vol. 95, p. 101851, 2020, doi: 10.1016/j.cose.2020.101851.

[10] Y. Novaria, S. Nurmaini, D. Stiawan, and B. Yudho, “Journal of Information Security and Applications Attack classification of an intrusion detection system using deep learning and hyperparameter optimization,” *J. Inf. Secur. Appl.*, vol. 58, no. March, p. 102804, 2021, doi: 10.1016/j.jisa.2021.102804.

[11] S. Al-emadi, A. Al-mohannadi, and F. Al-senaid, “Using Deep Learning Techniques for Network Intrusion Detection,” no. December 2019, 2020, doi: 10.1109/ICIoT48696.2020.9089524.

[12] I. Firat, F. Ertam, and A. Sengur, “Machine learning methods for cyber security intrusion detection : Datasets and comparative study,” *Comput. Networks*, vol. 188, no. December 2020, p. 107840, 2021, doi: 10.1016/j.comnet.2021.107840.

[13] M. U. Chowdhury, F. Hammond, G. Konowicz, C. Xin, H. Wu, and J. Li, “A Few-shot Deep Learning Approach for Improved Intrusion Detection,” pp. 456–462, 2017.

[14] W. Wang, X. Du, and N. A. Wang, “Building a Cloud IDS Using an Efficient Feature Selection Method and SVM,” *IEEE Access*, vol. 7, pp. 1345–1354, 2019, doi: 10.1109/ACCESS.2018.2883142.

[15] R. Hwang, M. Peng, P. Lin, V. Nguyen, and C. Huang, “An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection,” pp. 30387–30399, 2020.

[16] T. Lee, “Deep Learning enabled Intrusion Detection and Prevention System over SDN Networks,” pp. 2–7, 2020.

[17] Q. Zhou and D. Pezaros, “Evaluation of Machine Learning Classifiers for Zero-Day Intrusion Detection -- An Analysis on CIC-AWS-2018 dataset,” no. May 2019, 2019, [Online]. Available: http://arxiv.org/abs/1905.03685.

[18] S. Hosseini, B. Mohammad, and H. Zade, “New hybrid method for attack

detection using combination of evolutionary algorithms , SVM , and ANN," *Comput. Networks*, vol. 173, no. February 2019, p. 107168, 2020, doi: 10.1016/j.comnet.2020.107168.

[19] F. Meng, Y. Fu, F. Lou, and Z. Chen, "An effective network attack detection method based on kernel PCA and LSTM-RNN," *2017 Int. Conf. Comput. Syst. Electron. Control. ICCSEC 2017*, pp. 568–572, 2018, doi: 10.1109/ICCSEC.2017.8447022.

[20] S. H. Park and H. J. Park, "RNN-based Prediction for Network Intrusion Detection," *2020 Int. Conf. Artif. Intell. Inf. Commun.*, pp. 572–574, 2020.

[21] X. Zhang, "An Intrusion Detection System Based on Convolutional Neural Network for Imbalanced Network Traffic," pp. 456–460, 2019.

[22] H. S. Pratita and I. Sembiring, "Analisa Brute Force Attack menggunakan Scanning Aplikasi pada HTTP Attack Artikel Ilmiah," no. 672010194, 2016.

[23] A. Rahmad, "No Title," 2019.

[24] C. Applications, "A Proactive Framework for Capturing FTP Brute Force and Application Level Flood Attacks," no. June, 2012.

[25] J. K. Lee, S. J. Kim, C. Y. Park, T. Hong, and H. Chae, "Heavy-tailed distribution of the SSH Brute-force attack duration in a multi-user environment," *J. Korean Phys. Soc.*, vol. 69, no. 2, pp. 253–258, 2016, doi: 10.3938/jkps.69.253.

[26] "Apa Itu Brute Force? Apa Saja Metode yang Digunakan?" https://www.logique.co.id/blog/2020/02/12/apa-itu-brute-force/ (accessed Aug. 11, 2021).

[27] "Brute Force Attack: Pengertian, Tipe & Langkah Mencegahnya." https://www.goldenfast.net/blog/brute-force-adalah/ (accessed Aug. 11, 2021).

[28] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," no. Cic, pp. 108–116, 2018, doi: 10.5220/0006639801080116.

