

**Person-Job Fit Menggunakan Decision Trees Algorithm
(Studi Kasus PT PNR Perusahaan Jasa Pertambangan pada Divisi PL)**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Program Studi Teknik Industri - Fakultas Teknologi Industri
Universitas Islam Indonesia**



Nama : Jasmine Alifah Putri
No. Mahasiswa : 19522240

**PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2024**

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 10 Februari 2024



Jasmine Alifah Putri

19522240

SURAT BUKTI PENELITIAN



FAKULTAS
TEKNOLOGI INDUSTRI

Gedung KH. Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kalirejo km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. fti@uii.ac.id
W. fti.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 003/Ka.Lab.Datmin/70/Lab.Datmin/IV/2024

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Jasmine Alifah Putri
No. Mhs : 19522240
Dosen Pembimbing : Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.

Telah selesai melaksanakan penelitian yang berjudul "Person-Job Fit Menggunakan Decision Tree Algorithm (Studi Kasus Perusahaan Jasa Pertambangan)" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 09 Maret 2023 sampai dengan tanggal 12 April 2023

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 19 April 2024

Kepala Laboratorium
Data Mining

Annisa Uswatun Khasanah, ST., M.B.A., M.Sc.



LEMBAR PENGESAHAN PEMBIMBING**Person-Job Fit Menggunakan Decision Trees Algorithm
(Studi Kasus PT PNR Perusahaan Jasa Pertambangan pada Divisi PL)****Dosen Pembimbing****(Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.)**

LEMBAR PENGESAHAN DOSEN PENGUJI

**Person-Job Fit Menggunakan Decision Trees Algorithm
(Studi Kasus PT PNR Perusahaan Jasa Pertambangan pada Divisi PL)**

TUGAS AKHIR

Disusun Oleh :

Nama : Jasmine Alifah Putri

No. Mahasiswa : 19 522 240

Telah dipertahankan di depan sidang pengujian sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 20 - Juni - 2024

Tim Penguji

Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.

Ketua

Dr. Harwati, S.T., M.T.

Anggota I

Elanjati Worldailmi, S.T., M.Sc.

Anggota II

Mengetahui,

Ketua Program Studi Teknik Industri Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM.

015220101

HALAMAN PERSEMBAHAN

Tugas Akhir ini Saya persembahkan kepada ayah dan bunda, adik-adik, keluarga besar, teman, serta seluruh pihak yang telah memberi doa dan dukungan dalam perjalanan serta perjuangan yang Saya lalui selama masa perkuliahan selama ini.

MOTTO

The greatest glory in living lies not in never falling, but rising every time we fall.

– Nelson Mandela

KATA PENGANTAR

Puji dan syukur penulis panjatkan atas ke hadirat Allah SWT yang senantiasa memberikan rahmat serta karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “*Person-Job Fit Menggunakan Decision Trees Algorithm (Studi Kasus Perusahaan Pertambangan)*” sebagai salah satu syarat memperoleh gelar Sarjana Strata-1 Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Ucapan terima kasih tak luput penulis sampaikan kepada:

1. Bapak Prof., Dr., Ir., Hari Purnomo, M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM., selaku Ketua Program Studi Teknik Industri Program Sarjana dan dosen pembimbing yang telah memberikan bimbingan, ilmu, waktu dan dorongan kepada penulis selama penyusunan Tugas Akhir ini.
3. PT PNR yang telah memberikan kesempatan dan dukungan selama masa penelitian.
4. Keluarga tersayang, ayah Triawan Saleh dan bunda Ranti Puspitasari serta adik-adik tercinta Rayendra Rakha Pratama, Khayla Almira Maritza dan Fathan Raffasya Daniyal yang selalu memberikan kasih sayang, doa, dukungan dan semangat bagi penulis.
5. Teman-teman dekat saya, khususnya grup “Awtelo” dan Erhan yang senantiasa menjadi tempat penulis untuk bercerita dan berkeluh kesah, memberi semangat, dorongan, hiburan, maupun bantuan secara moril dan materiil.
6. Teman-teman Laboratorium Data Mining 2019 yang sudah memberikan wadah bagi penulis untuk belajar, berkembang semasa kuliah.

Penulis memahami sepenuhnya bahwa Tugas Akhir ini jauh dari kata sempurna dan penuh kekurangan. Kritik dan saran yang membangun akan senantiasa diterima untuk perbaikan di masa mendatang. Penulis berharap, Tugas Akhir ini dapat memberikan pengetahuan lebih kepada pembaca untuk terus memberikan manfaat dan pembaharuan yang lebih luas lagi.

ABSTRAK

PT PNR, sebagai salah satu perusahaan jasa pertambangan terbesar di Indonesia, menghadapi persaingan ketat dalam merekrut karyawan yang berkualitas. Dalam upaya mempertahankan keunggulan kompetitif, PT PNR berusaha menarik dan mempertahankan karyawan terbaik. Salah satu tahap krusial dalam proses rekrutmen adalah tes psikologi, yang memberikan gambaran karakteristik potensi calon karyawan. Namun, hasil dari tes psikologi ini belum sepenuhnya dimanfaatkan secara optimal oleh PT PNR. Oleh karena itu, penelitian ini dilakukan dengan fokus pada memprediksi person-job fit menggunakan dua model klasifikasi: Decision Tree dan Random Forest. Dalam penelitian ini kami memanfaatkan data dari 18 calon karyawan di divisi PL PT PNR. Karena ukuran sampel yang terbatas, kami melakukan penyesuaian dengan bantuan teknik augmentasi menggunakan machine learning. Hasil penelitian menunjukkan bahwa hyperparameter tuning dapat signifikan meningkatkan performa model. Model terbaik yang dihasilkan dapat menjadi alat bantu bagi PT PNR dalam proses rekrutmen, membantu memastikan person-job fit yang optimal.

Kata Kunci: *Person-Job Fit*, Klasifikasi, Industri Pertambangan, *Decision Trees Algorithm*

DAFTAR ISI

PERNYATAAN KEASLIAN.....	iii
SURAT BUKTI PENELITIAN	iv
LEMBAR PENGESAHAN PEMBIMBING	v
LEMBAR PENGESAHAN DOSEN PENGUJI.....	vi
HALAMAN PERSEMBAHAN	vii
MOTTO	viii
KATA PENGANTAR.....	ix
ABSTRAK	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Penelitian	5
BAB II TINJAUAN PUSTAKA	7
2.1 Kajian Literatur	7
2.2 <i>Research Gap</i>	11
2.3 Landasan Teori.....	15
2.3.1 War for Talents	15
2.3.2 Rekrutmen	16
2.3.3 Potensi Karyawan.....	17
2.3.4 Job Performance.....	18
2.3.5 Person-Job Fit.....	19
2.3.6 Algoritma Klasifikasi	20
2.3.7 Hyperparameter Optimization.....	22
2.3.8 Evaluasi Model Klasifikasi	24
BAB III METODE PENELITIAN.....	28
3.1 Subjek Penelitian.....	28
3.2 Objek Penelitian	28

3.3	Lokasi dan Waktu Penelitian	28
3.4	Alur Penelitian	28
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA.....		32
4.1	Pengumpulan Data	32
4.2	Augmentasi Data.....	32
4.3	<i>Pre-Processing</i>	32
4.3.1	Informasi Statistik dan Identifikasi Missing Value.....	33
4.3.2	Cleaning Data.....	37
4.3.3	Transformasi Data.....	41
4.3.4	Uji Korelasi	41
4.4	<i>Exploratory Data Analysis (EDA)</i>	44
4.5	Pembangunan Model.....	48
4.6	Pengolahan Model Klasifikasi dengan <i>Parameter Tuning</i>	51
4.7	<i>Features Importance</i>	53
BAB V PEMBAHASAN.....		56
5.1	Performa Model Klasifikasi <i>Decision Tree</i> dan <i>Random Forest</i>	56
5.2	Perbandingan Performa Model Klasifikasi	64
5.3	Analisis Perbaikan.....	65
BAB VI PENUTUP.....		68
6.1	Kesimpulan	68
6.2	Saran.....	68
DAFTAR PUSTAKA.....		70
LAMPIRAN		A-1

DAFTAR TABEL

Tabel 2.1.1 Kajian Literatur.....	7
Tabel 2.2.1 <i>Research Gap</i>	11
Tabel 2.3.1 Parameter Set <i>Decision Tree</i>	23
Tabel 2.3.2 Parameter Set <i>Random Forest</i>	24
Tabel 4.3.1 Informasi Statistik Data	34
Tabel 4.3.2 <i>Cleaning Data</i>	37
Tabel 4.3.3 Data Hasil <i>Cleaning</i>	38
Tabel 4.3.4 Atribut Transformasi	41
Tabel 4.3.5 Interpretasi Koefisien Korelasi	42
Tabel 4.3.6 Nilai Korelasi antar Atribut	42
Tabel 5.1.1 Perbandingan Performa model Klasifikasi Algoritma <i>Decision Tree</i>	56
Tabel 5.1.2 Perbandingan Performa model Klasifikasi Algoritma <i>Random Forest</i>	59
Tabel 5.1.3 Derajat Ketidakseimbangan.....	62
Tabel 5.2.1 Perbandingan Performa Model	64

DAFTAR GAMBAR

Gambar 1.1.1 Alur Rekrutmen PT PNR	3
Gambar 2.3.1 Struktur <i>Decision Tree</i>	21
Gambar 2.3.2 Struktur <i>Random Forest</i>	22
Gambar 2.3.3 Empat Konsep Utama Matriks Evaluasi	25
Gambar 2.3.4 <i>Confusion Matrix</i>	27
Gambar 3.4.1 Alur Penelitian	29
Gambar 4.3.1 <i>Correlation Matrix</i>	42
Gambar 4.4.1 Persentase Kelas Data <i>Person-Job Fit</i>	44
Gambar 4.4.2 Sebaran Pendidikan pada setiap Kelas Data	45
Gambar 4.4.3 Sebaran Asal Sekolah pada Divisi PL	46
Gambar 4.4.4 Sebaran Usia pada setiap Kelas Data.....	46
Gambar 4.4.5 Sebaran Skor IQ pada setiap Kelas Data	47
Gambar 4.4.6 Sebaran Skor Pauli pada setiap Kelas Data	47
Gambar 4.4.7 Sebaran <i>Gap</i> Skor Potensi pada setiap Kelas Data	47
Gambar 4.5.1 Pembangunan Model Menggunakan <i>Cross Validation</i>	48
Gambar 4.5.2 Tahap <i>Pre-processing</i> Data.....	49
Gambar 4.5.3 Pembangunan Model <i>Decision Tree</i> Menggunakan <i>Cross Validation</i>	50
Gambar 4.5.4 Pembangunan Model <i>Random Forest</i> Menggunakan <i>Cross Validation</i> ..	50
Gambar 4.6.1 Pembangunan Model Menggunakan <i>Parameter Tuning</i>	51
Gambar 4.6.2 <i>Parameter Tuning</i> pada Algoritma <i>Decision Tree</i>	52
Gambar 4.6.3 <i>Select parameters: Configure operator</i> Algoritma <i>Decision Tree</i>	52
Gambar 4.7.1 <i>Features Importance</i>	54
Gambar 4.7.2 Pareto <i>Chart</i> Fitur terhadap <i>Person-Job Fit</i>	55

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tingginya *demand* yang diinginkan oleh perusahaan ternyata juga diikuti dengan tingginya ekspektasi yang dimiliki oleh calon karyawan seperti yang dikatakan oleh Green (2022), seorang *expert* dalam *people analysis*. Kandidat di era ini cenderung lebih selektif dalam memilih pekerjaan mereka dengan mengharapkan sesuatu yang lebih dari perusahaan. Perusahaan pun menerapkan pendekatan yang sama dalam proses seleksi karyawan mereka.

Di tengah meningkatnya *war for talents*, perusahaan harus mencari dan memilih kandidat melalui tahapan rekrutmen yang bertujuan untuk merekrut individu dengan kualifikasi sesuai persyaratan (Zaky, 2021). Keberhasilan rekrutmen bergantung pada kemampuan perusahaan untuk memenuhi jumlah karyawan dengan kualifikasi yang sesuai serta memastikan kebutuhan tenaga kerja terpenuhi tepat waktu. Salah satu tahapan penting dalam mencari karyawan yang tepat adalah tes psikologi atau psikotes, yang digunakan untuk mengevaluasi potensi individu melalui berbagai dimensi psikologis, termasuk kepribadian, intelegensi, kecenderungan perilaku, dan sikap kerja (Susanto, Parmenas, Suryawan, & Apriyani, 2023).

Potensi individu adalah kapasitas seseorang untuk mencapai kinerja optimal di masa mendatang (Robledo-Ardila & Roman-Calderon, 2020). Penilaian potensi ini menjadi langkah krusial sebelum mengambil keputusan untuk menjalin hubungan kerja dengan seorang kandidat. Tindakan ini memberikan kesempatan bagi perusahaan untuk mengevaluasi apakah kandidat memiliki kapabilitas yang memadai untuk melaksanakan pekerjaannya secara efektif. Hal ini sangat signifikan, sebab investasi dalam kandidat yang *under performing* atau tidak bisa mencapai kinerja optimal dapat merugikan perusahaan (Susanto, Parmenas, Suryawan, & Apriyani, 2023). Oleh karena itu, sangat perlu mempertimbangkan dan menggunakan potensi kandidat sebagai salah satu dasar dalam pengambilan keputusan rekrutmen. Harapannya, dengan mengetahui potensi kandidat lebih awal, perusahaan dapat memilih kandidat yang paling sesuai dengan tuntutan pekerjaan dengan berdasarkan kapasitas dan karakteristik potensinya.

Kecocokan kandidat berdasarkan potensi terhadap tuntutan pekerjaannya sering juga disebut sebagai *person-job fit*.

Person-job fit merujuk pada kesesuaian karakteristik seseorang dengan tugas pekerjaan yang dilakukan. Konsep ini terdiri dari dua komponen dasar yaitu kesesuaian antara *job requirement* dan kemampuan seseorang (*demand – ability fit*), serta kesesuaian antara kebutuhan individu dengan hal yang terkait pekerjaannya seperti vakasi, sumber daya yang diberikan dan hubungan interpersonal (*needs - supplies fit*) (Xiao, et al., 2021). Potensi individu termasuk dalam kemampuan yang dimiliki kandidat. Penelitian oleh Chen dkk (2019) juga menyatakan bahwa pada saat karyawan diberi pelatihan dan potensi mereka berkembang, karyawan akan memiliki kecenderungan puas dengan pekerjaan mereka. Karyawan yang memiliki performa yang baik akan lebih efektif dalam berkontribusi untuk menciptakan lingkungan kerja yang lebih nyaman (Ali & Anwar, 2021) dan puas saat bekerja, sehingga memiliki kecil kemungkinan untuk *resign* (Chen, et al., 2019).

Ruang lingkup yang sempit dalam industri pertambangan di Indonesia meningkatkan persaingan antar perusahaan dalam mencari kandidat, sehingga seleksi menjadi lebih ketat dan spesifik. Banyak faktor yang menyebabkan perusahaan kesulitan memperoleh kandidat dalam industri ini, salah satunya yaitu tingginya kualifikasi yang dibutuhkan bagi kandidat, khususnya bagi divisi yang berhubungan langsung dengan proses penambangan, atau dalam PT PNR disebut juga sebagai *core business*. Sebagai salah satu perusahaan jasa pertambangan terbaik di Indonesia, PT PNR menjalani proses rekrutmen yang panjang untuk menemukan calon karyawan yang tepat dan berkualitas, yang sesuai dengan tuntutan pekerjaan berdasarkan kapasitas dan karakteristik potensinya.

