

**TESIS**

**Implementasi Metode Feature Selection  
pada Kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor  
dalam Melakukan Seleksi Klasifikasi Pegawai  
(Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)**



**Disusun oleh :**

Nama : Syaiful Ramadhan

NIM : 21.55.2169

Konsentrasi : Digital Transformation Intelligence

**PROGRAM STUDI S2 PJJ TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**TESIS**

**Implementasi Metode Feature Selection  
pada Kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor  
dalam Melakukan Seleksi Klasifikasi Pegawai  
(Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)**

*Implementation of the Feature Selection Method on The Combination of Support  
Vector Machine and K-Nearest Neighbor Algorithms in Conducting Employee  
Classification Selection  
(Case Study on Determining Access Rights for SAKTI Government Applications)*

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



**Disusun oleh :**

Nama : Syaiful Ramadhan  
NIM : 21.55.2169  
Konsentrasi : Digital Transformation Intelligence

**PROGRAM STUDI S2 PJJ TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**HALAMAN PENGESAHAN**

**Implementasi Metode Feature Selection  
pada Kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor  
dalam Melakukan Seleksi Klasifikasi Pegawai  
(Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)**

**Implementation of the Feature Selection Method on The Combination of  
Support Vector Machine and K-Nearest Neighbor Algorithms in Conducting  
Employee Classification Selection  
(Case Study on Determining Access Rights for SAKTI Government  
Applications)**

Dipersiapkan dan Disusun oleh

**Syaiful Ramadhan**

**21.55.2169**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari **Jumat**, tanggal **3 Nopember 2023**

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, tanggal 3 Nopember 2023  
**Rektor**

**Prof. Dr. M. Suvanto, M.M.**  
**NIK. 190302001**

**HALAMAN PERSETUJUAN**

**IMPLEMENTASI METODE FEATURE SELECTION  
PADA KOMBINASI ALGORITMA SUPPORT VECTOR MACHINE DAN K-  
NEAREST NEIGHBOR  
DALAM MELAKUKAN SELEKSI KLASIFIKASI PEGAWAI  
(Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)**

Dipersiapkan dan Disusun oleh

**Syaful Ramadhan**

**21.55.2169**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari **Jumat**, tanggal **3 Nopember 2023**

**Pembimbing Utama**

**Anggota Tim Penguji**

**Prof. Dr. Kusrini, M.Kom**

**Dhani Ariatmanto, M.Kom., Ph.D**

**NIK. 190302106**

**NIK. 190302197**

**Pembimbing Pendamping**

**Tonny Hidayat, M.Kom., Ph.D**

**NIK. 190302182**

**Kusnawi, S.Kom., M.Eng**

**Prof. Dr. Kusrini, M.Kom**

**NIK. 190302112**

**NIK. 190302106**

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, tanggal 3 Nopember 2023

**Direktur Program Pascasarjana**

**Prof. Dr. Kusrini, M.Kom.**

**NIK. 190302106**

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Syaiful Ramadhan  
NIM : 21.55.2169  
Konsentrasi : Digital Transformation Intelligence

Menyatakan bahwa Tesis dengan judul berikut:  
**Implementasi Metode Feature Selection pada Kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor dalam Melakukan Seleksi Klasifikasi Pegawai (Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)**

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M.Kom  
Dosen Pembimbing Pendamping : Kusnawi, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 3 Nopember 2023

Yang Menyatakan,



13A10C329297064

Syaiful Ramadhan

## HALAMAN PERSEMBAHAN

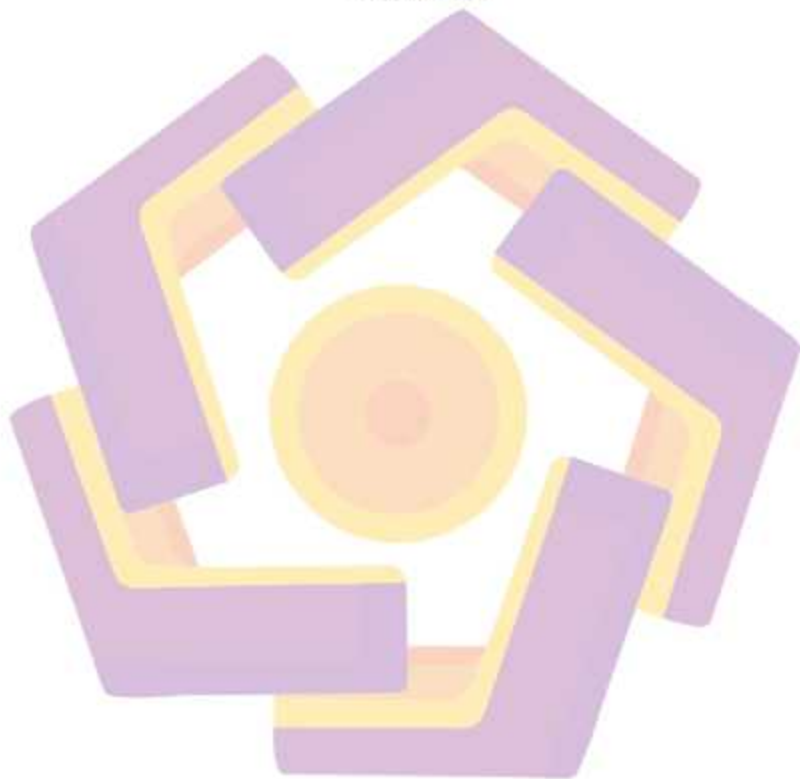
Dengan segala puji dan rasa syukur kepada Allah Subhanahuwataa'la atas berkah dan rahmah-Nya, serta dukungan doa dari orang-orang tercinta, akhirnya tesis ini dapat terselesaikan dengan baik. Oleh karena itu, Tesis ini penulis persembahkan kepada:

1. Adisti, istri sekaligus berperan dosen penulis di rumah, serta duoAL, Almahira dan Albirru yang selalu memberikan doa dan semangat yang luar biasa.
2. Kedua orang tua, Bpk Diswanto dan Ibu Sri Mulyani, atas doa dan dukungannya yang senantiasa diberikan selama penyusunan tesis ini.
3. Bpk Hedyana Yusuf dan Ibu Yani Kamasturyani selaku bapak ibu mertua sekaligus Ketua Yayasan dan Rektor Institut Mahardika, yang selalu menularkan semangat untuk terus menimba ilmu.
4. Fajar, Hasirun dan Lamsadi yang selalu support dan berlomba dalam progres penyusunan Tesis.
5. Teman-teman seperjuangan pada PJJ Magister Teknik Informatika angkatan 2021 atas kebersamaan dan semangat dalam menyelesaikan studi bersama-sama.
6. Pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah membantu hingga terselesaikannya penulisan tesis ini.



## HALAMAN MOTTO

“Menikmati proses dari setiap perubahan yang terjadi dengan menjadi lebih baik  
setiap hari nya”



## KATA PENGANTAR

Segala puji bagi Allah Subhanahuwataa'la atas berkah dan rahmah-Nya, sehingga penulis dapat menyelesaikan tesis dengan judul "Implementasi Metode Feature Selection pada Kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor dalam Melakukan Seleksi Klasifikasi Pegawai (Studi Kasus Penentuan Hak Akses Aplikasi Pemerintahan SAKTI)". Penelitian ini merupakan salah satu syarat untuk memperoleh gelar Magister Komputer pada Program Studi Teknik Informatika, Universitas AMIKOM Yogyakarta.

Penghormatan dan rasa terimakasih yang sebesar-besarnya penulis sampaikan kepada:

1. Prof. Dr. M.Suyanto, M.M selaku rektor Universitas AMIKOM Yogyakarta,
2. Prof. Dr. Kusriani, M.Kom dan Kusnawi, S.Kom., M.Eng selaku pembimbing utama dan pendamping, atas segala kesabaran dalam membimbing, dan meluangkan waktu yang diberikan serta dukungan kepada penulis dalam menyelesaikan tesis ini,
3. Yayasan Institut Mahardika Cirebon atas pemberian beasiswa S2 sehingga dapat menyelesaikan studi dengan baik, dan
4. Tim dosen Penguji atas ketersediaan waktunya untuk menjadi penguji dan memberikan koreksi dan evaluasi yang bermanfaat dari seminar proposal, sampai dengan seminar tesis.



Akhir kata, penulis menyadari masih terdapat kemungkinan kekurangan dalam penulisan tesis ini karena keterbatasan pengetahuan, serta luas dan berkembangnya keilmuan di bidang Teknik Informatika. Penulis berharap bahwa penelitian ini dapat memberikan manfaat bagi akademisi, praktisi, dan masyarakat dalam pengembangan ilmu informatika khususnya pada bidang Machine Learning.

Yogyakarta, 3 Nopember 2023

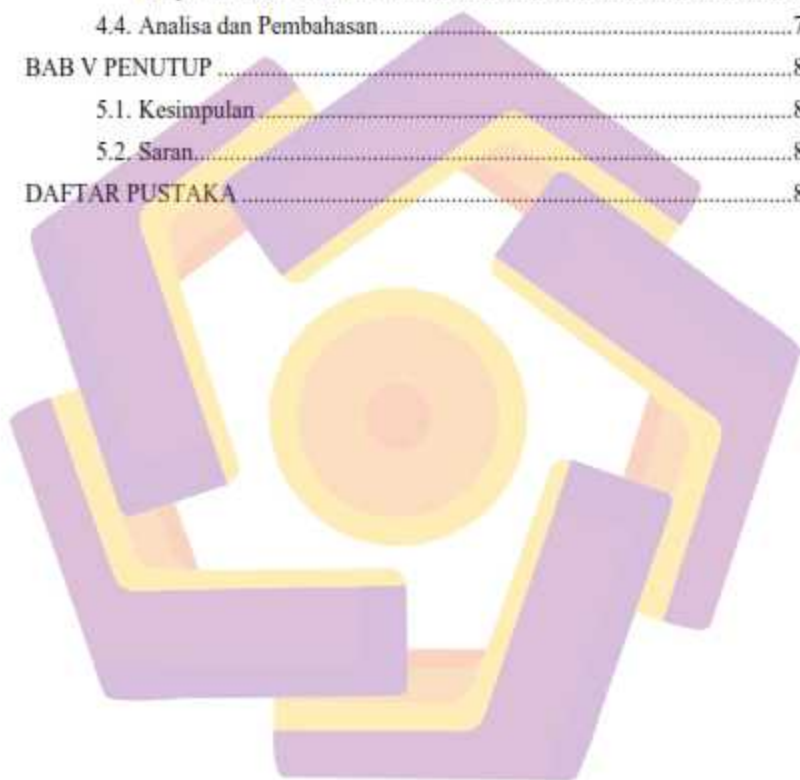
Penulis



## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS .....	v
HALAMAN PERSEMBAHAN .....	vi
HALAMAN MOTTO .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
INTISARI .....	xvi
<i>ABSTRACT</i> .....	xvii
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah .....	5
1.4. Tujuan Penelitian .....	5
1.5. Manfaat Penelitian .....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1. Tinjauan Pustaka .....	7
2.2. Keashian Penelitian.....	11
2.3. Landasan Teori.....	17
<b>BAB III METODE PENELITIAN</b> .....	29
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	29
3.2. Metode Pengumpulan Data .....	30
3.3. Metode Analisis Data.....	30
3.4. Alur Penelitian .....	32
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN</b> .....	40
4.1. Pengumpulan Data .....	40

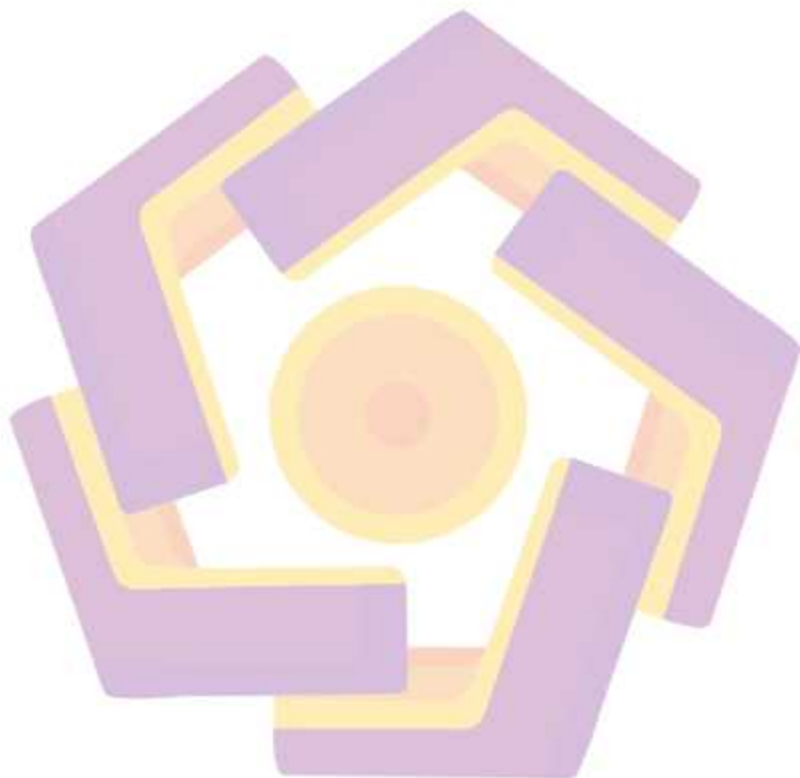
4.2. Preprocessing Data.....	50
4.3. Proses Data Mining.....	52
4.3.1. Pengujian Seleksi Fitur dengan Backward Elimination.....	53
4.3.2. Hasil Percobaan dan Pengujian Pada Kombinasi SVM-KNN..	55
4.3.3 Hasil Percobaan dan Pengujian Pada Kombinasi SVM-KNN Tanpa Seleksi Fitur .....	70
4.4. Analisa dan Pembahasan.....	74
BAB V PENUTUP .....	86
5.1. Kesimpulan.....	86
5.2. Saran.....	87
DAFTAR PUSTAKA.....	89



## DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian Implementasi Metode Feture Selection pada Kombinasi algoritma SVM dan KNN dalam melakukan Seleksi Klasifikasi Pegawai .....	11
Tabel 4.1. Parameter yang akan digunakan .....	41
Tabel 4.2. Set Parameter SVM-KNN tanpa seleksi Fitur .....	41
Tabel 4.3. Hasil Penilaian Kinerja Pegawai Periode 2022 .....	42
Tabel 4.4. Data Penempatan Pegawai pada Biro BMN .....	44
Tabel 4.5. Penempatan Pegawai Berdasarkan role SAKTI .....	46
Tabel 4.6. Dataset Utama .....	48
Tabel 4.7. Penggolongan Data .....	49
Tabel 4.8. Hasil Seleksi Fitur .....	54
Tabel 4.9. Tabel Evaluasi Kinerja pada percobaan ke1 .....	57
Tabel 4.10. Tabel Evaluasi Kinerja pada percobaan ke2 .....	60
Tabel 4.11. Tabel Evaluasi Kinerja pada percobaan ke3 .....	62
Tabel 4.12. Tabel Evaluasi Kinerja pada percobaan ke4 .....	64
Tabel 4.13. Tabel Evaluasi Kinerja pada percobaan ke5 .....	66
Tabel 4.14. Tabel Evaluasi Kinerja pada percobaan ke6 .....	67
Tabel 4.15. Tabel Evaluasi Kinerja pada percobaan ke7 .....	69
Tabel 4.16. Tabel Evaluasi Kinerja pada percobaan ke8 .....	70
Tabel 4.17. Tabel Evaluasi Kinerja .....	72
Tabel 4.18. Tabel Evaluasi Kinerja .....	74
Tabel 4.19. Tabel Atribut Utama .....	75

Tabel 4.20. Tabel Atribut Setelah Seleksi .....	76
Tabel 4.21. Hasil Pengujian dengan 10 kali kombinasi .....	77
Tabel 4.22. Tabel Perbandingan Hasil Seleksi Fitur dan Tanpa Seleksi Fitur .....	83
Tabel 4.23. Tabel Perbandingan Evaluasi Kinerja Percobaan ke3 dan ke8.....	83



## DAFTAR GAMBAR

Gambar 2.1. Pseudo Code Forward Selection .....	9
Gambar 2.2. Perbedaan Feature Selection .....	19
Gambar 2.3. Komponen Utama Metode Seleksi Fitur dengan penerapan Wrapper20	
Gambar 2.4. Komponen Pada Metode KDD .....	22
Gambar 2.5. Hyperplane .....	25
Gambar 3.1. Diagram Alur Penelitian .....	32
Gambar 4.1. Pembacaan Dataset .....	50
Gambar 4.2. Tipe Format Data .....	51
Gambar 4.3. Missing Value pada Atribut .....	52
Gambar 4.4. Transformasi Data .....	52
Gambar 4.5. Pembangunan Model Seleksi Fitur .....	53
Gambar 4.6. Atribut Independen dan Target .....	53
Gambar 4.7. Arsitektur Model Sekelsi Fitur .....	54
Gambar 4.8. Proporsi Data pada Kombinasi SVM-KNN .....	55
Gambar 4.9. Pengujian Model pada parameter $C=1$ dan $\text{gamma}=0.01$ .....	55
Gambar 4.10. Support Vector pada parameter $C=1$ dan $\text{gamma}=0.01$ .....	56
Gambar 4.11. Pengujian Model pada ALgoritma KNN nilai $K=9$ .....	57
Gambar 4.12. Evaluasi Kinerja pada Percobaan ke1 .....	57
Gambar 4.13. Pengujian Model pada parameter $C=5$ dan $\text{gamma}=0.01$ .....	58
Gambar 4.14. Support Vector pada parameter $C=5$ dan $\text{gamma}=0.01$ .....	59
Gambar 4.15. Evaluasi Kinerja pada Percobaan ke2 .....	59
Gambar 4.16. Classification Report Percobaan ke2 .....	60



Gambar 4.17. Pengujian Model pada parameter $C=1$ dan $\gamma=0.1$ .....	61
Gambar 4.18. Support Vector pada parameter $C=1$ dan $\gamma=0.1$ .....	61
Gambar 4.19. Evaluasi Kinerja pada Percobaan ke3 .....	62
Gambar 4.20. Support Vector pada parameter $C=5$ dan $\gamma=0.1$ .....	63
Gambar 4.21. Evaluasi Kinerja pada Percobaan ke4 .....	63
Gambar 4.22. Pengujian Model pada Algoritma KNN nilai $K=15$ .....	65
Gambar 4.23. Evaluasi Kinerja pada Percobaan ke5 .....	65
Gambar 4.24. Evaluasi Kinerja pada Percobaan ke6 .....	67
Gambar 4.25. Evaluasi Kinerja pada Percobaan ke7 .....	68
Gambar 4.26. Evaluasi Kinerja pada Percobaan ke8 .....	70
Gambar 4.27. Membangun Model SVM tanpa seleksi fitur .....	71
Gambar 4.28. Support Vector Tanpa Seleksi Fitur .....	71
Gambar 4.29. Mencari Nilai $K$ terbaik .....	71
Gambar 4.30. Evaluasi Kinerja .....	72
Gambar 4.31. Membangun Pemodelan KNN .....	73
Gambar 4.32. Evaluasi Kinerja .....	73
Gambar 4.33. Perbandingan Nilai Akurasi .....	82
Gambar 4.34. Perbandingan Nilai Presisi dan Recall .....	82

## INTISARI

SDM sangat penting bagi suatu organisasi, seperti halnya dengan Badan Riset dan Inovasi Nasional. Namun, agar kinerja organisasi menjadi efisien, perlu untuk memilih karyawan. Seleksi pegawai dilakukan karena jumlah pegawai di Badan Riset dan Inovasi Nasional (BRIN) saat ini +/- 14.000 pegawai berdasarkan Surat Keputusan Kepala BRIN yang tersebar di seluruh unit kerja di lingkungan BRIN, sehingga menjadi surplus. Selain itu, penyempurnaan sistem administrasi keuangan pemerintahan merupakan suatu keharusan yang diimbangi dengan pengembangan aplikasi terintegrasi Kementerian Keuangan yaitu Sistem Aplikasi Keuangan Tingkat Institusi (SAKTI).

Dalam mengelola aset di Biro BMN, setiap pegawai memiliki peran pengguna tingkat kewenangan SAKTI dengan ruang lingkup modul yang digunakan berbedabeda. Penelitian ini bertujuan untuk memilih klasifikasi pengguna berdasarkan hasil penilaian kinerja dengan menerapkan metode Feature Selection dan dilanjutkan dengan kombinasi algoritma SVM dan KNN menggunakan python.

Berdasarkan hasil pengujian pada 10 percobaan terhadap 321 data karyawan yang dinormalisasi menjadi 313 data karyawan, dengan 17 variabel karyawan, diperoleh hasil akurasi terbaik sebesar 94,1% dan F1-Score sebesar 94,4% dalam pemilihan fitur menjadi 13 variabel, dengan parameter algoritma SVM  $C=1$ ,  $\gamma=0,1$  dan parameter algoritma KNN pada nilai  $K=9$ , metrik Euclidean. Sedangkan hasil pengujian kombinasi algoritma SVM-KNN tanpa melalui seleksi fitur menghasilkan akurasi terbaik sebesar 88,2% dengan total 17 variabel data karyawan, parameter algoritma SVM pada nilai  $C$  dan  $\gamma$  terbaik, serta nilai  $K=8$  pada metrik Euclidean. Sehingga dalam melakukan seleksi klasifikasi pegawai yang memiliki spesifikasi rekapitulasi kehadiran (WFO dan WFH) serta penilaian kinerja dapat menerapkan Backward Elimination pada Kombinasi algoritma SVM dan KNN.

Kata kunci: Seleksi karyawan, Backward Elimination, algoritma SVM, algoritma KNN, kombinasi

## ABSTRACT

HR is very important for an organization, as is the case with the National Research and Innovation Agency. However, in order for organizational performance to be efficient, it is necessary to select employees. The selection of employees is carried out because the number of employees at the National Research and Innovation Agency (BRIN) is currently +- 14,000 employees based on the Decree of the Head of BRIN spread across work units within BRIN, so that it becomes a surplus. In addition, the improvement of the government financial administration system is a must which is balanced with the development of an integrated application of the Ministry of Finance, namely the Institution-Level Financial Application System (SAKTI). In managing assets at the BMN Bureau, each employee has a user role level of authority SAKTI with different scope of modules used. This study aims to select user classification based on the results of performance assessment by applying the Feature Selection method and continued with a combination of SVM and KNN algorithms using python.

