

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga laporan skripsi "ANALISIS SENTIMEN *TWITTER* TERHADAP *US ELECTION 2020* DENGAN METODE *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE (SVM)*" dapat diselesaikan dengan baik dan tepat waktu.

Skripsi ini disusun untuk memenuhi persyaratan bagi mahasiswa yang wajib ditempuh sesuai dengan kurikulum Program Studi Matematika Fakultas Sains dan Teknologi, Universitas Pelita Harapan, Tangerang.

Laporan skripsi ini diselesaikan dengan dukungan banyak pihak. Oleh karena itu, saya mengucapkan banyak terima kasih kepada:

1. Bapak Eric Jobiliong, Ph.D., selaku Dekan Fakultas Sains dan Teknologi
2. Ibu Dr. Nuri Arum Anugrahati, selaku Wakil Dekan Fakultas Sains dan Teknologi
3. Bapak Laurence, M.T., selaku Direktur Administrasi dan Kemahasiswaan Fakultas Sains dan Teknologi
4. Bapak Kie Van Ivanky Saputra, Ph.D., selaku Ketua Program Studi Matematika dan pembimbing skripsi yang telah memberikan bimbingan, arahan, dan dukungan dalam pengerjaan laporan skripsi.
5. Bapak Dion Krisnadi, S.Si., S.Inf., M.Kom., M.Act.Sci., selaku co-pembimbing yang telah memberikan saran-saran dan dukungan kepada saya dalam pengerjaan laporan skripsi.
6. Bapak Ukur Arianto Sembiring, S.Si., M.Si., selaku pembimbing akademik yang telah memberikan arahan selama perkuliahan.
7. Para dosen lainnya yang telah mendidik dan memberi dukungan moral selama perkuliahan.
8. Keluarga yang telah memberikan banyak dukungan
9. Vincent, yang memberikan semangat dan menjadi *support system* dalam penulisan skripsi.

10. Fatin Difa, yang telah membantu dalam proses pengolahan data.
11. Steven William, yang telah membantu dalam proses pembentukan PDF skripsi.
12. Teman-teman *Chaebol*, yang telah memberikan semangat bagi penulis untuk menyelesaikan skripsi.
13. Teman-teman *Yhabetedhe*, yang telah memberikan semangat dan dukungan selama masa perkuliahan.
14. Mahasiswa-mahasiswa Program Studi Matematika yang telah menemani penulis selama masa perkuliahan.
15. Semua pihak lainnya yang tidak dapat disebutkan satu per satu.

Akhir kata, penulis menyadari bahwa laporan skripsi ini masih sangat jauh dari kata sempurna. Oleh karena itu, penulis sangat terbuka akan kritik dan saran dari pembaca yang dapat membantu membuat laporan skripsi ini menjadi lebih baik lagi. Semoga laporan ini dapat bermanfaat bagi para pembacanya.

Tangerang, 28 Mei 2021

(Brigitta Monica Wiyono)

DAFTAR ISI

	halaman
HALAMAN JUDUL	i
PERNYATAAN DAN PERSETUJUAN UNGGAH TUGAS AKHIR	ii
PERSETUJUAN DOSEN PEMBIMBING SKRIPSI	iii
PERSETUJUAN TIM PENGUJI TUGAS AKHIR	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	3
1.5.1 Manfaat Teoretis	4
1.5.2 Manfaat Praktis	4
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	
2.1 Konsep Dasar Teori Peluang	5
2.1.1 Kejadian Saling Bebas (<i>Independent Events</i>)	5
2.1.2 Kejadian Saling Lepas (<i>Disjoint Events</i>)	5
2.1.3 Peluang Bersyarat (<i>Conditional Probability</i>)	6
2.1.4 <i>Bayes' Theorem</i>	6
2.2 Pengantar Aljabar Linier	7
2.2.1 Definisi Vektor	7
2.2.2 Konsep Panjang Vektor (<i>Norm of a Vector</i>)	8
2.2.3 Hasil Kali Dalam (<i>Dot Product</i>)	9
2.3 Optimisasi Fungsi	9
2.3.1 Metode <i>Lagrange Multipliers</i>	10
2.4 Analisis Prediksi	10
2.4.1 Masalah pada <i>Supervised Learning</i>	11
2.4.2 Evaluasi Model Klasifikasi	11
2.5 Model <i>Naive Bayes</i>	12
2.5.1 <i>Posterior Probabilities</i>	13
2.5.2 <i>Class-conditional Probabilities</i>	13
2.5.3 <i>Prior Probabilities</i>	14