Untuk mengetahui karakteristik potensi kandidat tersebut, dilakukan tes psikologi atau psikotes dalam tahap rekrutmen. Dari tahap tersebut, akan diperoleh laporan hasil psikotes yang berisi potensi kandidat. Hasil ini akan dijadikan sebagai bekal bagi perusahaan untuk dapat memberikan *coaching* dan pelatihan yang sesuai untuk dapat memaksimalkan potensi tersebut sehingga calon karyawan nantinya akan memiliki performa kerja yang tinggi. Selain itu, dalam melakukan seleksi kandidat yang cocok menggunakan hasil psikotes, tentu diperlukan standar khusus penilaian untuk masing-masing divisi tersebut. Saat ini, PT PNR sudah memiliki standar untuk hasil psikotes, namun standar yang ada masih belum seragam. Sedangkan pada konsep *person-job fit*

dengan pekerjaannya kelak, nyatanya masih terdapat kasus *resign* karyawan yang tergolong banyak khususnya pada tahun 2023-2024 pada divisi PL terdapat 352 kasus *resign* karyawan. Hal ini menunjukkan adanya tantangan dalam pengelolaan karyawan secara efektif dan efisien. Maksud pengelolaan karyawan dalam hal ini adalah memilih karyawan yang memiliki potensi yang sesuai dengan kebutuhan perusahaan untuk dapat berkontribusi secara positif dalam pekerjaannya sehingga nantinya mampu bertahan dalam jangka panjang.

Berdasarkan permasalahan di atas, penelitian ini dilakukan untuk memaksimalkan pemanfaatan data hasil psikotes dengan membuat sebuah model prediksi *person-job fit* menggunakan salah satu fungsi *machine learning* yaitu klasifikasi dan juga sebagai langkah preventif serta antisipatif dalam upaya mengurangi jumlah kasus ketidakcocokan karyawan dalam proses rekrutmen. Data yang akan digunakan pada penelitian ini yaitu *historical data* psikotes dan performansi karyawan *existing* yang akan dijadikan sebagai dasar pembuatan model klasifikasi. Pada penelitian ini, kelas data *person-job fit* digambarkan menggunakan nilai performansi kinerja yang akan dibagi kedalam dua kelas yaitu *highly fit* dan *moderately fit*. Dalam meneliti *person-job fit* menggunakan fungsi klasifikasi, metode yang digunakan yaitu *Decision Tree* dan *Random Forrest*. Adapun algoritma *Decision Tree* yang digunakan pada penelitian ini merupakan *Decision Tree* sederhana dan bukan merupakan turunan maupun algoritma perbaikan dari *Decision Tree*. Penggunaan *decision trees algorithm* dengan *Decision Tree* dan *Random Forrest* adalah karena banyaknya keunggulan dibanding metode klasifikasi lainnya yaitu *decision trees algorithm* umum digunakan dalam berbagai bidang seperti keuangan, kesehatan, studi lingkungan dan lain sebagainya. Mampu digunakan pada data kategorikal maupun numerikal dan pada metode *ensemble* seperti *Random Forrest* terbukti mampu untuk meningkatkan efektivitas dan akurasi model (Motaghifard, Omidvari, & Kazemi, 2023; Charbuty & Abdulazeez, 2021). Adapun hasil akhir yang diharapkan pada prediksi *person-job fit* yaitu sebuah model klasifikasi yang menggambarkan kondisi yang harus terpenuhi (berdasarkan variabel) dan kelas data dari sebuah data serta nilai akurasi dan evaluasi dari model yang digunakan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, berikut rumusan masalah pada penelitian ini:

1. Berapa kinerja model klasifikasi *person-job fit* calon karyawan PT PNR menggunakan algoritma *Decision Tree* dan *Random Forrest*?
2. Apa variabel potensi yang paling berpengaruh pada *person-job fit*?

1.3 Tujuan Penelitian

Tujuan dari dilaksanakannya penelitian ini yaitu:

1. Menghitung kinerja model klasifikasi *person-job fit* calon karyawan PT PNR menggunakan algoritma *Decision Tree* dan *Random Forrest*.
2. Mengidentifikasi variabel potensi yang paling berpengaruh pada *person-job fit*.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini bagi PT PNR yaitu:

1. Mengetahui penerapan *Machine Learning* khususnya algoritma *tree classifier* dalam melakukan proses rekrutmen.
2. Mengetahui urutan kepentingan dan hubungan dari variabel potensi yang memengaruhi *person-job fit*.
3. Merancang *tools* dan model prediksi *person-job fit*.
4. *Transfer knowledge* dan *improvement* mengenai prediktif analisis pada proses rekrutmen.

Manfaat yang diperoleh dari penelitian ini bagi Universitas yaitu:

1. Menerapkan keilmuan teknik industri terhadap prediktif analisis pada proses rekrutmen.
2. Sebagai referensi pembelajaran dalam penerapan dan pemanfaatan *data mining* dalam manajemen sumber daya manusia.

1.5 Batasan Penelitian

Batasan penelitian yang digunakan dalam penelitian ini yaitu:

1. Data psikotes karyawan yang digunakan dalam penelitian adalah dengan kriteria karyawan Divisi PL yang masuk pada tahun 2015-2019 dengan minimum pendidikan sarjana.
2. Variabel atau prediktor yang digunakan merupakan variabel psikologis yang diperoleh dari hasil psikotes karyawan.

3. Metode Klasifikasi yang digunakan pada penelitian ini *Decision Tree* dan *Random Forrest*.
4. Aplikasi yang digunakan untuk proses komputasi pembangunan model yaitu *Rapid Miner*..

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Penelitian terkait *person-job fit* sudah pernah dilakukan sebelumnya, Tabel 2.1.1 menunjukkan beberapa penelitian terdahulu yang mendukung penelitian ini:

Tabel 2.1.1 Kajian Literatur

Judul	Metode	Hasil
<i>An Enhanced Neural Network Approach to Person Job Fit in Talent Recruitment</i> (Qin, et al., 2020)	Recurrent Neural Network (RNN)	Penelitian ini dilakukan untuk meningkatkan efisiensi dan membuat perbaikan strategi pada prediksi <i>person-job fit</i> menggunakan RNN dan <i>hierarchical topic-based ability-aware</i> (TAPJFNN). Penelitian ini mengukur derajat kecocokan resume dan <i>job requirement</i> lalu membuat perbaikan model prediksi PJF dengan berdasarkan pada data historis yang ada. Dari hasil juga diketahui bahwa usulan metode ini memiliki hasil performa yang lebih tinggi dibanding metode lainnya.
<i>Person-Job Fit Estimation from Candidate Profile and Related Recruitment History With Co-Attention Neural Networks</i> (Wang, Wei, Xu, Xu, & Mao, 2022)	Graph Neural Network (GNN)	Penelitian ini dilakukan menggunakan GNN dengan usulan model baru PJFCANN yang memprediksi derajat kecocokan <i>job posting</i> dan resume kandidat tidak hanya berdasarkan isi konteks atau isi tetapi juga berdasarkan data historis rekrutmen yang berhasil mencocokkan postingan pekerjaan dan resume. Performa metode ini juga lebih tinggi dibandingkan metode lainnya. Peneliti juga menyatakan pentingnya melakukan pengembangan efisiensi <i>person-job fit</i> .
<i>Longitudinal Effects Of High-Performance Work Practices on Job Performance Via Person-Job Fit</i> (Kaur & Kaur, 2023)	Kuantitatif Longitudinal	Penelitian ini dilakukan untuk mengetahui peran <i>person-job fit</i> terhadap <i>job performance</i> . sebanyak 480 data diperoleh melalui kuesioner. Dari penelitian ini diperoleh hasil bahwa <i>person-job fit</i> memiliki pengaruh yang positif dan signifikan terhadap <i>employee performance</i> . Dari penelitian ini juga diketahui bahwa ketika terdapat kecocokan antara kebutuhan dan kompetensi terhadap suatu pekerjaan, produktivitas

Judul	Metode	Hasil
<p><i>The Impact Of Person Job Fit And Person Organization Fit On Employee Job Performance: A Study Among Employees Of Services Sector</i> (Rajper, Ghumro, & Mangi, 2020)</p>	<p>Kuantitatif - Kuesioner</p>	<p>karyawan akan semakin meningkat dan mereka cenderung menunjukkan peningkatan dalam kinerja.</p> <p>Penelitian ini dilakukan untuk mengetahui hubungan <i>person-job fit</i>, <i>person-organization fit</i> dan <i>employee job performance</i>. data sejumlah 381 diperoleh dengan menggunakan kuesioner. Dari hasil analisis diketahui bahwa <i>person job fit</i> memiliki hubungan yang positif dan signifikan terhadap <i>employee job performance</i> dengan persentase sebesar 78%. Peneliti juga mengatakan bahwa semakin baik kecocokan, akan mendorong <i>job performance</i> yang tinggi serta meningkatnya efektivitas perusahaan.</p>
<p><i>Classification Talent of Employee Using C4.5, KNN, SVM</i> (Stephanie & Sarno, 2019)</p>	<p>C4.5, KNN (K-Nearest Neighbor), SVM (Support Vector Machine)</p>	<p>Penelitian ini dilakukan untuk mengelompokkan karyawan kedalam 4-<i>box</i> manajemen talenta menggunakan C4.5, KNN dan SVM. Kriteria yang dipakai sebanyak 18 untuk mengukur performa dan potensi karyawan. Dari hasil diperoleh bahwa SVM memiliki akurasi tertinggi sebesar 94.62%, SVM sebesar 0.9462% dan KNN memiliki akurasi terendah sebesar 87.37%</p>
<p><i>Identifying High Potential Talent: A Neural Network Based Dynamic Social Profiling Approach</i> (Ye, et al., 2020)</p>	<p>Graph Neural Network (GNN)</p>	<p>Penelitian ini dilakukan untuk mengidentifikasi <i>high potential talents</i> (HIPOs) secara kuantitatif menggunakan <i>graph based neural network</i> pada karyawan baru dengan dengan memodelkan dinamika perilaku mereka dalam jaringan sosial organisasi. Hasil penelitian ini menunjukkan efektivitas GNN dalam mengidentifikasi HIPO berdasarkan pada kemampuan sosial mereka. Peneliti juga menyatakan bahwa HIPO sering dianggap sebagai pemimpin masa depan dalam organisasi. Dibandingkan rekannya, HIPO memiliki kemampuan kepemimpinan, ketajaman bisnis, dan keinginan untuk sukses dan biasanya berkembang dengan lebih cepat. Oleh karena itu penting untuk mengidentifikasi HIPO khususnya pada karyawan baru sehingga sehingga dapat diberikan perhatian khusus untuk membina calon pemimpin masa depan</p>
<p><i>Person or job? Change in person-job fit and its impact on employee work attitudes over time.</i></p>	<p>Uji Hipotesis</p>	<p>Penelitian ini meneliti hubungan yang dinamis antar <i>Person-Job Fit</i>, <i>demand-ability</i>, <i>needs-supply fit</i>, dan <i>employee work attitude</i>. Hasil penelitian menunjukkan bahwa perubahan pada <i>PJ fit</i> terkait secara signifikan dengan <i>job satisfaction</i></p>

Judul	Metode	Hasil
(Kim, Schuh, & Cal, 2019)		dan <i>effective organizational commitment</i> . Dari penelitian ini juga dapat diketahui bahwa <i>potential assessment</i> yang mengevaluasi faktor-faktor ini dapat digunakan untuk menilai kesesuaian orang-pekerjaan.
<i>Employee Perceptions on the Implementation of Talent Management at PT Pegadaian Office Region VIII Jakarta</i> (Ardhani & Indriati, 2023)	Kualitatif	Penelitian ini dilakukan untuk mendeskripsikan persepsi pegawai terhadap penerapan talent management di kantor wilayah. Hasil penelitian menunjukkan bahwa karyawan secara umum mempunyai persepsi positif terhadap penerapan manajemen talenta. Dalam jurnal ini dinyatakan juga bahwa organisasi yang mengidentifikasi potensi dan kemampuan yang luar biasa untuk mencapai tujuan organisasi. Talenta yang dimiliki oleh pegawai dapat dinilai atau diidentifikasi dan dikembangkan oleh organisasi.
<i>The Relationship Of Employee Potential, Employee Competence And Work Motivation With Performance Of Employees In The Public Works And People's Housing Department Jambi Province</i> (Mona, Syukri, & Risnita, 2022)	Uji Hipotesis	Penelitian ini dilakukan untuk menguji hipotesis hubungan antara potensi, kompetensi karyawan dan motivasi kerja terhadap kinerja karyawan. Hasil menunjukkan adanya hubungan yang positif dan signifikan antara potensi dan kinerja karyawan. Semakin baik potensi yang dimiliki oleh karyawan, maka kinerjanya juga akan semakin baik
<i>Person–Job Fit And Job Involvement: The Curvilinear Effect And The Moderating Role Of Goal Orientation</i> (Ju, Yao, & Ma, 2021)	Uji Hipotesis	Penelitian ini dilakukan menggunakan <i>multivariate regression</i> untuk untuk menguji <i>curvilinear model</i> . Penelitian ini juga menemukan Studi ini menemukan hubungan berbentuk U terbalik antara P–J <i>fit</i> dan <i>job involvement</i> . Dalam penelitian ini juga dapat diketahui bahwa penilaian potensial yang mempertimbangkan orientasi tujuan individu juga dapat berguna dalam mengevaluasi P-J <i>fit</i> .
<i>Identifying and assessing talent potential for future needs of a company</i> (Kabalina & Osipova, 2022)	Kualitatif Explanatory Research	Penelitian ini dilakukan dengan melakukan <i>deep interview</i> terhadap <i>experts</i> dari 45 perusahaan. Diketahui bahwa – potensi berpengaruh secara signifikan terhadap kinerja. Namun, besarnya pengaruh bergantung pada banyak faktor termasuk diantaranya metode yang digunakan untuk menilai potensi dan strategi perusahaan. Penelitian ini juga menyarankan dilakukannya mengidentifikasi dan menilai

Judul	Metode	Hasil
<i>The Relationship between Work Motivation, Job Satisfaction, and Employee Performance: The Moderating Role of Psychology Capital and The Mediating Role of Organizational Commitment</i> (Astuti W. & Amalia, 2021)	Kuantitatif – Structural Equation Modeling	potensi karyawan dengan kinerja yang baik untuk mencapai keberhasilan kerja di masa depan. Penelitian ini dilakukan untuk mengetahui apakah variabel motivasi kerja, komitmen organisasi, dan kepuasan kerja memiliki pengaruh terhadap kinerja pegawai. Hasil penelitian menunjukkan bahwa motivasi kerja, komitmen organisasi, dan kepuasan kerja berpengaruh positif dan signifikan terhadap kinerja pegawai. Dari penelitian ini juga dapat diketahui bahwa ketika penilaian potensi dilakukan dengan efektif dapat dengan positif memengaruhi kinerja karyawan.
<i>Linking Person Job Fit, Person Organization Fit and Organizational Culture to Employee Performance in Islamic Banks: the Mediating Role of Job Motivation</i> (Wenehenubun, 2023)	Uji Hipotesis	Penelitian ini dilakukan untuk mengetahui pengaruh <i>person-job fit</i> , <i>person-organization fit</i> , dan <i>organizational culture</i> terhadap <i>employee performance</i> . Data pada penelitian ini sebanyak 65 yang diperoleh dari hasil kuesioner. Hasil penelitian ini yaitu diketahui bahwa <i>person-job fit</i> memiliki pengaruh yang positif terhadap <i>employee performance</i> dengan nilai <i>coefficient</i> sebesar 0.278.
<i>The Effect of Job Characteristic, Person-Job Fit, Organizational Commitment on Employee Performance (Study of East Java BPJS Employees)</i> (Nurtjahjono, Nimran, Al Musadieg, & Utami, 2020)	Kuantitatif – Explanatory Research	Penelitian ini dilakukan untuk mengetahui pengaruh <i>Job Characteristic</i> , <i>Person-Job Fit</i> , dan <i>Organizational Commitment</i> terhadap <i>employee performance</i> . Responden pada penelitian ini adalah sebesar 238 orang yang diperoleh melalui kuesioner. Hasil penelitian ini menunjukkan bahwa semakin baik <i>person-job fit</i> , semakin baik pula <i>performance</i> dari karyawan. Dengan nilai <i>coefficient</i> sebesar 0.12, dapat dikatakan pula bahwa hubungan tersebut positif dan signifikan.

