

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa telah memberikan hikmat dan kekuatan kepada penulis untuk menyelesaikan skripsi ini dengan judul “PENGEMBANGAN SISTEM PAKAR MENGGUNAKAN METODE KLASIFIKASI *DATA MINING* PADA DATASET GEJALA PENYAKIT”. Skripsi ini ditulis sebagai salah satu persyaratan yang harus dipenuhi dalam menyelesaikan masa studi dan memperoleh gelar Sarjana Sistem Informasi Strata Satu Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pelita Harapan.

Penulis menyadari bahwa tanpa dukungan, doa, dan bantuan dari pihak – pihak lain, skripsi ini tidak akan dapat selesai tepat pada waktunya. Oleh sebab itu penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu proses penulisan skripsi ini. Penulis berterima kasih kepada:

- 1) Bapak Dr.Eng., Ir. Pujianto Yugopuspito, MSc., selaku Dekan Fakultas Ilmu Komputer.
- 2) Bapak Arnold Aribowo, S.T., M.T., selaku Ketua Program Studi Sistem Informasi.
- 3) Ibu Calandra Alencia Haryani, S.SI., S.E., M.T.I., sebagai dosen pembimbing penulis yang selalu memberikan arahan, saran, kritik, dan nasehat selama proses penulisan skripsi ini.
- 4) Seluruh dosen dan staff Fakultas Ilmu Komputer yang telah memberikan arahan dan ajaran selama masa pembelajaran penulis pada Program Studi Sistem Informasi, Universitas Pelita Harapan.
- 5) Orang tua dan adik – adik yang telah memberikan semangat dan doa dalam menyelesaikan skripsi ini.
- 6) Andy, Wivina, Amelia, Katrin, Lauren dan teman – teman Sistem Informasi angkatan 2017 lain yang telah memberikan semangat, doa, dan motivasi selama penulisan skripsi.

- 7) Teman – teman dari Himpunan Mahasiswa Program Studi Sistem Informasi periode 2019/2020 dan periode 2020/2021 yang mendoakan dan memberi semangat kepada penulis.
- 8) Mardhianto, Nicholas, Kevin, Dheo, dan Mario selaku teman SMA penulis yang tidak henti memberikan semangat dan dukungan kepada penulis.
- 9) Pihak – pihak lain yang tidak dapat disebutkan satu per satu yang membantu penulis secara langsung maupun tidak langsung.

Penulis menyadari bahwa skripsi ini masih penuh akan kekurangan. Oleh karena itu, penulis meminta maaf bila ada kekurangan yang kurang berkenan pada skripsi ini. Selain itu penulis juga menerima saran dan kritik yang dapat berguna bagi penulis untuk pengembangan dan pembuatan program selanjutnya. Semoga skripsi ini dapat bermanfaat bagi semua pihak yang membacanya.