Based on the test results on 10 experiments on 321 employee data normalized to 313 employee data, with 17 employee variables, obtained the best accuracy result of 94.1% and F-1 Score of 94.4% in feature selection into 13 variables, with SVM algorithm parameters  $C=1$ ,  $\gamma=0.1$  and KNN algorithm parameters at  $K=9$  values, Euclidean metrics. While the results of testing the SVM-KNN algorithm combination without going through feature selection resulted in the best accuracy of 88.2% with a total of 17 employee data variables, SVM algorithm parameters at the best  $C$  and  $\gamma$  values, and  $K = 8$  values on Euclidean metrics.

So that in selecting employee classifications that have attendance recapitulation specifications (WFO and WFH) and performance appraisals, they can apply Backward Elimination to the combination of SVM and KNN algorithms.

**Keywords:** Employee selection, Backward Elimination, SVM algorithm, KNN algorithm, combination

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

SDM merupakan hal yang sangat penting bagi suatu perusahaan. Dengan melakukan penyeleksian karyawan dapat meningkatkan produktivitas sehingga keberhasilan tujuan perusahaan dapat tercapai (Sunandar & Satar, 2020). Pada dasarnya, tujuan melakukan seleksi karyawan adalah untuk mendapatkan sumber daya manusia yang tepat untuk suatu jabatan tertentu sesuai dengan kompetensi sehingga orang tersebut mampu bekerja secara optimal (Suwinardi, 2017). Karena dengan kompetensi yang sesuai seorang karyawan akan lebih produktif dalam bekerja. Agar kinerja organisasi menjadi efisien, perlu dilakukan seleksi pegawai. Seleksi pegawai dilakukan karena jumlah pegawai pada Badan Riset dan Inovasi Nasional (BRIN) saat ini sebesar +- 14.000 pegawai berdasarkan SK Kepala BRIN yang tersebar pada unit kerja di lingkungan BRIN. Jumlah tersebut karena adanya integrasi dari 4 (empat) Lembaga Pemerintahan Non Kementrian dan 1 Kementrian Riset dan Teknologi menjadi BRIN sesuai amanat dari Peraturan Presiden No. 78 Tahun 2021. Pada saat terjadinya integrasi, tidak dilakukan pemetaan pegawai sehingga setiap pegawai bebas memilih unit kerja yang diinginkan, hal ini mengakibatkan pada beberapa unit kerja mengalami surplus jumlah pegawai, salah satunya pada Biro Manajemen BMN dan Pengadaan. Biro Manajeme



BMN dan Pengadaan BRIN adalah unit kerja Eselon II yang bertugas melakukan pengelolaan asset dan pengadaan.

Disamping itu, perbaikan sistem administrasi pemerintahan melalui pemanfaatan teknologi merupakan suatu tuntutan agar tercapai prinsip-prinsip pengelolaan keuangan yang tertib, efektif, efisien, ekonomis, transparan dan dapat dipertanggungjawabkan. Untuk memenuhi tuntutan tersebut, Direktorat Jenderal Perbendaharaan, Kementerian Keuangan telah mengembangkan suatu aplikasi terintegrasi yang diperuntukkan bagi seluruh unit kerja Instansi pemerintah Pusat di seluruh Indonesia yang diberi nama Sistem Aplikasi Keuangan Tingkat Instansi (SAKTI) (Amriani, 2019). Dalam melakukan kegiatan pengelolaan aset dan pengadaan pada Biro Manajemen BMN dan Pengadaan, setiap pegawai memiliki level kewenangan aplikasi SAKTI. Sistem Aplikasi Keuangan Tingkat Instansi (SAKTI) adalah aplikasi yang mendukung pengelolaan keuangan yang digunakan oleh instansi pemerintah pusat yang bersumber dari Anggaran Pendapatan Belanja Negara (APBN). Aplikasi ini mencakup seluruh proses pengelolaan keuangan negara pada Unit Kerja dimulai dari proses Penganggaran (modul penganggaran), Pelaksanaan (modul Komitmen, modul Bendahara, Modul Aset Tetap, Modul Persediaan, dan modul pembayaran), sampai dengan Pelaporan (modul pelaporan). Pemanfaatan aplikasi SAKTI pada Biro BMN mencakup sebagian kecil proses pengelolaan keuangan negara yaitu pada tahapan implementasi (modul komitmen, modul aset tetap dan modul persediaan) (Anwar, 2022). Selain itu pada aplikasi SAKTI juga merupakan transformasi dari manual ke digital.

Dimana pada proses bisnis sebelumnya agar asset dapat dicatat pada aplikasi, dari Pejabat Komitmen akan memberikan dokumen fisik ke operator aset. Dengan adanya aplikasi SAKTI maka penggunaan fisik kertas menjadi berkurang.

Implementasi aplikasi SAKTI sebagaimana diamanatkan dalam Peraturan Menteri Keuangan Nomor 223/PMK.05/2015 dan Nomor 131/PMK/05/2016 diawali dengan tahapan uji coba terbatas mulai tahun 2015 pada Kementerian Keuangan dan mulai diujicobakan pada seluruh Kementerian/Lembaga pada awal tahun 2022. Dalam melakukan kegiatan pengelolaan aset pada Biro BMN, setiap pegawai memiliki otorisasi aplikasi SAKTI dengan user kewenangan mulai dari admin, operator, validator, approver dan komitmen (Nugroho, 2020). Sehingga dengan adanya kebutuhan tingkatan level kewenangan pengguna saat penginputan data membuat proses kerja menjadi lebih panjang.

Agar tercapainya tujuan organisasi yang telah ditetapkan dalam pengelolaan keuangan yang efektif, efisien, dan akuntabel, untuk itu maka diperlukan pengelolaan Sumber Daya Manusia yang baik. Kegagalan dalam pengelolaan SDM akan berdampak buruk bagi suatu organisasi. Tidak terkecuali dalam hal penunjukan pegawai sebagai user aplikasi pada level kewenangan karena berkaitan dengan kompetensi dari pegawai itu sendiri dalam menentukan kinerja yang bersumber dari hasil kerjanya. Disamping itu diungkapkan berdasarkan artikel, "Hubungan Usia, Tingkat Pendidikan, Kemampuan Bekerja dan Masa Bekerja terhadap Kinerja Pegawai dengan menggunakan Metode Pearson Correlation", bahwa hubungan usia dan



tingkat Pendidikan dengan kinerja bersifat linier karena perkembangan karier meningkat selama faktor usia dan tingkat pendidikan juga meningkat. Sehingga dapat dikatakan dengan kompetensi pegawai (tingkat Pendidikan, usia), akan menentukan kelayakan klasifikasi pengguna dengan level kewenangan sebagai operator, validator, approver dan komitmen sehingga berdampak kinerja organisasi menjadi lebih efisien. Agar menghasilkan prediksi yang akurat dalam melakukan seleksi karyawan berdasarkan rekomendasi dari Kementerian PANRB dan kebutuhan di unit kerja, maka diperlukan suatu perencanaan model dan metode untuk memproses sekumpulan data agar dapat menghasilkan prediksi yang akurat sehingga menjadi lebih efisien.

### **1.2. Rumusan Masalah**

Berdasarkan Latar Belakang masalah, dapat disampaikan bahwa permasalahan yang menjadi perhatian dalam penelitian ini adalah :

- a. Berapa jumlah fitur dan nilai K agar menghasilkan akurasi tertinggi dalam melakukan seleksi klasifikasi pegawai?
- b. Seberapa besar peningkatan akurasi dan model terbaik sebelum dan setelah menggunakan metode Feature Selection dengan kombinasi algoritma klasifikasi dalam menyelesaikan permasalahan seleksi klasifikasi pegawai?

### 1.3. Batasan Masalah

Batasan Masalah pada penelitian ini agar permasalahan lebih terarah dan tidak meluas sesuai dengan judul yang telah ditetapkan, adalah sebagai berikut:

- a. Algoritma yang digunakan yaitu kombinasi algoritma SVM dan KNN dengan batasan variabel input nama, usia, jenis kelamin, tingkat pendidikan, pangkat/golongan, lokasi kerja, domisili, persentase kehadiran, level kewenangan aplikasi SAKTI, hasil penilaian Sistem Kinerja Pegawai 2022, dan kelayakan pegawai.
- b. Untuk meningkatkan optimasi algoritma, penulis membatasi menggunakan metode seleksi fitur (feature Selection) dengan teknik Wrapper Methods yaitu backward elimination.
- c. Pegawai berasal dari lingkup pengguna aplikasi SAKTI eks 4 Lembaga Pemerintah Non Kementerian yang diintegrasikan ke unit kerja Biro BMNP BRIN (Badan Riset dan Inovasi Nasional) sebanyak +-320 orang sesuai dengan SK penempatan pegawai.
- d. Pengolahan data menggunakan Bahasa pemrograman python.

### 1.4. Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah :

- a. Untuk mengetahui performa kombinasi algoritma SVM dan KNN
- b. Mengetahui variable yang tepat dalam melakukan optimasi dengan Teknik Feature Selection

- c. Mengukur dan membandingkan tingkat akurasi sebelum dan setelah hasil penerapan feature selection
- d. Melakukan prediksi hasil seleksi sehingga dapat memberikan rekomendasi dalam pemetaan pegawai agar kinerja organisasi lebih efisien

### **1.5. Manfaat Penelitian**

Penelitian ini diharapkan memberikan manfaat diantaranya:

- a. Menambah pengetahuan tentang teknik kombinasi algoritma SVM dan KNN dalam melakukan seleksi pegawai
- b. Memberikan pengetahuan tingkat akurasi sebelum dan setelah penerapan feature selection
- c. Dapat memberikan rekomendasi kepada pengambil kebijakan instansi dalam seleksi pegawai
- d. Meningkatkan efektivitas dan efisiensi organisasi dalam penempatan pegawai yang sesuai dengan kompetensi

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Teknik data mining dan machine learning dapat digunakan untuk melakukan prediksi berdasarkan data-data masa lalu yang digunakan sebagai data latih (Kotu & Deshpande, 2014). Salah satu metode yang dapat digunakan dalam melakukan prediksi seleksi karyawan adalah klasifikasi, seperti algoritma Naïve Bayes, KNN, Random Forest, Decision Stamp, Decision Tree, Rule Induction, Linear Regresi, Linear Discriminant Analysis, Neural Network, dan SVM dengan menghasilkan akurasi tertinggi pada algoritma SVM yaitu 65% namun masih jauh dari akurasi yang diinginkan (Saifudin, 2018). Dengan melakukan pembagian dataset pada tahap preprocessing menjadi data training dan data testing sebesar 70% 30%, dan atribut dataset yang berbeda menghasilkan akurasi yang berbeda (Reddy, Regella, Seelam, 2020). Melakukan normalisasi data pada tahap preprocessing juga penting dilakukan karena merupakan cara standarisasi data dengan memetakan sebuah nilai kedalam rentang 0 dan 1 dengan menggunakan metode min max normalization (Riani, Andreswari, Fauzi, 2021).

Penelitian lain dengan topik serupa, yaitu prediksi klasifikasi rekomendasi minat menggunakan algoritma KNN, dimana jarak antara data baru dengan data tetangga terdekat dikelompokkan dan dihitung dengan penggunaan Teknik metode Euclidean dan Hamming distance. (Prasetyo, Kusriani, Arief, 2019). Menentukan Parameter k yang terbaik pada algoritma KNN akan menghasilkan nilai akurasi yang

maksimal. Sedangkan pada algoritma SVM, dalam melakukan perhitungan klasifikasi menggunakan salah satu atau beberapa jenis Fungsi kernel seperti Kernel Kuadrat, Kubik, Kernel linier, dan Kernel Gaussian. Perbedaan skala kernel dilakukan verifikasi menggunakan fleksibilitas model yang berbeda. Kinerja dari fungsi kernel diverifikasi dengan menggunakan confusion matrix seperti presisi, sensitivity, specificity, dan F-measure selain akurasi. Kecuali kernel kubik, semua kernel lainnya menghasilkan akurasi lebih dari 85% (Karri.,dkk, 2021). Untuk mendapatkan performansi model klasifikasi yang optimal pada SVM terdapat banyak cara. Salah satunya yang dilakukan Michael, A (2022) pada penelitiannya dengan teknik kombinasi pre-trained model pada lapisan ekstraksi fitur dan SVM dengan tuning hyperparameter GridsearchCV mampu menghasilkan performa terbaik sebesar 96% pada nilai  $C=0.001$  dan kernel Linear. Penelitian serupa yang dilakukan Nica Astrianda (2020) dengan menggunakan ekstraksi fitur pada Support Vector Machine pada klasifikasi dua kelas dengan 54 data uji dan 20 data latih, menghasilkan akurasi sebesar 100%.

Selain itu dapat juga dilakukan dengan penerapan feature selection, dimana pada implementasinya dilakukan pemilihan atribut yang terbaik dari semua atribut yang ada. Dengan penerapan feature selection pada algoritma SVM dihasilkan peningkatan akurasi menjadi 95,17% (Sasongko, Arifin, 2019). Langkah awal dalam proses pemilihan atribut dengan feature selection dengan teknik forward selection diawali dengan empty model, selanjutnya tiap variable diinputkan hingga tercapainya kriteria kombinasi model yang terbaik dengan formulasi tahapan seperti pada gambar 1.



1. Membuat empty set:  $Y_k = \{\emptyset\}$ ,  $k = 0$
2. Memilih feature terbaik:  

$$X^+ = \arg \max_{x \in Y_k} [f(Y_k + X^+)]$$
3. Jika  $f((Y_k + x^+)) > f((Y_k))$ 
  - a. Update  $Y_{k+1} = Y_k + x^+$
  - b.  $k = k + 1$
  - c. Kembali ke step -2

Gambar 2.1. Pseudo Code Forward Selection

Sedangkan penerapan Feature selection pada algoritma Neighbor Weighted KNN yaitu dengan menggunakan n-jumlah fitur sebagai threshold. Fitur-fitur yang diujikan merupakan hasil seleksi fitur yang berurutan dari nilai yang paling besar ke paling kecil, dengan nilai tertinggi adalah 1 sedangkan yang berada dibawah n-jumlah fitur akan dieliminasi, selanjutnya dilakukan pengujian dengan K-fold cross validation dan didapatkan rata-rata nilai akurasi 60%, rata-rata presisi 72%, rata-rata recall 67%, rata-rata F-measure 66% dengan parameter K=45, fitur=9 dan nilai E (eksponen) = 6 (Azizah., dkk, 2022).

Performa model akurasi dikatakan sangat baik apabila grafik kurva ROC mendekati angka 1 berdasarkan standar katagori klasifikasi (Riani, Andreswari, Fauzi, 2021).

Penelitian lainnya yang dilakukan Paulina, dkk (2021) dengan melakukan seleksi fitur menggunakan K-Nearest Neighbor dalam klasifikasi kerusakan mutu tomat menghasilkan akurasi pada data latih sebesar 86,6% dan 70% pada data uji. Perbedaan tingkat akurasi dipengaruhi oleh data yang digunakan pada proses pelatihan dan pengujian, sehingga perlu diperhatikan data uji yang digunakan merepresentasikan data pelatihan.

Dalam meningkatkan akurasi pada klasifikasi data dengan teknik kombinasi KNN dan Relief-F yang dilakukan Yusra, dkk (2021) menghasilkan akurasi yang



maksimal dibandingkan hanya menggunakan KNN. Penggunaan seleksi fitur pada KNN sebagai solusi atas atribut yang kurang signifikan dan persentase pengaruh yang rendah dari suatu data dalam menentukan kelas pada data baru. Dengan melalui tahapan preprocessing yaitu cleansing, transformasi dan seleksi atribut mampu menghasilkan akurasi sebesar 95.63% dibanding hanya menggunakan KNN sebesar 85.31%. Selain itu pembagian data train test split serta jumlah data latih juga mempengaruhi performa dari akurasi seperti penelitian yang dilakukan oleh Fredilio (2023) dengan menggunakan algoritma KNN pada data latih sebanyak 500 data dan pembagian data 80:20 menghasilkan akurasi sebesar 86%.

Selanjutnya penelitian yang dilakukan Ginting, dkk (2018) dengan algoritma Support Vector Machines yang dikombinasikan dengan K-Nearest Neighbors dalam menentukan kelayakan nasabah Bank dimana tahapan penelitian yang digunakan yaitu SVM bekerja dengan cara menemukan hyperplane yang optimal dan support vector. Selanjutnya support vector digunakan sebagai data latih pada algoritma KNN untuk melakukan klasifikasi data nasabah bank. Dengan 2000 data latih dan 103 data uji serta melakukan tuning parameter yang optimal pada nilai  $C=0.1$ ,  $\gamma=2$ , sistem mengidentifikasi 1998 support vector, kemudian dengan nilai  $K=16$  sistem memberikan hasil performa sebesar 88,35%.

## 2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian

**Implementasi Metode Feature Selection pada kombinasi algoritma SVM dan KNN  
dalam melakukan Seleksi Klasifikasi Pegawai  
(studi kasus user aplikasi SAKTI pada Badan Riset dan Inovasi Nasional)**

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Penerapan Algoritma K Nearest Neighbor Untuk Rekomendasi Minat Konsentrasi Di Program Studi Teknik Informatika Universitas PGRI Yogyakarta	Adi Prasetyo, Kusri, M Rudyanto Arief; Jurnal Informatika dan Teknologi Informasi; Januari 2019	Penelitian ini bertujuan untuk pemilihan minat dengan menggunakan data siswa menggunakan Case Base Reasoning (CBR).	Dari Hasil prediction dengan Perhitungan algoritma K-NN yang diterapkan terhadap data mahasiswa sebanyak 100 mahasiswa diperoleh hasil klasifikasi dengan prosentase akurasi sebesar 95,89 %, dengan nilai K=7 sehingga dapat disimpulkan bahwa hasil klasifikasi ini dapat digunakan untuk merekomendasikan kepada universitas PGRI dalam menentukan konsentrasi yang sesuai untuk mahasiswa.	Peneliti selanjutnya mungkin bisa menggunakan algoritma lebih dari satu sebagai perbandingan	Penelitian yang dirujuk hanya menggunakan K-Nearest Neighbor sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
2	Classification and Prediction of Driving Behaviour at a Traffic Intersection Using SVM and KNN	Soni Lanka Karri, Lyanage C. De Silva, Daphne Teck Ching Lai, Shaiw Yin Yong; A Springer Nature Journal; April 2021	Penelitian ini bertujuan untuk menilai kinerja menggunakan teknik pembelajaran mesin SVM dan KNN diimplementasikan untuk memvalidasi klasifikasi perilaku mengemudi dalam hal berhenti aman/tidak aman di persimpangan bersinyal pada sinyal kuning	Pendekatan SVM diimplementasikan untuk mengklasifikasikan perilaku mengemudi dalam hal berhenti aman/berhenti tidak aman di persimpangan bersinyal pada permulaan lampu kuning. Kinerja klasifikasi dengan teknik pembelajaran mesin lainnya seperti teknik KNN dan Linear Diskriminan diberikan masing-masing 90,1% dan 89,4%. Hasil keseluruhan menunjukkan bahwa teknik SVM dapat menyimpulkan perilaku berhenti pengemudi dengan tingkat kinerja yang tinggi.	Penelitian selanjutnya diharapkan Dataset yang digunakan dapat lebih dikembangkan, dari segi jumlah data maupun parameter atau atributnya	Penelitian yang dirujuk menggunakan perbandingan K-Nearest Neighbor dan SVM sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor

Tabel 2.1 (Lanjutan)

3	Metode Data Mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di universitas Pamulang	Aries Saifudin; Jurnal Teknologi; 2017	Penelitian ini bertujuan untuk mendapatkan model prediksi ketepatan waktu lulus calon mahasiswa	Berdasarkan hasil implementasi dan model yang diuji diperoleh algoritma/model terbaik, yaitu Support Vector Machine (SVM) dengan akurasi 65,00%. Tetapi akurasi ini masih jauh dari nilai excellent (sangat baik).	Penelitian selanjutnya diharapkan Penentuan data training dapat mempengaruhi hasil pengujian. Pola dari data training tersebut digunakan sebagai rule dalam proses testing dikarenakan faktor penentu diterimanya calon mahasiswa memiliki nilai yang kurang konsisten.	Penelitian yang dirujuk hanya menggunakan SVM sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
4	Klasifikasi Kinerja Akademik Siswa Menggunakan Neighbor Weighted KNN dengan seleksi fitur Information Gain	Rizky Adinda Azizah, Fitra A. Bachtiar, Sigit Adinugroho; Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK); Juni 2022	Penelitian ini bertujuan untuk membantu dalam melakukan klasifikasi terhadap kinerja akademik siswa dengan menggunakan Neighbor Weighted KNN dengan seleksi fitur Information Gain	Berdasarkan pengujian dan analisis penelitian, didapatkan nilai akurasi terbaik sebesar 60,4%, dengan nilai <i>precision</i> adalah 0,719, nilai <i>recall</i> sebesar 0,676, dan nilai <i>f-measure</i> diperoleh adalah 0,661. Nilai tersebut dihasilkan saat menggunakan 9 fitur yaitu <i>VisitedResource</i> , <i>StudentAbsenceDay</i> , <i>RaisedHands</i> , <i>AnnouncementsView</i> , <i>Relation</i> , <i>ParentsAnsweringSurvey</i> , <i>Discussion</i> , <i>NationalITY</i> , dan <i>PlaceofBirth</i> dimana fitur tersebut memperoleh nilai <i>Gain</i> tertinggi dari urutan <i>Gain</i> keseluruhan fitur, dengan nilai $Gain \geq 0,1182$ dan menggunakan nilai parameter optimal yaitu nilai $E = 6$ , dan nilai $K = 45$	Pada penelitian selanjutnya dapat mempertimbangkan untuk menggunakan kombinasi dari algoritme seleksi fitur yang lain selain metode Information Gain dan bisa menerapkan algoritme klasifikasi lain yang lebih sesuai untuk menangani dataset yang memiliki atribut-atribut dengan sifat berbeda/mixed attribute dataset	Penelitian yang dirujuk menggunakan K-Nearest Neighbor dengan seleksi fitur information Gain sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor

Tabel 2.1 (Lanjutan)