Pada penelitian *person-job fit* terdahulu, pendekatan yang umum dilakukan adalah dengan melakukan *text matching* dengan melakukan kalkulasi derajat kemiripan dari resume kandidat dan *job posting* atau *job requirement* yang ada (Qin, et al., 2020; Wang, Wei, Xu, Xu, & Mao, 2022). Selain itu, penelitian terkait *person-job fit* juga umum dilakukan menggunakan uji hipotesis dan kuesioner (Kaur & Kaur, 2023; Rajper, Ghumro, & Mangi, 2020). Oleh karena itu, penelitian ini menggunakan pendekatan yang berbeda dengan memanfaatkan fungsi klasifikasi dari *machine learning* dalam *person-job fit* dengan menggunakan prediktor berupa nilai potensi karyawan. Hal ini sesuai dengan penelitian yang menyatakan bahwa penilaian potensial yang mempertimbangkan

orientasi tujuan individu juga dapat berguna dalam mengevaluasi P-J *fit* (Kim, Schuh, & Cal, 2019; Ju, Yao, & Ma, 2021). Adapun kelas data yang akan digunakan pada penelitian ini berasal dari nilai *job performance* karyawan yang akan dibagi menjadi *highly* dan *moderate fit*. Penggunaan *job performance* sebagai kelas data didasarkan pada penelitian yang mengatakan bahwa adanya pengaruh signifikan *person-job fit* terhadap *job performance* (Rajper, Ghumro, & Mangi, 2020; Kaur & Kaur, 2023; Nurtjahjono, Nimran, Al Musadieq, & Utami, 2020; Wulandari, 2021), sehingga dapat dikatakan bahwa kecocokan dengan pekerjaan dapat diukur dari *job performance* karyawan. *Potential assessment* juga dinilai penting, dalam beberapa penelitian terdahulu juga mendukung terkait pentingnya *potential assessment* dan menyatakan bahwa potensi memiliki pengaruh terhadap performansi kerja (Kabalina & Osipova, 2022; Mona, Syukri, & Risnita, 2022). Mengidentifikasi karyawan yang berpotensi tinggi di awal karir merupakan salah satu kepentingan strategis dalam manajemen sumber daya manusia, selain itu perlu juga menjadikan potensi sebagai indikator untuk membantu perusahaan dalam mencari dan mengenali calon pemimpin perusahaan di masa depan (Kabalina & Osipova, 2022; Ye, et al., 2020).

2.2 Research Gap

Perbedaan penelitian ini dengan penelitian terdahulu dapat dilihat pada Tabel 2.1.1 Kajian Literatur.

Tabel 2.2.1 *Research Gap*

Penelitian	<i>Person-Job Fit</i>	<i>Talent Management</i>	Industri Pertambangan	Klasifikasi	<i>Decision Tree</i>	<i>Random Forest</i>
<i>An Enhanced Neural Network Approach to Person Job Fit in Talent Recruitment</i> (Qin, et al., 2020)	✓	✓		✓		
<i>Person-Job Fit Estimation from Candidate Profile and Related Recruitment History With Co-Attention Neural Networks</i>	✓	✓		✓		

Penelitian	Person-Job Fit	Talent Management	Industri Pertambangan	Klasifikasi	Decision Tree	Random Forest
(Wang, Wei, Xu, Xu, & Mao, 2022) <i>Longitudinal Effects Of High-Performance Work Practices on Job Performance Via Person-Job Fit</i>	✓	✓				
(Kaur & Kaur, 2023) <i>The Impact Of Person Job Fit And Person Organization Fit On Employee Job Performance: A Study Among Employees Of Services Sector</i>	✓	✓				
(Rajper, Ghumro, & Mangi, 2020) <i>Classification Talent of Employee Using C4.5, KNN, SVM</i>	✓	✓		✓		
(Stephanie & Sarno, 2019) <i>Identifying High Potential Talent: A Neural Network Based Dynamic Social Profiling Approach</i>	✓	✓				
(Ye, et al., 2020) <i>Person or job? Change in person-job fit and its impact on employee work attitudes over time.</i>	✓	✓				
(Kim, Schuh, & Cal, 2019) <i>Employee Perceptions on</i>		✓				

Penelitian	Person-Job Fit	Talent Management	Industri Pertambangan	Klasifikasi	Decision Tree	Random Forest
<i>the Implementation of Talent Management at PT Pegadaian Office Region VIII Jakarta (Ardhani & Indriati, 2023)</i>						
<i>The Relationship Of Employee Potential, Employee Competence And Work Motivation With Performance Of Employees In The Public Works And People's Housing Department Jambi Province (Mona, Syukri, & Risnita, 2022)</i>		✓				
<i>Person–Job Fit And Job Involvement: The Curvilinear Effect And The Moderating Role Of Goal Orientation (Ju, Yao, & Ma, 2021)</i>	✓	✓				
<i>Identifying and assessing talent potential for future needs of a company (Kabalina & Osipova, 2022)</i>		✓				
<i>The Relationship between Work Motivation, Job Satisfaction, and Employee Performance: The</i>		✓				

Penelitian	Person-Job Fit	Talent Management	Industri Pertambangan	Klasifikasi	Decision Tree	Random Forest
<i>Moderating Role of Psychology Capital and The Mediating Role of Organizational Commitment (Astuti W. & Amalia, 2021)</i>						
<i>Linking Person Job Fit, Person Organization Fit and Organizational Culture to Employee Performance in Islamic Banks: the Mediating Role of Job Motivation (Wenehenubun, 2023)</i>	✓	✓				
<i>The Effect of Job Characteristic, Person-Job Fit, Organizational Commitment on Employee Performance (Study of East Java BPJS Employees) (Nurtjahjono, Nimran, Al Musadieg, & Utami, 2020)</i>	✓	✓				
<i>Analisis Peran Person Job-Fit Terhadap Human Resources Development Practices Dalam Meningkatkan Employee Job Performance (Wenehenubun, 2023)</i>	✓	✓	✓			
Usulan	✓	✓	✓	✓	✓	✓

Celah dan keterbaruan pada penelitian ini yang membedakan dengan penelitian terdahulu yaitu penelitian ini berfokus pada *talent management* khususnya terkait klasifikasi *person-job fit* karyawan industri pertambangan. Seperti yang dipaparkan pada Tabel 2.2.1, penelitian terdahulu yang membahas *person-job fit* pada industri pertambangan masih sangat kurang. Meskipun masih terdapat penelitian terdahulu yang membahas *person-job fit*, namun penelitian tersebut umumnya berfokus pada kecocokan antara resume dengan *job requirement* yang ada (*text matching*) atau hanya sebatas penelitian kualitatif dengan uji hipotesis. Berdasarkan hal tersebut, berbeda dengan penelitian *person-job fit* sebelumnya, penelitian ini akan membangun model klasifikasi *person-job fit* dengan menggunakan data hasil tes psikologi yang berisi atribut potensi dan kecerdasan sebagai prediktor atau atribut klasifikasi dan nilai performa karyawan sebagai label atau target dalam model klasifikasi yang akan dibagi menjadi dua kelas yaitu yaitu *highly* dan *moderate fit*.

2.3 Landasan Teori

2.3.1 War for Talents

Sekelompok konsultan McKinsey pada tahun 1998 memperkenalkan istilah “*war for talent*” dan menyatakan bahwa talenta adalah kunci keunggulan organisasi (Michaels dkk, 2001 dalam (Montero Guerra, Danvila-del-Valle, & Méndez-Suárez, 2023)). *War for talents* mencerminkan persaingan antar perusahaan untuk menarik dan mempertahankan “*the most ables employees*” atau karyawan terbaik. Seperti yang diungkapkan oleh Biscoe (Montero Guerra, Danvila-del-Valle, & Méndez-Suárez, 2023), saat ini jumlah orang yang siap dan aktif bekerja lebih meningkat, lebih bervariasi, lebih mudah dalam berganti pekerjaan dan bekerja jarak jauh, dan memiliki keterampilan lebih tinggi. Dapat juga disimpulkan bahwa saat ini tenaga kerja tidak hanya bertambah dari segi jumlah, namun juga dari sisi keberagaman, fleksibilitas kerja dan juga kompetensi. Selain itu, tingginya permintaan dari perusahaan menyebabkan tenaga kerja menjadi lebih selektif dalam memilih pekerjaan dan memiliki ekspektasi tinggi terhadap perusahaan (Green, 2022). Ketika pekerja bersaing untuk mendapatkan pekerjaan yang diinginkan, perusahaan juga harus bersaing keras untuk menemukan talenta yang tepat (Anonim, 2015).

War for talents juga dirasakan oleh PT PNR, sebagai salah satu perusahaan jasa pertambangan terbesar di Indonesia. Dalam menjalankan bisnisnya, PT PNR selalu menerapkan

kaidah teknik pertambangan yang baik dan benar, terlebih karena karakteristik bidang pertambangan yang berisiko tinggi. Untuk mendorong penerapan kaidah tersebut secara lebih maksimal, tenaga kerja yang bermutu dan memenuhi kualifikasi akan sangat diperlukan (Kemnaker, 2019). Oleh karena itu untuk memenangkan *war for talents*, perusahaan harus mampu bersaing dalam menemukan talenta yang paling tepat dan sesuai dengan kebutuhan perusahaan (Anonim, 2015).

2.3.2 Rekrutmen

Rekrutmen merupakan serangkaian proses untuk mengidentifikasi dan memilih kandidat yang berkualitas sesuai dengan kebutuhan perusahaan (Zaky, 2021). Proses rekrutmen yang komprehensif melibatkan adanya analisis kebutuhan, *screening curriculum vitae* atau resume, dan wawancara kandidat (Tsai, Kao, & Kuo, 2023). Dengan prosesnya yang kompleks, tujuan utama dari rekrutmen yaitu menarik dan menemukan talenta yang potensial (Abbasi, Tahir, Abbas, & Shabbir, 2020). Namun, untuk mengembangkan semua kandidat jika hanya dilihat dari potensi yang dimiliki tentu akan mahal. Melatih dan mengembangkan kandidat memakan waktu yang tergolong panjang dan membutuhkan prediksi yang akurat terkait ketrampilan yang diperlukan. Menurut Buhler (Anonim, 2015, p. 670), untuk mengatasi hal tersebut terdapat sebuah metode pada *Human Capital* yaitu *make and buy*. Dalam hal ini, “*make*” memiliki arti untuk mempekerjakan calon tenaga kerja yang belum memiliki semua ketrampilan yang diperlukan untuk mengisi posisi di perusahaan, dan dapat dilatih (“*made*”) untuk mengembangkan mereka. Kunci kesuksesan metode ini berada pada bagaimana perusahaan mampu untuk membedakan tenaga kerja *high-potential* (belum memiliki ketrampilan tapi dapat dilatih), dan *mediocre employee* (tidak memiliki ketrampilan). Metode ini sudah dilakukan oleh PT PNR dengan merekrut *fresh graduate* dan tenaga kerja yang memiliki pengalaman kurang dari tiga tahun untuk mengikuti program pengembangan yang ada. Tenaga kerja yang dapat mengikuti program pengembangan ini tentunya sudah lolos pada tahap seleksi administrasi, tes psikologi serta wawancara. Selanjutnya metode “*buy*” berarti mempekerjakan tenaga kerja yang sudah memiliki semua ketrampilan dan pengalaman untuk mengisi posisi tersebut sejak hari pertama. Pada PT PNR, penerapan metode ini dilakukan untuk lowongan pada level *expert* dengan minimum pengalaman tenaga kerja adalah 3 tahun bekerja.

2.3.3 Potensi Karyawan

Potensi merujuk pada kemampuan untuk melakukan kinerja secara maksimal di masa yang akan mendatang (Robledo-Ardila & Roman-Calderon, 2020). Potensi seorang tenaga kerja dapat dilihat melalui hasil *psychological assessment* atau sering pula disebut sebagai tes psikologi (Rebele, 2019), yang berarti seseorang dapat dideskripsikan dan dibedakan berdasarkan serangkaian karakteristik psikologis (Sartori, Costantini, & Ceschi, 2020). Adapun manfaat dari penilaian potensi yaitu untuk mengelola kinerja di masa depan (Fernandes França, São Mamede, Manuel Pereira Barroso, & Manuel Pereira Duarte dos Santos, 2023). Menurut Fernández-Aráoz dkk (2011), konteks potensi dalam sebuah organisasi mengacu pada kemampuan individu untuk maju dan sukses dalam posisi yang lebih tinggi, atau dengan kata lain untuk berkembang dan menangani lebih banyak tanggung jawab yang lebih luas. Alziari (Fernandes França, São Mamede, Manuel Pereira Barroso, & Manuel Pereira Duarte dos Santos, 2023) juga mengatakan, bahwa dengan melakukan *potential assessment*, perusahaan mendapatkan pengetahuan penting untuk selanjutnya dapat digunakan sebagai tujuan yang realistis berdasarkan kemampuan tim, mencegah ketidakpuasan dalam bekerja dan kinerja buruk.