Karawaci, 8 Januari 2021

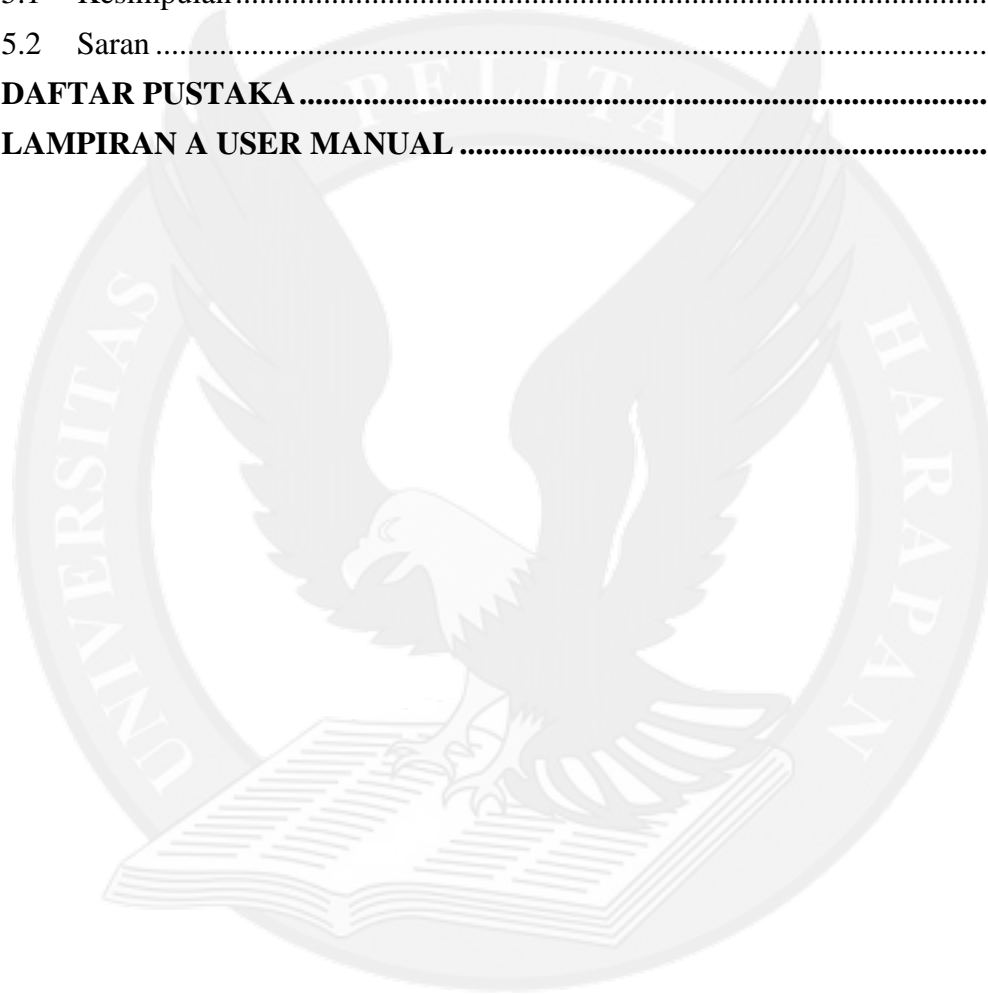
Penulis

DAFTAR ISI

Pernyataan dan Persetujuan Unggah Tugas Akhir	ii
PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR	iii
PERSETUJUAN TIM PENGUJI TUGAS AKHIR	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xv
DAFTAR RUMUS	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	2
1.3 Tujuan dan Manfaat Penelitian	2
1.3.1 Tujuan Penelitian	2
1.3.2 Manfaat Penelitian	3
1.4 Ruang Lingkup Penelitian	3
1.5 Sistematika Penulisan	3
BAB II LANDASAN TEORI	5
2.1 Data Mining	5
2.2 Klasifikasi (Classification)	6
2.2.1 Support Vector Machine	7
2.2.2 Logistic Regression.....	9
2.2.3 Naïve Bayes	10
2.3 Evaluasi Algoritma	11
2.4 Medical Data Mining.....	12
2.5 Klasifikasi Penyakit.....	15
2.6 Sistem Pakar (Expert System)	15
2.7 BPMN.....	17
2.8 Penelitian Terdahulu	19
2.9 Ringkasan Penelitian Terdahulu	28

BAB III METODOLOGI PENELITIAN	30
3.1 <i>Theoretical Framework</i>	30
3.2 Tahapan Penelitian.....	32
3.3 Metode Pengumpulan Data.....	34
3.4 Metode Analisis Data	34
BAB IV IMPLEMENTASI	40
4.1 Tahap Pemahaman Data	40
4.1.1 Penjelasan Data	40
4.1.2 Sumber Data.....	45
4.2 Tahap Data Cleaning	46
4.3 Tahap Data Integration	49
4.4 Tahap Data Selection.....	50
4.4.1 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas.....	50
4.4.2 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid	51
4.4.3 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara.....	53
4.4.4 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum	55
4.5 Tahap Data Transformation.....	65
4.5.1 Tahap Pengubahan Nilai Data Kategorial dan Label Diagnosis	66
4.5.2 Tahap Pengubahan Bahasa Kolom Gejala dan Label Diagnosis Dataset Penyakit	70
4.6 Tahap Data Mining	83
4.6.1 <i>Preprocessing</i>	84
4.6.2 Pemisahan Dataset	84
4.6.3 Pemodelan Algoritma Klasifikasi	84
4.7 Tahap Pattern Evaluation.....	85
4.8 Analisa Data.....	86
4.8.1 Eksplorasi Data	87
4.8.2 Analisis Perbandingan Algoritma Klasifikasi.....	88
4.8.3 Analisa Hasil Penelitian	93
4.9 Tampilan Sistem Pakar (<i>Expert System</i>)	97
4.9.1 Tampilan Sistem Pakar Penyakit Obesitas.....	99