5	Implementasi Metode Forward Selection pada algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier Kernel Density (studi Kasus Klasifikasi Jalur Minat SMA)	Theopilus Sasongko, Bayu Oki Arifin, Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK); Mei 2019	Penelitian ini bertujuan untuk membandingkan metode <i>forward selection</i> pada Algoritma SVM dan Naive Bayes Kernel Density.	Hasil dari penelitian ini diantaranya adalah <i>feature subset dataset</i> SMA ABC yang terpilih yaitu nilai IPA, tes akademik, abstrak konseptual, Analisa sintesa, dan logika numerik, sedangkan <i>feature subset</i> SMA XYZ yaitu nilai IPA, logika numerik, dan Analisa sintesa. Hasil pengujian <i>dataset</i> SMA ABC menggunakan algoritma FS-SVM berbasis <i>kernel anova</i> parameter C=10,0 sebesar 99,29%. Sedangkan hasil pengujian <i>dataset</i> SMA XYZ menggunakan algoritma FS-SVM berbasis <i>kernel anova</i> parameter C=10,0 sebesar 95,17%.	Pada penelitian selanjutnya diharapkan dapat menggunakan <i>automatic kernel selection</i> pada SVM. Kemudian dapat menggunakan <i>automatic parameter selection</i> pada parameter factor pinalti (C) di algoritma SVM dan <i>minimumbandwidth</i> pada algoritma <i>Naive bayes kernel density</i> .	Penelitian yang dirujuk menggunakan perbandingan SVM dan Naive Bayes Classifier Kernel Density dengan <i>forward selection</i> sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
6	Recruitment prediction using machine Learning	Jagan Mohan Reddy, Sirisha Regella; IEEE; 2020	Tujuan penelitian ini adalah untuk memprediksi bergabungnya kandidat yang efisien sebelum pemilihan resume dan total proses harus dilakukan dengan cara yang efisien dengan biaya minimal dan waktu minimal	Prediksi perekrutan karyawan penting untuk pertumbuhan perusahaan. Membuat keputusan perekrutan yang baik tidak hanya meningkatkan retensi karyawan tetapi juga mengurangi biaya yang terkait dengan perekrutan yang buruk. Kompetensi yang efektif memastikan rekrutmen berkualitas dari bakat yang tepat untuk mendorong kinerja bisnis organisasi. Dalam penelitian ini kami melakukan model pembelajaran mesin klasifikasi biner: peluang atau tingkat kandidat mundur setelah seleksi dan membuat prediksi yang efektif untuk proses rekrutmen di masa mendatang. Faktor-faktor inilah yang dapat membuat kandidat mundur dari pekerjaan yang dipilihnya ditentukan.	Penelitian ini pada hasil akhirnya akan membantu memprediksi kandidat yang nantinya bergabung	Penelitian yang dirujuk mengenai menggunakan perbandingan Decision Tree, Random Forest, G-Naive Bayes dan KNN sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor



Tabel 2.1 (Lanjutan)

7	Kombinasi K-Means Dan Support Vector Machine (SVM) Untuk Memprediksi Unsur Sara Pada Tweet	Wiga Maulana Baihaqi, Mulasari Pinilih, Miftakhul Rohmah; Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK); Juni 2020	Tujuan penelitian ini adalah mengusulkan metode <i>k-means</i> untuk proses <i>labeling</i> dan SVM dalam proses klasifikasi apakah konten pada twitter memiliki unsur SARA atau tidak.	Berdasarkan hasil dari penelitian yang telah dilakukan, Kombinasi <i>k-means</i> dan <i>support vector machine</i> (SVM) sudah berhasil untuk menganalisis <i>tweet</i> yang positif mengandung SARA dan <i>tweet</i> yang negatif mengandung SARA. Hasil yang diperoleh berdasarkan <i>k-means</i> antara lain 118 <i>tweet</i> positif SARA dan 83 <i>tweet</i> negatif SARA. Dalam proses klasifikasi menggunakan dua metode validasi, yaitu <i>5-fold cross validation</i> yang dibandingkan dengan <i>10-fold cross validation</i> , hasil akurasi dari kedua metode validasi tersebut yaitu masing-masing 64,18% dan 63,68%. Berdasarkan hasil akurasi yang diperoleh untuk meningkatkan hasil akurasi, data hasil proses <i>k-means</i> diolah kembali dengan validasi pakar bahasa, hasil yang diperoleh menjadi 139 <i>tweet</i> positif SARA dan 62 <i>tweet</i> negatif SARA, hasil akurasi meningkat menjadi 70,15% dan 71,14%	Perlu adanya perbaikan metode yang disulkan untuk memperoleh hasil yang lebih akurat, baik dalam mengelompokan dan klasifikasi data twitter yang mengandung unsur SARA dan tidak mengandung unsur SARA. Selain itu, sebaiknya dibangun sistem yang dapat diterapkan untuk menganalisis konten twitter.	Penelitian yang dirujuk menggunakan Kombinasi K-Means dan SVM, sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
---	--	--	---	--	--	---

Tabel 2.1 (Lanjutan)

8	Hybrid Classifier System: Support Vector Machines Dikombinasikan dengan K-Nearest Neighbors untuk Menentukan Kelayakan Nasabah Bank dalam Pengajuan Kredit	S L Br Ginting, A A Permana; Jurnal Sistem Komputer ; 2018	Penelitian ini bertujuan membangun aplikasi yang dapat menganalisis data nasabah bank kemudian menentukan kelayakan nasabah tersebut dalam hal pemberian pinjaman, agar terhindar dari masalah kredit macet dikemudian hari.	SVM bekerja dengan cara menemukan hyperplane yang optimal dan support vector. Lebih lanjut, algoritma KNN akan melakukan klasifikasi data nasabah bank berdasarkan pengidentifikasian support vector tersebut. Dengan 2000 data latih dan 103 data uji, nilai parameter $cost=0.1$ , $gamma=2$ , sistem mengidentifikasi 1998 support vector, kemudian dengan nilai $K=16$ sistem memberikan hasil 88,35% data yang cocok (91 data dari 103). Dapat disimpulkan bahwa aplikasi ini bekerja dengan cukup baik dan dapat membantu credit analyst dalam merekomendasikan nasabah yang layak memperoleh pinjaman	Dari penelitian ini ditemukan faktor-faktor yang mempengaruhi hasil dari klasifikasi yaitu parameter cost dan gamma yang merupakan parameter pada metode SVM dan parameter K pada metode KNN.	Penelitian yang dirujuk menggunakan Kombinasi K-Nearest Neighbor dan SVM sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
9	Predicting Employee Attrition Using Machine Learning Techniques	Francesca Fallucchi, Marco Coldangelo dan Ernesto William De Luca : MDPI November 2020	Tujuan dari studi ini adalah untuk menganalisis bagaimana factor obyektif mempengaruhi gesekan karyawan, untuk mengidentifikasi penyebab utama yang berkontribusi pada keputusan pekerja untuk meninggalkan perusahaan, dan untuk dapat memprediksi apakah karyawan tertentu akan meninggalkan perusahaan.	Hasilnya dinyatakan dalam bentuk metrik evaluasi menggunakan Confusion Matrix yang membandingkan algoritma Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Logistic Regression, KNN, Decision Tree, Random Forest, SVC dan Linear SVC. Berdasarkan tujuannya untuk memprediksi jumlah pegawai yang dapat meninggalkan perusahaan dengan meminimalkan jumlah <i>False Negative</i> , maka Gaussian Naïve Bayes diidentifikasi sebagai algoritma klasifikasi terbaik yang mampu mencapai tujuan analisis.	Pada penelitian selanjutnya untuk meningkatkan analisis dengan mempertimbangkan peluang karyawan baru serta kondisi kerja yang merugikan (misalnya, bahaya dan bahaya) dan prospek promosi yang buruk, diskriminasi dan dukungan sosial yang rendah, yang berhubungan positif dengan niat keluar karyawan	Penelitian yang dirujuk menggunakan perbandingan Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Logistic Regression, KNN, Decision Tree, Random Forest, SVC dan Linear SVC sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor.



Tabel 2.1 (Lanjutan)

10	A Hybrid Convolutional Neural Network And Support Vector Machine For Dysarthria Speech Classification	Hanifa Dyoniputri dan Afiabayati; International Journal of Innovative Computing, Information and Control; Februari 2021	Tujuan penelitian ini adalah untuk memperkenalkan model hybrid CNN-SVM untuk mengenali angka 10 digit yang diucapkan oleh penderita disartria.	Penelitian ini menunjukkan bahwa menggunakan CNN dengan SVM sebagai pengklasifikasi mencapai skor rata-rata akurasi klasifikasi yang lebih baik (94,29%), skor 7,5% lebih tinggi dari CNN sederhana dan 3,64% lebih tinggi dari CNN penelitian sebelumnya.	merekomendasikan untuk menggunakan kata yang bergantung pada pembicara dan terisolasi untuk mengklasifikasikan ucapan disartria di masa mendatang	Penelitian yang dirujuk menggunakan kombinasi CNN dan SVM sedangkan penelitian ini menggunakan metode seleksi fitur Backward Elimination, dan kombinasi SVM-K-Nearest Neighbor
----	---	---	--	--	---	--

### 2.3. Landasan Teori

Pada penelitian ini, diperlukan definisi dengan tujuan untuk melakukan pendalaman materi dan mempermudah dalam penulisan sehingga penyelesaian permasalahan lebih terarah, diantaranya definisi yang digunakan sebagai berikut:

#### 2.3.1 Aplikasi SAKTI

Sistem Aplikasi Keuangan Tingkat Instansi (SAKTI) adalah aplikasi yang digunakan sebagai sarana bagi unit kerja/ satuan kerja yang berada baik dibawah Pengguna Barang maupun Pengelola Barang yang dibuat oleh Kementerian Keuangan dalam mendukung implementasi Sistem Perbendaharaan dan Anggaran Negara (SPAN) untuk melakukan pengelolaan keuangan. Tahapan dalam melakukan pengelolaan keuangan negara dimulai dari tahapan perencanaan, pelaksanaan hingga pertanggungjawaban anggaran.

SAKTI mengintegrasikan seluruh aplikasi satker yang ada. Aplikasi SAKTI merupakan aset sistem informasi yang penting bagi Kementerian Keuangan selaku Pengelola Barang dalam menyajikan Laporan Keuangan Pemerintah Pusat (LKPP). Tidak berfungsinya aplikasi SAKTI dapat memberikan implikasi tersendat bahkan tidak cairnya dana APBN, seperti operasional kementerian, pembiayaan proyek, sampai kepada pembayaran gaji Pegawai Negeri yang merupakan hajat hidup banyak orang (Supriestiwadi, 2017). Single Database, single entry point dan akutansi berbasis akrual merupakan konsep dari aplikasi SAKTI. Aplikasi SAKTI digunakan oleh entitas akuntansi dan entitas pelaporan Kementerian Negara/Lembaga. Seluruh

Transaksi entitas akuntansi dan entitas pelaporan dilakukan secara sistem elektronik. Periodeisasi pelaporan dan rekonsiliasi data atas transaksi dalam SAKTI dilakukan sebanyak 14 kali meliputi Januari sampai dengan Desember, unaudited, dan audited.

Aplikasi SAKTI mencakup seluruh proses pengelolaan keuangan negara pada Unit Kerja di setiap Kementerian/Lembaga dimulai dari proses Penganggaran, Pelaksanaan, sampai dengan Pelaporan. Masing-masing proses pengelolaan keuangan diperankan oleh modul-modul aplikasi sebagai berikut :

1. Proses penganggaran diperankan oleh modul Penganggaran.
2. Proses pelaksanaan diperankan oleh beberapa modul, yaitu modul Komitmen (meliputi sub-modul Manajemen Supplier dan sub-modul Manajemen Komitmen), modul Bendahara, modul Aset Tetap, modul Persediaan, dan modul Pembayaran.
3. Proses pelaporan diperankan oleh modul GL dan Pelaporan.

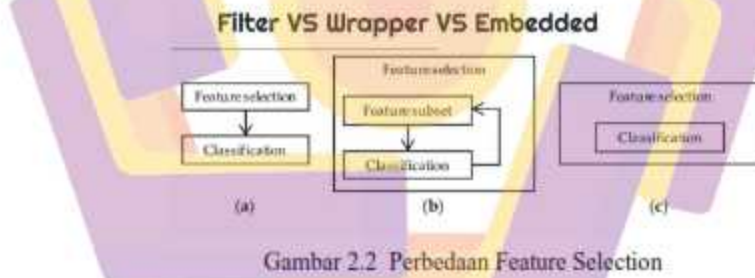
### **2.3.2 Feature Selection**

Dalam penerapan algoritma machine learning, hal paling mendasar yang diperlukan adalah ketersediaan dataset. Agar dataset yang tersedia memiliki atribut yang relevan dan tidak memperburuk data sehingga bernilai negative maka perlu dilakukan feature selection. Feature selection adalah suatu proses untuk menghapus fitur yang berlebihan dan tidak relevan dari dataset awal, sehingga waktu yang digunakan dalam melakukan eksekusi proses data menjadi lebih singkat dan adanya peningkatan akurasi karena features yang tidak relevan

sudah tereliminasi (Rahmansyah,dkk,2018). Seleksi fitur merupakan cara efektif untuk melakukan reduksi data dan merupakan langkah penting yang perlu dilakukan agar dapat memperoleh kinerja yang baik (Yusra, 2021). Tujuan dilakukannya seleksi fitur adalah:

- a. Menghasilkan model yang lebih sederhana
- b. Mengurangi waktu training
- c. Menghindari overfitting sehingga akurasi meningkat

Menurut Suppers et al., (2018) terdapat tiga katagori metode seleksi fitur tergantung pada interaksinya dengan model pembelajaran, yaitu metode Filter, metode wrapper dan metode embedded. Secara bagan, perbedaan dari ketiga jenis kelompok dapat terlihat pada gambar 2.



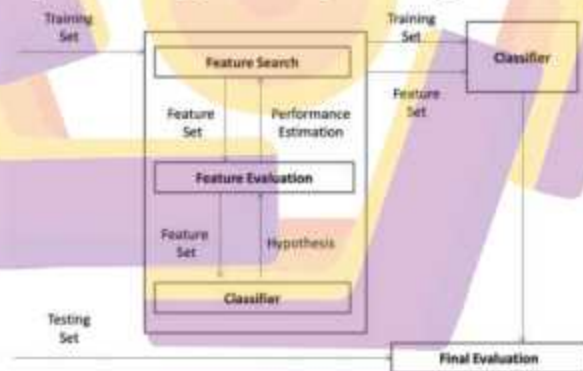
Berdasarkan gambar 2 (a) Metode filter melakukan pemilihan fitur secara independen tanpa konstruksi model klasifikasi; (b) Metode Wrapper secara iteratif menambah atau menghilangkan sekumpulan fitur menggunakan akurasi prediksi dari model klasifikasi; dan (c) Metode Embedded melakukan seleksi fitur yang menjadi bagian tidak terpisahkan dari model klasifikasi.

### 2.3.2 Wrapper Method

Menurut Malekipirbazari, et al., (2021) pada Metode seleksi fitur wrapper membutuhkan sebuah algoritma Machine Learning yang digunakan sebagai black box evaluator dalam proses seleksi atau menemukan subset fitur terbaik sehingga metode ini sangat bergantung pada jenis classifier. Secara umum model pada metode wrapper akan bekerja seperti berikut:

1. Mencari subset fitur,
2. Mengevaluasi subset fitur yang dipilih dengan mengandalkan kinerja Classifier,
3. ulangi langkah 1 dan 2 sampai kualitas yang diinginkan tercapai.

Terdapat tiga komponen utama dalam melakukan seleksi fitur dengan penerapan metode wrapper dan ditunjukkan oleh gambar 3.



Gambar 2.3. Komponen Utama Metode Seleksi Fitur dengan penerapan Wrapper



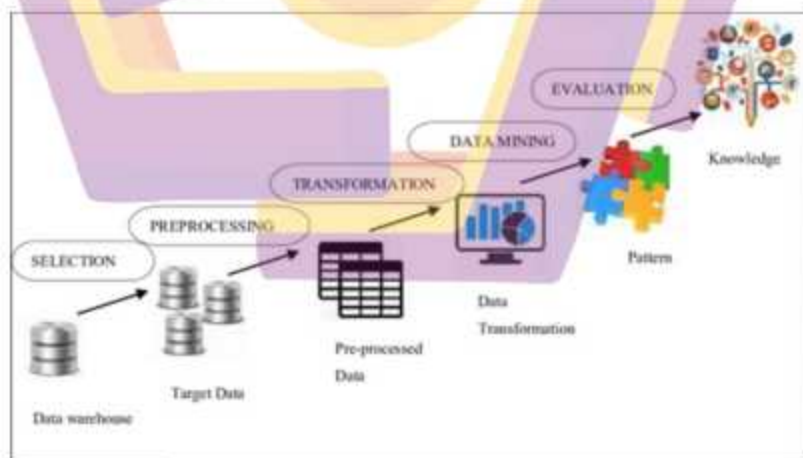
### 2.3.2.1 Backward Elimination

Backward Elimination merupakan suatu metode yang termasuk kedalam feature selection tipe wrapper yang memiliki fungsi untuk mengoptimalkan kinerja suatu model. Pemilihan variabel dilakukan dengan cara pemilihan kedepan yakni menguji semua variabel kemudian menghapus variabel-variabel yang dianggap tidak signifikan. (Bode, 2017). Setiap fitur akan diproses satu per satu, jika fitur dianggap tidak berpengaruh atau berpengaruh tapi tidak signifikan dalam model maka akan dihapus. Iterasi akan diulangi sampai penghapusan fitur tidak lagi memberikan perbaikan terhadap model. (Ary dan Rismati, 2019). Langkah-langkah dalam melakukan penerapan Backward Elimination adalah sebagai berikut:

1. Untuk setiap  $n$  fitur, akan ada sebuah Predefined Criterion Function.
2. Pada setiap Langkah, satu fitur akan dibuang dan fungsi criterion akan ditemukan untuk semua subset yang berisi  $n-1$  fitur.
3. Berdasarkan fungsi criterion function tersebut, fitur yang mempunyai performa paling rendah akan dibuang.
4. Prosedur ini akan terus diulang sampai tersisa fitur yang telah ditentukan sebelumnya.

### 2.3.3. Data Mining

*Data Mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. *Data Mining* didefinisikan sebagai proses penemuan pola dalam data. Beberapa aplikasi *data mining* fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Adnyana, 2019). Data mining merupakan bagian dari proses *Knowledge Discovery from Data (KDD)*. proses penjelajahan pengetahuan dimulai dari beberapa database dilakukan proses *cleaning* dan *integration* sehingga menghasilkan data *warehouse*. Selanjutnya dilakukan proses *selection* dan *transformation* yang kemudian disebut sebagai data mining hingga menemukan pola dan memperoleh pengetahuan dari data (*knowledge*). Untuk lebih jelasnya, berikut gambaran dari metode KDD (Gullo,2015).



Gambar 2.4. Komponen Pada Metode KDD

### 1. Selection

Dalam tahapan ini dilakukan seleksi variable yang memiliki tujuan utama adalah untuk membuat dataset target dari data asli, yaitu memilih subset dari variabel atau sampel data.

### 2. Preprocessing

Tahapan Preprocessing ini digunakan untuk "membersihkan" data sehingga data yang nantinya akan diolah adalah data yang tepat. Banyak cara yang dapat dilakukan seperti pemodelan kebisingan dan penghapusan, maupun mendefinisikan strategi yang tepat untuk menangani bidang data yang hilang.

### 3. Transformation

Untuk dapat mengurangi dan memproyeksikan data dalam rangka memperoleh representasi yang sesuai terhadap target yang akan dilakukan maka dilakukan tahapan transformasi ini, karena dengan Teknik transformasi ini biasanya dapat menemukan representasi invarian dari data.

### 4. Data mining

Data mining merupakan tahapan yang berkaitan dengan melakukan ekstraksi pola menarik dalam pemilihan:

- Metode atau tugas data-mining tertentu (klasifikasi, pengelompokan, regresi, dan sebagainya),
- Algoritma yang tepat untuk melakukan tugas yang dihadapi, dan
- Representasi yang sesuai.

## 5. Evaluation/ Interpretation

Pada tahapan ini merupakan tahapan dimana suatu data dapat memberikan pengetahuan bagi pengguna sehingga memberikan manfaat untuk menafsirkan dan mengekstrak pengetahuan dari pola yang ditambang, dengan memvisualisasikan pola; interpretasi ini biasanya dilakukan dengan memvisualisasikan pola, model, atau data yang diberikan.

### 2.3.4. Supervised Learning

Supervised Learning atau Pembelajaran dengan pengawasan adalah algoritma pembelajaran mesin yang mempelajari suatu fungsi berdasarkan contoh yang sudah diajarkan sebelumnya. Algoritma yang termasuk kedalam Supervised learning agar dapat bekerja perlu dilatih terlebih dahulu sehingga dapat melakukan prediksi, klasifikasi dengan menggunakan data latih (Tinuk, 2021).

### 2.3.5 Algoritma SVM

Support Vector Machine atau lebih dikenal SVM merupakan salah satu metode Machine Learning dengan pendekatan Supervised Learning yang digunakan untuk melakukan klasifikasi. Algoritma ini mengelaskan data baru dengan memisahkan berdasarkan hyperplane dalam ruang N-dimensi. *Hyperplane* merupakan sebuah fungsi yang digunakan sebagai pemisah antar kelas yang satu dengan yang lain. SVM dikembangkan berdasarkan Statistical Learning Theory.

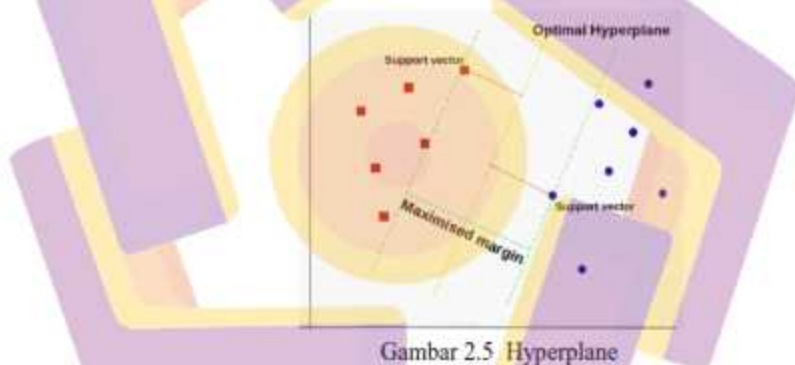
Beberapa karakteristik SVM diantaranya:

- 1.1. Bersifat non probabilistic

- 1.2. Membagi data menjadi dua kategori (binary)
- 1.3. Masing-masing kelompok data dibatasi oleh hyperplane.

Tujuan algoritma SVM adalah menemukan optimal hyperplane (fungsi pemisah) yang memisahkan dua buah kelas, dengan cara memaksimalkan margin antara dua class. Margin adalah jarak antara hyperplane dengan input vectors (data) terdekat (Desiani, 2022). Input vectors yang bersentuhan dengan boundary disebut support vector.