Dalam penelitiannya, Künneke (Fernandes França, São Mamede, Manuel Pereira Barroso, & Manuel Pereira Duarte dos Santos, 2023) menegaskan pentingnya membedakan tenaga kerja *high potential* dan *high performer*. Tenaga kerja *high potential* dapat menangani tugas yang membingungkan dan menantang, sehingga membuat mereka menjadi tenaga kerja yang baik untuk menduduki posisi senior dan sebagai pemimpin dalam perusahaan. Selain itu, *high potential* mampu beradaptasi dengan lebih cepat terhadap lingkungan dan aktivitas baru, bahkan beberapa di antara mereka juga unggul di bidang selain pekerjaan mereka saat ini. Sedangkan tenaga kerja *high performer* memiliki pengetahuan yang mendalam terkait keahlian mereka dan lebih unggul dari rekan-rekan mereka. Kandidat yang tergolong *high potential* juga cenderung menunjukkan perkembangan kemampuannya yang lebih cepat dan lebih sukses dibandingkan dengan rekan kerjanya. Mereka juga akan lebih menunjukkan kinerja yang sangat baik dalam mengerjakan tugas yang diberikan, sering menyelesaikan tugas lebih awal, konsisten dan *detail-oriented* dalam melakukan pekerjaan dan memiliki kemampuan yang menonjol dalam hal kepemimpinan (Beaumont-Oates, 2023). Oleh karena itu, penting untuk mempertimbangkan potensi seorang karyawan untuk mengetahui apa yang bisa seorang karyawan capai di masa depan. Tentu hal tersebut harus didukung dengan terbukanya peluang

untuk terus berkembang bagi karyawan. Hal tersebut sesuai dengan apa yang disampaikan oleh *Assessment Associates International* (Fernandes França, São Mamede, Manuel Pereira Barroso, & Manuel Pereira Duarte dos Santos, 2023) bahwa untuk menentukan karyawan berbakat, dibutuhkan dua kriteria utama yaitu performa (efisiensi individu dan hasil akhir dalam peran mereka saat ini) dan potensi (kinerja individu yang diantisipasi di masa depan jika mereka memiliki peluang untuk berkembang dan tanggung jawab yang lebih besar).

2.3.4 *Job Performance*

Job performance atau dalam dunia kerja sering juga disebut sebagai *employee performance* (Iqbal, Khan, Mohmand, & Mujtaba, 2020) menurut Robbins (Nurtjahjono, Nimran, Al Musadieg, & Utami, 2020) adalah hasil kerja secara kualitas dan kuantitas yang dicapai oleh karyawan dalam menjalankan fungsi dan perannya sesuai dengan tanggung jawab yang diberikan. Adapun komponen krusial yang menjadi dasar *job performance* menurut Conway (Sørlie, Hetland, Bakker, Espevik, & Olsen, 2022) dibagi menjadi dua yaitu *task performance* dan *contextual performance*. Secara lebih lanjut Sørlie (2022) dalam penelitiannya menerangkan *task performance* sebagai kemahiran karyawan dalam melaksanakan tugas aktual yang sesuai *job description* atau tugas lainnya yang memberikan kontribusi baik secara langsung maupun tidak langsung terhadap *technical core* perusahaan. Sedangkan *contextual performance* mengacu pada perilaku kerja vital bagi dalam jalannya fungsi organisasi secara efisien, dan biasanya berada di luar *job description*. Dari hal tersebut dapat disimpulkan pula bahwa *job performance* merupakan sebuah konsep perilaku dan tindakan seorang karyawan yang berkontribusi untuk mencapai tujuan baik individu maupun tujuan perusahaan (Iqbal, Khan, Mohmand, & Mujtaba, 2020). Sehingga dapat dikatakan bahwa *job performance* adalah salah satu penentu keberhasilan suatu perusahaan.

Dalam industri pertambangan, talenta dan performa karyawan dinilai sebagai penggerak *core value* perusahaan. Misalnya, seorang *mine planner* yang bekerja secara efektif (seorang profesional yang merancang dan mengoordinasikan penggalian atau letak lokasi tambang) dapat dengan cepat memberikan dampak yang signifikan terhadap *value* perusahaan. Pada saat yang bersamaan, ketidakpastian energi global serta kenaikan harga dan permintaan komoditas (seperti litium, bijih besi, dan tembaga), serta meningkatnya ekspektasi dalam hal lingkungan, dan sosial (ESG), semakin meningkatkan tekanan bagi karyawan pertambangan untuk bekerja dengan performa yang lebih tinggi (Abenov, Grabbert, Franklin-Hensler, & Larrat, 2023). Oleh

karena itu, perlu diterapkannya mekanisme untuk memastikan ketercapaian performa kerja yang optimal dari karyawan. Khususnya dalam bidang pertambangan, untuk meningkatkan efektivitas operasional dapat dilakukan dengan investasi yang relatif kecil melalui pengoptimalan keterampilan, pengetahuan, dan efisiensi karyawan dan dalam interval waktu yang lebih singkat dibandingkan harus melakukan upaya perubahan atribut fisik suatu aset (Abenov, Grabbert, Franklin-Hensler, & Larrat, 2023).

2.3.5 *Person-Job Fit*

Topik kecocokan dalam rekrutmen dan seleksi kandidat sudah sejak lama menjadi landasan dari manajemen sumber daya manusia dan psikologi industri/organisasi menurut Saks & Ashford, 1997 (Kristof-Brown & Guay, 2011). Bentuk kecocokan tersebut adalah *person-job fit* (PJF) yang merupakan salah satu dimensi dari teori *person-environment fit* (PEF) (Edwards, 2008; Chi, 2014; Chuang, Shen, & Judge, 2015; Goetz & Wals, 2022). Menurut (Goetz & Wals, 2022), secara umum PEF adalah kecocokan antara seseorang dengan lingkungannya yang akan memberi efek positif bagi individu (contohnya *job satisfaction*) dan lingkungan (contohnya *job performance*). PEF memiliki banyak dimensi lainnya dan tiga diantaranya sering dipelajari dalam literatur yaitu *person-job fit*, *person-environment fit*, dan *person-organization fit*. Pada penelitian ini akan berfokus pada *person-job fit* yang didefinisikan sebagai kecocokan karakteristik seseorang dengan pekerjaan atau tugas yang dilakukan (Xiao, et al., 2021). Hal mendasar dibalik kecocokan tersebut adalah dengan berdasarkan pada sifat karakteristiknya, beberapa orang akan cenderung lebih cocok untuk posisi tertentu dibandingkan dengan posisi atau orang lainnya (De Cooman & Vleugels, 2022). Jika dibandingkan dengan tipe kecocokan lainnya, penelitian terdahulu menyatakan bahwa *person-job fit* merupakan prediktor terbaik terhadap hasil akhir yang berkaitan dengan pekerjaan dan karir seperti *job satisfaction*, *work commitment* dan *job performance* (Flatau-Harrison, Wilson, & Vleugels, 2023). Oleh karena itu Wenehenubun (2023) merekomendasikan penerapan *person-job fit* sebagai dasar dalam seleksi dan pemilihan serta pengembangan dan pelatihan karyawan di bidang pertambangan.

Person-job fit terdiri atas dua tipe kecocokan (Xiao, et al., 2021), pertama yaitu *demand-abilities fit* yang berfokus kepada kecocokan antara kapabilitas individu dengan pekerjaannya. Kapabilitas dalam hal ini menggambarkan kombinasi individu dan teknis antara keahlian, pengetahuan dan perilaku yang dapat membantu individu dalam bekerja secara lebih efektif. Atau dengan kata lain terkait potensi untuk menerapkan keahlian dan pengetahuannya di

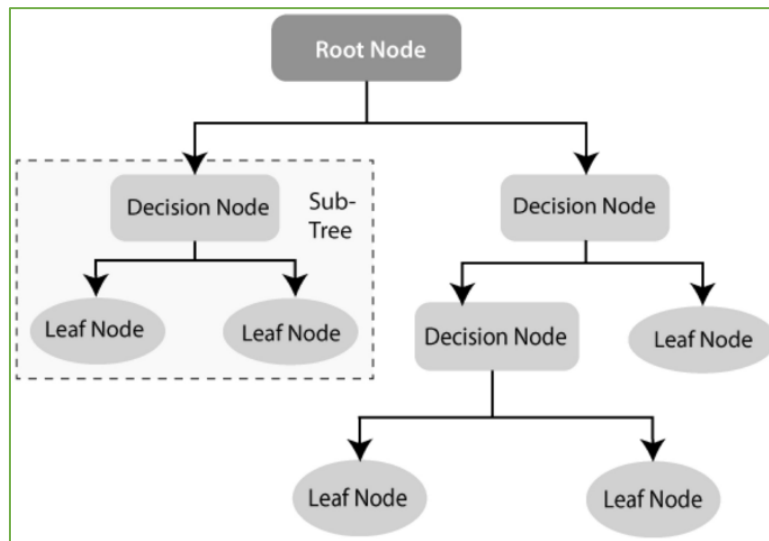
berbagai situasi (Rusconi, 2023). Menurut Kim dkk (2019), *demand-abilities fit* bersifat dinamis, terlebih karena elemen dalam pekerjaan seperti keahlian maupun tugas selalu berubah seiring berjalannya waktu. Melalui pelatihan serta pembelajaran berkelanjutan dan dengan meningkatnya kompleksitas tuntutan dalam pekerjaan, karyawan dapat lebih mudah untuk mengembangkan potensi yang mereka miliki dan dapat mengekspresikan diri melalui pekerjaan mereka yang mana akan meningkatkan *reward* untuk memenuhi kebutuhan mereka (Guo & Hou, 2022). Pemenuhan *reward* termasuk kedalam tipe kecocokan yang kedua yaitu (*needs - supplies fit*) yang merupakan kecocokan antara kebutuhan individu dengan hal yang terkait pekerjaannya seperti vakasi, sumber daya yang diberikan dan hubungan interpersonal (Xiao, et al., 2021).

2.3.6 Algoritma Klasifikasi

Algoritma klasifikasi adalah salah satu teknik dalam *machine learning* dan *data mining* untuk mengategorikan atau mengklasifikasikan data berdasarkan fitur atau atribut. Algoritma ini dapat digunakan pada beberapa data set termasuk, gambar, teks, angka, untuk mengidentifikasi informasi atau pola tersembunyi dan melakukan prediksi (Sarker, 2021). Terdapat banyak algoritma klasifikasi seperti *naïve bayes*, *K-neural network* dan lain sebagainya, namun metode yang akan digunakan pada penelitian ini adalah sebagai berikut:

a) *Decision Tree*

Decision tree dan metode *ensemble*-nya (*forest*) merupakan salah satu algoritma yang paling banyak dipelajari dan digunakan dalam *machine learning* dan *data mining* (Blockeel, Devos, Frénay, Nanfack, & Nijssen, 2023). *Decision tree* merupakan salah satu metode *supervised learning* non-parametrik (Sarker, 2021) yang umum digunakan dalam membangun model klasifikasi yang memiliki kemiripan dengan penalaran manusia dan mudah dimengerti. *Decision tree* digambarkan dalam bentuk pohon di mana setiap node mewakili fitur atau atribut, setiap tepi mewakili nilai yang mungkin dari atribut tersebut, daun yang berisi target label kelas dan jalur atau bentuk dari akar (*root node*) hingga ke daun (*leaf node*) menggambarkan proses prediksi (Vanneschi & Silva, 2023).



Gambar 2.3.1 Struktur *Decision Tree*

Sumber: Sarker, 2021

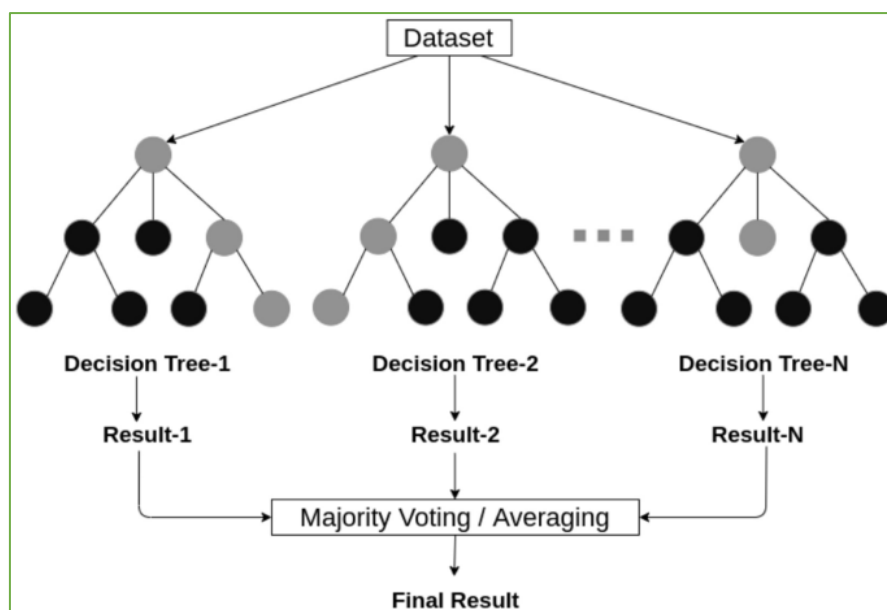
b) *Random Forest*

Random Forest merupakan suatu teknik *ensemble learning* yang menggunakan kombinasi dari beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat. Pendekatan ini didasarkan pada konsep *bagging*, di mana beberapa model individual dilatih menggunakan subset acak dari data yang tersedia. Dengan menggabungkan prediksi dari model-model tersebut, *random forest* dapat meningkatkan akurasi dan mengurangi risiko *overfitting* dibandingkan dengan penggunaan *decision tree* tunggal (Solorio-Ramírez, Jiménez-Cruz, Villuendas-Rey, & Yáñez-Márquez, 2023).

Algoritma *random forest* mengaplikasikan teknik *ensemble learning* dengan melatih sekelompok pohon keputusan pada subset acak dari data yang tersedia. Setiap pohon keputusan selanjutnya dilatih menggunakan subset acak dari fitur-fitur yang ada. Proses pengenalan unsur keacakan ini memiliki peran yang signifikan dalam mengurangi *overfitting* dan meningkatkan kemampuan generalisasi model. Prediksi akhir dihasilkan melalui agregasi prediksi individual dari seluruh pohon keputusan yang tergabung dalam hutan, sehingga menghasilkan estimasi yang lebih akurat dibandingkan dengan penggunaan pohon keputusan tunggal (Solorio-Ramírez, Jiménez-Cruz, Villuendas-Rey, & Yáñez-Márquez, 2023).

Random forest menawarkan beberapa keunggulan dibandingkan metode lain. Pertama, teknik ini secara inheren membatasi *overfitting* melalui agregasi prediksi dari beberapa pohon keputusan, yang masing-masing dilatih pada subset data yang berbeda. Kedua, *random forest*

memiliki kemampuan untuk mendeteksi efek non-linier dan interaksi antara variabel prediktor, yang merupakan tantangan bagi model linier tradisional (Fife & D’Onofrio, 2023). Ketiga, bekerja dengan baik bahkan untuk data dengan jumlah sampel kecil dan mampu menangani data berdimensi tinggi dengan banyak fitur, menjadikannya pilihan yang populer untuk aplikasi seperti klasifikasi gambar, klasifikasi teks, dan bio-informatika. Keempat, algoritma ini tahan terhadap nilai yang hilang dan *outlier* dalam data, sehingga dapat diterapkan secara efektif pada *dataset* yang memiliki karakteristik tersebut (Solorio-Ramírez, Jiménez-Cruz, Villuendas-Rey, & Yáñez-Márquez, 2023).