2.5.4	<i>Evidence</i>	14
2.5.5	Kasus <i>Smoothing</i>	15
2.6	Model <i>Support Vector Machine</i> (SVM)	15
2.6.1	<i>Decision Rule</i>	15
2.6.2	Pengadaan <i>Constraint</i>	16
2.6.3	Penghitungan Margin	17
2.6.4	<i>Masalah Optimasi</i>	19
2.6.5	<i>Polynomial Kernel</i>	20
2.7	Tinjauan Pustaka	21
BAB III METODE PENELITIAN		
3.1	Pengumpulan Data	24
3.1.1	<i>US Election 2020</i>	24
3.1.2	Proses Pengumpulan Data	26
3.2	Pelabelan Data	27
3.3	<i>Pre-processing Data</i>	28
3.3.1	Algoritma TF-IDF	29
3.4	<i>Dataset Splitting</i>	30
3.5	<i>K-Fold Cross-Validation</i>	30
3.6	Pengujian Model dengan Keseluruhan Data	31
3.7	Analisis Hasil dan Perbandingan Akurasi <i>Naive Bayes</i> dan SVM	31
BAB IV ANALISIS DAN PEMBAHASAN		
4.1	Data Sentimen <i>Twitter</i>	32
4.2	Hasil Penerapan <i>Support Vector Machine</i> (SVM)	38
4.2.1	Parameter <i>C</i> Terbaik untuk Model SVM	38
4.2.2	Pemilihan Metode Kernel	42
4.2.3	Akurasi pada Model SVM	43
4.3	Hasil Penerapan <i>Naive Bayes</i>	44
4.3.1	Akurasi pada Model <i>Naive Bayes</i>	45
4.4	Prediksi Hasil pada SVM dan <i>Naive Bayes</i>	46
4.5	Pengujian Model menggunakan <i>Variance Inflation Factor</i> (VIF)	50
4.6	Prediksi Hasil Eleksi pada Sentimen dan Penghitungan Asli	52
BAB V KESIMPULAN DAN SARAN		
5.1	Kesimpulan	53
5.2	Saran	54