Support vector adalah input vector paling luar dari masing-masing kelas, seperti pada gambar 4.



Gambar 2.5 Hyperplane

Dari gambar, kita dapat melihat support vector, maximized margin, dan hyperplane, dimana terlihat support vectors menyentuh boundary (garis putus-putus) setiap kelas. Sedangkan hyperplane berada tepat di tengah dua buah boundary ini.

Dalam algoritma Support Vector Machine, terdapat beberapa jenis

Kernel, yaitu:

- a. Kernel Linear, digunakan untuk dataset linear.
- b. Kernel Polynomial, digunakan untuk dataset normal.



- c. Kernel Radial Basis Function (RBF) atau Gaussian, digunakan untuk dataset yang tidak terpisah secara linear. Jenis ini merupakan kernel yang paling banyak digunakan karena akurasi yang tinggi.
- d. Kernel Sigmoid, merupakan pengembangan dari jaringan saraf tiruan.

### 2.3.6 Algoritma KNN

K-nearest neighbour atau biasa disingkat KNN merupakan salah satu algoritma populer dan termasuk kedalam kategori supervised learning. KNN adalah suatu metode yang bertujuan membuat klasifikasi suatu data yang dilihat dari lokasi terdekat k dengan suatu objek yang didasarkan pada data training. Data dibagi menjadi data training dan data testing, setelah itu dilakukan penghitungan jarak dari data testing ke data training. Metode algoritma KNN banyak dipilih karena lebih sederhana, mudah dipahami, data input tersedia, parameter yang diperlukan sedikit, yaitu jumlah tetangga yang terdekat (K). Terdapat beberapa metode dalam melakukan perhitungan jarak ke tetangga terdekat (K) yaitu:

- a. Euclidean Distance
- b. Manhattan Distance
- c. Hamming Distance
- d. Minkowski Distance

Cara kerja pada algoritma KNN baik dalam penentuan klasifikasi maupun regresi adalah sebagai berikut:

1. Menentukan jumlah tetangga terdekat (K)
2. Menghitung jarak antara data baru terhadap semua titik data dalam data set
3. Melakukan urutan jarak dari terdekat sampai yang terjauh dalam data set, kemudian ambil titik data dengan jarak terdekat
4. Melakukan perhitungan jumlah titik data K pada setiap kategori
5. Untuk penentuan klasifikasi, masukan data baru ke kelas dengan jumlah tetangga terdekat (K) terbanyak; pada penentuan regresi, ambil nilai rata-rata dari semua nilai pada tetangga terdekat yang ditentukan (K) sebagai output prediksi.

### 2.3.7 Evaluasi dan Validasi Model

Confusion matrix adalah suatu metode yang sering digunakan dalam melakukan perhitungan akurasi pada pembelajaran data mining. Metode ini melakukan perhitungan dengan menghasilkan 4 keluaran, yaitu: recall, precision, accuracy dan error rate. Cara kerja confusion matrix yaitu menilai kinerja model klasifikasi sesuai jumlah objek yang diprediksi dengan benar dan salah (Dewi, 2019).

Untuk menghitung nilai akurasi, recall, precision dan error rate dapat dilakukan dengan rumus perhitungan seperti pada rumus 1.

$$Precision = \frac{TP}{TP+FP} \quad (a)$$

$$Recall = \frac{TP}{TP+FN} \quad (b)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (c)$$

$$F1 - Score = \frac{2 \times recall \times precision}{recall + precision} \quad (d)$$

dimana:

- True positif (TP) merupakan nilai positif, baik nilai prediksi maupun nilai aktualnya.
- False positif (FP) merupakan nilai yang prediksinya positif dan nilai aktualnya negatif
- False negative (FN) merupakan nilai yang prediksinya negative dan nilai aktualnya positif
- True negative (TN) merupakan nilai negative, baik nilai prediksinya maupun nilai aktualnya.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1. Jenis, Sifat dan Pendekatan Penelitian**

Penelitian yang merupakan penelitian eksperimental, dimana penelitian dilakukan dengan melakukan kombinasi antara metode algoritma SVM dan KNN dengan penerapan feature selection dalam melakukan seleksi klasifikasi pegawai. Peneliti juga melakukan pengujian dengan membandingkan nilai akurasi yang didapatkan antara hasil kombinasi algoritma SVM dan KNN dengan implemmentasi feature selection dan kombinasi algoritma SVM dan KNN tanpa feature selection.

Dilihat dari segi sifatnya, penelitian ini merupakan penelitian deskriptif yang mana penelitian ini menjelaskan dampak implementasi feature selection pada kombinasi algoritma SVM dan KNN dalam melakukan seleksi klasifikasi pegawai layak atau tidaknya untuk menunjang organisasi berdasarkan kebutuhan. Sedangkan pendekatan yang dilakukan adalah pendekatan kuantitatif. Data kuantitatif didapatkan berupa hasil observasi, arsip data dan bersifat obyektif. Penelitian dilakukan terhadap variabel data pegawai yang dikumpulkan seperti variabel input nama, usia, jenis kelamin, tingkat pendidikan, pangkat/golongan, lokasi kerja, domisili, persentasi kehadiran, level kewenangan aplikasi SAKTI, hasil penilaian Kinerja Pegawai 2022, dan kelayakan pegawai menjadi database dan kemudian dilakukan menjadi data latih dan data uji pada kombinasi algoritma SVM dan KNN.

### 3.2. Metode Pengumpulan Data

Dalam penelitian ini dilakukan pengambilan data yang akan digunakan sebagai sumber data dengan dua buah metode yaitu:

1. Pengumpulan data primer, didasarkan atas SK Kepala BRIN tentang penempatan pegawai pada Biro Manajemen BMN, SK Sekretaris Utama BRIN tentang Penetapan User Aplikasi SAKTI, selanjutnya data pegawai didapatkan dengan permohonan data ke internal Biro SDM BRIN berupa persantese kehadiran WFO dan WFH, dan penilaian kinerja masing-masing pegawai pada Biro Manajemen BMN selama periode 2022. Disamping itu, dengan pemanfaatan web Intra BRIN digunakan untuk pengumpulan data pangkat/golongan, usia, tingkat pendidikan, lokasi kerja, lokasi tinggal dari masing-masing pegawai dan variabel-variabel lain yang dibutuhkan.
2. Pengumpulan data sekunder, berupa studi pustaka untuk mendapatkan hasil yang maksimal dan lebih detail dalam melakukan penelitian melalui buku ilmiah, referensi jurnal, buku-buku ilmiah dan informasi lainnya yang sesuai dengan judul penelitian.

### 3.3. Metode Analisa Data

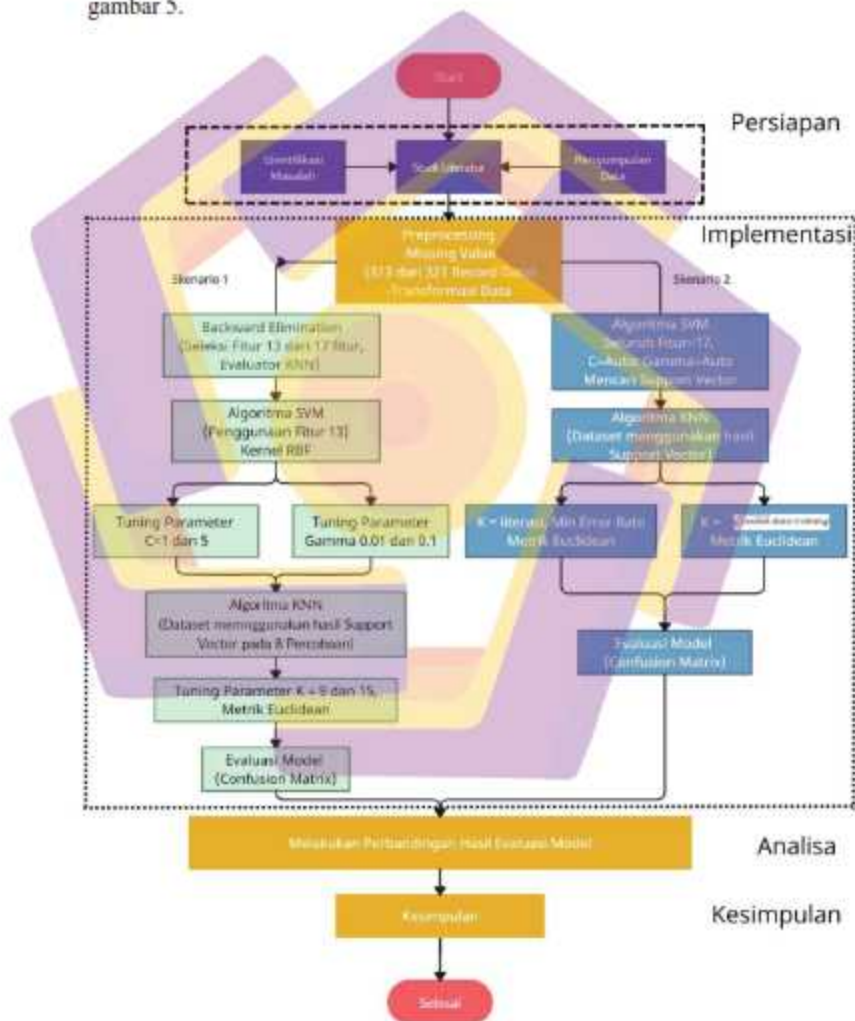
Dalam melakukan seleksi klasifikasi pegawai digunakan model Machine Learning dengan beberapa tahapan yang harus dilakukan. Langkah yang dilakukan adalah pengumpulan sumber data yang menjadi input yaitu data pegawai dengan file berformat .csv. Sumber data yang telah dibentuk selanjutnya melalui tahapan pre processing dengan tujuan menyederhanakan dan memilah data yang tidak memiliki nilai/null. Hal ini dilakukan agar tidak



mempengaruhi performa saat proses klasifikasi dilakukan. Berikutnya adalah pengujian data, dari keseluruhan dataset dibagi menjadi data training dan data testing, Proses selanjutnya adalah melakukan seleksi fitur menggunakan teknik Wrapper Method, Backward Elimination dengan tujuan untuk menemukan fitur atau variabel terbaik yang akan mempengaruhi tingkat akurasi. Pada tahap ini, peneliti melakukan perbandingan dataset pengurangan variabel hasil seleksi fitur dan dataset tanpa seleksi fitur untuk selanjutnya diproses dengan kombinasi SVM dan KNN dengan tujuan untuk mengetahui tingkat akurasi terbaik. Penggunaan kombinasi SVM-KNN untuk mempercepat proses komputasi. Dalam melakukan kombinasi algoritma, langkah awal adalah dengan menggunakan algoritma SVM yaitu mencari support vector yang berada dekat dengan hyperplane terbaik yang memisahkan 2 buah class. Hasil Support vector ini selanjutnya digunakan sebagai dataset untuk di proses menggunakan algoritma KNN. Dalam melakukan proses menggunakan algoritma KNN, peneliti akan menentukan nilai K (tetangga terdekat) terlebih dahulu, untuk selanjutnya melakukan proses klasifikasi data. Dalam menentukan nilai K akan didasarkan pada jumlah dataset yang akan diuji. Setelah pengujian menggunakan data training, kemudian dengan menggunakan data testing untuk menghasilkan peramalan seleksi klasifikasi pegawai. Hasil dari peramalan seleksi selanjutnya dilakukan evaluasi hasil dengan penerapan confusion matrix untuk memberikan pengetahuan model terbaik hasil perbandingan akurasi model kombinasi algoritma SVM dan KNN dengan dan tanpa implementasi metode Backward Elimination sehingga dapat disajikan menjadi sebuah kesimpulan yang lebih mudah untuk dipahami.

### 3.4. Alur Penelitian

Secara garis besar, dalam melakukan suatu penelitian maka diperlukan kerangka kerja sebagai upaya sistematis yang dilakukan peneliti agar dapat menjawab permasalahan-permasalahan yang terjadi seperti tampak pada gambar 5.



Gambar 3.1 Diagram Alur Penelitian

Terdapat Empat besaran proses diantaranya:

a. Persiapan

- Identifikasi Masalah, Studi Literatur, dan Pengumpulan Data (Dataset)

Tahap ini merupakan Langkah dalam menentukan rumusan masalah, batasan masalah dan latar belakang dari penelitian. Dalam hal ini mengamati variabel yang berhubungan dengan level kewenangan pengguna aplikasi SAKTI baik di level operator, validator, approver dan komitmen. Selanjutnya mempelajari dasar teori dari berbagai literatur melalui jurnal-jurnal mengenai penerapan kombinasi algoritma machine learning SVM dan KNN dan cara meningkatkan performanya dengan implementasi feature selection. Pada Tahap pengumpulan data merupakan prosedur sistematis yang digunakan untuk mengumpulkan dataset yaitu dengan melakukan observasi dan pengumpulan data pegawai yang didapatkan dari SK Penempatan Pegawai, SK Penetapan User Aplikasi SAKTI, penilaian kinerja dari Biro SDM, Intra BRIN Serta Biro Manajemen BMN dan Pengadaan yang nantinya digunakan untuk proses analisis data.

b. Implementasi

Sesuai dengan Sub bab Metode Analisa Data, pada tahap ini ditekankan kepada bagaimana proses pengolahan datanya. Data yang telah dikumpulkan, diolah menggunakan algoritma Machine learning dengan dengan Teknik kombinasi yaitu SVM dan KNN setelah melalui

proses seleksi fitur untuk menentukan seleksi pegawai pengguna aplikasi SAKTI. Alur kerja pada tahap ini terdiri dari:

### 1. Pra Proses Data

- Preprocessing

Pada tahap ini, dataset yang sudah terbentuk dilakukan pembersihan data dari data yang bisa mengganggu (noise) dan data yang tidak konsisten saat pemrosesan data. Artinya dataset yang telah diperoleh perlu diseleksi kembali terutama pada data yang tidak lengkap, tidak relevan, maupun duplikasi.

- Train Test Split

Selanjutnya dataset dibagi menjadi data latih dan data uji dengan pembagian data secara acak menggunakan `train_test_split` yang disediakan pada modul Sklearn dari library Python. Pada penelitian ini pembagian data training dan data latih adalah 70:30.

- Variabel Independen dan Dependen

Setelah dilakukan pembagian data, pada tahap ini digunakan untuk mendeklarasikan variabel yang digunakan sebagai input maupun variabel target.

### 2. Proses Data

- Skenario Penelitian

Karena jenis penelitian ini bersifat eksperimen, maka alur penelitian disediakan langkah skenario. Selain itu pada tahap ini juga

merupakan langkah dalam menentukan jumlah fitur dan nilai parameter-parameter yang diperlukan, diantaranya:

a) Seleksi Fitur Backward Elimination

Pada tahap ini, peneliti tidak melakukan tuning pembatasan fitur yang akan dicapai sehingga hasil eliminasi fitur yang akan digunakan pada proses selanjutnya merupakan hasil dari perhitungan kombinasi fitur yang memiliki performa terbaik. Algoritma Classifier yang digunakan sebagai evaluator adalah algoritma KNN.

b) Tuning Parameter SVM

Pada algoritma SVM, untuk mendapatkan optimal hyperplane diperlukan tuning pada parameter-parameter yang berpengaruh terhadap tingkat akurasi klasifikasi yaitu Kernel, parameter C dan parameter Gamma. Modul yang digunakan pada python untuk menjalankan fungsi SVM adalah scikit learn dengan menggunakan fungsi SVC (Support Vector Classifier). Pada Penelitian ini akan dilakukan tuning kombinasi parameter Kernel RBF pada nilai  $C = 1$  dan 5; dan nilai Gamma = 0.01 dan 0.1; untuk mendapatkan performa model yang terbaik. Pemilihan nilai C dan gamma didasarkan pada penelitian terdahulu, salah satunya yang dilakukan (Putri dkk, 2018), dari pengujian nilai C dari rentang 0.001-1 menghasilkan akurasi tertinggi pada  $c=1$ , sedangkan nilai gamma menghasilkan akurasi tertinggi pada 0.01, dilanjutkan 0,1 dan semakin menurun sampai dengan  $\text{gamma}=10$ . Pada penelitian yang dilakukan Nurkholis, dkk (2021) pada algoritma SVM memiliki akurasi terbaik pada nilai C



dengan rentang 1,0-3,0. Parameter  $C$  sangat penting untuk mengontrol kompleksitas model. Pemilihan nilai  $C$  didasarkan pada tujuan meminimalkan nilai error dan keseimbangan bias-variance. Nilai  $C$  yang lebih kecil cenderung memberikan model yang lebih toleran terhadap kesalahan klasifikasi dan dapat menghasilkan margin keputusan yang lebih luas. Sedangkan nilai  $C$  yang tinggi akan memberikan model yang vokal terhadap kesalahan klasifikasi sehingga margin keputusan lebih ketat. Parameter Gamma berfungsi untuk mengontrol seberapa jauh pengaruh sebuah titik data dapat mencapai dalam membangun batas keputusan (hyperplane). Rentang nilai yang kecil menunjukkan pengaruh yang lebih luas sehingga titik data yang lebih jauh dari margin tetap memiliki pengaruh sehingga dapat menghasilkan batas keputusan yang kompleks, sedangkan nilai gamma yang tinggi menghasilkan pengaruh yang lebih terlokalisasi dimana titik data yang paling dekat yang akan memiliki pengaruh sehingga menghasilkan batas keputusan yang sederhana namun berpotensi underfitting.

#### c) Tuning Parameter $K$ pada KNN dan Metriks Perhitungan Jarak

Dalam Algoritma KNN perlu ditentukan nilai  $K$  (tetangga terdekat) yang tepat agar menghasilkan akurasi yang baik pada saat pemodelan dengan mempertimbangkan pada besarnya jumlah data yang tersedia karena waktu pemrosesan yang lama jika dataset sangat besar (Ginting, Permana, 2018). Pada penelitian ini, penentuan parameter nilai  $K$  menggunakan iterasi untuk mencari error rate

minimal pada kombinasi SVM-KNN tanpa seleksi fitur, dan dengan tuning parameter berdasarkan record dataset yaitu

$$K = \sqrt{\text{jumlah data training.}}$$

Metriks perhitungan jarak digunakan untuk menentukan tingkat kesamaan maupun ketidaksamaan data baru terhadap data training. Untuk menghitung jarak data baru dengan nilai K (tetangga terdekat) menggunakan metriks Euclidean Distance. Pemilihan metriks didasarkan atas beberapa penelitian terdahulu dimana euclidean menghasilkan performa yang lebih baik (Nishom, 2019).

Dalam melakukan seleksi pegawai digunakan model Machine Learning dimana peneliti membagi menjadi 2 tahapan yaitu:

#### 1. Kombinasi SVM dan KNN tanpa seleksi fitur

Pada penelitian tahap pertama, setelah melakukan tahapan preprocessing data dan split data, proses selanjutnya adalah menyiapkan library yang diperlukan pada pemrograman python. Penggunaan kombinasi SVM-KNN untuk mempercepat proses komputasi.

Dalam melakukan kombinasi algoritma, langkah awal adalah dengan menggunakan algoritma SVM yaitu mencari support vector yang berada dekat dengan hyperplane terbaik yang memisahkan 2 buah class. Hasil Support vector ini selanjutnya digunakan sebagai dataset untuk di proses menggunakan algoritma KNN. Dalam

melakukan proses menggunakan algoritma KNN, peneliti akan menentukan nilai K (tetangga terdekat) terlebih dahulu, untuk selanjutnya melakukan proses klasifikasi data. Dalam menentukan nilai K akan didasarkan pada jumlah dataset yang akan diuji. Setelah pengujian menggunakan data training, kemudian dengan menggunakan data testing (uji) untuk menghasilkan peramalan seleksi klasifikasi pegawai.

## 2. Implementasi feature selection pada kombinasi SVM dan KNN

Pada penelitian tahap 2 proses yang dilakukan hampir serupa dengan penelitian tahap 1, namun yang membedakan pada tahap ini sebelum dilakukan proses kombinasi SVM dan KNN perlu dilakukan seleksi fitur dengan metode Backward Elimination.

Dengan melakukan seleksi fitur metode Backward Elimination memiliki tujuan untuk menemukan fitur atau variabel terbaik yang akan mempengaruhi tingkat akurasi. Pada tahap ini, peneliti melakukan perbandingan performa dari masing-masing kombinasi dataset melalui pengurangan fitur pada setiap iterasi. Pemilihan Fitur dilakukan dengan cara pemilihan kedepan yaitu menguji semua fitur kemudian dilakukan eliminasi fitur-fitur yang dianggap tidak signifikan. Selanjutnya hasil seleksi fitur dilakukan penerapan kombinasi dengan algoritma SVM dan KNN. Setelah pengujian menggunakan data training, kemudian dengan menggunakan data testing (uji) untuk menghasilkan peramalan seleksi klasifikasi pegawai.

### 3. Evaluasi Model

Tahap ini bertujuan memberikan pengetahuan dengan melakukan perhitungan akurasi, presisi, recall dan F1-Score pada semua skenario penelitian yang telah dilakukan dengan menggunakan confusion matrix.

#### c. Analisa

Tahap ini bertujuan untuk melakukan analisa mendalam terhadap hasil evaluasi model untuk dilakukan perbandingan dari semua percobaan dan pengujian, sehingga dapat membantu dalam penarikan kesimpulan terkait skenario terbaik pada percobaan ke berapa, apakah dengan penerapan seleksi fitur pada kombinasi algoritma SVM dan KNN atau tanpa seleksi fitur dalam melakukan seleksi klasifikasi Pegawai.

#### d. Kesimpulan

Tahap ini menyajikan hasil dari penelitian. Hasil penelitian berupa data fakta yang dihasilkan dari nilai perhitungan akurasi pada pembelajaran machine learning seleksi pegawai dengan metode kombinasi SVM dan KNN serta metode implementasi Feature Selection pada kombinasi SVM dan KNN, sehingga dapat disajikan menjadi sebuah kesimpulan yang lebih mudah untuk dipahami.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Pengumpulan Data

Data yang digunakan adalah data yang didapatkan dari Surat Keputusan Kepala BRIN tentang penempatan dan homebase pegawai, Surat Keputusan Sekretariat Utama BRIN tentang User Aplikasi SAKTI, rekapitulasi kehadiran setiap pegawai selama periodesitas 2022, penelitian atas lokasi homebase dan tempat tinggal pegawai pada Biro BMN melalui aplikasi intra BRIN, serta rekapitulasi penilaian kinerja yang kemudian dikompilasi menjadi satu dataset utama, yang akan digunakan sebagai data penelitian. Penelitian terbagi menjadi 2 tahap dengan tahap awal menggunakan backward elimination untuk melakukan seleksi fitur pada tahapan preprocessing sehingga data yang tidak diperlukan direduksi untuk menghasilkan performa yang baik. Selanjutnya dataset hasil seleksi fitur dilakukan klasifikasi menggunakan kombinasi algoritma SVM dan KNN dengan setting nilai  $C$ ,  $\gamma$  dan  $K$  yang bervariasi yang menghasilkan 8 kali percobaan seperti pada tabel 4.1.



