

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan segala berkah dan karunia-Nya, memberikan kekuatan dan kesabaran serta memberikan jalan menuju kemudahan sehingga penyusunan tugas akhir yang berjudul **PENGARUH SELEKSI FITUR BERBASIS *RECURSIVE FEATURE ELIMINATION (RFE)* TERHADAP KLASIFIKASI KINERJA KARYAWAN MENGGUNAKAN *RANDOM FOREST*** telah selesai. Adapun tugas akhir ini merupakan salah satu syarat akademik bagi seluruh mahasiswa Jurusan Informatika di Universitas Siliwangi.

Dalam penyusunan laporan ini penulis banyak dibantu oleh berbagai pihak, baik langsung maupun tidak langsung. Untuk itu penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Prof. Dr. Eng. Ir. H. Aripin selaku Dekan Fakultas Teknik Universitas Siliwangi Tasikmalaya.
2. Bapak Ir. Nur Widiyasono, M.Kom. selaku Ketua Jurusan Teknik Informatika sekaligus Dosen Penguji pada sidang tugas akhir penulis.
3. Bapak Ir. Acep Irham Gufroni, S.Kom., M.Eng., dan Bapak Cecep Muhamad Sidik R., S.T., M.T., selaku Dosen Pembimbing yang senantiasa sabar memberikan bimbingan, arahan dan meluangkan waktu serta pikirannya dalam menyempurnakan tugas akhir ini.
4. Bapak Alam Rahmatulloh, S.T., M.T., selaku Dosen Penguji pada sidang tugas akhir penulis.

5. Seluruh staf dosen pengajar serta segenap karyawan di lingkungan Fakultas Teknik Universitas Siliwangi Tasikmalaya.
6. Kedua orang tua, dan adik penulis yang menjadi motivasi terbesar penulis untuk segera menyelesaikan laporan ini.
7. Rekan-rekan mahasiswa Universitas Siliwangi khususnya Fakultas Teknik.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah memberi bantuan dan dorongan baik moril maupun materil.

Semoga laporan tugas akhir penulis dapat memberikan informasi dan memberikan inspirasi bagi pembaca. Tanpa melupakan banyaknya kekurangan yang ada dalam laporan, penulis menerima kritik serta saran dari para pembaca.

Bandung, 01 Desember 2021

Penulis

DAFTAR ISI

ABSTRACT	v
ABSTRAKSI.....	vi
MOTTO DAN PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN.....	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-5
1.3 Batasan Masalah.....	I-6
1.4 Tujuan Penelitian	I-7
1.5 Manfaat Penelitian	I-7
1.6 Metodologi Penelitian	I-8
1.7 Sistematika Penulisan.....	I-10
BAB II LANDASAN TEORI.....	II-1
2.1 Teori Terkait.....	II-1
2.1.1 Kecerdasan Buatan Untuk Evaluasi Kinerja	II-1
2.1.2 <i>Machine Learning</i>	II-3
2.1.3 Klasifikasi	II-5

2.1.4	<i>Random Forest</i>	II-6
2.1.5	<i>Naïve Bayes</i>	II-7
2.1.6	<i>Support Vector Machine (SVM)</i>	II-7
2.1.7	<i>K-Nearest Neighbors</i>	II-8
2.1.8	<i>Logistic Regression</i>	II-8
2.1.9	<i>Recursive Feature Elimination (RFE)</i>	II-9
2.1.10	<i>Confusion Matrix Pada Klasifikasi Biner</i>	II-11
2.1.11	<i>Area Under The Receiver Operating Characteristcs (AUC-ROC)</i>	II-12
2.1.12	<i>Python</i>	II-16
2.1.13	<i>Google Colab</i>	II-16
2.2	Studi Pustaka	II-17
2.2.1	<i>State Of The Art</i>	II-17
2.2.2	Matrik Penelitian	II-23
2.2.3	Penelitian Terdekat	II-25
BAB III METODOLOGI		III-1
3.1	Metode Penelitian	III-1
3.2	Peta Jalan (<i>Road Map</i>) Penelitian	III-2
3.3	Alat Penelitian	III-3
3.4	Tahapan Penelitian	III-4
3.4.1	Studi Pustaka	III-6
3.4.2	Pengumpulan Data	III-7
3.4.3	Proses Awal Pengolahan Data	III-9
3.4.4	Eksperimen	III-17

