

DAFTAR ISI

	Halaman
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiv
INTISARI.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	4
1.5 Manfaat.....	4
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI.....	12
3.1. URL (<i>Uniform Resource Locator</i>).....	12
3.2. Analisa Leksikal	13
3.3. Penyamaran (<i>Obfuscating</i>) <i>Phishing Website</i>	19
3.4. Pendeteksian <i>Phishing website</i>	20
3.4.1. Algoritma <i>Deep Learning</i>	22
3.4.2. Algoritma <i>Machine Learning</i>	31
3.5. Matrik Evaluasi.....	40
3.5.1. Evaluasi Akurasi.....	40
3.5.2. Evaluasi Waktu.....	42
BAB IV ANALISA DAN PERANCANGAN SISTEM.....	44
4.1 Studi Literatur.....	44
4.2 Alat & Bahan.....	45
4.3 Alat	45
4.4 Pengumpulan Data.....	45
4.5 Kerangka Kerja.....	48

4.6	Analisa dan Perancangan Sistem.....	50
4.6.1	<i>Data Preprocessing</i>	50
4.6.2	Featuring.....	54
4.6.3	Pemodelan <i>Machine Learning</i>	59
4.6.4	Pemodelan <i>Deep Learning</i>	61
4.7	Skenario Pengujian.....	63
4.7.1	Pengujian Akurasi Deteksi Phising Website dengan ML vs DL.	63
4.7.2	Pengujian <i>Previous RNN vs Proposed RNN</i>	65
4.7.3	Penilaian Metrik	67
BAB V IMPLEMENTASI		70
5.1	Implementasi <i>Data Preprocessing</i>	70
5.2	Implementasi <i>Featuring</i>	72
5.2.1	Implementasi Ekstraksi	72
5.2.2	Implementasi EDA & analisa <i>Outlier</i>	75
5.2.3	Implementasi <i>Converting & Scaling</i>	78
5.2.4	Implementasi <i>Dimension Reduction</i> dengan PCA	80
5.3	Implementasi Model Machine Learning	83
5.3.1	Implementasi Model SVM	83
5.3.2	Implementasi Model RF	85
5.4	Implementasi Model <i>Deep Learning</i>	87
5.4.1	Implementasi Model GRU	87
5.4.2	Implementasi Model LSTM	94
BAB VI HASIL PENELITIAN DAN PEMBAHASAN		102
6.1	<i>Data Preprocessing</i>	102
6.2	<i>Data Featuring</i>	103
6.2.1	<i>Extracted dataset cleaned</i>	103
6.2.2	EDA dan analisa <i>outlier</i>	104
6.2.3	Analisa PCA	113

6.3	Hasil Pengujian Model	121
6.1.1	Hasil Pengujian Akurasi ML dan DL	121
6.1.2	Hasil Pengujian <i>Based RNN</i> dan <i>Proposed RNN</i>	123
6.1.3	Pengukuran Waktu ML dan DL	125
BAB VII KESIMPULAN DAN SARAN		128
7.1	Kesimpulan.....	128
7.2	Saran.....	128
DAFTAR PUSTAKA		130
LAMPIRAN.....		134

DAFTAR GAMBAR

Gambar 3. 1 URL dalam alfabet non-latin.....	13
Gambar 3. 2 Teknik <i>Phishing Website Detection</i>	21
Gambar 3. 3 Ilustrasi sel GRU	22
Gambar 3. 4 Ilustrasi sel LSTM.....	27
Gambar 3. 5 Ilustrasi <i>Random Forest(RF)</i>	32
Gambar 3. 6 Ilustrasi Model SVM.....	37
Gambar 4. 1 Metodologi Penelitian	44
Gambar 4. 2 Spesifikasi <i>Google Colabs</i>	45
Gambar 4. 3 Proses Persiapan <i>Raw Dataset</i>	46
Gambar 4. 4 Sampel <i>Raw Dataset Preprocessing</i>	47
Gambar 4. 5 Distribusi <i>Top 10 Raw dataset</i> berdasarkan <i>Domain</i>	48
Gambar 4. 6 Kerangka Kerja Penelitian	48
Gambar 4. 7 Proses <i>Preprocessing Dataset</i>	50
Gambar 4. 8 <i>Pseudocode</i> Fungsi Ekstrasi Fitur	52
Gambar 4. 9 Sampel Hasil ekstrasi <i>Raw Dataset</i>	53
Gambar 4. 10 <i>Pseudocode</i> Fungsi deteksi <i>outlier</i> fitur.....	54
Gambar 4. 11 Proses <i>Featured Dataset</i>	54
Gambar 4. 12 <i>Pseudocode</i> Fungsi <i>Cleansing</i>	55
Gambar 4. 13 Aktivitas <i>Cleansing</i>	56
Gambar 4. 14 Aktivitas <i>Converting</i> dan <i>Scaling</i>	56
Gambar 4. 15 Aktivitas PCA	57
Gambar 4. 16 <i>Pseudocode</i> Fungsi PCA.....	58
Gambar 4. 17 Ilustrasi Tokenisasi.....	59
Gambar 4. 18 Parameter & <i>Hyperparameter</i> Model SVM.....	59
Gambar 4. 19 Parameter & <i>Hyperparameter</i> Model RF.....	60
Gambar 4. 20 Model GRU yang akan digunakan.....	61
Gambar 4. 21 Model LSTM yang akan digunakan.....	62