Tabel 4.1 Parameter yang akan digunakan

Percobaan	Parameter SVM, Kernel RBF		Parameter KNN, Metrik Eudidean
	c	Gamma	K
1	1	0.01	9
2	5	0.01	9
3	1	0.1	9
4	5	0.1	9
5	1	0.01	15
6	5	0.01	15
7	1	0.1	15
8	5	0.1	15

Hasil penelitian yang telah diuji dievaluasi menggunakan confusion matrix untuk melihat seberapa baik akurasi yang didapat dari masing-masing percobaan. Penelitian tahap kedua adalah melakukan klasifikasi kelayakan pegawai dengan kombinasi algoritma SVM dan KNN tanpa menggunakan seleksi fitur dimana parameter yang digunakan seperti pada tabel 4.2 dibawah. Dalam melakukan penelitian, langkah awal adalah pemilahan variable dari data mentah awal hasil penilaian kinerja pegawai selama periode 2022. Data tersebut dapat dilihat pada tabel 4.3 berikut ini.

Tabel 4.2 Set Parameter SVM-KNN tanpa Seleksi Fitur

Algoritma	Parameter Percobaan 1	Parameter Percobaan 2
SVM	Kernel RBF	Kernel RBF
KNN	K = iterasi ; metrik = euclidean	$K = \sqrt{\text{jumlah data training}}$ ; metrik = euclidean

Tabel 4.3 Hasil Penilaian Kinerja Pegawai Periode 2022

NO	IDP	NIP	NAMA	FANGKAT GOL	JABATAN	PENUGASAN	PERIODE MULAI	PERIODE SELESAI	NAMA ATASAN LANGSUNG	PANGGOL ATASAN LANGSUNG	JABATAN ATASAN LANGSUNG	NILAI AKHIR HASIL KERJA	KELAYAKAN	NILAI AKHIR PERILAKU KERJA	RATING PERILAKU KERJA	PREDIKAT KINERJA
1	adir001	196408111992031001	Adi Rahardjo	Pembina Tingkat I IV/b	Pengelola Pengadaan Barang/Jasa Ahli Madya	Pengelola Pengadaan Barang/Jasa Ahli Madya	2022-01-01	2022-12-31	Tri Wahyu Widodo	Pembina IV/a	Pengelola Pengadaan Barang/Jasa Ahli Madya	106	DIPERTAHANKAN	100	SESUAI EKSPEKTASI	Baik
2	sury012	197111262007011004	Suryana	Pengatur Tingkat I II/d	Teknisi Pemeliharaan Sarana Dan Prasarana	Teknisi Pemeliharaan Sarana dan Prasarana	2022-01-01	2022-12-31	Yunita Budiarti	Penata Tingkat I III/d	Analisis Pengelolaan Keuangan APBN Ahli Muda	100	DIPERTIMBANGKAN	100	SESUAI EKSPEKTASI	Baik
3	murg001	196309121980021001	Murglentoro	Penata III/c	Penata Laksana Barang Penyedia	Pelaksana Fungsi BMN Eks LIP	2022-01-01	2022-12-31	Yunita Budiarti	Penata Tingkat I III/d	Analisis Pengelolaan Keuangan APBN Ahli Muda	103	DIPERTIMBANGKAN	103	SESUAI EKSPEKTASI	Baik
320	mbal001	197209111999031003	M. Baidilah	Penata Muda Tingkat I III/b	Pengelola Barang Milik Negara	Pelaksana Fungsi BMN Eks LIP	2022-01-01	2022-12-31	Yunita Budiarti	Penata Tingkat I III/d	Analisis Pengelolaan Keuangan APBN Ahli Muda	99	DIPERTIMBANGKAN	100	SESUAI EKSPEKTASI	Baik
321	triw007	196801091994021002	Tri Wahyu Widodo	Pembina IV/a	Pengelola Pengadaan Barang/Jasa Ahli Madya	Koordinator Pelaksana Fungsi Pembinaan dan Advokasi	2022-01-01	2022-12-31	Arywarti Merganingstih	Pembina Tingkat I IV/b	Pengelola Pengadaan Barang/Jasa Ahli Madya	108	DIPERTAHANKAN	108	SESUAI EKSPEKTASI	Baik

Pada tabel 4.3 dapat terlihat data pegawai dan hasil penilaian kinerja pegawai pada Biro BMN selama periode 2022 yang berjumlah 321 record data.

Data kedua adalah data yang didapatkan dari lampiran Surat Keputusan Kepala BRIN tentang penempatan dari masing-masing pegawai. Dikarenakan pada SK Kepala memuat semua penempatan pegawai, maka peneliti melakukan pemilahan disesuaikan dengan batasan pada penelitian ini yaitu pada Biro BMN seperti pada Tabel 4.4.

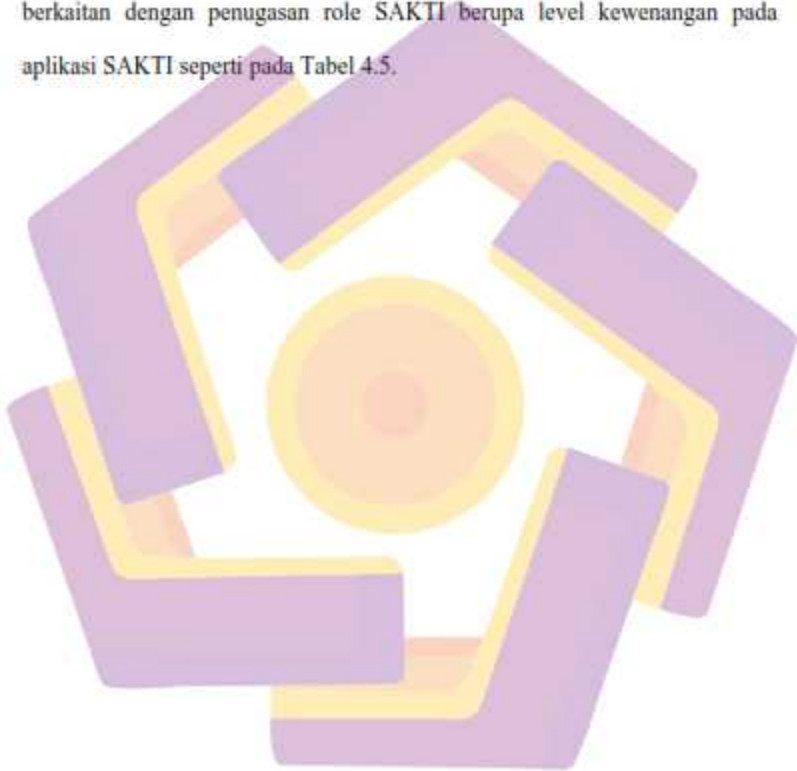


Tabel 4.4 Data Penempatan Pegawai pada Biro BMN

No	Nama	NIP	Pangkat/ Golongan	Kel. Fungsi	Jabatan	Pengawasan/ Penempatan	Unit Kerja Induk	Hombase Unit
1	2	3	4	5	6	7	8	9
2431	Kusni Eddy, SE	196706251993 021001	Penata-III/c	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Muda	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KA Thamrin (Bachrudin Jusuf Habibie)
2432	Ramanda Riuki A. S.E.	199107182019 021001	Penata Muda-III/a	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Pertama	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KA Thamrin (Bachrudin Jusuf Habibie)
2433	Ir. Novita Fitriyanti	197211082009 112001	Penata-III/c	Pengadaan Barang dan Jasa	Penata Barang dan Jasa	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KS Ancol (Aprilani Soegiarto)
2434	Caang Nurjanto, S.E.	198101042008 101001	Penata Muda-III/a	Pengadaan Barang dan Jasa	Penelaah Kebijakan Barang dan Jasa	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KS Gatot Subroto (Sarwono Prawirohardjo)
2435	Fitri Yendra, S.E.	198411122008 012003	Penata Mula Tk.1 - III/b	Pengadaan Barang dan Jasa	Penelaah Kebijakan Barang dan Jasa	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KST Serpong (Bachrudin Jusuf Habibie)
2436	Tri Pangkat Soesiawan, ST	197006171994 111001	Penata-III/c	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Pertama	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KST Serpong (Bachrudin Jusuf Habibie)
2437	Toto Prihantoro, SE	196911171994 011002	Penata Tk.1-III/d	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Muda	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KST Serpong (Bachrudin Jusuf Habibie)
2438	Rd. Nenny Gunawati, SE	196511041986 032004	Penata Tk.1-III/d	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Muda	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KST Serpong (Bachrudin Jusuf Habibie)
2439	Sulastri, SE	196712301988 032001	Penata Tk.1-III/d	Pengadaan Barang dan Jasa	Pengelola Pengadaan Barang/Jasa Ahli Muda	Pelaksana Fungsi Pengelolaan dan Pengadaan Barang/Jasa	Biro Manajemen Barang Milik Negara dan Pengadaan	KST Serpong (Bachrudin Jusuf Habibie)

Dari Tabel 4.4, terlihat beberapa atribut yang serupa dengan tabel 4.3 diantaranya nama, NIP, pangkat golongan, jabatan, sehingga nantinya pada saat pembentukan dataset utama perlu dilakukan penyesuaian.

Data ketiga yang digunakan adalah data penempatan pegawai yang berkaitan dengan penugasan role SAKTI berupa level kewenangan pada aplikasi SAKTI seperti pada Tabel 4.5.



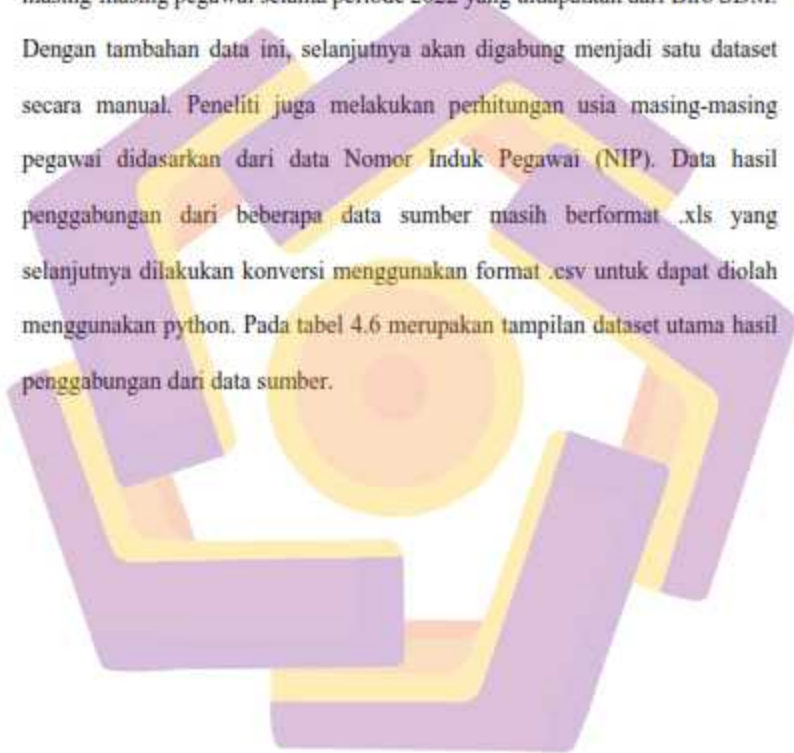


Tabel 4.5 Penempatan Pegawai berdasarkan role SAKTI

NO	NAMA	NIP	PANGKAT/ GOLONGAN	PENUGASAN/PENEMPATAN	KEWENANGAN PENGGUNA	JABATAN PERBENDAHARAAN
1	Dadang Sundara, S.Hut., M.AP	197509052008111001	Penata Tingkat I - III/d	Subkoordinator Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Fasilitas Riset dan Inovasi	OPERATOR	PKPB Setmana cq Deputi Fasilitas Riset dan Inovasi
2	Budhi Subakti, A.Md.	197404301998031004	Penata Tingkat I - III/d	Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Fasilitas Riset dan Inovasi	KOMITMEN	PKPB Setmana cq Deputi Fasilitas Riset dan Inovasi
3	Agus Sudaryanto, S.Kom	196608261986021001	Penata Tingkat I - III/d	Subkoordinator Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Pemanfaatan Riset dan Inovasi	OPERATOR	PKPB Setmana cq Deputi Pemanfaatan RI BRIN
4	Harry Hermawan Hambap	197309212012121001	Pengatur - II/c	Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Pemanfaatan Riset dan Inovasi	KOMITMEN	PKPB Setmana cq Deputi Pemanfaatan RI BRIN
5	Erwan Bastumi	197412132007101001	Penata - III/e	Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Pemanfaatan Riset dan Inovasi	OPERATOR	PKPB Setmana cq Deputi Pemanfaatan RI BRIN
6	Tukimin	197208162012121001	Pengatur - II/c	Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Pemanfaatan Riset dan Inovasi	KOMITMEN	PKPB Setmana cq Deputi Pemanfaatan RI BRIN
7	Anna Kristina Tampubolon, M.A.B	197512012002122001	Pembina - IV/a	Subkoordinator Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Kebijakan Pembangunan	OPERATOR	PKPB Setmana cq Deputi Kebijakan Pembangunan Riset BRIN
8	Ari Pratama, SH.	198907252019031007	Penata Muda - III/a	Pelaksana Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat Deputi Bidang Kebijakan Pembangunan	KOMITMEN	PKPB Setmana cq Deputi Kebijakan Pembangunan Riset BRIN
...	...	...	...	...	...	...
320	Daniarto	198012022008101001	Pengatur Muda Tingkat I - II/b	Pelaksana Fungsi BMN Eks LIPI	OPERATOR	PKPB Setmana cq BIT
321	Hadna Wiguna	197207162007011006	Pengatur - II/c	Pelaksana Fungsi BMN Eks LIPI	VALIDATOR	PKPB Setmana cq Informatika

Tabel 4.5 menunjukkan beberapa data penempatan pegawai berdasarkan level kewenangan yang berbeda-beda yaitu operator, komitmen, validator maupun sebagai approver.

Data selanjutnya yang digunakan adalah rekapitulasi kehadiran masing-masing pegawai selama periode 2022 yang didapatkan dari Biro SDM. Dengan tambahan data ini, selanjutnya akan digabung menjadi satu dataset secara manual. Peneliti juga melakukan perhitungan usia masing-masing pegawai didasarkan dari data Nomor Induk Pegawai (NIP). Data hasil penggabungan dari beberapa data sumber masih berformat .xls yang selanjutnya dilakukan konversi menggunakan format .csv untuk dapat diolah menggunakan python. Pada tabel 4.6 merupakan tampilan dataset utama hasil penggabungan dari data sumber.



Tabel 4.6 Dataset Utama

NO	NAMA	JENIS KELAMIN	PENDIDIKAN	GOLONGAN	PENUGASAN	MODUL SAKTI	USER	HOME BASE UNIT	DOMISILI	HOME BASE AREA	WFO	WFH	USIA	NILAI	RATING KERJA	PREDIKAT KINERJA	KELAYAKAN
1	ID001	F	Magister	Pembina Tingkat I - IV/b	Kepala Biro Manajemen Barang Milik Negara dan Pengadaan	Administrator	Admin	KA Thamrin (Bacharuddin Jusuf Habibie)	Jabode tabek	Jabode tabek	133	111	40	114	DI ATAS EKSPEKTASI	Sangat Baik	Dipertahankan
2	ID002	L	Magister	Penata Tingkat I - III/d	Subkoordinat or Pelaksana Fungsi Pengadaan Barang/Jasa Pelaksana	Komitmen	Approver	KA Thamrin (Bacharuddin Jusuf Habibie)	Jabode tabek	Jabode tabek	44	104	48	106	SESUAI EKSPEKTASI	Baik	Dipertimbangkan
3	ID003	L	Diploma	Penata Tingkat I - III/d	Fungsi Pengadaan Barang/Jasa dan BMN Sekretariat	Komitmen	Operator	KA Thamrin (Bacharuddin Jusuf Habibie)	Luar jabode tabek	Jabode tabek	158	88	49	112	SESUAI EKSPEKTASI	Baik	Dipertahankan
319	ID319	L	SMA	Penata Muda Tingkat I - III/b	Pelaksana Fungsi BMN Eks LPI	Aset Tetap	Operator	KKI Kebun Raya Bogor	Jabode tabek	Jabode tabek	33	207	52	104	SESUAI EKSPEKTASI	Baik	Dipertimbangkan
320	ID320	L	Magister	Pembina Tingkat I - IV/b	Pelaksana Fungsi BMN Eks LPI	Aset Tetap	Operator	KST Cisit (Samaun Samadikun)	Luar jabode tabek	Luar jabode tabek	71	171	58	104	SESUAI EKSPEKTASI	Baik	Dipertahankan
321	ID321	L	SMA	Pengatur Muda Tingkat I - II/b	Pelaksana Fungsi BMN Eks LPI	Persediaan	Operator	KST Cisit (Samaun Samadikun)	Luar jabode tabek	Luar jabode tabek	165	77	43	99	BIASA	Baik	Dipertimbangkan

Pada tabel 4.5 merupakan dataset hasil penggabungan dari banyak data sumber dimana pada dataset menunjukkan beragam atribut mulai dari nomor, nama, jenis kelamin, homebase, jumlah kehadiran WFO – WFH, nilai sampai dengan faktor kelayakan yang berjumlah sebanyak 18 atribut. Dari tabel diatas, kemudian dilakukan penggolongan data secara manual mengingat terlalu heterogennya data set seperti pada tabel 4.7 berikut ini.

Tabel 4.7 Penggolongan Data

No	Pangkat/Golongan	Penggolongan Data
1	Juru Tingkat I - I/d	1
2	Pengatur Muda - II/a	2
3	Pengatur Muda Tingkat I - II/b	3
4	Pengatur - II/c	4
5	Pengatur Tingkat I - II/d	5
6	Penata Muda - III/a	6
7	Penata Muda Tingkat I - III/b	7
8	Penata - III/c	8
9	Penata Tingkat I - III/d	9
10	Pembina - IV/a	10
11	Pembina Tingkat I - IV/b	11
MODUL SAKTI		
No		Penggolongan Data
1	Persediaan	1
2	Aset Tetap	2
3	Komitmen	3
4	Administrator	4
Role User		
No		Penggolongan Data
1	Administrator	1
2	Operator	2
3	Validator	3
4	Approver	4

Dari data pada tabel 4.6 yang merupakan dataset utama, kemudian di simpan pada penyimpanan yang disediakan pada jupyter notebook seperti pada gambar 4.1 dibawah ini.

```
df.head(10)
```

NO	NAMA	JENIS KELAKSA	PENDIDIKAN	LOKASI	REKAM JEKAL	UMUR	JENIS BANGSA	DOMISILI	AREA	WFO	WTH	USA	ML
0	1	IC001	F	Magyar	11	1	1	IA Trianon (Berkehadiah Jaku) HESDA	Jakarta	Jakarta	125	111	-45
1	2	IC002	L	Magyar	8	2	4	IA Trianon (Berkehadiah Jaku) HESDA	Jakarta	Jakarta	125	111	-45
2	3	IC003	L	Duncker	8	2	2	IA Trianon (Berkehadiah Jaku) HESDA	Jakarta	Jakarta	125	111	-45

Gambar 4.1 Pembacaan dataset

Gambar 4.1 merupakan pembacaan data set yang sudah di ekspor ke penyimpanan Jupyter Notebook setelah dilakukan penggolongan data secara manual dengan menggunakan modul pandas.

#### 4.2. Preprocessing Data

Tahapan berikutnya setelah melakukan pengumpulan data adalah tahapan preprocessing. Pada tahapan ini dilakukan proses cleaning, yaitu proses pembersihan data yang tidak relevan salah satunya terhadap data yang bersifat null. Selanjutnya dilakukan transformasi data sehingga diperoleh pola yang lebih mudah dipahami dan agar memudahkan dalam proses pengolahan nantinya. Dari dataset pada gambar 4.1 ditunjukkan jenis format tipe data yang



berbeda pada setiap atribut. Hasil penunjukan mempengaruhi dan menentukan langkah yang akan dilakukan selanjutnya.

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 343 entries, 0 to 342
Data columns (total 18 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   NO              343 non-null    int64
 1   NAMA            343 non-null    object
 2   JENJANGKELAHN  343 non-null    object
 3   PENDIDIKAN     343 non-null    object
 4   GOLONGAN       343 non-null    int64
 5   PEMERIKSAAN   343 non-null    object
 6   PRODUKSAKTI   343 non-null    int64
 7   USIA           343 non-null    int64
 8   HOPERASE      343 non-null    object
 9   DORISILI      343 non-null    object
10   AREA          343 non-null    object
11   WAO           343 non-null    int64
12   WEH           343 non-null    int64
13   USZA         343 non-null    int64
14   RILAI        343 non-null    int64
15   RATINGKENDI  343 non-null    object
16   PREDIKATHEK  343 non-null    int64
17   KELANJARAN   343 non-null    int64
dtypes: int64(8), object(10)
memory usage: 44.1+ KB
```

Gambar 4.2 Tipe format data

Berdasarkan gambar 4.2 mengenai tipe format data, selanjutnya peneliti memastikan kembali apakah terdapat data yang bernilai null (missing value) seperti ditunjukkan pada gambar 4.3 dan selanjutnya melakukan transformasi data menjadi kategorikal agar proses pengolahan nantinya dapat berjalan yang diharapkan dengan memanggil fitur `LabelEncoder` pada modul `scikitLearn` yang ditunjukkan pada gambar 4.4 berikut ini. Pada Tahapan ini, peneliti juga menghapus atribut 'Nama' karena pada saat dilakukan transformasi menjadi kategorikal, diketahui memiliki hasil kategorikal yang sejenis dengan atribut 'No.'