Gambar 2.3.2 Struktur *Random Forest*

Sumber: Sarker, 2021

2.3.7 *Hyperparameter Optimization*

Hyperparameter Optimization merupakan langkah penting dalam *machine learning* untuk memperoleh performa model yang optimal dengan pencarian sistematis untuk kombinasi *hyperparameter* yang juga dikenal sebagai *hyper-parameter tuning* (Khan, Kanwal, Alamri, & Mumtaz, 2020). Setiap metode klasifikasi tentu memiliki fitur yang berbeda-beda yang perlu dilakukan optimasi (Tantithamthavorn, McIntosh, Hassan, & Matsumoto, 2019). Selain itu, menjaga agar proses pembelajaran algoritma klasifikasi tetap terkendali merupakan tantangan yang kompleks. Hal ini dikarenakan *hyperparameter* berasal dari formulasi dasar model

pembelajaran mesin. Akibatnya, performa model klasifikasi sangat bergantung pada penyesuaian *hyperparameter* yang tepat (Shekar & Dagneu, 2019).

Dengan asumsi bahwa kombinasi hyperparameter terbaik akan menghasilkan performa terbaik untuk model klasifikasi tertentu, maka optimasi *hyperparameter* menjadi langkah krusial dalam mencapai performa model yang optimal. Oleh karena itu, pada penelitian ini dilakukan *hyperparameter optimization* pada metode yang digunakan.

1. *Decision Tree*

Parameter yang akan digunakan dalam membangun model klasifikasi pada algoritma *decision tree* menggunakan *rapid miner* dapat dilihat pada Tabel 2.3.1.

Tabel 2.3.1 Parameter Set *Decision Tree*

Parameter	Keterangan	Nilai
<i>criterion</i>	Untuk memilih kriteria yang akan digunakan dalam membagi data	<i>gain_ratio</i> , <i>information_gain</i> , <i>gini_index</i>
<i>max_depth</i>	Parameter ini digunakan untuk membatasi kedalaman pohon keputusan	5, 6, 7, 8, 9, 10, 11, 12, ..., 20
<i>confidence</i>	Parameter ini digunakan untuk menentukan tingkat kepercayaan yang digunakan dalam tahap pemangkasan	0.1, 0.2, 0.3, 0.4, 0.5, ..., 0.95
<i>min_gain</i>	Untuk mencari nilai <i>gain</i> atau keuntungan dari setiap <i>node</i> akan dihitung sebelum melakukan pemangkasan	0.0, 0.1, 0.2, 0.3, 0.4, 0.5
<i>min_leaf_size</i>	Untuk menentukan jumlah daun dalam data.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

2. *Random Forest*





Parameter yang digunakan dalam membangun model klasifikasi pada algoritma *Random Forest* dapat dilihat pada Tabel 2.3.2.

Tabel 2.3.2 Parameter Set *Random Forest*

Parameter	Keterangan	Nilai
<i>num_of_trees</i>	Parameter ini digunakan untuk menentukan banyaknya jumlah pohon keputusan yang akan dibentuk	1, 2, 3, 4, 5, 6, 7, 8, 9, ..., 400
<i>criterion</i>	Untuk memilih kriteria yang akan digunakan dalam membagi data	<i>gain_ratio</i> , <i>information_gain</i> , <i>gini_index</i>
<i>max_depth</i>	Parameter ini digunakan untuk membatasi kedalaman pohon keputusan	5, 6, 7, 8, 9, 10, 11, 12, ..., 20
<i>confidence</i>	Parameter ini digunakan untuk menentukan tingkat kepercayaan yang digunakan dalam tahap pemangkasan	0.1, 0.2, 0.3, 0.4, 0.5, ..., 0.95
<i>min_gain</i>	Untuk mencari nilai <i>gain</i> atau keuntungan dari setiap <i>node</i> akan dihitung sebelum melakukan pemangkasan	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, ..., 1
<i>min_leaf_size</i>	Untuk menentukan jumlah daun dalam data.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
<i>subset_ratio</i>	Parameter ini digunakan untuk menentukan ratio dari attribute yang diambil secara acak untuk diuji.	0.0, 0.1, 0.2, 0.3, 0.4, 0.5
<i>voting_strategy</i>	Untuk menentukan strategi prediksi jika terdapat perbedaan pendapat pada model pohon keputusan	<i>confidence_vote</i> , <i>majority_vote</i>

2.3.8 Evaluasi Model Klasifikasi

Dalam membangun model klasifikasi, diperlukan investigasi lebih lanjut terkait efektivitas dari model yang sudah dibangun. Evaluasi dari performa model *machine learning* dapat dilakukan menggunakan *evaluation matrix* atau matriks evaluasi. Dalam menilai performa model klasifikasi, terdapat empat konsep penting yang mana seluruh matriks evaluasi nantinya akan sangat bergantung. Empat konsep tersebut yaitu *true positive*, *true negative*, *false positive*, dan *false negative* (Ahad, Antar, & Ahmed, 2021). Adapun penjelasan terkait hal tersebut ditunjukkan pada Gambar 2.3.3.

<p style="text-align: center;">True Positive (T_p)</p> <p style="text-align: center;"></p> <p>True outcome: A thief arrives. Predicted outcome: Thief (alarm on). Comment: Good performing robot.</p>	<p style="text-align: center;">False Positive (F_p)</p> <p style="text-align: center;"></p> <p>True outcome: No thief arrives. Predicted outcome: Thief (alarm on). Comment: False alarm waking the owner.</p>
<p style="text-align: center;">False Negative (F_n)</p> <p style="text-align: center;"></p> <p>True outcome: A thief arrives. Predicted outcome: No thief (no alarm). Comment: Security breached. Bad performing robot.</p>	<p style="text-align: center;">True Negative (T_n)</p> <p style="text-align: center;"></p> <p>True outcome: No thief arrives. Predicted outcome: No thief (no alarm). Comment: System is fine.</p>

Gambar 2.3.3 Empat Konsep Utama Matriks Evaluasi

Sumber: Ahad, Antar, & Ahmed, 2021

Konsep pertama adalah *true positive* yang menunjukkan jumlah ketika model berhasil melakukan prediksi label kelas positif dengan benar. Selanjutnya adalah *true negatif* yang menunjukkan jumlah ketika model berhasil melakukan prediksi kelas dengan label negatif dengan benar. *False positive* menunjukkan jumlah hasil ketika model gagal untuk melakukan prediksi dengan benar pada label kelas positif yang juga dikenal sebagai *type-1 error*. Konsep keempat dan terakhir yaitu *false negative* yang menunjukkan jumlah ketika model gagal untuk memprediksi dengan benar pada kelas dengan label negatif dan konsep ini dikenal juga dengan *type-2 error* (Ahad, Antar, & Ahmed, 2021). Dengan mengacu pada keempat konsep tersebut, matriks evaluasi yang akan digunakan pada penelitian ini yaitu:

1.3.8.1. Akurasi

Akurasi diartikan sebagai nilai rasio dari jumlah prediksi yang berhasil atau benar oleh model dengan jumlah total prediksi yang dibuat. Dan berdasarkan dengan keempat konsep sebelumnya, dapat disimpulkan bahwa jumlah prediksi benar adalah hasil penjumlahan dari *true positive* dan *true negative*.

$$\begin{aligned}
 Accuracy &= \frac{\text{Jumlah prediksi benar}}{\text{Jumlah prediksi yang dibuat}} \\
 &= \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}
 \end{aligned}
 \tag{1.1}$$

1.3.8.2. *Precision*

Presisi merupakan nilai prediksi yang berlabel positif yang menggambarkan proporsi dari nilai prediksi berlabel positif yang positif.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1.2)$$

1.3.8.3. *Recall*

Matriks *recall* sering juga disebut sebagai sensitivitas atau *rate* dari *true positive*. *Recall* didefinisikan sebagai proporsi dari nilai yang positif aktual yang diprediksikan dengan label positif. Artinya nilai persentase total dari data yang berhasil diprediksi dengan benar oleh model.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1.2)$$

1.3.8.4. *F1 Score*

F1 score sering didefinisikan sebagai nilai rata-rata atau perpaduan antara metrik *precision* dan *recall*. Metrik ini mempertimbangkan *false positive* dan *false negative* secara bersamaan. Untuk data dengan distribusi kelas yang tidak sama besar atau *imbalanced data set*, metrik ini merupakan matriks yang tepat dan baik untuk digunakan.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1.2)$$

1.3.8.5. *Confusion Matrix*

Metrik ini merupakan salah satu metode evaluasi model klasifikasi yang umum digunakan. Metrik ini digambarkan dalam bentuk tabel yang berisi hasil prediksi dari model klasifikasi. Melalui metrik ini, dapat dilakukan analisis rasionalitas kesalahan akurasi pada setiap label kelas (Ahad, Antar, & Ahmed, 2021).

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

Gambar 2.3.4 *Confusion Matrix*

Sumber: Goyal, 2021

1.3.8.6. AUC (*Area Under Curve*)

Matriks ini menggambarkan tingkat atau derajat *separability* pada kelas. Jika diperoleh nilai AUC tinggi, artinya model akan sangat baik dalam memisahkan kelas dengan akurat. Dan jika AUC hanya sebesar 0.5, dapat dikatakan bahwa model hanya akan melakukan prediksi secara random karena kurangnya kemampuan dalam membedakan kelas dalam data (Ahad, Antar, & Ahmed, 2021).

BAB III

METODE PENELITIAN

3.1 Subjek Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari *database* internal PT PNR dan terfokus pada hasil tes psikologi dan performa karyawan Divisi PL PT PNR. Adapun kriteria data yang akan digunakan yaitu data hasil psikotes karyawan Divisi PL yang masuk pada tahun 2015-2019 atau memiliki masa kerja lebih dari 3 tahun.

3.2 Objek Penelitian

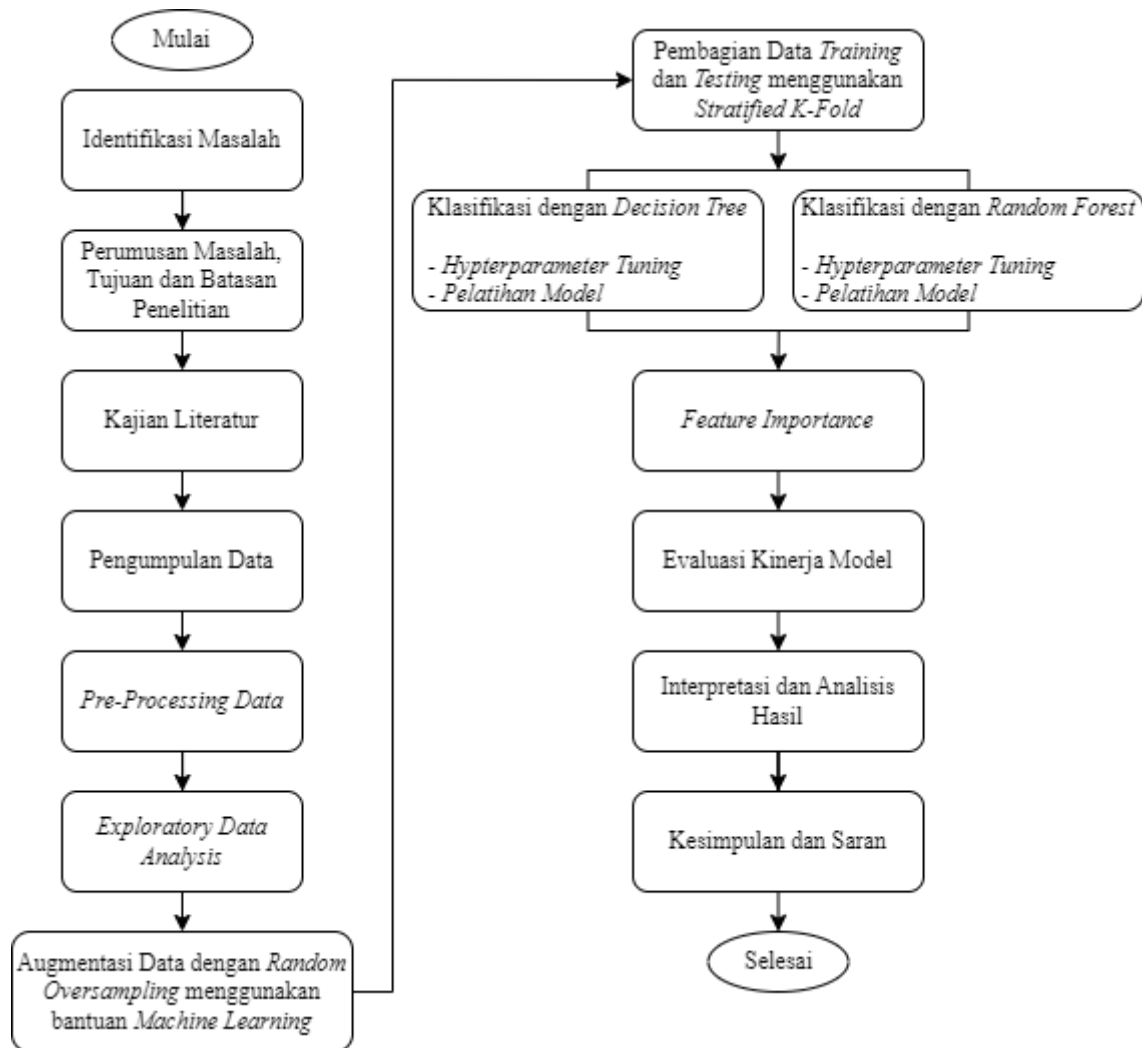
Penelitian ini terfokus pada pembangunan model klasifikasi dalam melakukan seleksi karyawan menggunakan hasil tes psikologi menggunakan *decision trees algorithm* yaitu *Decision Tree* dan *Random Forest*.

3.3 Lokasi dan Waktu Penelitian

Penelitian dilakukan di PT PNR, sebuah perusahaan jasa pertambangan yang terletak di Jakarta. Penelitian berlangsung selama 3 bulan pada tahun 2023 di bagian *section Recruitment and Selection*, divisi *Human Capital and Learning*.

3.4 Alur Penelitian

Tahapan dari penelitian yang dilakukan digambarkan menggunakan *flowchart* yang ditunjukkan pada Gambar 3.4.1.



Gambar 3.4.1 Alur Penelitian

1. Identifikasi Masalah.

Permasalahan yang digunakan pada penelitian ini diperoleh melalui observasi langsung dan diskusi dengan manajer dan pihak terkait PT PNR. Diketahui bahwa PT PNR belum melakukan pemanfaatan hasil tes psikologi karyawan secara optimal dan mengharapkan dilakukannya tindakan preventif dan antisipatif dalam tahap rekrutmen khususnya dalam hal kecocokan seseorang dengan pekerjaannya, dengan menggunakan algoritma pada *machine learning*.

2. Perumusan Masalah, Tujuan dan Batasan Penelitian.

Rumusan masalah yang digunakan pada penelitian ini didasarkan pada hasil observasi dan diskusi yang telah disusun sebelumnya dan selanjutnya akan digunakan dalam penentuan tujuan penelitian. Selain itu, dilakukan pula penentuan batasan dari masalah guna membantu dalam penentuan metode dan proses analisis.

3. Kajian Literatur.

Tahap selanjutnya yaitu mengkaji sumber-sumber penelitian terdahulu yang mendukung teori penelitian guna mencari *research gap* terkait *person-job fit* untuk selanjutnya dijadikan acuan dalam memperkuat teori yang digunakan.