3.4.5	Penarikan Kesimpulan	III-25
BAB IV HASIL DAN PEMBAHASAN		IV-1
4.1	Implementasi Proses Awal Pengolahan Data.....	IV-1
4.1.1	<i>Data Cleansing</i>	IV-8
4.1.2	<i>Label Encoding</i>	IV-13
4.1.3	<i>Categorical Binning</i>	IV-14
4.1.4	<i>Dummy Encoding</i>	IV-20
4.1.5	<i>Date Feature Engineering</i>	IV-21
4.2	Implementasi dan Pembahasan Hasil Eksperimen.....	IV-26
4.2.1	<i>Processed Dataset</i>	IV-27
4.2.2	Implementasi Parameter <i>K-Fold Cross Validation</i>	IV-31
4.2.3	Implementasi Parameter <i>Recursive Feature Elimination (RFE)</i>	IV-32
4.2.4	Implementasi dan Hasil Evaluasi <i>Random Forest</i> Dengan <i>RFE</i>	IV-34
4.2.5	Implementasi dan Hasil Evaluasi Model Tanpa <i>RFE</i>	IV-44
4.2.6	Pembahasan Hasil Eksperimen	IV-50
BAB V KESIMPULAN DAN SARAN		V-1
5.1	Kesimpulan	V-1
5.2	Saran.....	V-3
DAFTAR PUSTAKA		1

DAFTAR TABEL

Tabel II-1 <i>Confusion Matrix</i>	II-12
Tabel II-2 Ambang Batas <i>AUC-ROC</i>	II-14
Tabel II-3 Kalkulasi <i>AUC-ROC</i>	II-15
Tabel II-4 Matrik Penelitian.....	II-23
Tabel II-5 Matrix Penelitian (Lanjutan).....	II-24
Tabel II-6 Penelitian Terdekat	II-25
Tabel II-7 Penelitian Terdekat (Lanjutan 1)	II-26
Tabel III-1 Atribut-Atribut Duplikat.....	III-10
Tabel III-2 Atribut-Atribut yang Bersifat Privasi	III-11
Tabel IV-1 Deskripsi Dataset.....	IV-5
Tabel IV-2 Deskripsi Dataset (Lanjutan)	IV-6
Tabel IV-3 Deskripsi Kategori Kinerja Pada Atribut Kelas	IV-7
Tabel IV-4 Deskripsi Atribut, Variabel Prediktor dan Kelas.....	IV-28
Tabel IV-5 Deskripsi Atribut, Variabel Prediktor dan Kelas (Lanjutan 1).....	IV-29
Tabel IV-6 Deskripsi Atribut, Variabel Prediktor dan Kelas (Lanjutan 2).....	IV-30
Tabel IV-7 Fungsi yang Diperlukan Untuk Evaluasi Model dengan <i>RFE</i>	IV-34
Tabel IV-8 Fungsi yang Diperlukan Untuk Evaluasi Model Tanpa <i>RFE</i>	IV-44
Tabel IV-9 Fungsi yang Diperlukan Untuk Evaluasi Model Tanpa <i>RFE</i>	IV-45
Tabel IV-10 Rata-rata <i>AUC-ROC</i> sebelum dan sesudah penerapan <i>RFE</i>	IV-51
Tabel IV-11 Komparasi Rata-Rata <i>AUC-ROC</i> Antara Model <i>Random Forest</i> dengan Model <i>Supervised Learning</i> Lain.	IV-54