```

# Menghilangkan missing value
df.dropna(inplace=True)

# Menghitung jumlah data
df.info()

```

Gambar 4.3 Missing Value pada Atribut

```

from sklearn.preprocessing import LabelEncoder

encoder = LabelEncoder()

df['MATERIAL'] = encoder.fit_transform(df['MATERIAL'])
df['WISATA'] = encoder.fit_transform(df['WISATA'])
df['DOKUMEN'] = encoder.fit_transform(df['DOKUMEN'])
df['MATA'] = encoder.fit_transform(df['MATA'])
df['PENGALAMAN'] = encoder.fit_transform(df['PENGALAMAN'])
df['KELOMPOK'] = encoder.fit_transform(df['KELOMPOK'])
df['MATAKULIAH'] = encoder.fit_transform(df['MATAKULIAH'])
df['MATAKULIAH'] = encoder.fit_transform(df['MATAKULIAH'])

```

NO	JERIBELAHM	PERAKARAN	COLONGAN	PENGADAN	MOOLSAKTI	EKER	HEMBERKIL	QOMBI	AREG	WFO	WFO	USM	MIAI	RATINGER
0	1	1	1	1	0	0	0	0	0	1	0	101	111	41
1	2	0	0	0	0	0	0	0	2	1	0	100	100	40
2	0	0	0	0	0	0	0	0	2	2	0	100	100	40
3	0	0	0	0	0	0	0	0	1	1	0	100	100	40
4	0	0	0	0	0	0	0	0	2	0	0	100	100	40
5	0	0	0	0	0	0	0	0	0	0	0	100	100	40
6	0	0	0	0	0	0	0	0	0	0	0	100	100	40
7	0	0	0	0	0	0	0	0	0	0	0	100	100	40
8	0	0	0	0	0	0	0	0	0	0	0	100	100	40
9	0	0	0	0	0	0	0	0	0	0	0	100	100	40
10	0	0	0	0	0	0	0	0	0	0	0	100	100	40
11	0	0	0	0	0	0	0	0	0	0	0	100	100	40
12	0	0	0	0	0	0	0	0	0	0	0	100	100	40
13	0	0	0	0	0	0	0	0	0	0	0	100	100	40
14	0	0	0	0	0	0	0	0	0	0	0	100	100	40
15	0	0	0	0	0	0	0	0	0	0	0	100	100	40
16	0	0	0	0	0	0	0	0	0	0	0	100	100	40
17	0	0	0	0	0	0	0	0	0	0	0	100	100	40
18	0	0	0	0	0	0	0	0	0	0	0	100	100	40
19	0	0	0	0	0	0	0	0	0	0	0	100	100	40

Gambar 4.4 Transformasi Data

Berdasarkan data yang ditunjukkan pada gambar 4.4, maka sudah dipastikan dataset sudah siap untuk dilakukan pengolahan untuk selanjutnya digunakan dalam melakukan prediksi klasifikasi. Pada dataset diketahui terdapat 313 record data yang ditunjukkan pada baris dan jumlah atribut sebanyak 17 yang ditunjukkan pada kolom.

### 4.3. Proses Data Mining

Proses Mining merupakan tahap dalam mengaplikasikan metode-metode untuk menemukan informasi penting dan tersembunyi dari data

sehingga hasilnya memberikan pengetahuan. Proses ini menjadi tahap utama dalam melakukan data mining. Metode yang digunakan dalam proses ini adalah tahapan pertama adalah melakukan seleksi fitur dengan backward elimination. Dataset hasil seleksi fitur selanjutnya digunakan untuk melakukan klasifikasi dengan teknik kombinasi algoritma SVM dan KNN pada setting C, gamma dan nilai K yang ditentukan.

#### 4.3.1 Pengujian Seleksi Fitur dengan Backward Elimination

Pada tahapan pengujian seleksi fitur, perlu membangun model dengan modul-modul yang diperlukan. Untuk seleksi fitur membutuhkan fitur Sequential Feature Selector dari modul Machine Learning Extended (mlxtend) dan juga membutuhkan evaluator Algoritma Classifier, dalam hal ini yang digunakan adalah algoritma KNN seperti ditunjukkan pada gambar 4.5.

```
import pandas as pd
import numpy as np
import sklearn.metrics
from sklearn.feature_selection import SequentialFeatureSelector
from sklearn.neighbors import KNeighborsClassifier
```

Gambar 4.5 Pembangunan Model Seleksi Fitur

```
X = df[['L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'L9', 'L10', 'L11', 'L12', 'L13']]
y = df['L14']
sfs = SequentialFeatureSelector(KNeighborsClassifier(),
                               scoring='accuracy',
                               cv=5,
                               n_jobs=-1)
```

Gambar 4.6 Atribut Independen dan Target

Pada gambar 4.6 menunjukkan penentuan atribut independent (x) dan atribut target berupa atribut kelayakan sebagai target (y). Dalam membangun seleksi fitur, peneliti yang menentukan jumlah atribut yang akan digunakan yaitu sejumlah 13 atribut seperti pada gambar 4.7.

```

Berkas_Feature_Selector = Instance(FeatureSelector)(n_neighbors=Classifier(1, 200),
                                                  fit_params={'C': 0.1, 'max_iter': 1000, 'tol': 0.0001, 'verbose': 0},
                                                  scoring='accuracy')

Actual = Instance(AdaBoostClassifier)(n_estimators=100, learning_rate=0.01,
                                     fit_params={'C': 0.1, 'max_iter': 1000, 'tol': 0.0001, 'verbose': 0})

fit_params = {'C': 0.1, 'max_iter': 1000, 'tol': 0.0001, 'verbose': 0}

Berkas_Feature_Selector = Instance(FeatureSelector)(n_neighbors=Classifier(1, 200),
                                                  fit_params={'C': 0.1, 'max_iter': 1000, 'tol': 0.0001, 'verbose': 0},
                                                  scoring='accuracy')

```

Gambar 4.7 Arsitektur Model Seleksi Fitur

Dari percobaan atas penerapan yang dilakukan dengan menentukan hasil atribut sebanyak 13 memberikan hasil pengujian berupa Fitur yang tereliminasi dalam melakukan seleksi fitur dari 17 atribut menjadi 13 atribut independent dapat dituliskan kedalam bentuk tabel 4.8.

Tabel 4.8 Hasil Seleksi Seleksi Fitur

No	Atribut Sebelum Seleksi	Atribut Setelah Seleksi
1	NO	NO
2	NAMA	-
3	JENIS KELAMIN	-
4	PENDIDIKAN	-
5	GOLONGAN	GOLONGAN
6	PENUGASAN	PENUGASAN
7	MODUL SAKTI	MODUL SAKTI
8	USER	USER
9	HOME BASE	HOME BASE
10	DOMISILI	DOMISILI
11	AREA	-
12	WFO	WFO
13	WFH	WFH
14	USIA	USIA
15	NILAI	NILAI
16	RATING KERJA	RATING KERJA
17	PREDIKAT KINERJA	PREDIKAT KINERJA
18	KELAYAKAN (Target)	KELAYAKAN (Target)

### 4.3.2 Hasil Percobaan dan Pengujian Pada Kombinasi SVM-KNN

#### a. Pada Nilai $C = 1$ ; $\gamma = 0.01$ dan Nilai $K=9$

Percobaan pertama adalah melakukan percobaan dengan menggunakan parameter algoritma SVM berupa nilai  $C=1$  dan  $\gamma = 0.01$  pada kernel Radial Base Function (RBF) dengan proporsi data training dan data testing adalah 70:30 seperti pada gambar 4.8. Hasil support vector dari algoritma SVM digunakan sebagai data set baru pada Algoritma KNN dengan melakukan pemilihan nilai  $K=9$  pada metric Euclidean.

```

membagi data training dan testing dengan import dari library sklearn
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)

```

Gambar 4.8 Proporsi Data pada Kombinasi SVM-KNN

Pada gambar 4.8 dapat dijelaskan dengan nilai `test_size` sebesar 0.3 dapat diartikan bahwa proporsi data test adalah sebesar 30%, dengan begitu data latih adalah sebesar 70% dari keseluruhan dataset sebanyak 313 record data. Setelah dilakukan pembagian dataset langkah selanjutnya dilakukan membangun dan melakukan pengujian model berdasarkan parameter yang telah ditentukan.

```

from sklearn.svm import SVC
classifier_svm=SVC(kernel="rbf", random_state=0, C=1, gamma=0.01)
classifier_svm.fit(x_train, y_train)

SVC(C=1, gamma=0.01, random_state=0)

```

Gambar 4.9 Pengujian Model pada parameter  $C=1$  dan  $\gamma=0.01$



Berdasarkan proses pengujian yang telah dilakukan pada gambar 4.9 dengan menggunakan algoritma SVM maka menghasilkan support vector sejumlah 114 record data. Pencarian support vector dilakukan untuk kebutuhan dataset baru pada tahapan selanjutnya menggunakan algoritma KNN. Support Vector adalah kumpulan data terdekat dari hyperplane, artinya data outlier yang akan mempengaruhi kinerja dari algoritma KNN dapat diminimalisir dan optimasi dari kinerja KNN dapat tercapai seperti pada gambar 4.10.

```

j>classifier_support_1
array([[ 5, 14,  8, ..., 148,  1,  3],
       [ 1,  5, 28, ..., 208, 17,  0],
       [ 8,  8, 47, ..., 100, 12,  0],
       ...,
       [216,  8, 41, ..., 390,  2,  0],
       [210,  8, 22, ..., 282,  2,  0],
       [218,  8, 10, ..., 308,  1,  0]])

j>classifier_support_2
array([[ 8,  8, 1,  9,  9, 1, 1,  9,  8, 1,  8,  8,  8,  8, 1,  9,  9,  8,  1,
        4,  8,  8,  8,  1, 1,  8, 1,  8, 1,  8,  0, 1,  8,  8,  8,  8,  8,
        2,  1,  8, 1,  8, 1,  8, 1,  8,  8,  8,  8,  8,  8,  8,  8,  8,  8,
        2,  8,  8, 1,  8,  8, 1,  8,  8,  8,  8,  1,  8,  8,  1,  8,  8,  8,
        8, 1,  8, 1,  8, 1,  8,  8,  8,  8, 1, 1, 1, 1, 1, 1,  8, 1,  8,  8,
        8, 1, 1,  8], dtype=int64)

#stage
114, 20
  
```

Gambar 4.10 Support Vector pada parameter  $C=1$  dan  $\text{gamma}=0.01$

Langkah berikutnya sebagai bagian dari kombinasi SVM-KNN adalah data support vector yang menjadi dataset baru kemudian dilatih dan diuji menggunakan algoritma KNN dengan melakukan pembangunan model seperti pada gambar 4.11 yang didasarkan parameter yang telah ditentukan yaitu pada nilai tetangga terdekat,  $K=9$  dengan perhitungan jarak ke tetangga terdekat menggunakan metric Euclidean. Prediksi akan dilakukan pada data training ( $x_{test}$ ) dan hasil prediksi akan disimpan pada variabel  $y_{pred}$ .

```
x = x[classifier_svm.support_]
y = y[classifier_svm.support_]

from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=9, metric = "euclidean")
classifier.fit(x,y)

kneighborsclassifier(metric='euclidean', n_neighbors=9)
```

Gambar 4.11 Pengujian Model pada Algoritma KNN nilai K=9

Hasil prediksi yang telah dilakukan selanjutnya perlu dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN dengan membandingkan data hasil uji ( $y_{test}$ ) dengan data hasil prediksi ( $y_{pred}$ ) menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar berikut:

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)
print(cm)

[[ 0  2]
 [ 2 15]]

from sklearn.metrics import classification_report

akurasi = classification_report(y_test,y_pred)
print(akurasi)
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	22
1	0.00	0.00	0.00	17
accuracy			0.00	39
macro avg	0.00	0.00	0.00	39
weighted avg	0.00	0.00	0.00	39

Gambar 4.12 Evaluasi Kinerja pada percobaan ke 1

Proses evaluasi kinerja dari pemodelan algoritma SVM-KNN dengan parameter  $C=1$ ,  $\text{Gamma}=0.01$  dan  $K=9$  seperti pada gambar 4.12 dapat dituliskan dalam bentuk tabel 4. 9 berikut.

Tabel 4.9 Tabel Evaluasi Kinerja Percobaan ke 1

	Prediksi Positif	Prediksi Negatif
Aktual Positif	16	2
Aktual Negatif	2	15

Dari Tabel 4.9 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN pada nilai  $C=1$ ,  $\text{Gamma}=0.01$  dan  $K=9$ , sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+15}{16+15+2+2} = 0.885 \text{ atau } 88.5\%$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{16}{16+2} = 0.889 \text{ atau } 88.9\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{16}{16+2} = 0.889 \text{ atau } 88.9\%$$

Berdasarkan perhitungan nilai kinerja model menggunakan confusion matrix menghasilkan nilai yang sama dengan membangun model evaluasi menggunakan fitur `classification_report` pada library `metric` `scikit learn`.

**b. Pada Nilai C = 5; gamma = 0.01 dan Nilai K=9**

Pada percobaan kedua, dengan kombinasi algoritma SVM-KNN dilakukan perbedaan nilai parameter algoritma SVM pada nilai  $C=5$  dan parameter lain yang sama yaitu pada  $\text{gamma} = 0.01$  dan nilai  $K = 9$  pada metric Euclidean. Setelah dilakukan inisialisasi dan pembagian data `train_test` 70:30, berikutnya melakukan pengujian model berdasarkan parameter yang telah ditentukan seperti pada gambar 4.13 dibawah ini.

```
from sklearn.svm import SVC
classifier_svm=SVC(kernel="rbf", random_state=0, C=5, gamma=0.01)
classifier_svm.fit(x_train, y_train)
SVC(C=5, gamma=0.01, random_state=0)
```

Gambar 4.13 Pengujian Model pada parameter  $C=5$  dan  $\text{gamma} = 0.01$

Berdasarkan proses pengujian yang telah dilakukan pada gambar 4.13 dengan menggunakan algoritma SVM maka menghasilkan support vector sejumlah 76 record data seperti ditunjukkan pada gambar 4.14. Pencarian support vector dilakukan untuk kebutuhan dataset baru pada tahapan selanjutnya menggunakan algoritma KNN. Support Vector adalah kumpulan data terdekat dari hyperplane, artinya data outlier pada dataset baru tidak akan digunakan sehingga tidak akan memberikan pengaruh kinerja dari algoritma KNN dan optimasi dari kinerja KNN dapat tercapai.

```

>>>X_train_svm_support_1
array([[0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1,
       1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0,
       1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1,
       1, 1, 0, 1, 0, 0, 0, 1, 1, 0], dtype=int64)

>>>X_test
(76, 24)

>>>y_test
(76,)

```

Gambar 4.14 Support Vector pada parameter  $C=5$  dan  $\gamma=0.01$

Tahapan berikutnya yang dilakukan seperti pada percobaan ke 1 yaitu dengan melakukan pembangunan model KNN seperti pada gambar 4.11. Hasil prediksi yang telah dilakukan selanjutnya perlu dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN dengan membandingkan data hasil uji ( $y_{test}$ ) dengan data hasil prediksi ( $y_{pred}$ ) menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar 4.15:

```

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)
print(cm)

[[25  0]
 [ 4  4]]

```

Gambar 4.15 Evaluasi Kinerja pada percobaan ke2

Proses evaluasi kinerja pada gambar 4.15 dari pemodelan algoritma SVM-KNN dengan parameter  $C=5$ ,  $\text{Gamma}=0.01$  dan  $K=9$  dapat dituliskan dalam bentuk tabel 4. 10 berikut.

Tabel 4.10 Tabel Evaluasi Kinerja Percobaan ke2

	Prediksi Positif	Prediksi Negatif
Aktual Positif	15	0
Aktual Negatif	4	4

Dari Tabel 4.10 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN pada nilai  $C=5$ ,  $\text{Gamma}=0.01$  dan  $K=9$ , sebagai berikut:

$$\text{Akurasi} = \frac{TP+T}{TP+TN+FP+FN} = \frac{15+4}{15+4+4+0} = 0.826 \text{ atau } 82.6\%$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{15}{15+4} = 0.789 \text{ atau } 78.9\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{15}{15+0} = 1 \text{ atau } 100\%$$

Berdasarkan perhitungan nilai kinerja model menggunakan confusion matrix menghasilkan nilai yang sama dengan membangun model evaluasi menggunakan fitur `classification report` pada `library metric scikit learn` seperti pada gambar 4.16.

	precision	recall	f1-score	support
0	0.79	1.00	0.88	15
1	1.00	0.50	0.67	8
accuracy			0.83	23
macro avg	0.89	0.75	0.77	23
weighted avg	0.86	0.83	0.81	23

Gambar 4.16 Classification Report Percobaan ke 2



**c. Pada Nilai C = 1; gamma = 0.1 dan Nilai K=9**

Percobaan ketiga adalah melakukan percobaan dengan menggunakan parameter algoritma SVM berupa nilai C =1 dengan nilai parameter **gamma menjadi = 0.1** pada kernel Radial Base Function (RBF) dengan proporsi data training dan data testing yang sama 70:30 seperti langkah pada gambar 4.8 diatas. Setelah dilakukan pembagian dataset, langkah selanjutnya dilakukan membangun dan melakukan pengujian model berdasarkan parameter yang telah ditentukan.

```
from sklearn.svm import SVC
classifier_svm=SVC(kernel="rbf", random_state=0, C=1, gamma=0.1)
classifier_svm.fit(x_train, y_train)
svc(C=1, gamma=0.1, random_state=0)
```

Gambar 4.17 Pengujian Model pada parameter C=1 dan gamma =0.1

Berdasarkan proses pengujian yang telah dilakukan pada gambar 4.17 dengan menggunakan algoritma SVM, menghasilkan support vector sejumlah 111 record data seperti pada gambar 4.18.

```
(classifier.support_)
array([[0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0,
        0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1,
        0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
        1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0,
        0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1,
        0], dtype=int64)

(111,)
```

Gambar 4.18 Support Vector pada parameter C=1 dan gamma =0.1

Tahapan berikutnya yang dilakukan seperti pada percobaan ke 1 yaitu dengan melakukan pembangunan model KNN seperti pada gambar 4.11. Hasil prediksi yang telah dilakukan selanjutnya perlu dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN dengan membandingkan data hasil

uji ( $y_{test}$ ) dengan data hasil prediksi ( $y_{pred}$ ) menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar 4.19:

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)
print(cm)
[[17  0]
 [ 2 15]]
```

Gambar 4.19 Evaluasi Kinerja pada percobaan ke3

Proses evaluasi kinerja pada gambar 4.19 dari pemodelan algoritma SVM-KNN dengan parameter  $C=1$ ,  $\text{Gamma}=0.1$  dan  $K=9$  dapat dituliskan dalam bentuk tabel 4. 11 berikut.

Tabel 4.11 Tabel Evaluasi Kinerja Percobaan ke3

	Prediksi Positif	Prediksi Negatif
Aktual Positif	17	0
Aktual Negatif	2	15

Dari Tabel 4.11 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN pada nilai  $C=1$ ,  $\text{Gamma}=0.1$  dan  $K=9$ , sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+F} = \frac{17+}{17+15+2+0} = 0.941 \text{ atau } 94.1\%$$

$$\text{Presisi} = \frac{TP}{TP+} = \frac{17}{17+2} = 0.894 \text{ atau } 89.4\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{17}{17+0} = 1 \text{ atau } 100\%$$

**d. Pada Nilai C = 5; gamma = 0.1 dan Nilai K=9**

Percobaan selanjutnya membangun model dengan nilai parameter algoritma SVM pada nilai  $C=5$ , nilai  $\gamma = 0.1$  dan nilai  $K = 9$ . seperti pada gambar 4.20. Berdasarkan proses pengujian yang telah dilakukan pada gambar 4.20 dengan menggunakan algoritma SVM, menghasilkan support vector sejumlah 85 record data seperti pada gambar 4.20.

```
SVC(C=5, gamma=0.1, random_state=0)
x[('support_vectors_')]
array([[0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0,
       0, 0, 0, 1, 1, 0, 0, 0, 2, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1,
       dtype=int64])
y_support
(25,)
```

Gambar 4.20 Support Vector pada parameter  $C=5$  dan  $\gamma=0.1$

Tahapan berikutnya yang dilakukan seperti pada percobaan ke 1 yaitu dengan melakukan pembangunan model KNN seperti pada gambar 4.11. Hasil prediksi selanjutnya dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar 4.21.

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
[[12  2]
 [ 1 11]]
```

Gambar 4.21 Evaluasi Kinerja pada percobaan ke4

Proses evaluasi kinerja pada gambar 4.21 dari pemodelan algoritma SVM-KNN dengan parameter  $C=5$ ,  $\text{Gamma}=0.1$  dan  $K=9$  dapat dituliskan dalam bentuk tabel 4. 12 berikut.

Tabel 4.12 Tabel Evaluasi Kinerja Percobaan ke4

	Prediksi Positif	Prediksi Negatif
Aktual Positif	12	2
Aktual Negatif	1	11

Dari Tabel 4.12 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN pada nilai  $C=5$ ,  $\text{Gamma}=0.1$  dan  $K=9$ , sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+F} = \frac{12+11}{12+11+1+2} = 0.884 \text{ atau } 88.4\%$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{12}{12+1} = 0.923 \text{ atau } 92.3\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{12}{12+2} = 0.857 \text{ atau } 85.7\%$$

**e. Pada Nilai  $C = 1$ ;  $\text{gamma} = 0.01$  dan Nilai  $K=15$**

Pada percobaan ke 5, dengan nilai parameter  $C=1$ ,  $\text{gamma}=0.01$  untuk algoritma SVM maka dapat dikatakan pembangunan model algoritma kembali seperti pada percobaan ke 1, yang membedakan adalah nilai parameter tetangga terdekat  $K=15$  pada penerapan algoritma KNN. Dengan proporsi data training dan data testing adalah 70:30, dan pembangunan model algoritma SVM yang sama dengan gambar 4.9, sehingga menghasilkan support vector sejumlah 114 record data. Langkah berikutnya sebagai bagian dari kombinasi SVM-KNN adalah

data support vector yang menjadi dataset baru kemudian dilatih dan diuji menggunakan algoritma KNN dengan melakukan pembangunan model seperti pada gambar 4.22 yang didasarkan parameter yang telah ditentukan yaitu pada nilai tetangga terdekat,  $K = 15$  dengan perhitungan jarak ke tetangga terdekat menggunakan metric Euclidean. Prediksi akan dilakukan pada data training ( $x_{test}$ ) dan hasil prediksi akan disimpan pada variabel  $y_{pred}$ .