4.9.2	Tampilan Sistem Pakar Penyakit Tiroid	102
4.9.3	Tampilan Sistem Pakar Penyakit Jantung	105
4.9.4	Tampilan Sistem Pakar Penyakit Kanker Payudara.....	108
4.9.5	Tampilan Sistem Pakar Penyakit Umum	110
4.9.6	Tampilan Sistem Pakar Penyakit COVID-19	113
BAB V KESIMPULAN DAN SARAN		116
5.1	Kesimpulan	116
5.2	Saran	117
DAFTAR PUSTAKA		118
LAMPIRAN A USER MANUAL		A-1



DAFTAR GAMBAR

Gambar 2.1 Proses Data Mining	6
Gambar 2.2 Proses Menentukan <i>Hyperplane</i>	8
Gambar 2.3 Contoh Kurva Hubungan Antara Variabel Hasil Bersifat Dikotomis dan Variabel Prediktor Bersifat Continuous	9
Gambar 2.4 <i>Confusion Matrix</i>	11
Gambar 2.5 Framework Medical Data Mining	13
Gambar 2.6 <i>Medical Data Mining Framework</i> Untuk Mengembangkan <i>Expert System</i> .	14
Gambar 2.7 Diagram <i>Expert System</i>	16
Gambar 3.1 Theoretical Framework	30
Gambar 3.2 Tahapan Penelitian	33
Gambar 3.3 Contoh <i>Medical Transcription Digital</i>	34
Gambar 3.4 Metode Analisis Data	35
Gambar 3.5 Alur Penelitian	37
Gambar 3.6 Alur Model Sistem Pakar	39
Gambar 4.1 Bentuk Data Awal dengan Format CSV	40
Gambar 4.2 Bentuk Data Awal dengan Format .data	41
Gambar 4.3 Hasil Penyalinan File .data di File Excel	47
Gambar 4.4 Menu Text to Columns Untuk Memisahkan Data	47
Gambar 4.5 Tampilan Pemisah Data Berdasarkan Tanda Baca	47
Gambar 4.6 Tampilan Pemisah Data Menggunakan Tanda Baca Koma	48
Gambar 4.7 Perbandingan Hasil Pembersihan Data Penyakit Obesitas, Tiroid, dan Kanker Payudara.....	48
Gambar 4.8 Perbandingan Hasil Pembersihan Data Penyakit Umum, Jantung, dan COVID-19.....	49
Gambar 4.9 Tampilan Awal Kolom Diagnosis Dataset Penyakit Tiroid.....	66
Gambar 4.10 Tampilan Kolom Diagnosis Dataset Penyakit Tiroid Setelah Proses Transformasi	67
Gambar 4.11 Kode Untuk Pelabelan Nilai Kolom Gejala yang Bernilai Kategorial.....	70
Gambar 4.12 Baris Kode untuk Melakukan Preprocessing	84
Gambar 4.13 Baris Kode untuk Memisahkan Data <i>Training</i> dan Data <i>Testing</i>	84
Gambar 4.14 Kode untuk Memodelkan Algoritma Klasifikasi	85
Gambar 4.15 Baris Kode untuk Menunjukkan <i>Confusion Matrix</i>	85
Gambar 4.16 Baris Kode untuk Menunjukkan Plot <i>Confusion Matrix</i>	86
Gambar 4.17 Baris Kode untuk Menunjukkan Nilai Evaluasi Algoritma	86
Gambar 4.18 Grafik Perbandingan Jumlah Data Penyakit	88
Gambar 4.19 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Obesitas..	89
Gambar 4.20 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Tiroid.....	90
Gambar 4.21 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Jantung ...	91
Gambar 4.22 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Umum	92
Gambar 4.23 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Kanker Payudara.....	92