DAFTAR GAMBAR

Gambar II-1 Konsep <i>AIHRM</i>	II-2
Gambar II-2 Metode <i>Wrapper</i>	II-10
Gambar II-3 Prosedur <i>RFE</i>	II-11
Gambar III-1 Peta Jalan Penelitian KK ISI Unsil	III-2
Gambar III-2 Tahapan Penelitian.....	III-4
Gambar III-3 <i>Fishone</i> Ruang Lingkup Penelitian	III-5
Gambar III-4 <i>Human Resource Dataset</i> di <i>Kaggle</i>	III-7
Gambar III-5 Prosedur Pengumpulan Data Tanpa Pemrograman <i>Python</i>	III-8
Gambar III-6 Prosedur Pengumpuldan Data Dengan Pemrograman <i>Python</i> ..	III-8
Gambar III-7 Prosedur Pengolahan Awal Data	III-9
Gambar III-8 Ilustrasi Prosedur <i>Label Encoding</i>	III-11
Gambar III-9 Ilustrasi Prosedur <i>Categorical Binning</i>	III-13
Gambar III-10 Ilustrasi Prosedur <i>Dummy Encoding</i>	III-14
Gambar III-11 Prosedur <i>Feature Engineering</i> Pada <i>DOB</i>	III-15
Gambar III-12 Prosedur <i>Feature Engineering</i> Pada <i>DateofHire</i>	III-16
Gambar III-13 Prosedur <i>Feature Engineering</i> Pada <i>LastPerformanceReviewDate</i>	III-16
Gambar III-14 Prosedur Eksperimen	III-17
Gambar III-15 Ilustrasi Strategi <i>One vs One</i> dan <i>One vs Rest</i>	III-19
Gambar III-16 Proses Dalam <i>5-Fold Cross Validation</i>	III-21
Gambar III-17 Prosedur Evaluasi Model <i>Random Forest</i> Dengan <i>RFE</i>	III-22
Gambar III-18 Prosedur Evaluasi Model Tanpa <i>RFE</i>	III-23
Gambar III-19 Pembahasan Pada Hasil Eksperimen	III-24
Gambar IV-1 <i>Login</i> atau <i>Register</i> akun <i>Kaggle</i>	IV-1
Gambar IV-2 Unduh Token Akun <i>Kaggle</i>	IV-2
Gambar IV-3 Tampilan <i>Notebook Google Colab</i>	IV-2
Gambar IV-4 Token Akun <i>Kaggle</i> Berhasil Diunggah.....	IV-3
Gambar IV-5 File Dataset Siap Digunakan Untuk Penelitian.....	IV-4
Gambar IV-6 Informasi Dasar <i>Human Resource Dataset</i>	IV-4

Gambar IV-7 Informasi Dasar <i>Human Resource Dataset</i> (Lanjutan).....	IV-5
Gambar IV-8 Rasio Kategori Kinerja Pada Atribut Kelas	IV-7
Gambar IV-9 Membaca dan Menampilkan Informasi <i>Human Resource Dataset</i>	IV-9
Gambar IV-10 Dataset Tanpa Atribut Duplikat	IV-10
Gambar IV-11 Dataset Tanpa Atribut yang Bertentangan Dengan Etika Data .IV- 11	
Gambar IV-12 Status Kerja Karyawan di <i>Human Resource Dataset</i>	IV-12
Gambar IV-13 Atribut-Atribut Terkait Pemecatan dan Pengunduran Diri	IV-12
Gambar IV-14 Informasi Dataset Setelah Melalui Pembersihan Data.....	IV-13
Gambar IV-15 Implementasi <i>Label Encoding</i>	IV-14
Gambar IV-16 Rasio Kelas <i>Position</i>	IV-15
Gambar IV-17 Visualisasi Hasil <i>Binning</i> Pada Atribut <i>Position</i>	IV-16
Gambar IV-18 Visualisasi Rasio Kelas Atribut <i>Department</i>	IV-17
Gambar IV-19 Implementasi dan Hasil <i>Binning</i> pada Atribut <i>Department</i> ...	IV-18
Gambar IV-20 Visualisasi Kelas Pada Atribut <i>RecruitmentSource</i>	IV-19
Gambar IV-21 Implementasi <i>Dummy Encoding</i>	IV-20
Gambar IV-22 Hasil <i>Dummy Encoding</i>	IV-20
Gambar IV-23 Atribut Tanggal yang Memerlukan <i>Feature Engineering</i>	IV-21
Gambar IV-24 Konversi Data Pada Atribut <i>DOB</i>	IV-22
Gambar IV-25 Hasil Implementasi <i>Feature Engineering</i> pada Atribut <i>DOB</i> ...	IV-23
Gambar IV-26 Hasil Implementasi <i>Feature Engineering</i> pada Atribut <i>DateofHire</i>	IV-24
Gambar IV-27 Konversi Data dan <i>Rename</i> Atribut <i>LastPerformanceReviewDate</i>	IV-25
Gambar IV-28 Hasil Implementasi <i>Feature Engineering</i> pada Atribut <i>LastPerformanceReviewDate</i>	IV-26
Gambar IV-29 <i>Procesed Human Resource Dataset</i>	IV-28
Gambar IV-30 Variabel Prediktor dan Variabel Kelas	IV-30
Gambar IV-31 Implementasi Parameter <i>Scoring</i> dan nilai K	IV-31
Gambar IV-32 Implementasi <i>RFE</i> untuk Evaluasi <i>Random Forest</i>	IV-32