```
from sklearn.neighbors import KNeighborsClassifier
Classifier = KNeighborsClassifier(n_neighbors = 15, metric = "euclidean")
Classifier.fit(x,y)

KNeighborsClassifier(metric='euclidean', n_neighbors=15)
```

Gambar 4.22 Pengujian Model pada Algoritma KNN nilai  $K=15$

Hasil prediksi yang telah dilakukan selanjutnya perlu dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN dengan membandingkan data hasil uji ( $y_{test}$ ) dengan data hasil prediksi ( $y_{pred}$ ) menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar berikut:

```
from sklearn.metrics import confusion_matrix
CM = confusion_matrix(y_test,y_pred)
print(CM)

[[17  1]
 [ 4 15]]
```

Gambar 4.23 Evaluasi Kinerja pada percobaan ke5

Proses evaluasi kinerja dari pemodelan algoritma SVM-KNN dengan parameter  $C=1$ ,  $\text{Gamma}=0.01$  dan  $K=15$  seperti pada gambar 4.23 dapat dituliskan dalam bentuk tabel 4. 13 berikut.



Tabel 4.13 Tabel Evaluasi Kinerja Percobaan ke5

	Prediksi Positif	Prediksi Negatif
Aktual Positif	17	1
Aktual Negatif	2	15

Dari Tabel 4.13 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN pada nilai  $C=1$ ,  $\text{Gamma}=0.01$  dan  $K=15$ , sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{17+15}{17+15+2+1} = 0.914 \text{ atau } 91.4\%$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{17}{17+2} = 0.894 \text{ atau } 89.4\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{17}{17+1} = 0.944 \text{ atau } 94.4\%$$

#### f. Pada Nilai $C = 5$ ; $\text{gamma} = 0.01$ dan Nilai $K=15$

Percobaan berikutnya yaitu percobaan ke 6, dengan nilai parameter  $C=5$ ,  $\text{gamma}=0.01$  maka pembangunan model algoritma kembali seperti pada percobaan ke 2, dengan perbedaan nilai parameter tetangga terdekat  $K=15$  pada penerapan algoritma KNN. Dengan split data yang sama 70:30, dan pembangunan model algoritma SVM yang sama dengan gambar 4.15, sehingga menghasilkan support vector sejumlah 76 record data. Langkah berikutnya dengan melakukan pembangunan model Algoritma KNN seperti pada gambar 4.22 yang didasarkan parameter yang telah ditentukan yaitu pada nilai tetangga terdekat,  $K = 15$  dengan perhitungan jarak ke tetangga terdekat menggunakan metric Euclidean. Hasil prediksi selanjutnya dilakukan

evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar 4.24.

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
[[14  1]
 [ 3  5]]
```

Gambar 4.24 Evaluasi Kinerja pada percobaan ke6

Proses evaluasi kinerja dari pemodelan algoritma SVM-KNN dengan parameter  $C=5$ ,  $\text{Gamma}=0.01$  dan  $K=15$  seperti pada gambar 4.24 dapat dituliskan dalam bentuk tabel 4.14 berikut.

Tabel 4.14 Tabel Evaluasi Kinerja Percobaan ke6

	Prediksi Positif	Prediksi Negatif
Aktual Positif	14	1
Aktual Negatif	3	5

Dari Tabel 4.14 memudahkan peneliti dalam melakukan perhitungan nilai akurasi, presisi dan recall untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{14+5}{14+5+3+1} = 0.826 \text{ atau } 82.6\%$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{14}{14+3} = 0.823 \text{ atau } 82.3\%$$

$$\text{Recall} = \frac{TP}{TP+F} = \frac{14}{14+1} = 0.933 \text{ atau } 93.3\%$$

**g. Pada Nilai  $C = 1$ ;  $\text{gamma} = 0.1$  dan Nilai  $K=15$**

Pada Percobaan ke7, dengan nilai parameter  $C=1$ ,  $\text{gamma}=0.1$  maka pembangunan model algoritma kembali seperti pada percobaan ke 3, dengan perbedaan nilai parameter tetangga terdekat  $K=15$  pada penerapan algoritma KNN. Dengan split data yang sama 70:30, dan pembangunan model algoritma SVM yang sama dengan gambar 4.17, sehingga menghasilkan support vector sejumlah 111 record data. Langkah berikutnya dengan melakukan pembangunan model Algoritma KNN seperti pada gambar 4.22 yang didasarkan parameter yang telah ditentukan yaitu pada nilai tetangga terdekat,  $K = 15$  dengan perhitungan jarak ke tetangga terdekat menggunakan metric Euclidean. Hasil prediksi dilakukan evaluasi kinerja pemodelan untuk mengukur seberapa baik kinerja kombinasi algoritma SVM dan KNN menggunakan confusion matrix untuk menghasilkan nilai TP, TN, FP dan FN sehingga dapat dilakukan perhitungan akurasi, presisi, dan recall seperti pada tampilan gambar 4.25.

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

[[10  1]
 [ 3 14]]
```

Gambar 4.25 Evaluasi Kinerja pada percobaan ke7

Proses evaluasi kinerja dari pemodelan algoritma SVM-KNN dengan parameter  $C=1$ ,  $\text{Gamma}=0.1$  dan  $K=15$  seperti pada gambar 4.25 dapat dituliskan dalam bentuk tabel 4.15 berikut.

Tabel 4.15 Tabel Evaluasi Kinerja Percobaan ke7

	Prediksi Positif	Prediksi Negatif
Aktual Positif	16	1
Aktual Negatif	3	14

Selanjutnya peneliti melakukan perhitungan nilai akurasi, presisi dan recall berdasarkan nilai TP, TN, FP dan FN yang ditampilkan pada tabel 4.15 untuk mengetahui kinerja dari penerapan seleksi fitur dengan kombinasi algoritma SVM-KNN sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+14}{16+14+3+1} = 0.882 \text{ atau } 88.2\%$$

$$Presisi = \frac{TP}{TP+FP} = \frac{16}{16+3} = 0.842 \text{ atau } 84.2\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{16}{16+1} = 0.941 \text{ atau } 94.1\%$$

**h. Pada Nilai C = 5; gamma = 0.1 dan Nilai K=15**

Percobaan selanjutnya merupakan percobaan terakhir untuk penggunaan dataset yang telah dilakukan seleksi fitur yaitu pada nilai  $c=5$ ,  $\text{gamma}=0.1$  dan nilai  $K = 15$ . Dalam pemodelan SVM dilakukan seperti pada percobaan ke 4 yang ditunjukkan pada gambar 4.20, dengan melakukan split data yang sama 70:30, sehingga menghasilkan support vector sejumlah 85 record data. Selanjutnya membangun model Algoritma KNN seperti pada gambar 4.22 dengan nilai  $K = 15$ , dan metric Euclidean. Untuk melakukan evaluasi kinerja pemodelan SVM-KNN pada akurasi, presisi, dan recall menggunakan confusion matrix seperti pada tampilan gambar 4.26.

```

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
[[13  1]
 [ 1 11]]

```

Gambar 4.26 Evaluasi Kinerja pada percobaan ke8

Proses evaluasi kinerja dari pemodelan algoritma SVM-KNN dengan parameter  $C=5$ ,  $\text{Gamma}=0.1$  dan  $K=15$  seperti pada gambar 4.26 dapat dituliskan dalam bentuk tabel 4.16 berikut.

Tabel 4.16 Tabel Evaluasi Kinerja Percobaan ke8

	Prediksi Positif	Prediksi Negatif
Aktual Positif	13	1
Aktual Negatif	1	11

Untuk memudahkan perhitungan berdasarkan nilai TP, TN, FP dan FN yang dihasilkan dari Confusion Matrix pada tabel 4.16, adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{13+11}{13+11+1+1} = 0.923 \text{ atau } 92.3\%$$

$$\text{Presisi} = \frac{TP}{TP+} = \frac{13}{13+1} = 0.928 \text{ atau } 92.8\%$$

$$\text{Recall} = \frac{TP}{TP+} = \frac{13}{13+} = 0.928 \text{ atau } 92.8\%$$

#### 4.3.3 Hasil Percobaan dan Pengujian Pada Kombinasi SVM-KNN Tanpa Seleksi Fitur

##### a. Pada Nilai K dengan Minimum Error Rate

Penelitian lain yang dilakukan percobaan adalah melakukan kombinasi algoritma SVM dan KNN tanpa dilakukan seleksi fitur terlebih dahulu. Setting parameter pada pengujian ini adalah



penggunaan Kernel yang sama yaitu RBF (Radial Base Function), namun tidak dilakukan penentuan nilai C dan gamma. Dalam membangun model SVM, dataset yang digunakan adalah dataset awal dengan split data yang sama 70:30 seperti terlihat pada gambar 4.27.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)

from sklearn.svm import SVC
classifier = SVC(kernel='rbf', random_state=0)
classifier.fit(x_train, y_train)

sv = classifier.predict(x_test)

# shape
(111, 17)
```

Gambar 4.27 Membangun Model SVM tanpa seleksi fitur

```
[[classifier.support_]]
array([[0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0,
        0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0,
        1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0,
        0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,
        0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1,
        0], dtype=int64)

[[classifier.support_].shape
(111,)]
```

Gambar 4.28 Support Vector Tanpa Seleksi Fitur

Dalam pemodelan SVM yang ditunjukkan pada gambar 4.27, menghasilkan support vector sejumlah 111 record data seperti pada gambar 4.28. Selanjutnya membangun model Algoritma KNN dengan nilai parameter K didapatkan dari nilai error rate terendah dengan melakukan iterasi (perulangan) pada nilai K, dari K=1 sampai dengan nilai K=20 seperti yang ditunjukkan pada gambar 4.29.

```
from sklearn.neighbors import KNeighborsClassifier
import numpy as np

error_rate = []
for i in range(1, 21):
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(x_train, y_train)
    pred_i = knn.predict(x_test)
    error_rate.append(np.mean(pred_i != y_test))

print('Nilai error =', min(error_rate), 'at k=', error_rate.index(min(error_rate)))
# Output error= 0.000000000000000000 at k= 0
```

Gambar 4.29 Mencari nilai K terbaik

Untuk melakukan evaluasi kinerja pemodelan SVM-KNN dilakukan menggunakan confusion matrix dengan parameter Nilai K=8 berdasarkan hasil error rate terendah seperti pada gambar 4.30 dan dapat dituliskan dalam bentuk tabel 4.17.

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

[[21  0]
 [ 4  9]]
```

Gambar 4.30 Evaluasi Kinerja

Tabel 4.17 Tabel Evaluasi Kinerja

	Prediksi Positif	Prediksi Negatif
Aktual Positif	21	0
Aktual Negatif	4	9

Agar memudahkan perhitungan berdasarkan nilai TP, TN, FP dan FN yang dihasilkan dari Confusion Matrix pada tabel 4.17, adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP+T}{TP+TN+FP+FN} = \frac{21+9}{21+9+4+0} = 0.882 \text{ atau } 88.2\%$$

$$\text{Presisi} = \frac{TP}{TP+F} = \frac{21}{21+4} = 0.84 \text{ atau } 84\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{21}{21+0} = 1 \text{ atau } 100\%$$

- b. Pada Nilai  $K = \sqrt{\text{jumlah data training}}$

Percobaan terakhir yang dilakukan adalah melakukan kombinasi algoritma SVM dan KNN tanpa dilakukan seleksi fitur dengan setting parameter penggunaan kernel RBF tanpa menentukan nilai C dan gamma, dan dengan parameter nilai  $K = \sqrt{\text{jumlah data training}}$ . Dalam membangun model SVM, split data dan pemodelan yang digunakan

sama dengan percobaan **Pada Nilai K dengan Minimum Error Rate** seperti pada gambar 4.27 sebelumnya.

Sehingga support vector yang dihasilkan sama yaitu sejumlah 111 record data. Support Vector ini kemudian digunakan sebagai dataset baru untuk kemudian dilakukan pemodelan dengan algoritma KNN dengan penggunaan parameter nilai tetangga terdekat sejumlah  $K=15$  dengan perhitungan sebagai berikut:

$$\text{Data } K = \sqrt{\text{jumlah train data}} = \sqrt{(70\% \times 313 \text{ record data})} = \sqrt{219} = 15$$

Dengan menggunakan nilai  $K=15$ , selanjutnya membangun pemodelan KNN pada jupyter seperti pada gambar 4.31.

```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=15, metric='euclidean')
classifier.fit(x,y)
KNeighborsClassifier(metric='euclidean', n_neighbors=15)
```

Gambar 4.31 Membangun Pemodelan KNN

Agar kinerja pemodelan dapat diukur maka dilakukan evaluasi menggunakan confusion matrix, seperti pada gambar 4.32 menghasilkan nilai TP, TN, FP dan FN untuk kemudian dapat dilakukan perhitungan akurasi, presisi dan recall, dan dapat dituliskan dalam bentuk tabel 4.18.

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)
print(cm)
[[10  1]
 [ 4  8]]
```

Gambar 4.32 Evaluasi Kinerja

Tabel 4.18 Tabel Evaluasi Kinerja

	Prediksi Positif	Prediksi Negatif
Aktual Positif	20	1
Aktual Negatif	4	9

Agar memudahkan perhitungan berdasarkan nilai TP, TN, FP dan FN yang dihasilkan dari Confusion Matrix pada tabel 4.18, adalah sebagai berikut:

$$Akurasi = \frac{TP+}{TP+TN+FP+F} = \frac{20+9}{20+9+4+1} = 0.853 \text{ atau } 85.3\%$$

$$Presisi = \frac{TP}{TP+F} = \frac{20}{20+4} = 0.833 \text{ atau } 83.3\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{20}{20+} = 0.952 \text{ atau } 95.2\%$$

#### 4.4. Analisa dan Pembahasan

Berdasarkan hasil percobaan terhadap penerapan seleksi fitur menggunakan backward elimination, dengan parameter cross validation=5, algoritma evaluator menggunakan algoritma KNN dan fitur yang diinginkan menjadi 13 atribut, python melakukan pengujian kombinasi 13 fitur terhadap 17 fitur sebelumnya, dan memisahkan 1 atribut target (kelayakan) seperti tampak pada tabel 4.19 berikut.

Tabel 4.19 Tabel Atribut Utama

No	Atribut Sebelum Seleksi
1	NO
2	NAMA
3	JENIS KELAMIN
4	PENDIDIKAN
5	GOLONGAN
6	PENUGASAN
7	MODUL SAKTI
8	USER
9	HOME BASE
10	DOMISILI
11	AREA
12	WFO
13	WFH
14	USIA
15	NILAI
16	RATING KERJA
17	PREDIKAT KINERJA
Target	KELAYAKAN (Target)

Selanjutnya dengan menggunakan rumus matematika untuk menghitung jumlah kombinasi yang dilakukan python adalah sebagai berikut:

$$C(n, k) = \frac{n!}{k!(n-k)!}$$

$$C(17,13) = \frac{17!}{13!(17-13)!} = \frac{17 \times 16 \times 15 \times 14 \times 13!}{13!(4 \times 3 \times 2 \times 1)}$$

$$= 2380 \text{ kombinasi fitur}$$

Dengan menggunakan python, kombinasi terbaik dalam pemodelan seleksi fitur backward elimination memberikan hasil pengujian sebesar 89% dan disajikan pada tabel 4.20. Sehingga dapat dikatakan pemodelan seleksi fitur telah siap digunakan untuk proses



selanjutnya yaitu dipergunakan untuk pengujian klasifikasi menggunakan kombinasi algoritma SVM dan KNN.

Tabel 4.20 Tabel Atribut Setelah Seleksi

No	Atribut Setelah Seleksi
1	NO
2	GOLONGAN
3	PENUGASAN
4	MODUL SAKTI
5	USER
6	HOME BASE
7	DOMISILI
8	WFO
9	WFH
10	USIA
11	NILAI
12	RATING KERJA
13	PREDIKAT KINERJA
Target	KELAYAKAN

Selanjutnya pada percobaan yang telah dilakukan sebanyak 10 kali kombinasi dengan penerapan seleksi fitur pada 8 kali percobaan, dan 2 kali percobaan tanpa penerapan seleksi fitur dengan atribut maksimal 17 atribut, dimana setiap skenario pengujian yang telah dilakukan menggunakan proporsi data yang sama yaitu dengan pembagian data latih 70% dan data uji 30% pada dataset klasifikasi kinerja pegawai berdasarkan user SAKTI, dengan penggunaan kernel RBF menghasilkan nilai evaluasi yang beragam seperti ditunjukkan pada tabel 4.21.

Tabel 4.21. Hasil Pengujian dengan 10 kali kombinasi

Percobaan ke-	Dataset Awal	Jumlah Atribut	Parameter SVM, Kernel RBF		Dataset terbentuk (Support Vector)	Parameter KNN, Metrik Euclidean	Confusion Matrix		
			C	Gamma		K	Akurasi(%)	Presisi (%)	Recall (%)
1	313	13	1	0.01	114	9	88.5	88.9	88.9
2		13	5	0.01	76	9	82.6	78.9	100
3		13	1	0.1	111	9	<b>94.1</b>	<b>89.4</b>	<b>100</b>
4		13	5	0.1	85	9	88.4	92.3	85.7
5		13	1	0.01	114	15	91.4	89.4	94.4
6		13	5	0.01	76	15	82.6	82.3	93.3
7		13	1	0.1	111	15	88.2	84.2	94.1
8		13	5	0.1	85	15	92.3	92.8	92.8
9		17	Auto	Auto	111	min error rate, pada K=8	88.2	84	100
10		17	Auto	Auto	111	K=jumlah train data, K=15	85.3	83.3	95.2

Pada percobaan pertama dan kedua berdasarkan tabel 4.21, dengan menggunakan Kombinasi SVM-KNN dengan seleksi fitur pada parameter  $C=1$ ,  $\gamma=0.01$ , dan Nilai  $K=9$  mampu menghasilkan support vector sebagai dataset baru paling banyak yaitu 114 data, sedangkan pada parameter  $C=5$ ,  $\gamma=0.01$ , dan Nilai  $K=9$  menghasilkan support vector 76 record data. Pada percobaan pertama, pemodelan menghasilkan dengan tingkat akurasi dalam melakukan klasifikasi sebesar 88.5%, presisi 88.9% dan recall 88.9%, yang artinya penerapan seleksi fitur pada kombinasi SVM-KNN bekerja dengan cukup baik dan algoritma mampu melakukan klasifikasi yang benar, ketepatan model dalam identifikasi pengukuran dalam mengidentifikasi dengan benar instance positif dari prediksi bernilai positif, dan kemampuan mendeteksi benar positif dari semua data positif. Sedangkan pada percobaan kedua, nilai Akurasi

yang dihasilkan sebesar 82.6% adalah nilai akurasi terkecil dari 10 kali percobaan yang dilakukan. Hal ini dipengaruhi oleh faktor support vector yang dihasilkan tidak terlalu besar mengakibatkan variasi data dalam melakukan pelatihan dan pengujian terbatas sehingga hasil model algoritma berpengaruh. Nilai Presisi 78.9% menunjukkan kombinasi parameter yang dipilih masih kurang sesuai karena pemodelan mengidentifikasi beberapa false positif. Dari percobaan pertama dan kedua, dapat dipahami bahwa nilai parameter C yang lebih kecil mampu menghasilkan kinerja yang lebih baik.

Pada percobaan ketiga dan keempat dengan nilai  $C=1$  dan  $5$ ,  $\text{gamma}=0.1$  dan Nilai  $K=9$ , masing-masing menghasilkan support vector 111 pada nilai  $C=1$  dan support vector sebesar 85 record data pada nilai  $C=5$ . Hasil dari percobaan ketiga, pemodelan menghasilkan akurasi sebesar 94.1%, presisi 89.4% dan recall 100%, yang artinya kinerja pemodelan semakin baik pada setting parameter  $\text{gamma} 0.1$ , hal ini menunjukkan pemodelan mampu melakukan klasifikasi yang tepat, dan kemampuan dalam mendeteksi benar positif dari semua data positif (recall) adalah 100%. Sedangkan pada percobaan keempat, masing-masing nilai akurasi, presisi dan recall adalah 88.4%, 92.3%, dan 85.7%. Dengan nilai presisi 92,3% artinya pemodelan mampu menghasilkan false positif yang rendah yaitu mengklasifikasi data negatif terklasifikasi menjadi positif. Namun nilai akurasi yang tidak cukup baik dari percobaan lainnya dipengaruhi oleh jumlah support vector yang tidak terlalu banyak. Dari percobaan ketiga dan keempat, dapat dipahami bahwa nilai parameter C yang lebih kecil dengan nilai gamma yang lebih tinggi dari percobaan pertama dan kedua mampu menghasilkan kinerja yang lebih baik.

Pada percobaan kelima dan keenam dengan nilai  $C=1$  dan  $5$ ,  $\gamma=0.01$  menghasilkan dataset yang sama dengan percobaan pertama dan kedua yaitu 114 dan 76 record data. Dengan perubahan nilai tetangga terdekat menjadi  $K=15$ , masing-masing pemodelan menghasilkan akurasi sebesar 91.4%, presisi 89.4% dan recall 94.4%, pada nilai  $C=1$ ; dan akurasi 82.6%, presisi 82.3% dan recall 92.3% pada nilai  $C=5$ , yang artinya hasil kinerja pemodelan cukup baik dalam menggunakan nilai  $K$  lebih tinggi. Pada percobaan kelima dengan akurasi 91.4% artinya kemampuan pemodelan dalam melakukan klasifikasi memiliki tingkat keakuratan yang lebih baik pada 114 record data dan kemampuan dalam mendeteksi nilai yang diprediksi negatif namun aktual positif dari semua data positif (recall) adalah 94.4%. Sedangkan pada percobaan kelima, menghasilkan akurasi 82.6% yang artinya tingkat keakuratan pemodelan dalam melakukan klasifikasi masih perlu ditingkatkan. Dengan nilai presisi yaitu kemampuan pemodelan dalam mengidentifikasi klasifikasi dengan benar data positif diklasifikasi sebagai positif sebesar 82.3% disebabkan prediksi data positif namun aktual negatif yang masih tinggi. Dari percobaan kelima dan keenam, nilai  $K$  yang berbeda berpengaruh terhadap nilai akurasi pada parameter  $C$  dan  $\gamma$  yang sudah sebelumnya dilakukan percobaan.