Gambar 4. 22 Skenario 1 -Pengujian Akurasi PWD dengan ML vs DL	64
Gambar 4. 23 Skenario 2 -Pengujian waktu pada RNN.....	65
Gambar 5. 1 Arsitektur Aplikasi Antarmuka	70
Gambar 5. 2 Implementasi Fungsi <i>Cleansing</i>	71
Gambar 5. 3 Implementasi Fungsi Ekstraksi Fitur.....	73
Gambar 5. 4 Implementasi Fungsi EDA & Deteksi Outlier (1).....	75
Gambar 5. 5 Implementasi Fungsi EDA & Deteksi Outlier (2).....	76
Gambar 5. 5 Implementasi Fungsi <i>Converting & Scaling</i>	79
Gambar 5. 6 Implementasi Fungsi Dimension Reduction dengan PCA.....	81
Gambar 5. 8 Implementasi Fungsi Model SVM (1)	83
Gambar 5. 9 Implementasi Fungsi Model SVM (2)	84
Gambar 5. 10 Implementasi Fungsi Model RF (1)	85
Gambar 5. 11 Implementasi Fungsi Model RF (2)	86
Gambar 5. 12 Implementasi Fungsi Model GRU (1).....	87
Gambar 5. 13 Implementasi Fungsi Model GRU (2).....	88
Gambar 5. 14 Implementasi Fungsi Model GRU (3).....	89
Gambar 5. 15 Implementasi Fungsi Model GRU (4).....	90
Gambar 5. 16 Implementasi Fungsi Model LSTM (1)	94
Gambar 5. 17 Implementasi Fungsi Model LSTM (2)	95
Gambar 5. 18 Implementasi Fungsi Model LSTM (3)	96
Gambar 5. 19 Implementasi Fungsi Model LSTM (4)	97
Gambar 6. 1 <i>Raw dataset</i> hasil <i>preprocessing</i>	102
Gambar 6. 2 <i>Extracted Dataset</i>	103
Gambar 6. 3 Analisa fitur <i>Number of Dots in Domain</i>	104
Gambar 6. 4 Analisa fitur <i>URL Length</i>	105
Gambar 6. 5 Analisa fitur <i>Domain Length</i>	106
Gambar 6. 6 Analisa fitur <i>Filename Length</i>	106
Gambar 6. 7 Analisa fitur <i>Path URL Ratio</i>	107

Gambar 6. 8 Analisa fitur <i>Number of Dots in URL</i>	107
Gambar 6. 9 Analisa fitur <i>Digits in Query</i>	108
Gambar 6. 10 Analisa fitur <i>Longest path Token Length</i>	109
Gambar 6. 11 Analisa fitur <i>Symbol count in Domain</i>	110
Gambar 6. 12 Analisa fitur <i>Delimiter count in Domain</i>	110
Gambar 6. 13 Analisa fitur <i>Delimiter count in Path</i>	111
Gambar 6. 14 Analisa fitur <i>Entropy of Domain</i>	111
Gambar 6. 15 Matrik kovarian dari fitur	115
Gambar 6. 16 Nilai-Eigen dan Eigen-Vektor	117
Gambar 6. 17 Analisa <i>Variance (Elbow) Principal Component</i>	118
Gambar 6. 18 Pemetaan <i>Principal Component</i> terhadap Fitur	121
Gambar 6. 19 Grafik Tingkat Deteksi <i>Phishing</i> dan <i>Benign</i>	122
Gambar 6. 20 Detail Hasil Pengujian <i>Based RNN</i> vs <i>Proposed RNN</i>	123

DAFTAR TABEL

Tabel 3. 1 Tabel Karakterisasi URL	16
Tabel 3. 2 Tabel Ekstrasi Fitur	17
Tabel 4. 1 Tabel <i>Raw Dataset</i>	47
Tabel 4. 2 Tabel <i>Raw Dataset duplicated</i>	51
Tabel 4. 3 Tabel Fitur Ekstraksi	52
Tabel 4. 4 Paramater Model pada Skenario 1	64
Tabel 4. 5 Paramater Model pada Skenario 2 – <i>Input Dimension</i>	65
Tabel 4. 6 Parameter <i>output</i> model pada Skenario 2 – <i>Input Dimension</i>	66
Tabel 4. 7 Paramater Model pada Skenario 2 – <i>Learning Rate</i>	66
Tabel 4. 8 Parameter Output Model pada Skenario 2 – <i>Learning Rate</i>	67
Tabel 4. 9 Parameter Model Rujukan (<i>Based Model</i>)	67
Tabel 6. 1 Hasil Pengujian Skenario 1– Akurasi PWD dengan <i>ML vs DL</i> ...	122
Tabel 6. 2 Hasil Pengujian Skenario 2– <i>Input Dimension</i>	123
Tabel 6. 3 Hasil Pengujian Skenario 2 – <i>Learning Rate</i>	124
Tabel 6. 4 Ringkasan Hasil Pengujian Skenario 2 – <i>Input Dimension</i>	124
Tabel 6. 5 Ringkasan Hasil Pengujian Skenario 2 – <i>Learning Rate</i>	125
Tabel 6. 6 Ringkasan Hasil Pengujian Skenario 2 – Rata-rata	125