Gambar 4.24 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit COVID-19	93
Gambar 4.25 Baris Kode <i>Import Ipywidgets</i> untuk Tampilan Awal Sistem Pakar.....	97
Gambar 4.26 Baris Kode Tampilan Awal Sistem Pakar	97
Gambar 4.27 Baris Menginisialisasi <i>Button, Inttext</i> , dan Fungsi Saat <i>Button</i> Diklik dalam Menu Sistem Pakar	98
Gambar 4.28 Baris Kode untuk Menjalankan Sistem Pakar.....	98
Gambar 4.29 Tampilan Awal Sistem Pakar.....	98
Gambar 4.30 Baris Kode Penggunaan <i>Classifier</i> Sebagai <i>Knowledge Base</i> Sistem Pakar Penyakit Obesitas.....	99
Gambar 4.31 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian I ...	100
Gambar 4.32 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian II ..	100
Gambar 4.33 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian III.	101
Gambar 4.34 Tampilan Saat Menunggu Diagnosis Penyakit Obesitas Dihasilkan	101
Gambar 4.35 Tampilan Saat Diagnosis Penyakit Obesitas Dihasilkan.....	101
Gambar 4.36 Baris Kode Penggunaan <i>Classifier</i> Sebagai <i>Knowledge Base</i> Sistem Pakar Penyakit Tiroid.....	102
Gambar 4.37 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian I	103
Gambar 4.38 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian II.....	103
Gambar 4.39 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian III.....	104
Gambar 4.40 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian IV	104
Gambar 4.41 Tampilan Saat Menunggu Diagnosis Penyakit Tiroid Dihasilkan	105
Gambar 4.42 Tampilan Saat Diagnosis Penyakit Tiroid Dihasilkan	105
Gambar 4.43 Baris Kode Penggunaan <i>Classifier</i> Sebagai <i>Knowledge Base</i> Sistem Pakar Penyakit Jantung	106
Gambar 4.44 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian I.....	106
Gambar 4.45 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian II ...	107
Gambar 4.46 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian III ..	107
Gambar 4.47 Tampilan Saat Menunggu Diagnosis Penyakit Jantung Dihasilkan.....	107
Gambar 4.48 Tampilan Saat Diagnosis Penyakit Jantung Dihasilkan	108
Gambar 4.49 Baris Kode Penggunaan <i>Classifier</i> Sebagai <i>Knowledge Base</i> Sistem Pakar Penyakit Kanker Payudara.....	108
Gambar 4.50 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Kanker Payudara	109
Gambar 4.51 Tampilan Saat Menunggu Diagnosis Penyakit Kanker Payudara Dihasilkan	109
Gambar 4.52 Tampilan Saat Diagnosis Penyakit Kanker Payudara Dihasilkan.....	109
Gambar 4.53 Baris Kode Penggunaan <i>Classifier</i> Sebagai <i>Knowledge Base</i> Sistem Pakar Penyakit Umum	110
Gambar 4.54 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian I	111
Gambar 4.55 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian II.....	111
Gambar 4.56 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian III ...	112
Gambar 4.57 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian IV ...	112
Gambar 4.58 Tampilan Saat Menunggu Diagnosis Penyakit Umum Dihasilkan	112
Gambar 4.59 Tampilan Saat Diagnosis Penyakit Umum Dihasilkan	113

Gambar 4.60 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit COVID-19..... 113
Gambar 4.61 Tampilan Sistem Pakar untuk Mendeteksi Penyakit COVID-19 Bagian I 114
Gambar 4.62 Tampilan Sistem Pakar untuk Mendeteksi Penyakit COVID-19 Bagian II 114
Gambar 4.63 Tampilan Saat Menunggu Diagnosis Penyakit COVID-19 Dihasilkan 115
Gambar 4.64 Tampilan Saat Diagnosis Penyakit COVID-19 Dihasilkan 115