Pada percobaan ketujuh dan kedelapan adalah percobaan dengan penerapan nilai  $C=1$  dan  $5$ ,  $\gamma=0.1$  pada setting parameter  $K=15$ . Dengan perubahan nilai tetangga terdekat lebih tinggi menjadi  $K=15$ , masing-masing pemodelan menghasilkan akurasi sebesar 88.2%, presisi 84.2% dan recall 94.1%, pada nilai  $C=1$ ; dan akurasi 92.3%, presisi 92.8% dan recall 92.8% pada



nilai  $C=5$ . Pada percobaan ketujuh, dapat dijelaskan bahwa kemampuan pemodelan dalam melakukan klasifikasi dengan tepat menghasilkan akurasi 88.2%. pada 111 record data dan kemampuan dalam mendeteksi nilai yang diprediksi negatif namun aktual positif dari semua data positif (recall) adalah 94.1%. Dengan nilai presisi 84.2% artinya masih terdapat identifikasi beberapa false positif yang dilakukan pemodelan. Pada percobaan kedelapan, menghasilkan akurasi 92.3% yang artinya tingkat keakuratan pemodelan dalam melakukan klasifikasi meningkat dengan penggunaan  $K=15$  namun masih lebih rendah dari akurasi yang dihasilkan pada percobaan ketiga pada  $K=9$ . Dengan nilai presisi 92.8% artinya hasil pemodelan menunjukkan sedikitnya false positif. Dari percobaan ketujuh dan kedelapan, penggunaan nilai  $K$  yang tinggi memiliki pengaruh terhadap nilai evaluasi kinerja pada penggunaan parameter  $C$  dan  $\gamma$  yang sudah sebelumnya dilakukan percobaan.

Pada percobaan berikutnya yaitu percobaan kesembilan, penelitian dilakukan tanpa menggunakan seleksi fitur pada Kombinasi SVM-KNN mampu menghasilkan support vector sebagai dataset baru sebanyak 111 data, dengan Nilai  $K$  yang didapatkan dari nilai error rate terendah pada iterasi yang dilakukan sehingga mendapatkan nilai  $K=8$ , hasil pengujian dari pemodelan yang dibuat menghasilkan akurasi 88.2%, presisi 84% dan recall 100%. Artinya kemampuan pemodelan dalam melakukan ketepatan klasifikasi adalah sebesar 88.2%. Nilai presisi yang dihasilkan menunjukkan pemodelan masih mengidentifikasi beberapa false positif yaitu pada FP sejumlah 4. Nilai recall 100% menunjukkan kemampuan pemodelan dalam menemukan dan mengklasifikasi positif instance dengan benar. Pada percobaan ini



membuktikan kemampuan kombinasi algoritma SVM-KNN tanpa seleksi fitur sudah cukup baik namun tidak lebih baik dari pemodelan SVM-KNN dengan seleksi fitur dalam melakukan seleksi klasifikasi pegawai pada user SAKTI di Biro BMN.

Pada percobaan terakhir, dimana penelitian dilakukan tanpa menggunakan seleksi fitur pada Kombinasi SVM-KNN dengan Nilai  $K=15$  yang didapatkan dari  $\sqrt{\text{jumlah train data}}$ , hasil pengujian mampu menghasilkan nilai akurasi 85.3%, presisi 83.3% dan recall 95.2%. Artinya kemampuan pemodelan dalam melakukan ketepatan klasifikasi adalah sebesar 85.3%. Nilai presisi yang dihasilkan menunjukkan kemampuan pemodelan masih mengidentifikasi beberapa false positif yaitu pada FP sejumlah 4. Nilai recall 95.2% menunjukkan kemampuan pemodelan dalam menemukan dan mengklasifikasi positif instance dengan benar. Pada percobaan ini membuktikan kemampuan kombinasi algoritma SVM-KNN tanpa seleksi fitur dengan nilai  $K=15$  tetap tidak lebih baik dari pemodelan SVM-KNN dengan seleksi fitur dalam melakukan seleksi klasifikasi pegawai pada user SAKTI di Biro BMN.

Perbandingan nilai akurasi dari 10 bentuk kombinasi pengujian yang telah dilakukan dapat diinterpretasikan kedalam diagram seperti pada gambar 4.33.



Gambar 4.33 Perbandingan Nilai Akurasi

Sedangkan visualisasi dari perbandingan nilai Presisi dan Recall dapat ditunjukkan pada gambar 4.34.



Gambar 4.34 Perbandingan Nilai Presisi dan Recall

Hasil penelitian kami menunjukkan bahwa dengan dilakukan penerapan seleksi fitur backward elimination pada kombinasi algoritma SVM-KNN mampu menghasilkan akurasi, presisi dan recall yang lebih baik dari penerapan kombinasi algoritma SVM-KNN tanpa seleksi fitur sehingga dapat dikatakan

berhasil menaikkan kinerja klasifikasi ditunjukkan dengan nilai percobaan terbaik seperti pada tabel 4.22 berikut.

Tabel 4.22 Tabel Perbandingan Hasil Seleksi Fitur dan Tanpa Seleksi Fitur

Percobaan Terbaik	Jumlah Atribut	Confusion Matrix		
		Akurasi(%)	Presisi (%)	Recall (%)
Dengan Seleksi Fitur (Perc ke 3 dan 8)	13	94.1	89.4	100
	13	92.3	92.8	92.8
Tanpa Seleksi Fitur (Perc ke 9)	17	88.2	84	100

Namun perlu diperhatikan penggunaan parameter  $C$  dan  $\gamma$  pada SVM serta nilai tetangga terdekat  $K$  pada algoritma KNN turut berpengaruh terhadap hasil kinerja pemodelan. Selanjutnya dengan penelaahan lebih jauh, ditemukan hasil evaluasi pada percobaan ke 3 dan percobaan ke 8 dengan penerapan seleksi fitur pada kombinasi algoritma SVM-KNN menghasilkan nilai akurasi, presisi dan recall yang mendekati seperti dijelaskan pada tabel 4.23 berikut.

Tabel 4.23 Tabel Perbandingan Evaluasi Kinerja Percobaan ke3 dan ke8

Tabel Evaluasi	Percobaan ke 3	Percobaan ke 8
TP	17	13
TN	15	11
FP	2	1
FN	0	1
Acc, Presisi, Recall	94,1; 89,4; 100	92,3; 92,8; 92,8

Berdasarkan tabel 4.23 diatas, selanjutnya peneliti melakukan perhitungan nilai Kesalahan Klasifikasi dan F-1 Score sebagai berikut:

➤ Percobaan 3

$$\text{Kesalahan Klasifikasi} = \frac{FP + FN}{TP + TN + FP + FN} = \frac{2 + 0}{17 + 15 + 2 + 0} = 0.059 \text{ atau } 5,9\%$$

$$F - 1 \text{ Score} = \frac{2 \times \text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} = \frac{2 \times 1 \times 0.894}{1 + 0.894} = 0.944 \text{ atau } 94,4\%$$

➤ Percobaan 8

$$\text{Kesalahan Klasifikasi} = \frac{FN + FP}{TP + TN + FP + FN} = \frac{1 + 1}{13 + 11 + 1 + 1} = 0,077 \text{ atau } 7,7\%$$

$$F - 1 \text{ Score} = \frac{2 \times \text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} = \frac{2 \times 0.928 \times 0.928}{0.928 + 0.928} = 0.928 \text{ atau } 92,8\%$$

Dalam melakukan pemilihan acuan performa hasil evaluasi terhadap kedua percobaan, bergantung terhadap kebutuhan yang ingin diperoleh. Jika yang diperlukan adalah performa seberapa akurat pemodelan dalam mengklasifikasi dengan benar, maka peneliti dapat menggunakan Akurasi sebagai acuan. Dari kedua percobaan (ke 3 dan ke 8), **akurasi terbaik** didapatkan pada percobaan ke 3 dengan nilai 94,1% dibandingkan 92,3%, didukung dengan hasil perhitungan kesalahan klasifikasi sebesar 5,9% dan nilai F-1 Score sebesar 94,4% yang mana menggambarkan perbandingan rata-rata dari nilai presisi dan recall yang dibobotkan. Sedangkan pada percobaan ke 8 hanya mampu menghasilkan nilai kesalahan klasifikasi terbaik pada 7,7% dengan nilai F-1 Score sebesar 92,8%.

Sedangkan Jika yang diperlukan adalah seberapa sering pemodelan tepat dalam melakukan prediksi, maka percobaan ke 8 yang memiliki nilai terbaik yaitu pada skor presisi 92,8% dibandingkan percobaan ke 3 yang menghasilkan skor presisi 89,4%.

Performa Recall akan dipilih sebagai acuan jika pada kondisi kebutuhan pemilihan False Positif lebih baik terjadi daripada false negatif atau rasio prediksi benar positif dibandingkan dengan keseluruhan data benar positif, maka dalam hal ini nilai FP pada percobaan 3 = 2, sedangkan nilai FN = 0 sehingga menghasilkan Recall 100%. sedangkan pada percobaan 8 menghasilkan recall 92.8%.

Berdasarkan pengerucutan pengujian dan didasarkan pada rumusan masalah, maka dapat dikatakan percobaan ke 3 mampu menghasilkan akurasi terbaik.

Kombinasi algoritma SVM dan KNN dengan Penerapan seleksi fitur backward elimination dan setting parameter yang tepat dapat diterapkan pada pengolahan dataset dalam melakukan seleksi klasifikasi pegawai yang memiliki rekapitulasi kehadiran (WFO dan WFH) dan penilaian kinerja.

Seperti halnya pada pemodelan ini mampu digunakan dalam melakukan seleksi klasifikasi pegawai dengan parameter nilai C yang rendah  $C=1$ , nilai gamma yang tinggi yaitu pada 0.1 dan nilai  $K=9$ , pengujian mampu menghasilkan akurasi terbaik yaitu pada **94.1%**, dan **F1 score sebesar 94.4%**.



## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan serangkaian tahapan pengujian dalam melakukan seleksi klasifikasi pegawai menggunakan penerapan Seleksi Fitur Backward Elimination pada kombinasi Algoritma Support Vector Machine dan K-Nearest Neighbor, maka dapat disimpulkan bahwa:

1. Penerapan seleksi fitur pada kombinasi algoritma SVM-KNN dalam melakukan seleksi klasifikasi pegawai mampu menghasilkan perbaikan peningkatan akurasi lebih baik sebesar 94.1% dibandingkan tanpa seleksi fitur sebesar 88.2%. Hal ini sesuai dengan 2 penelitian terdahulu yang dilakukan Ahmad Bode (2017) dimana penerapan algoritma KNN dengan seleksi fitur Backward Elimination menghasilkan tingkat error paling kecil 0.093 dan akurasi yang lebih baik 95,3% dibandingkan tingkat error tanpa seleksi fitur 0.098 dan penelitian berikutnya yang dilakukan Pham, B dkk (2021) dengan penerapan backward elimination mampu menghasilkan perbaikan akurasi 91.31% dari 76,4%. Sehingga penerapan backward elimination dapat dijadikan pertimbangan dalam pemodelan klasifikasi penelitian lainnya untuk kebutuhan peningkatan akurasi.
2. Dari hasil perbandingan perhitungan performa pemodelan yang didasarkan pada rumusan masalah kebutuhan keakuratan pemodelan dalam mengklasifikasi dengan benar atas serangkaian pengujian didapatkan

akurasi terbaik sebesar 94,1% dengan nilai F1 Score sebesar 94,4% pada percobaan ke 3 penerapan seleksi fitur dengan menggunakan parameter  $C = 1$ , nilai  $\text{Gamma} = 0.1$  dan Nilai tetangga terdekat  $K$  pada  $K=9$ . Akurasi terendah pada percobaan ke 2 dengan parameter  $C=5$ ,  $\text{Gamma}=0.01$  dan Nilai  $K=9$ , sebesar 82.6%.

3. Hasil serangkaian pengujian yang dilakukan, akurasi akan lebih baik pada nilai parameter  $C=1$  daripada Nilai parameter  $C$  yang tinggi,  $C=5$  sedangkan pada parameter  $\text{Gamma}$  akan menghasilkan pemodelan yang lebih baik pada nilai  $\text{Gamma}=0.1$  dibanding pada nilai parameter  $\text{Gamma}=0.01$ .
4. Penerapan seleksi fitur backward elimination pada kombinasi algoritma SVM dan KNN dapat diterapkan dalam melakukan seleksi klasifikasi pegawai yang memiliki rekapitulasi kehadiran (WFO dan WFH) dan penilaian kinerja.

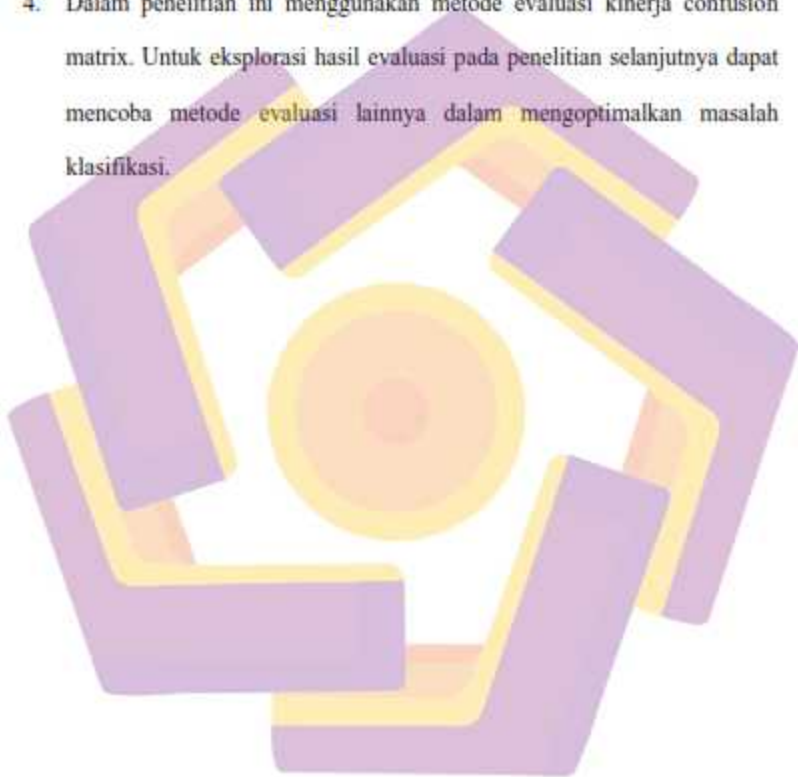
## 5.2. Saran

Untuk mendapatkan hasil pengujian yang lebih baik, berdasarkan kesimpulan diatas, terdapat beberapa saran untuk penelitian selanjutnya, antara lain:

1. Menambah jumlah data yang lebih besar dengan atribut yang lebih banyak sehingga hasil pengujian didapatkan lebih baik lagi.
2. Penelitian ini menggunakan proporsi data 70:30, diharapkan penelitian selanjutnya dapat melakukan perbandingan train-test-split yang berbeda -

beda, karena hal ini dapat memberikan pengaruh hasil pemodelan yang berbeda.

3. Perlu menambah variasi nilai parameter baik parameter  $C$ ,  $\gamma$  maupun nilai  $K$  sehingga dapat mengetahui kinerja pemodelan yang lebih beragam
4. Dalam penelitian ini menggunakan metode evaluasi kinerja confusion matrix. Untuk eksplorasi hasil evaluasi pada penelitian selanjutnya dapat mencoba metode evaluasi lainnya dalam mengoptimalkan masalah klasifikasi.



## DAFTAR PUSTAKA

### PUSTAKA BUKU

Kotu, V., & Deshpande, B. *Predictive analytics and data mining: concepts and practice with rapidminer*. Waltham, USA: Morgan Kaufmann, 2014

Kusrini and E. T. Luthfi, *ALGORITMA DATA MINING*. Yogyakarta: ANDI, 2009

R. Primartha, *algoritma machine learning*. Bandung: Informatika Bandung, 2021

### PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

Adnyana, I. M. B. (2019). Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa. *Jurnal Sistem dan Informatika (JSI)*, 13(2), 72-76.

Amriani, T. N., & Iskandar, A. (2019). Analisis Kesuksesan Implementasi Sistem Aplikasi Keuangan Tingkat Instansi (SAKTI) pada Satuan Kerja di Lingkungan Badan Pendidikan dan Pelatihan Keuangan (BPPK). *Kajian Ekonomi dan Keuangan*, 3(1), 54-74.

Anwar, A. I., & Hadi, M. (2022). Implementasi aplikasi SAKTI dan SPAN dalam penyusunan laporan keuangan. *Jurnal Informatika*, 1(2), 32-55.

Astrianda, N. (2020). Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine. *VOCATECH: Vocational Education and Technology Journal*, 1(2), 110-117.

Azizah, R. A., Bachtiar, F., & Adinugroho, S. (2022). Klasifikasi Kinerja Akademik Siswa Menggunakan Neighbor Weighted K-Nearest Neighbor dengan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(4), 655-664.

Baihaqi, W. M., Pinilih, M., & Rohmah, M. (2020). Kombinasi K-MEANS dan Support Vector Machine (SVM) Untuk Memprediksi Unsur Sara pada Tweet. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), 501-510.

Bode, A. (2017). K-nearest neighbor dengan feature selection menggunakan backward elimination untuk prediksi harga komoditi kopi arabika. *ILKOM Jurnal Ilmiah*, 9(2), 188-195.

Desiani, A., Akbar, M., Irmeilyana, & Amran, A. (2022). Implementasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. *Jurnal Teknik Elektro Dan Komputasi (ELKOM)*, 4(2), 207-214

Dyoniputri, H. (2021). A hybrid convolutional neural network and support vector machine for dysarthria speech classification. *International Journal of Innovative Computing, Information and Control*, 17(1), 111-123.

Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting employee attrition using machine learning techniques. *Computers*, 9(4), 86.

Fredilio, F., Rahmad, J., Sinurat, S. H., Sitompul, D. R. H., Ziegel, D. J., & Indra, E. (2023). Perbandingan Algoritma K-Nearest Neighbors (K-NN) dan Random forest terhadap Penyakit Gagal Jantung. *Jurnal Teknologi Informatika dan Komputer*, 9(1), 471-486.



- Ginting, S. L. B., & Permana, A. A. (2018). Hybrid Classifier System: Support Vector Machines Dikombinasikan dengan K-Nearest Neighbors untuk Menentukan Kelayakan Nasabah Bank dalam Pengajuan Kredit. *Komputika: Jurnal Sistem Komputer*, 7(1), 9-16.
- Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, 62, 18-22.
- Kari, S. L., De Silva, L. C., Lai, D. T. C., & Yong, S. Y. (2021). Classification and prediction of driving behaviour at a traffic intersection using SVM and KNN. *SN computer science*, 2(3), 1-11.
- Michael, A. (2022). Komparasi Kombinasi Pre-trained Model dengan SVM pada Klasifikasi Kematangan Kopi Berbasis Citra. *Journal Dynamic saint E-ISSN*, 2722, 5364.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika*, 4(01), 20-24.
- Nugroho, H. P., & Lestiyowati, J. (2020). Analisis Tingkat Kepuasan dan Kepentingan Pengguna Aplikasi SAKTI dengan PIECES Framework. *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara dan Kebijakan Publik*, 5(2), 93-104.
- Nurkholis, A., Abidin, Z., & Sulistiani, H. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 904-910.
- PAULINA, N. E., FITRI, Z. E., MADJID, A., & IMRON, A. M. N. (2021). Klasifikasi Kerusakan Muta Tomat Berdasarkan Seleksi Fitur Menggunakan K-Nearest Neighbor. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, 6(2), 144-154.
- Pham, B. T., Van Dao, D., Acharya, T. D., Van Phong, T., Costache, R., Van Le, H., ... & Prakash, I. (2021). Performance assessment of artificial neural network using chi-square and backward elimination feature selection methods for landslide susceptibility analysis. *Environmental Earth Sciences*, 80, 1-13.
- Prasetyo, A., Kusriani, K., & Arief, M. R. (2019). PENERAPAN ALGORITMA K NEAREST NEIGHBOR UNTUK REKOMENDASI MINAT KONSENTRASI DI PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS PGRI YOGYAKARTA. *Informasi Interaktif*, 4(1), 1-6
- Putri, D. D., Furqon, M. T., & Perdana, R. S. (2018). Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDTSVM)(Studi Kasus: Puskesmas Dinoyo Kota Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(5), 1912-1920.
- Rahmansyah, A., Dewi, O., Andini, P., Ningrum, T. H. P., & Suryana, M. E. (2018, August). Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- Reddy, D. J. M., Regella, S., & Seelam, S. R. (2020, October). Recruitment prediction using machine learning. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-4). IEEE.



- Riani, N. A., Andreswari, R., & Fauzi, R. (2021). IMPLEMENTASI ALGORITMA ARTIFICIAL NEURAL NETWORK PADA PROSES SELEKSI PILIH PEMINATAN (STUDI KASUS: PRODI SISTEM INFORMASI UNIVERSITAS TELKOM). *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, 4(3), 241-247
- Saifudin, A. (2018). Metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di Universitas Pamulang. *Jurnal Teknologi*, 10(1), 25-36
- Sasongko, T. B., & Arifin, O. (2019). Implementasi Metode Forward Selection pada Algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier Kernel Density (Studi Kasus Klasifikasi Jalur Minat SMA). *J. Teknol. Inf. dan Ilmu Komput*, 6(4), 383-388.
- Sunandar, H. S., & Satar, M. (2020). Tinjauan tentang proses recruitment tenaga kerja di kantor pusat PT. Y Bandung. *Jurnal Industri Elektro dan Penerbangan*, 6(1)
- Suppers, A., van Gool, A. J., & Wessels, H. J. C. T. (2018). Integrated chemometrics and statistics to drive successful proteomics biomarker discovery. *Proteomes*, 6(2). <https://doi.org/10.3390/PROTEOMES6020020>
- Supristiowadi, E., & Sucahyo, Y. G. (2018). Manajemen Risiko Keamanan Informasi Pada Sistem Aplikasi Keuangan Tingkat Instansi (SAKTI) Kementerian Keuangan. *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara dan Kebijakan Publik*, 3(1), 23-33.
- Suwinardi, S. (2017). Profesionalisme dalam bekerja. *Orbith: Majalah Ilmiah Pengembangan Rekayasa dan Sosial*, 13(2)
- Yusra, R. N., Sitompul, O. S., & Sawaluddin, S. (2021). Kombinasi K-Nearest Neighbor (KNN) dan Relief-F untuk Meningkatkan Akurasi Pada Klasifikasi Data. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, 6(1), 16-21.

#### **PUSTAKA LAPORAN PENELITIAN**

- Agustin, T (2021). Analisis Kombinasi Convolution Neural Network(CNN) dan Support Vector Machine (SVM) dalam deteksi otomatis Non-Proliferative Diabetic Retinopathy
- Dharmawan, S (2018). Feature Selection Berbasis ABC-SVM dan PSO-SVM dalam masalah klasifikasi