4. Pengumpulan Data.

Guna menyelesaikan permasalahan yang sudah dirumuskan, dilakukan tahap pengumpulan data hasil tes psikologi dan performa karyawan dengan masa kerja di atas 3 tahun pada Divisi PL yang diperoleh dari *database* Divisi Human Capital & Learning Division, PT PNR. Data dikumpulkan dalam *file Ms. Excel* dengan total data sebanyak 18 baris dengan 32 atribut dan 1 target.

5. *Pre-Processing Data*.

Merupakan sebuah tahap awal dalam upaya untuk meningkatkan pemahaman terhadap data dan membuat data lebih siap serta sesuai untuk selanjutnya masuk ke tahap pembangunan model dan analisis. Adapun pada penelitian ini, tahap *pre-processing* terdiri dari *cleaning data* untuk memilah atribut yang akan digunakan, transformasi data untuk menyamakan tipe dari data yang memudahkan proses komputasi model, dan uji korelasi guna mengetahui tingkat hubungan antar variabel.

6. *Exploratory Data Analysis*.

Merupakan sebuah tahap untuk melakukan analisis secara statistik dari data yang akan digunakan pada penelitian.

7. Augmentasi Data dengan *Random Oversampling* menggunakan bantuan *Machine Learning*.

Tahap ini dilakukan karena kecilnya ukuran data yang diperoleh dengan hanya 18 data. Langkah ini bertujuan untuk membantu model dalam meningkatkan pemahaman terhadap data untuk membantu generalisasi model dalam melakukan klasifikasi dan meningkatkan akurasi model.

8. Pembagian Data *Training* dan *Data Testing* Menggunakan *Stratified K-Fold*.

Tahap pembagian data dilakukan dengan menggunakan salah satu metode pembagian data yaitu *stratified K-fold cross validation*. *Cross validation* melakukan pembagian data dengan 'K' lipatan yang sama besar.

9. Klasifikasi dengan *Decision Tree* dan *Random Forest*.

Pada tahap pembangunan model klasifikasi menggunakan bantuan algoritma *Decision Tree* dan *Random Forest* dengan bantuan aplikasi *RapidMiner*. Pada tahap

pembangunan model ini pula dilakukan proses parameter tuning untuk menemukan nilai terbaik pada masing-masing parameter-nya dalam membangun model untuk selanjutnya digunakan dalam tahap pelatihan model.

10. *Feature Importance*.

Tahap ini dilakukan untuk melihat tingkat kepentingan setiap atribut dalam membangun model prediksi.

11. Evaluasi Kinerja Model.

Tahap evaluasi dilakukan menggunakan *confusion matrix* untuk mengetahui kemampuan dan performa model dalam melakukan klasifikasi. Adapun metrik evaluasi yang digunakan pada penelitian ini antara lain: tingkat akurasi, AUC (*Area Under Curve*), *Precision*, *Recall*, *F1 Score*, *Class Precision* dan *Class Recall*, serta *Classification Error*.

12. Interpretasi dan Analisis Hasil.

Pada bagian ini, skor dari masing-masing metrik evaluasi akan diinterpretasi untuk memudahkan dan meningkatkan pemahaman atas performa model klasifikasi menggunakan *Decision Tree* dan *Random Forest* yang selanjutnya dilakukan analisis terhadap hasil yang diperoleh.

13. Kesimpulan dan Saran

Setelah dilakukan pengolahan dan analisis data, dibuat kesimpulan hasil penelitian yang menjawab rumusan masalah dan tujuan penelitian serta pemberian saran pengembangan.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Penelitian ini dilakukan dengan menggunakan data historis hasil tes psikologi dan nilai performa karyawan Divisi PL PT PNR dengan tahun masuk perusahaan antara 2015 hingga 2019. Adapun untuk pemilihan rentang waktu masa kerja karyawan mempertimbangkan adanya pandemi *Covid-19* yang menyebabkan tidak adanya rekrutmen pada Divisi PL tahun 2020-2022. Selain itu, pada PT PNR karyawan baru pada umumnya akan menjalani masa pelatihan dan pengembangan selama kurang lebih 1 tahun. Sehingga pemilihan masa kerja karyawan yang akan digunakan pada penelitian ini adalah karyawan yang memiliki masa kerja di atas 3 tahun karena karyawan dinilai cenderung lebih stabil dan lebih mengenal pekerjaan atau tugas yang dilakukannya. Data dikumpulkan dalam *file Ms. Excel* dengan total data sebanyak 18 baris dengan 32 atribut dan 1 label.

4.2 Augmentasi Data

Dalam pembangunan model, dikarenakan kecilnya ukuran data, dilakukan proses augmentasi data menggunakan bantuan *machine learning* yang merupakan sebuah proses untuk membuat data baru dari data yang sudah ada dan secara artifisial menambahkan jumlah data dengan tujuan untuk membantu model dalam mempelajari pola dalam data (Nanthini, et al., 2023). Dalam dalam kasus ini, hasil data data augmentasi akan disatukan dengan data yang ada dan di acak untuk selanjutnya dilakukan pembagian data menjadi data pelatihan dan data uji. Adapun hasil data augmentasi yang diperoleh adalah sebanyak 22 data, sehingga total data yang akan digunakan pada penelitian yaitu sebanyak 40 data.

4.3 Pre-Processing

Tahap *pre-processing* merupakan langkah awal dalam pengolahan dan analisis data untuk membantu memahami data dengan lebih baik, membuat data siap untuk analisis lebih lanjut dan interpretasi serta meningkatkan efisiensi dan efektivitas pengolahan data.

4.3.1 Informasi Statistik dan Identifikasi *Missing Value*

Tahap identifikasi *missing value* dilakukan untuk mengetahui *attributes* yang memiliki *instance* tidak lengkap atau hilang. *Missing value* pada data akan sangat mengganggu proses pengolahan dan analisis sehingga perlu adanya penghapusan atau perbaikan pada data dengan *missing value*. Adapun hasil identifikasi *missing value* pada data dalam penelitian ini dapat dilihat pada Tabel 4.3.1.

Tabel 4.3.1 memberikan informasi statistik umum dari data yang terdiri dari nama *feature* seperti *pj_fit*, *nama*, *pendidikan*, *asal_sekolah* hingga *gap*. Lalu tipe data pada masing-masing *attributes*, diketahui pada data ini diketahui terdapat empat jenis tipe data yaitu binomial, polinomial, *integer*, dan *real*. Selanjutnya informasi jumlah *missing value* yang dapat disimpulkan bahwa tidak ada *attributes* yang memiliki *missing value* dikarenakan semua *attributes* memiliki jumlah *missing value* sebesar 0. Terdapat pula informasi statistik seperti *value*, *instance* dengan jumlah terkecil (*least*) dan terbesar (*most*), nilai terkecil (*min*) dan terbesar (*max*), kelas positif dan negatif pada *label*, nilai rata-rata dan juga nilai standar deviasi. *Output* ini juga memberikan informasi terkait jumlah *attributes* yaitu sebanyak 33, lalu jumlah data pada keterangan *examples* sebanyak 40, *special attributes* yang dalam hal ini merupakan *label* sebanyak 1, dan informasi *regular attributes* sebanyak 32. Dari *output* tersebut juga dapat diketahui bahwa dalam data terdapat 5 atribut kategorikal dan sisanya merupakan atribut nominal.

Tabel 4.3.1 Informasi Statistik Data

Nama	Type	Missing	Statistics			
Label			<i>Negative</i>	<i>Positive</i>	<i>Value</i>	
<i>pj_fit</i>	<i>Binomial</i>	0	<i>Highly Fit</i>	<i>Moderately Fit</i>	<i>Highly Fit (24), Moderately Fit (16)</i>	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
nama	<i>Polynomial</i>	0	K40 (1)	K1 (1)	K1 (1), K2 (1), K3 (1), K4 (1), K5 (1), K6 (1), K7 (1), K8 (1), ..., K40 (1)	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
pendidikan	<i>Polynomial</i>	0	S1 Teknik Lingkungan (2)	S1 Teknik Geologi (11)	S1 Teknik Geologi (11), S1 Teknik Pertambangan (6), S1 Teknik Sipil (11), S1 Teknik Elektro (4), S1 Pend. Teknik Informatika (2), S1 Pend. Teknik Mesin (4), S1 Teknik Lingkungan (2)	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
asal_sekolah	<i>Polynomial</i>	0	Universitas Trisakti (1)	UPN Yogyakarta (7)	UPN Yogyakarta (7), ITB (3), UGM (4), Universitas Negeri Malang (6), ITP (2), Politeknik Negeri Padang (2), Sekolah Tinggi Teknologi Mineral (4), UNY (2), Univ. Bina Nusantara (5), Univ. Pendidikan Indonesia (UPI) (2), Universitas Riau (2), Universitas Trisakti (1)	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
jenis_kelamin	<i>Polynomial</i>	0	Laki-laki (40)	Laki-laki (40)	Laki-laki (40)	
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
usia	<i>Integer</i>	0	21	25	23.125	1.324
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>

Nama	Type	Missing	Statistics			
iq	<i>Integer</i>	0	88	145	119.250	14.723
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
pauli	<i>Integer</i>	0	2025	4581	3210.625	691.260
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
keterampilan_sosial	<i>Integer</i>	0	2	3	2.950	0.221
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kepekaan_sosial	<i>Integer</i>	0	2	3	2.725	0.452
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
komitmen_terhadap_tugas	<i>Integer</i>	0	3	3	3	0
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kerjasama	<i>Integer</i>	0	3	3	3	0
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kepatuhan_terhadap_aturan	<i>Integer</i>	0	2	3	2.850	0.362
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
ketelitian_kerja	<i>Integer</i>	0	3	4	3.050	0.221
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
hasrat_berprestasi	<i>Integer</i>	0	2	3	2.775	0.423
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
daya_juang	<i>Integer</i>	0	2	3	2.775	0.423
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
Inisiatif	<i>Integer</i>	0	2	3	2.525	0.506
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
penyesuaian_diri	<i>Integer</i>	0	2	3	2.900	0.304
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
fleksibilitas_berpikir	<i>Integer</i>	0	2	4	3.025	0.480
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
pengelolaan_emosi	<i>Integer</i>	0	2	3	2.850	0.362
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kepemimpinan	<i>Integer</i>	0	2	3	2.900	0.304

Nama	Type	Missing	Statistics			
			Min	Max	Average	Deviation
asertivitas	Integer	0	2	3	2.375	0.490
kepercayaan_diri	Integer	0	2	3	2.950	0.221
perencanaan	Integer	0	2	3	2.675	0.474
antisipasi_resiko	Integer	0	2	4	3	0.320
sistematika_kerja	Integer	0	2	3	2.725	0.452
kemampuan_komunikasi	Integer	0	3	3	3	0
taraf_kecerdasan	Integer	0	2	5	3.975	0.800
kemampuan_analisis	Integer	0	2	4	3.100	0.441
kemampuan_berpikir_komprehensif	Integer	0	2	4	3	0.320
pengambilan_keputusan	Integer	0	2	3	2.475	0.506
skor_potensi	Real	0	133.330	160	148.001	5.712
gap	Integer	0	1	9	3.850	2.402
<i>Showing attributes 1 - 33</i>		<i>Examples: 40</i>	<i>Special Attributes: 1</i>		<i>Regular Attributes: 32</i>	

4.3.2 *Cleaning Data*

Pada tahap ini, *attributes* yang tidak dipakai akan dihapus. Adapun *attributes* yang akan dihapus diantaranya adalah *attributes* yang memiliki *instance* homogen dan *irrelevant feature*. Adapun beberapa atribut yang dihapus karena tidak dapat digunakan dapat dilihat pada Tabel 4.3.2.

Tabel 4.3.2 *Cleaning Data*

Nama Atribut	Tipe Data	Value			
nama	<i>Polynomial</i>	K1 (1), K2 (1), K3 (1), K4 (1), K5 (1), K6 (1), K7 (1), K8 (1),, K40 (1)			
jenis_kelamin	<i>Polynomial</i>	Laki-laki (40)			
komitmen_terhadap_tugas	<i>Integer</i>	<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
		3	3	3	0
kerjasama	<i>Integer</i>	<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
		3	3	3	0
kemampuan_komunikasi	<i>Integer</i>	<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
		3	3	3	0

Atribut dengan *instance* yang sangat mirip atau homogen seringkali tidak berkontribusi secara signifikan terhadap diferensiasi antar kelas dikarenakan variasi atribut yang sangat rendah (Pranto, 2020). Disisi lain, *irrelevant feature* merupakan atribut yang juga tidak memberikan kontribusi terhadap hasil analisis, tidak membantu dan tidak mempengaruhi pemahaman model dalam melakukan klasifikasi. Selain itu, atribut tersebut meningkatkan dimensi data tanpa memberikan pengaruh signifikan terhadap performa model karena tidak mampu memberikan informasi tambahan yang membantu membedakan kelas-kelas data yang berbeda.(Pranto, 2020). Adapun hasil data setelah dilakukan *cleaning* dapat dilihat pada Tabel 4.3.3 Data Hasil *Cleaning*.

Tabel 4.3.3 Data Hasil *Cleaning*

Nama	Type	Missing	Statistics			
Label			<i>Negative</i>	<i>Positive</i>	<i>Value</i>	
<i>pj_fit</i>	<i>Binomial</i>	0	<i>Highly Fit</i>	<i>Moderately Fit</i>	<i>Highly Fit (24), Moderately Fit (16)</i>	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
pendidikan	<i>Polynomial</i>	0	S1 Teknik Lingkungan (2)	S1 Teknik Geologi (11)	S1 Teknik Geologi (11), S1 Teknik Pertambangan (6), S1 Teknik Sipil (11), S1 Teknik Elektro (4), S1 Pend. Teknik Informatika (2), S1 Pend. Teknik Mesin (4), S1 Teknik Lingkungan (2)	
			<i>Least</i>	<i>Most</i>	<i>Value</i>	
asal_sekolah	<i>Polynomial</i>	0	Universitas Trisakti (1)	UPN Yogyakarta (7)	UPN Yogyakarta (7), ITB (3), UGM (4), Universitas Negeri Malang (6), ITP (2), Politeknik Negeri Padang (2), Sekolah Tinggi Teknologi Mineral (4), UNY (2), Univ. Bina Nusantara (5), Univ. Pendidikan Indonesia (UPI) (2), Universitas Riau (2), Universitas Trisakti (1)	
usia	<i>Integer</i>	0	<i>Min</i> 21	<i>Max</i> 25	<i>Average</i> 23.125	<i>Deviation</i> 1.324
iq	<i>Integer</i>	0	<i>Min</i> 88	<i>Max</i> 145	<i>Average</i> 119.250	<i>Deviation</i> 14.723
pauli	<i>Integer</i>	0	<i>Min</i> 2025	<i>Max</i> 4581	<i>Average</i> 3210.625	<i>Deviation</i> 691.260
keterampilan_sosial	<i>Integer</i>	0	<i>Min</i> 2	<i>Max</i> 3	<i>Average</i> 2.950	<i>Deviation</i> 0.221

Nama	Type	Missing	Statistics			
			Min	Max	Average	Deviation
kepekaan_sosial	Integer	0	2	3	2.725	0.452
kepatuhan_terhadap_aturan	Integer	0	2	3	2.850	0.362
ketelitian_kerja	Integer	0	3	4	3.050	0.221
hasrat_berprestasi	Integer	0	2	3	2.775	0.423
daya_juang	Integer	0	2	3	2.775	0.423
Inisiatif	Integer	0	2	3	2.525	0.506
penyesuaian_diri	Integer	0	2	3	2.900	0.304
fleksibilitas_berpikir	Integer	0	2	4	3.025	0.480
pengelolaan_emosi	Integer	0	2	3	2.850	0.362
kepemimpinan	Integer	0	2	3	2.900	0.304
asertivitas	Integer	0	2	3	2.375	0.490
kepercayaan_diri	Integer	0	2	3	2.950	0.221
perencanaan	Integer	0	2	3	2.675	0.474
antisipasi_resiko	Integer	0	2	4	3	0.320

Nama	Type	Missing			Statistics	
sistematika_kerja	<i>Integer</i>	0	2	3	2.725	0.452
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
taraf_kecerdasan	<i>Integer</i>	0	2	5	3.975	0.800
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kemampuan_analisis	<i>Integer</i>	0	2	4	3.100	0.441
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
kemampuan_berpikir_komprehensif	<i>Integer</i>	0	2	4	3	0.320
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
pengambilan_keputusan	<i>Integer</i>	0	2	3	2.475	0.506
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
skor_potensi	<i>Real</i>	0	133.330	160	148.001	5.712
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
gap	<i>Integer</i>	0	1	9	3.850	2.402

ExampleSet (40 examples, 1 special attribute, 27 regular attributes)

4.3.3 Transformasi Data

Tahap transformasi data adalah sebuah proses yang melibatkan konversi format, struktur, atau nilai data untuk mempersiapkannya pada tahap analisis. Transformasi merupakan sebuah aspek penting pada data analisis dan *machine learning* karena dapat secara signifikan memberikan pengaruh pada performa model. Selain itu, algoritma klasifikasi biasanya memerlukan *input* numerik, sehingga data kategorik harus di transformasi ke bentuk numerik sebelum membangun model (Yan, Chen, & Guo, 2022).