DAFTAR TABEL

Tabel 2.1 Pengukuran Evaluasi.....	12
Tabel 2.2 Elemen Desain BPMN yang Digunakan.....	17
Tabel 2.2 Elemen Desain BPMN yang Digunakan (lanjutan)	18
Tabel 2.2 Elemen Desain BPMN yang Digunakan (lanjutan)	19
Tabel 2.3 Penelitian Terdahulu Bagian I	21
Tabel 2.3 Penelitian Terdahulu Bagian I (lanjutan)	22
Tabel 2.3 Penelitian Terdahulu Bagian I (lanjutan)	23
Tabel 2.4 Penelitian Terdahulu Bagian II	25
Tabel 2.4 Penelitian Terdahulu Bagian II (lanjutan).....	26
Tabel 2.4 Penelitian Terdahulu Bagian II (lanjutan).....	26
Tabel 4.1 Tabel Jumlah Data	41
Tabel 4.2 Pertanyaan Kuesioner, Kemungkinan Jawaban, dan Kodifikasi Pertanyaan dalam Dataset Penyakit Obesitas	42
Tabel 4.2 Pertanyaan Kuesioner, Kemungkinan Jawaban, dan Kodifikasi Pertanyaan dalam Dataset Penyakit Obesitas (lanjutan).....	43
Tabel 4.3 Tabel Sumber Data	45
Tabel 4.3 Tabel Sumber Data (lanjutan).....	46
Tabel 4.4 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas	50
Tabel 4.4 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas (lanjutan)	51
Tabel 4.5 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid 52	
Tabel 4.5 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid (lanjutan)	53
Tabel 4.6 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara.....	54
Tabel 4.6 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara (lanjutan)	55
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum57	
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	58
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	59
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	60
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	61
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	62
Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)	63

Tabel 4.8 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Jantung	63
Tabel 4.8 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Jantung (lanjutan)	64
Tabel 4.9 Tabel Pemilihan Kolom Gejala Dataset Penyakit COVID-19	65
Tabel 4.10 Tabel Kode Gejala Virus COVID-19.....	68
Tabel 4.11 Tabel Kode Hasil Keputusan Diagnosis	68
Tabel 4.12 Tabel Kombinasi Kode Gejala dan Hasil Kode Keputusan	68
Tabel 4.13 Tabel Pengubahan Bahasa dalam Dataset Penyakit Obesitas	71
Tabel 4.13 Tabel Pengubahan Bahasa dalam Dataset Penyakit Obesitas (lanjutan)	72
Tabel 4.14 Tabel Pengubahan Bahasa dalam Dataset Penyakit Tiroid.....	73
Tabel 4.14 Tabel Pengubahan Bahasa dalam Dataset Penyakit Tiroid (lanjutan)	74
Tabel 4.15 Pengubahan Bahasa dalam Dataset Penyakit Kanker Payudara	74
Tabel 4.15 Pengubahan Bahasa dalam Dataset Penyakit Kanker Payudara (lanjutan).....	75
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum.....	76
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)	77
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)	78
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)	79
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)	80
Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)	81
Tabel 4.17 Pengubahan Bahasa dalam Dataset Penyakit Jantung	82
Tabel 4.18 Pengubahan Bahasa dalam Dataset Penyakit COVID-19.....	82
Tabel 4.18 Pengubahan Bahasa dalam Dataset Penyakit COVID-19 (lanjutan)	83
Tabel 4.19 Tabel Rekapitulasi Data Penelitian	87

DAFTAR RUMUS

Rumus 2.1 Rumus Support Vector Machine	8
Rumus 2.2 Rumus Logistic Regression	10
Rumus 2.3 Rumus Naïve Bayes.....	10



DAFTAR LAMPIRAN

LAMPIRAN A USER MANUAL.....	A-1
-----------------------------	-----