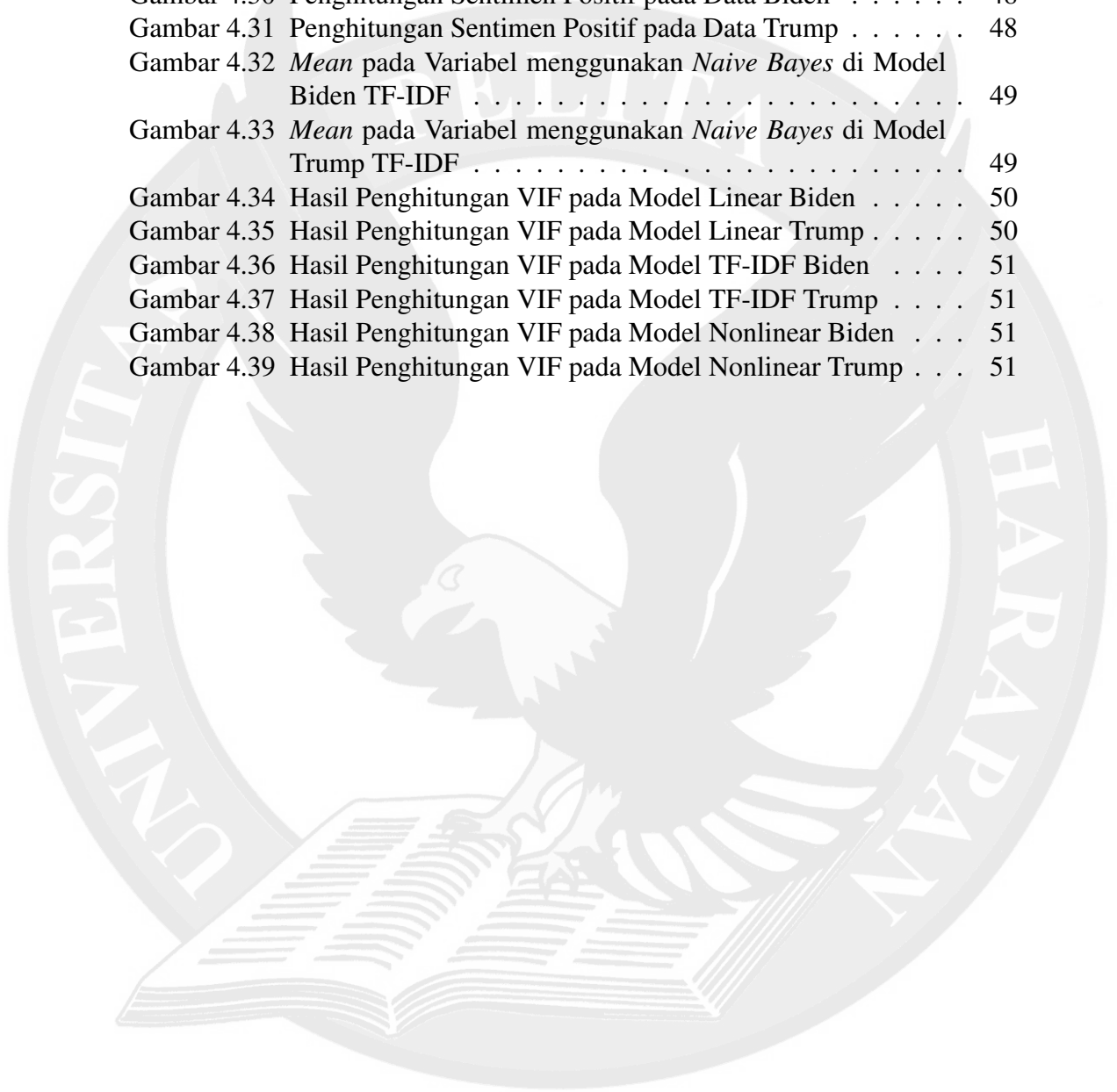
DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