[29] M. Scholarworks, “Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks Advisor : James Deverick,” 2019.

[30] A. H. Lashkari, G. D. Gil, M. Saiful, I. Mamun, and A. A. Ghorbani, “Characterization of Tor Traffic using Time based Features,” no. Cic, pp. 253–262, 2017, doi: 10.5220/0006105602530262.

[31] S. Chormunge and S. Jena, “Correlation based feature selection with clustering for high dimensional data,” *J. Electr. Syst. Inf. Technol.*, vol. 5, no. 3, pp. 542–549, 2018, doi: 10.1016/j.jesit.2017.06.004.

[32] S. Moon and Y. Kim, “An improved forecast of precipitation type using correlation-based feature selection and multinomial logistic regression,” *Atmos. Res.*, vol. 240, no. February, p. 104928, 2020, doi: 10.1016/j.atmosres.2020.104928.

[33] S. Adi and A. Sunyoto, “The Effect of Feature Selection on Classification Algorithms in Credit Approval,” pp. 451–456, 2020.

[34] N. P. and K. M. Te-Shun Chou, Kang K. Yen, Jun Luo, “CORRELATION-BASED FEATURE SELECTION FOR INTRUSION DETECTION DESIGN Te-Shun Chou, Kang K. Yen, and Jun Luo,” *Mil. Commun. Conf.*, pp. 1–7, 2007.

[35] “What is Artificial Intelligence (AI)? | IBM.” https://www.ibm.com/cloud/learn/what-is-artificial-intelligence (accessed Sep. 16, 2021).

[36] A. L. Samuel, “Some Studies in Machine Learning,” *IBM J. Res. Dev.*, vol. 3, no. 3, pp. 210–229, 1959, [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5392560.

[37] R. Konieczny and R. Idczak, “Supervised machine learning: A review of classification techniques,” *Hyperfine Interact.*, vol. 237, no. 1, pp. 1–8, 2016, doi: 10.1007/s10751-016-1232-6.

[38] “What are Neural Networks? | IBM.” https://www.ibm.com/cloud/learn/neural-networks (accessed Sep. 16,

2021).

[39] “What is Deep Learning? | IBM.” https://www.ibm.com/cloud/learn/deep-learning (accessed Sep. 16, 2021).

[40] R. Zhao, Y. Yin, Y. Shi, and Z. Xue, “Jou rna lP pro of,” *Phys. Commun.*, p. 101157, 2020, doi: 10.1016/j.phycom.2020.101157.

[41] V. Nourani and N. Behfar, “Multi-station runoff-sediment modeling using seasonal LSTM models,” *J. Hydrol.*, vol. 601, no. April, p. 126672, 2021, doi: 10.1016/j.jhydrol.2021.126672.

[42] S. Ameur, A. Ben, and M. Salim, “A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion,” *Entertain. Comput.*, vol. 35, no. August 2019, p. 100373, 2020, doi: 10.1016/j.entcom.2020.100373.

[43] L. Frassineti, C. Barba, F. Melani, F. Piras, R. Guerrini, and C. Manfredi, “Automatic detection and soni fi cation of nonmotor generalized onset epileptic seizures : Preliminary results,” vol. 1721, no. June, 2019, doi: 10.1016/j.brainres.2019.146341.

[44] M. Kabir *et al.*, “Chemometrics and Intelligent Laboratory Systems Improving prediction of extracellular matrix proteins using evolutionary information via a grey system model and asymmetric under-sampling technique,” *Chemom. Intell. Lab. Syst.*, vol. 174, no. February, pp. 22–32, 2018, doi: 10.1016/j.chemolab.2018.01.004.

[45] M. Bach *et al.*, “PT US CR,” 2016, doi: 10.1016/j.ins.2016.09.038.

[46] D. Ding and S. Han, “Predictive biomarkers of colorectal cancer,” vol. 83, no. January, 2019, doi: 10.1016/j.compbiolchem.2019.107106.

[47] S. Su, Y. Sun, X. Gao, J. Qiu, and Z. Tian, “applied sciences A Correlation-Change Based Feature Selection Method for IoT Equipment Anomaly Detection,” 2019, doi: 10.3390/app9030437.