Gambar IV-33 Rancangan Fungsi Evaluasi Model Dengan <i>RFE</i>	IV-36
Gambar IV-34 Impor Pustaka yang Dibutuhkan untk Evaluasi Model dengan <i>RFE</i>	IV-36
Gambar IV-35 Pengkodean Fungsi Evaluasi Model <i>Random Forest</i> dengan <i>RFE</i>	IV-37
Gambar IV-36 Pengkodean Fungsi Evaluasi Model <i>Random Forest</i> dengan <i>RFE</i> (Lanjutan).....	IV-38
Gambar IV-37 Eksekusi Fungsi Evaluasi Model <i>Random Forest</i> dengan <i>RFE</i>	IV-38
Gambar IV-38 <i>List</i> Nama Atribut-Atribut Berdasarkan Jumlah yang Terseleksi	IV-39
Gambar IV-39 <i>List</i> Nama Atribut-Atribut Berdasarkan Jumlah yang Terseleksi (Lanjutan 1).....	IV-40
Gambar IV-40 <i>List</i> Nama Atribut-Atribut Berdasarkan Jumlah yang Terseleksi (Lanjutan 2).....	IV-41
Gambar IV-41 <i>List</i> Nama Atribut-Atribut Berdasarkan Jumlah yang Terseleksi (Lanjutan 3).....	IV-42
Gambar IV-42 Rata-Rata dan Nilai <i>AUC-ROC RF-RFE</i> secara <i>OVO</i>	IV-42
Gambar IV-43 Rata-Rata dan Nilai <i>AUC-ROC RF-RFE</i> secara <i>OVO</i> (Lanjutan)	IV-43
Gambar IV-44 Rata-Rata dan Nilai <i>AUC-ROC RF-RFE</i> secara <i>OVR</i>	IV-43
Gambar IV-45 Rata-Rata dan Nilai <i>AUC-ROC RF-RFE</i> secara <i>OVR</i> (Lanjutan)	IV-44
Gambar IV-46 Rancangan Fungsi Evaluasi Model Tanpa <i>RFE</i>	IV-46
Gambar IV-47 Impor Pustaka yang Dibutuhkan untk Evaluasi Model Tanpa <i>RFE</i>	IV-47
Gambar IV-48 Pengkodean Fungsi Evaluasi Model Tanpa <i>RFE</i>	IV-47
Gambar IV-49 Pengkodean Fungsi Evaluasi Model Tanpa <i>RFE</i>	IV-48
Gambar IV-50 Eksekusi Fungsi Evaluasi Model Tanpa <i>RFE</i>	IV-49
Gambar IV-51 Rata-Rata Nilai <i>AUC-ROC RF,SVM, Logistic Regression, Naïve Bayes</i> serta <i>KNN</i> secara <i>OVO</i> dari setiap <i>fold</i>	IV-49

Gambar IV-52 Rata-Rata dan Nilai <i>AUC-ROC</i> <i>RF</i> , <i>SVM</i> , <i>Logistic Regression</i> , <i>Naïve Bayes</i> serta <i>KNN</i> secara <i>OVR</i> dari setiap <i>fold</i>	IV-49
Gambar IV-53 Visualisasi Antara Jumlah Atribut Hasil Seleksi Fitur dan Rata- Rata <i>AUC- ROC</i> <i>Random Forest</i>	IV-53