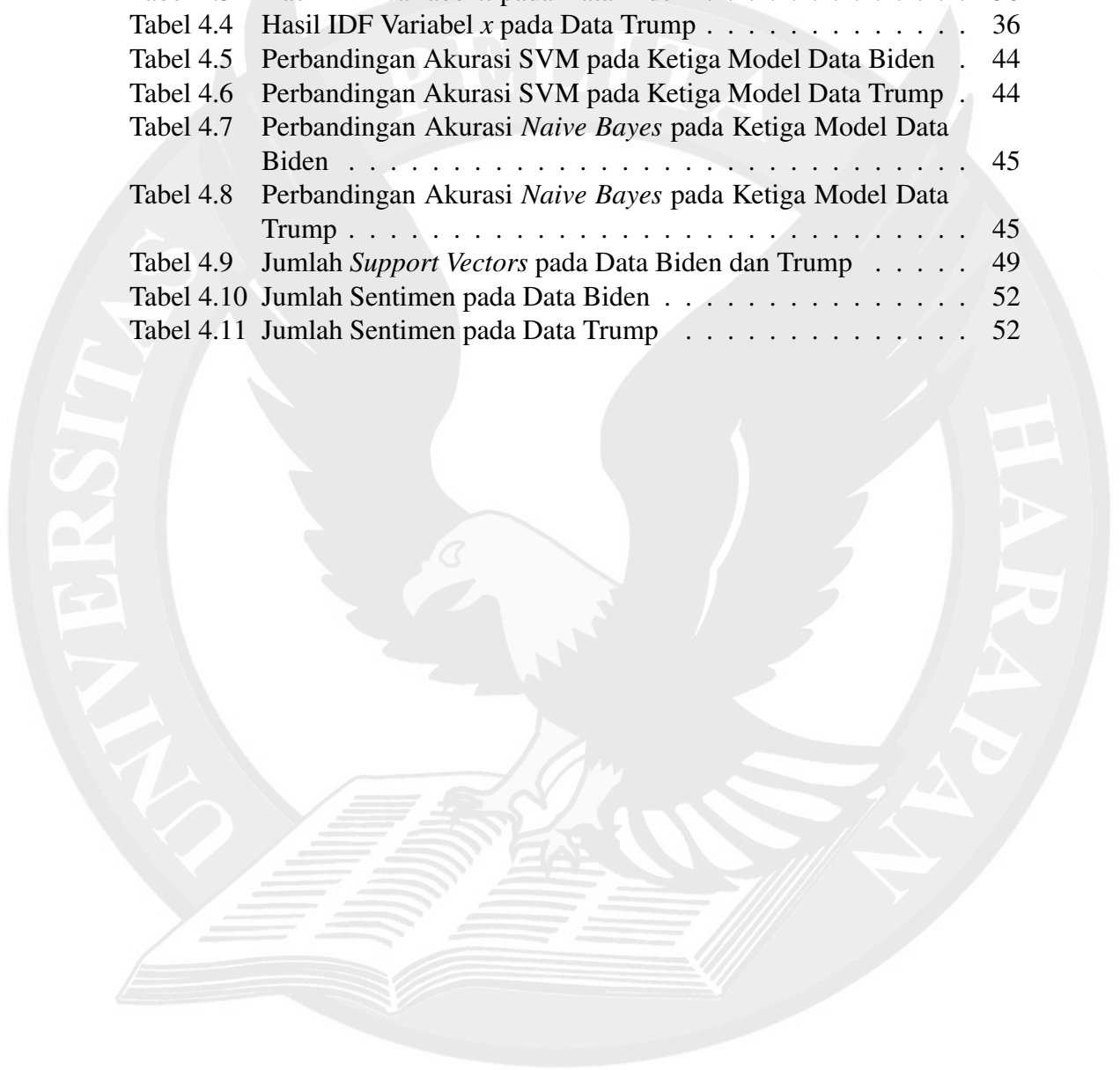
	halaman
Gambar 1.1 Persentase Pengguna <i>Twitter</i> berdasarkan Usia di Amerika Serikat tahun 2019	1
Gambar 2.1 Kejadian Saling Lepas	6
Gambar 2.2 Vektor \vec{AB}	7
Gambar 2.3 Vektor <i>Equivalent</i>	7
Gambar 2.4 Jarak Antar Vektor	8
Gambar 2.5 Sudut θ antara \mathbf{u} dan \mathbf{v} yang memenuhi $0 \leq \theta \leq \pi$	9
Gambar 2.6 <i>Linear problems</i>	12
Gambar 2.7 Grafik <i>hyper-plane</i>	16
Gambar 2.8 Klasifikasi dua kelas menggunakan SVM	18
Gambar 3.1 <i>Flowchart</i> Langkah-Langkah Pengerjaan	23
Gambar 3.2 Proses Pemilihan Presiden di Amerika Serikat	24
Gambar 3.3 Jumlah anggota <i>electoral college</i> di setiap negara bagian di Amerika Serikat	26
Gambar 3.4 Tahapan Pengumpulan Data dengan <i>Rapid Miner</i>	27
Gambar 3.5 Tahapan Pelabelan Data pada <i>Rapid Miner</i>	28
Gambar 4.1 Data dengan kata kunci Biden	32
Gambar 4.2 Data dengan kata kunci Trump	32
Gambar 4.3 <i>Pre-processing</i> Data Biden	33
Gambar 4.4 <i>Pre-processing</i> Data Trump	33
Gambar 4.5 Dua Kalimat yang mengandung kata <i>Trump</i> pada Model TF-IDF Trump	34
Gambar 4.6 Model Linear TF Biden	37
Gambar 4.7 Model Linear TF Trump	37
Gambar 4.8 Model Nonlinear TF Biden	37
Gambar 4.9 Model Nonlinear TF Trump	37
Gambar 4.10 Model Linear TF-IDF Biden	37
Gambar 4.11 Model Linear TF-IDF Trump	38
Gambar 4.12 Parameter C Terbaik pada Data Linear TF Biden	39
Gambar 4.13 Parameter C Terbaik pada Data Linear TF Trump	39
Gambar 4.14 Parameter C Terbaik pada Data Nonlinear TF Biden	40
Gambar 4.15 Parameter C Terbaik pada Data Nonlinear TF Trump	40
Gambar 4.16 Parameter C Terbaik pada Data TF-IDF Biden	41
Gambar 4.17 Parameter C Terbaik pada Data TF-IDF Trump	41
Gambar 4.18 <i>Radial Kernel</i> pada Model TF-IDF Biden	42
Gambar 4.19 <i>Radial Kernel</i> pada Model TF-IDF Trump	42
Gambar 4.20 <i>Linear Kernel</i> pada Model TF-IDF Biden	42
Gambar 4.21 <i>Linear Kernel</i> pada Model TF-IDF Trump	42