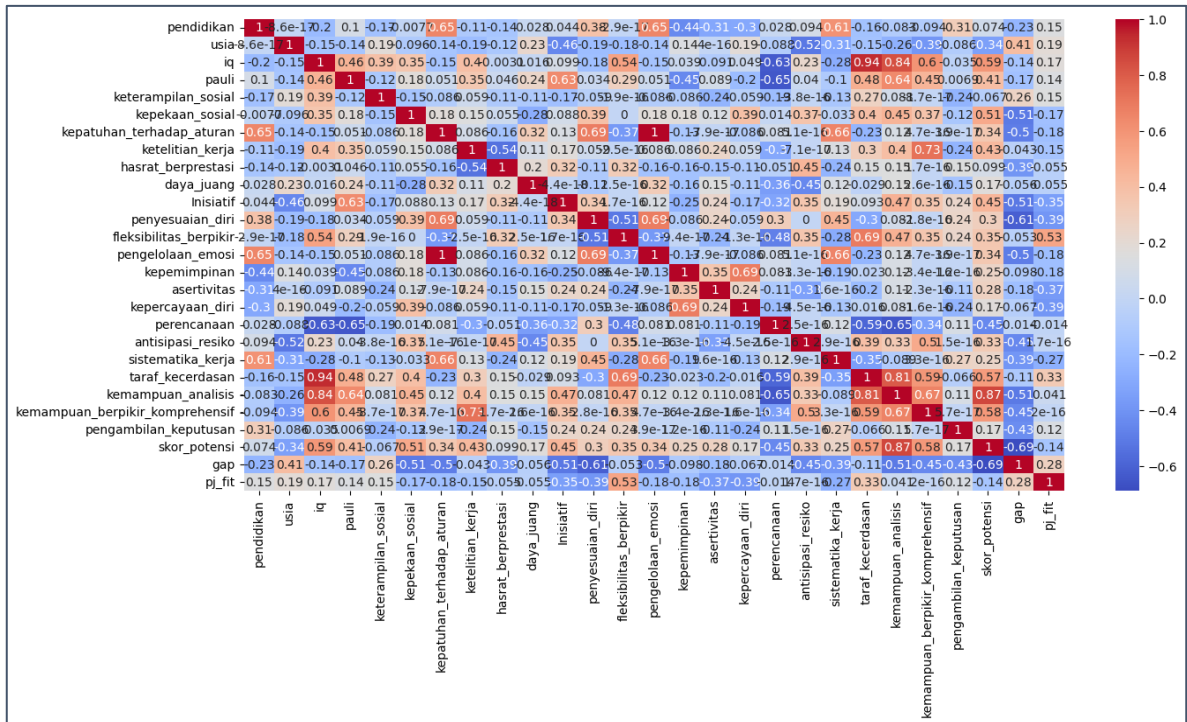
Tahap transformasi data pada penelitian ini dapat dilihat pada Tabel 4.3.4. Sebagian besar data yang digunakan dalam penelitian ini sudah dalam tipe data *integer* dan bilangan *real*, sehingga hanya terdapat 2 atribut yang ditransformasi yaitu pendidikan dan asal_sekolah dikarenakan keduanya masih dalam tipe data nominal. Selain itu, pada tahap transformasi ini, dilakukan juga *remap binominal*, dimana dilakukan modifikasi kelas positif dan negatif dari label.

Tabel 4.3.4 Atribut Transformasi

Nama	Type	Missing	Statistics			
Label			<i>Negative</i>		<i>Positive</i>	<i>Value</i>
<i>pj_fit</i>	<i>Binomial</i>	0	<i>Moderately Fit</i>		<i>Highly Fit</i>	<i>Highly Fit (13), Moderately Fit (5)</i>
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
pendidikan	<i>Numeric</i>	0	0	6	2.556	2.121
			<i>Min</i>	<i>Max</i>	<i>Average</i>	<i>Deviation</i>
asal_sekolah	<i>Integer</i>	0	0	11	4.333	3.481

4.3.4 Uji Korelasi

Uji korelasi pada penelitian ini menggunakan operator *correlation matrix* seperti yang ditunjukkan pada Gambar 4.3.1 *Correlation Matrix*.



Gambar 4.3.1 Correlation Matrix

Hasil uji korelasi ditunjukkan pada Gambar 4.3.1 Correlation Matrix yang menunjukkan nilai korelasi antar atribut. Berdasarkan koefisien korelasi oleh Sugiyono (2006, p. 214), tingkat hubungan pada korelasi dapat digambarkan pada Tabel 4.3.5.

Tabel 4.3.5 Interpretasi Koefisien Korelasi

Internal Koefisien	Tingkat Hubungan
0,00 – 0,199	Sangat Rendah
0,20 – 0,399	Rendah
0,40 – 0,599	Sedang
0,60 – 0,799	Kuat
0,80 – 1,00	Sangat Kuat

Berdasarkan Gambar 4.3.1, nilai korelasi antar variabel yang termasuk kedalam tingkat hubungan kuat dan sangat kuat dapat dilihat pada Tabel 4.3.6 Nilai Korelasi antar Atribut.

Tabel 4.3.6 Nilai Korelasi antar Atribut

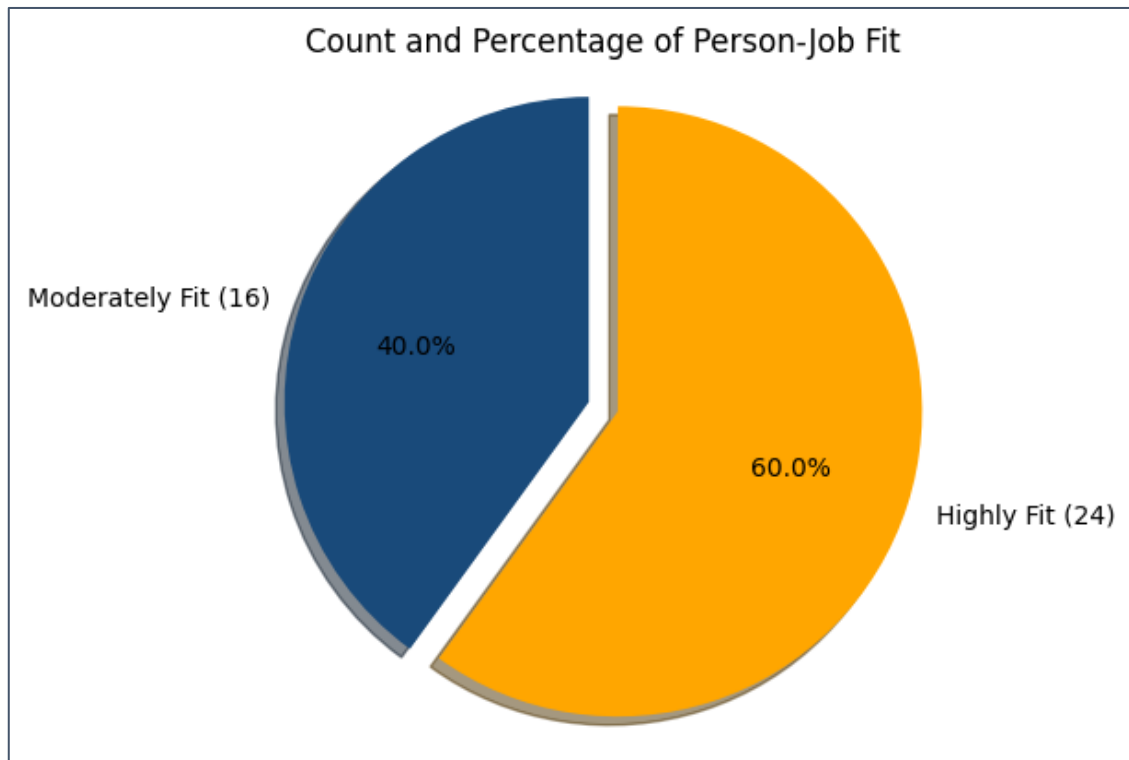
Atribut 1	Atribut 2	Nilai Korelasi	Tingkat Korelasi
kemampuan_analisis	skor_potensi	0.86188	Sangat Kuat

Atribut 1	Atribut 2	Nilai Korelasi	Tingkat Korelasi
iq	taraf_kecerdasan	0.93720	Sangat Kuat
iq	kemampuan_analisis	0.83848	Sangat Kuat
kepatuhan_terhadap_aturan	pengelolaan_emosi	1.0	Sangat Kuat
kemampuan_analisis	taraf_kecerdasan	0.80798	Sangat Kuat
gap	skor_potensi	-0.83589	Sangat Kuat
pendidikan	pengelolaan_emosi	0.65059	Kuat
pendidikan	sistematika_kerja	0.60850	Kuat
iq	kemampuan_berpikir_komprehensif	0.60075	Kuat
gap	penyesuaian_diri	-0.61451	Kuat
pengelolaan_emosi	sistematika_kerja	0.66143	Kuat
iq	perencanaan	-0.63017	Kuat
kemampuan_analisis	kemampuan_berpikir_komprehensif	0.66666	Kuat
kepemimpinan	kepercayaan_diri	0.68599	Kuat
kemampuan_analisis	perencanaan	-0.64577	Kuat
kemampuan_analisis	pauli	0.64266	Kuat
kepatuhan_terhadap_aturan	sistematika_kerja	0.66143	Kuat
pengelolaan_emosi	penyesuaian_diri	0.68599	Kuat
Inisiatif	pauli	0.63324	Kuat
kepatuhan_terhadap_aturan	pendidikan	0.65059	Kuat
kepatuhan_terhadap_aturan	penyesuaian_diri	0.68599	Kuat
fleksibilitas_berpikir	taraf_kecerdasan	0.69486	Kuat
kemampuan_berpikir_komprehensif	ketelitian_kerja	0.72760	Kuat
kemampuan_berpikir_komprehensif	skor_potensi	0.68821	Kuat
pauli	perencanaan	-0.64845	Kuat
pendidikan	pengelolaan_emosi	0.65059	Kuat
pendidikan	sistematika_kerja	0.60850	Kuat
iq	kemampuan_berpikir_komprehensif	0.60075	Kuat
gap	penyesuaian_diri	-0.61451	Kuat
pengelolaan_emosi	sistematika_kerja	0.66143	Kuat
iq	perencanaan	-0.63017	Kuat

Dari Tabel 4.3.6 di atas, dapat diketahui bahwa data pada penelitian ini tergolong data berkorelasi tinggi. Data dapat dikatakan berkorelasi tinggi jika terdapat hubungan yang kuat antar dua atau lebih atribut dalam data (Biswas, Fole, Khare, & Agrawal, 2023). Korelasi ini umum terjadi pada aspek potensi dalam tes psikologi (Grailey, Murray, Reader, & Brett, 2021; Fenn, Tan, & George, 2020). Tes psikologi dirancang untuk mengukur berbagai aspek dari kondisi psikologis seseorang seperti kemampuan, kepribadian, dan fungsi kognitif. Berbagai aspek ini seringkali saling terkait. Misalnya, kemampuan kognitif seseorang mungkin memengaruhi kepribadiannya, atau sebaliknya (Fenn, Tan, & George, 2020). Oleh karena itu, dalam kasus kali ini *feature selection* tidak akan dilakukan karena dibutuhkan pemetaan yang lebih jelas pada keseluruhan fitur hasil tes psikologi karyawan untuk memprediksi kecocokan seseorang dengan pekerjaannya.

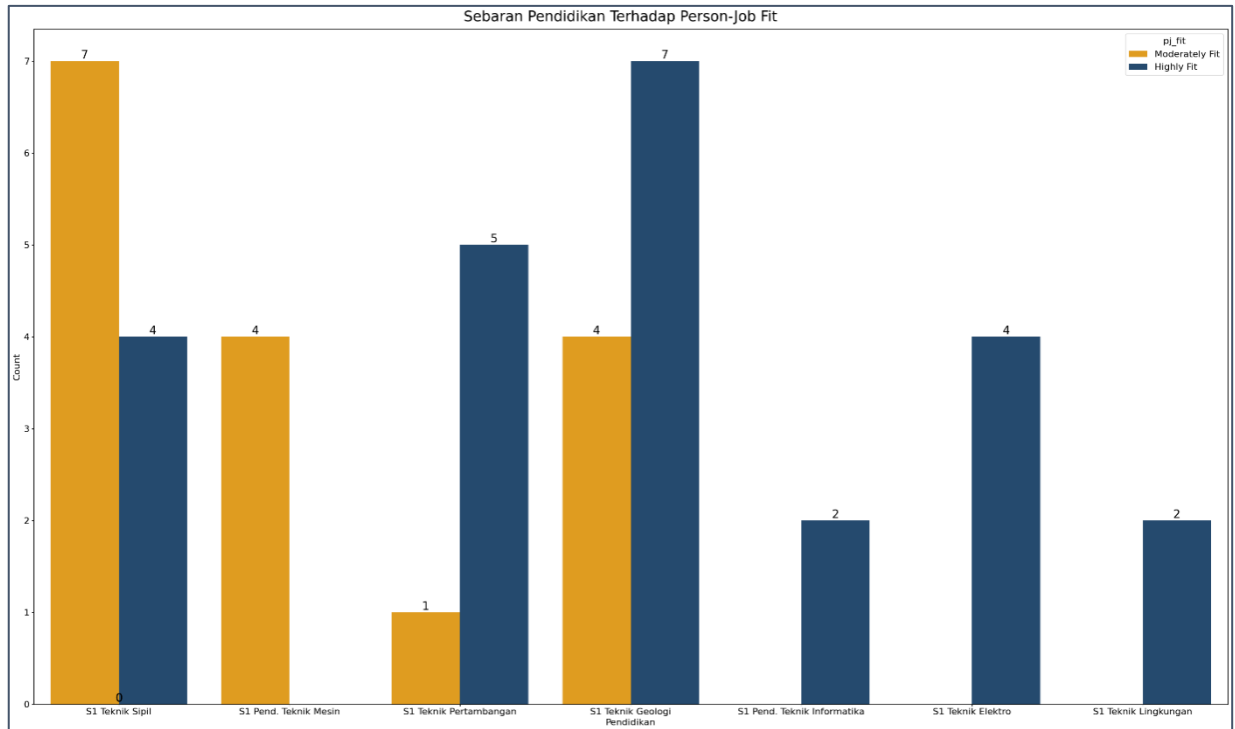
4.4 Exploratory Data Analysis (EDA)

EDA dilakukan untuk mengetahui sebaran data yang akan digunakan dalam penelitian, data yang digunakan adalah data hasil *cleaning data*.