Gambar 4.22	<i>Sigmoid Kernel</i> pada Model TF-IDF Biden	43
Gambar 4.23	<i>Sigmoid Kernel</i> pada Model TF-IDF Trump	43
Gambar 4.24	<i>Polynomial Kernel</i> pada Model TF-IDF Biden	43
Gambar 4.25	<i>Polynomial Kernel</i> pada Model TF-IDF Trump	43
Gambar 4.26	Perbandingan Sentimen pada Variabel Data Biden	46
Gambar 4.27	Perbandingan Sentimen pada Variabel Data Trump	47
Gambar 4.28	Plot pada Data Biden dengan Variabel x_1 dan x_3	47
Gambar 4.29	Plot pada Data Trump dengan Variabel x_1 dan x_3	47
Gambar 4.30	Penghitungan Sentimen Positif pada Data Biden	48
Gambar 4.31	Penghitungan Sentimen Positif pada Data Trump	48
Gambar 4.32	<i>Mean</i> pada Variabel menggunakan <i>Naive Bayes</i> di Model Biden TF-IDF	49
Gambar 4.33	<i>Mean</i> pada Variabel menggunakan <i>Naive Bayes</i> di Model Trump TF-IDF	49
Gambar 4.34	Hasil Penghitungan VIF pada Model Linear Biden	50
Gambar 4.35	Hasil Penghitungan VIF pada Model Linear Trump	50
Gambar 4.36	Hasil Penghitungan VIF pada Model TF-IDF Biden	51
Gambar 4.37	Hasil Penghitungan VIF pada Model TF-IDF Trump	51
Gambar 4.38	Hasil Penghitungan VIF pada Model Nonlinear Biden	51
Gambar 4.39	Hasil Penghitungan VIF pada Model Nonlinear Trump	51



DAFTAR TABEL

	halaman
Tabel 2.1 Tabel <i>Confusion Matrix</i>	11
Tabel 2.2 Rangkuman Tinjauan Pustaka	22
Tabel 4.1 Variabel x pada Data Biden	34
Tabel 4.2 Variabel x pada Data Trump	34
Tabel 4.3 Hasil IDF Variabel x pada Data Biden	36
Tabel 4.4 Hasil IDF Variabel x pada Data Trump	36
Tabel 4.5 Perbandingan Akurasi SVM pada Ketiga Model Data Biden	44
Tabel 4.6 Perbandingan Akurasi SVM pada Ketiga Model Data Trump	44
Tabel 4.7 Perbandingan Akurasi <i>Naive Bayes</i> pada Ketiga Model Data Biden	45
Tabel 4.8 Perbandingan Akurasi <i>Naive Bayes</i> pada Ketiga Model Data Trump	45
Tabel 4.9 Jumlah <i>Support Vectors</i> pada Data Biden dan Trump	49
Tabel 4.10 Jumlah Sentimen pada Data Biden	52
Tabel 4.11 Jumlah Sentimen pada Data Trump	52



DAFTAR LAMPIRAN

	halaman
Lampiran A Hasil Penghitungan Akurasi	
A.1 <i>Confussion Matrix</i> dan Akurasi dengan Metode <i>Naive Bayes</i> pada <i>Test Data</i> Trump	A-1
A.2 <i>Confussion Matrix</i> dan Akurasi dengan Metode <i>Naive Bayes</i> pada <i>Test Data</i> Biden	A-2
A.3 <i>Confussion Matrix</i> dan Akurasi dengan Metode <i>Support Vector Machine</i> (SVM) pada <i>Test Data</i> Trump	A-3
A.4 <i>Confussion Matrix</i> dan Akurasi dengan Metode <i>Support Vector Machine</i> (SVM) pada <i>Test Data</i> Biden	A-4
Lampiran B Hasil <i>Mean</i> pada <i>Naive Bayes</i>	
B.1 <i>Mean</i> pada Data Trump	B-1
B.2 <i>Mean</i> pada Data Biden	B-2
Lampiran C <i>R-Code</i>	