Gambar 4.4.1 Persentase Kelas Data *Person-Job Fit*

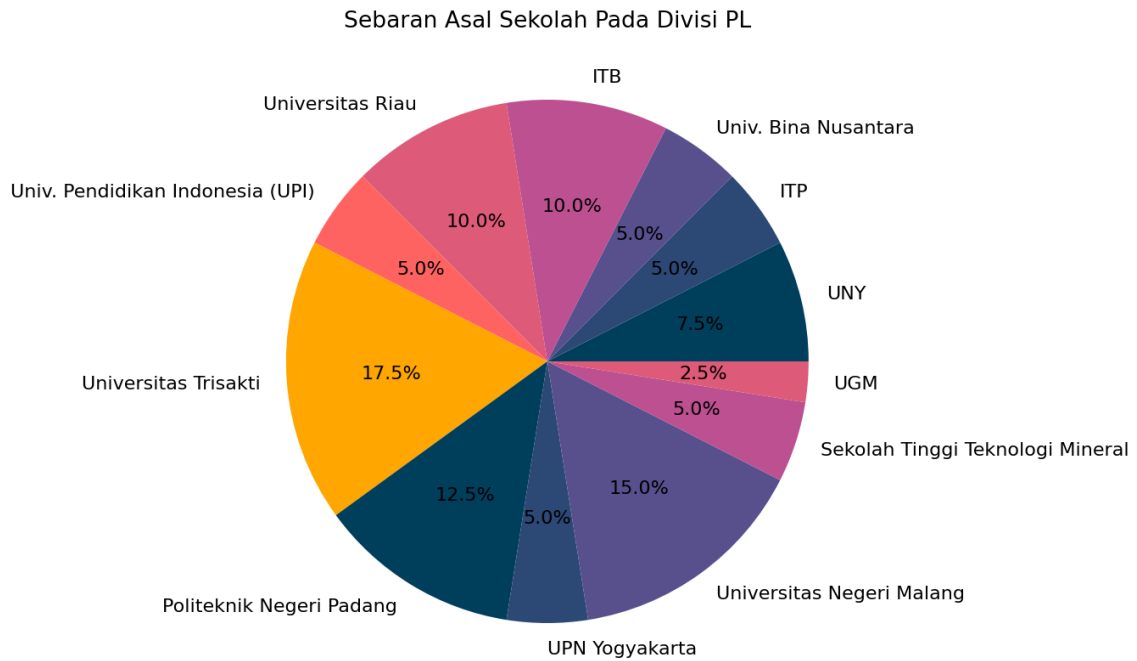
Dari Gambar 4.4.1 dapat diketahui bahwa jumlah data untuk label kelas *Highly Fit* adalah sebanyak 24 data atau 60% dari keseluruhan data dan untuk jumlah label kelas *Moderately Fit* adalah sebanyak 16 data atau 40% dari keseluruhan data.



Gambar 4.4.2 Sebaran Pendidikan pada setiap Kelas Data

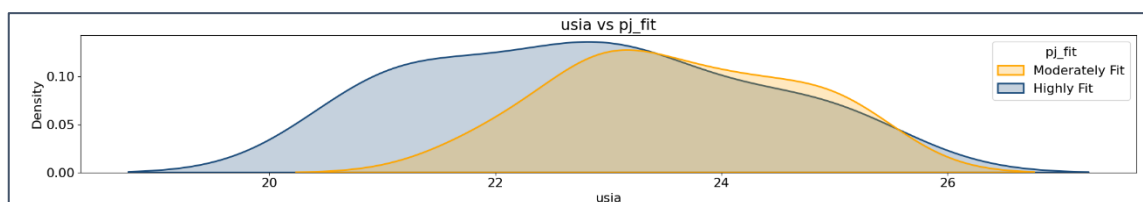
Dari Gambar 4.4.2 menunjukkan sebaran pendidikan pada kelas data. Dapat diketahui bahwa lulusan S1 Teknik Geologi memiliki jumlah karyawan yang masuk kedalam kategori *highly fit* terbanyak dibandingkan jenis pendidikan lain dengan jumlah karyawan sebanyak 7 orang. Kemudian S1 Teknik Pertambangan memiliki jumlah karyawan terbanyak kedua sebanyak 5 orang. Sedangkan untuk S1 Teknik Lingkungan dan S1 Teknik Lingkungan memiliki jumlah karyawan dengan kelas *highly fit* terkecil dengan jumlah 2 orang. Hal ini dapat disimpulkan bahwa urutan pendidikan yang memiliki jumlah terbanyak yang masuk kedalam kelas *highly fit* yaitu S1 Teknik Geologi, S1 Teknik Pertambangan, S1 Teknik Elektro, S1 Teknik Sipil, S1 Pend. Teknik Informatika dan Teknik Lingkungan.

Kelas *moderately fit*, pendidikan yang memiliki jumlah terbanyak yaitu S1 Teknik Sipil dengan jumlah 7 orang dan S1 Teknik Pertambangan dengan jumlah terkecil sebanyak 1 orang. Dengan demikian urutan pendidikan yang termasuk kedalam kelas *moderately fit* yaitu S1 Teknik Sipil, S1 Pend. Teknik Mesin, S1 Teknik Geologi dan terakhir S1 Teknik Pertambangan. Dari penjelasan tersebut dapat disimpulkan pada divisi PL pendidikan yang paling cocok adalah S1 Teknik Geologi.



Gambar 4.4.3 Sebaran Asal Sekolah pada Divisi PL

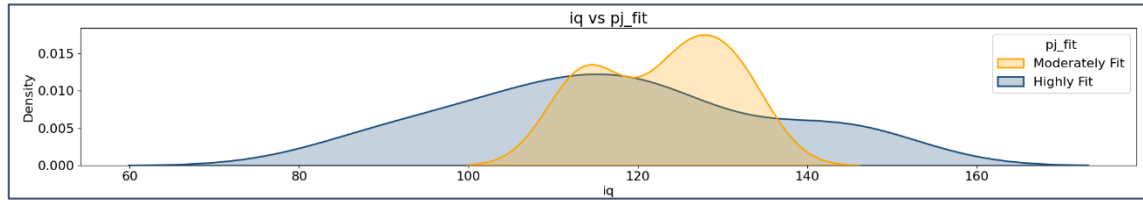
Gambar 4.4.3 menunjukkan bahwa Universitas Trisakti memiliki jumlah karyawan terbanyak pada Divisi PL dengan persentase sebanyak 17,5% diikuti oleh Politeknik Negeri Padang dengan persentase 12,5%. UGM merupakan universitas dengan persentase jumlah karyawan terkecil pada divisi PL dengan 2,5%.



Gambar 4.4.4 Sebaran Usia pada setiap Kelas Data

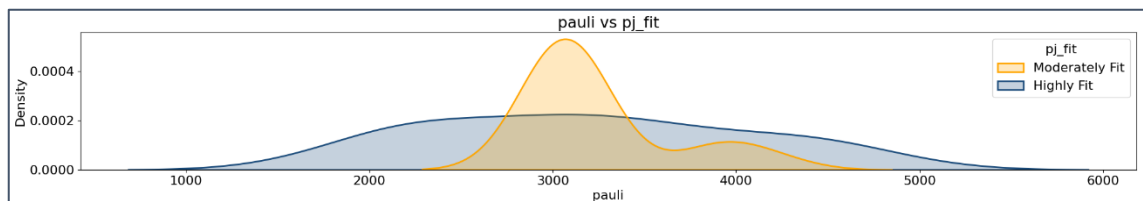
Gambar 4.4.4 merupakan grafik KDE (*Kernel Density Estimation*) yang menunjukkan tingkat *density* atau kepadatan pada atribut usia. Pada **Error! Reference source not found.** menunjukkan bahwa seorang karyawan yang berusia direntang 22-24 tahun memiliki kepadatan puncak pada *person-job fit*. Untuk kelas *moderately fit* pada atribut usia 23 merupakan puncak kepadatan. Sedangkan pada kelas *highly fit* memiliki distribusi yang lebih luas dengan puncak kepadatan yang berada di sekitar usia 22-23 tahun. Dengan

luasnya distribusi pada kelas *highly fit* ini menandakan bahwa individu yang masuk kedalam kategori ini memiliki rentang usia yang lebih besar.



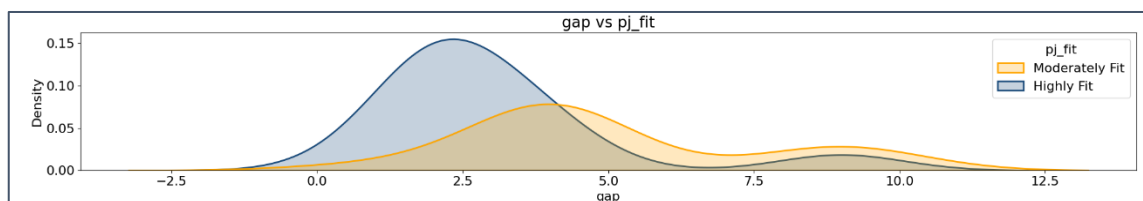
Gambar 4.4.5 Sebaran Skor IQ pada setiap Kelas Data

Berdasarkan **Error! Reference source not found.**, pada kelas *moderately fit* untuk skor antara 120-140 memiliki kepadatan puncak yang artinya bahwa pada kelas ini sebagian besar diisi dengan oleh karyawan dengan skor iq yang berada diantara 120-140. Sedangkan pada kelas *highly fit* memiliki distribusi yang lebih luas dengan puncak kepadatan berada pada rentang skor iq 100-120.



Gambar 4.4.6 Sebaran Skor Pauli pada setiap Kelas Data

Pada Gambar 4.4.6 untuk kelas data *moderately fit* memiliki puncak kepadatan pada skor 3000 yang berarti banyak individu atau karyawan yang memiliki skor pauli 3000 pada kelas data ini. Selanjutnya pada kelas data *highly fit* memiliki sebaran data yang lebih luas atau dengan kata lain berarti lebih banyak variasi data pada kelas data ini dengan puncak kepadatan yang sedikit menonjol sekitar skor 3000.

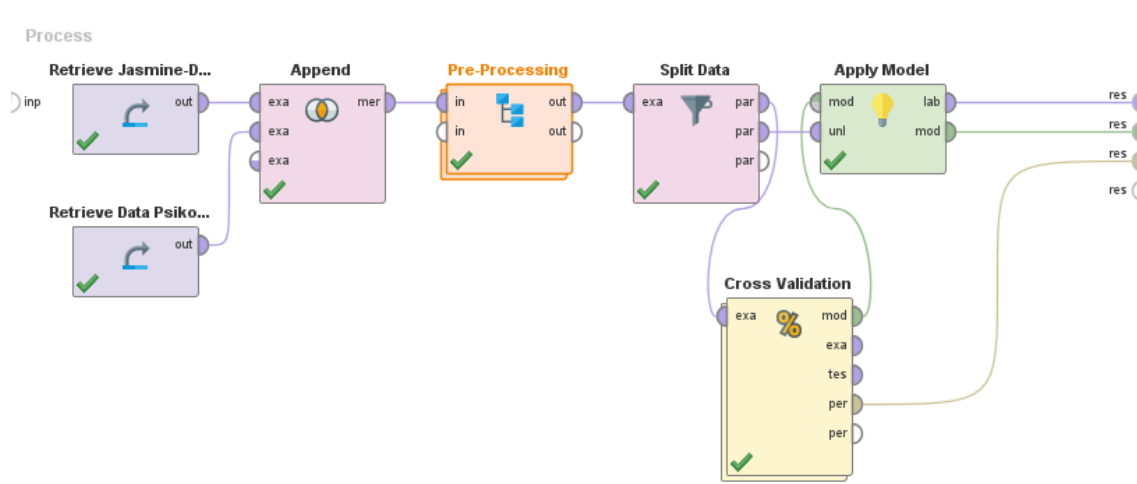


Gambar 4.4.7 Sebaran *Gap* Skor Potensi pada setiap Kelas Data

Gambar 4.4.7 menunjukkan sebaran *gap* skor potensi pada kelas *moderately fit* yang cenderung lebih luas dan menunjukkan sebaran dengan lebih banyak variasi pada data dengan puncak kepadatan berada diantara skor 2,5 hingga 5. Sedangkan pada kelas *highly fit* menunjukkan sebaran yang cenderung lebih tinggi dan sempit dengan puncak kepadatan pada skor sekitar 2,5.

4.5 Pembangunan Model

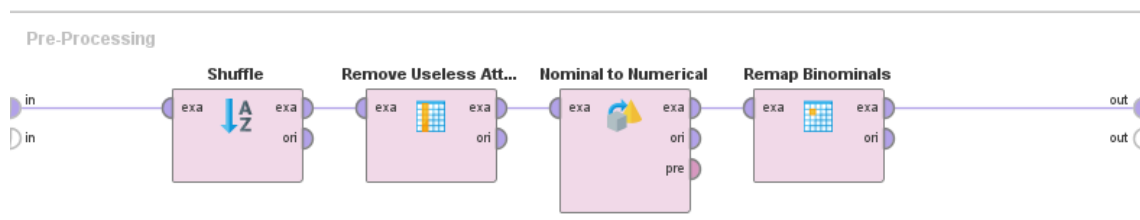
Setelah dilakukan tahap *pre-processing*, diketahui bahwa total data yang akan digunakan pada pembangunan model klasifikasi adalah sebanyak 40 data dan 28 atribut dengan 1 label atau target dan 27 atribut lainnya merupakan prediktor. Terdapat beberapa aspek penting tentang *tree-based models* dan bagaimana model tersebut berinteraksi dengan fitur yang berkorelasi, dimana *tree-based models* menunjukkan tingkat ketahanan tertentu terhadap fitur-fitur yang berkorelasi (Schweimer & Scher, 2022). Berdasarkan pendapat tersebut, pembangunan model klasifikasi pada penelitian ini dilakukan dengan menggunakan algoritma *Decision Tree* dan *Random Forest* dengan bantuan aplikasi *Rapid Miner* 9.10.011.



Gambar 4.5.1 Pembangunan Model Menggunakan *Cross Validation*

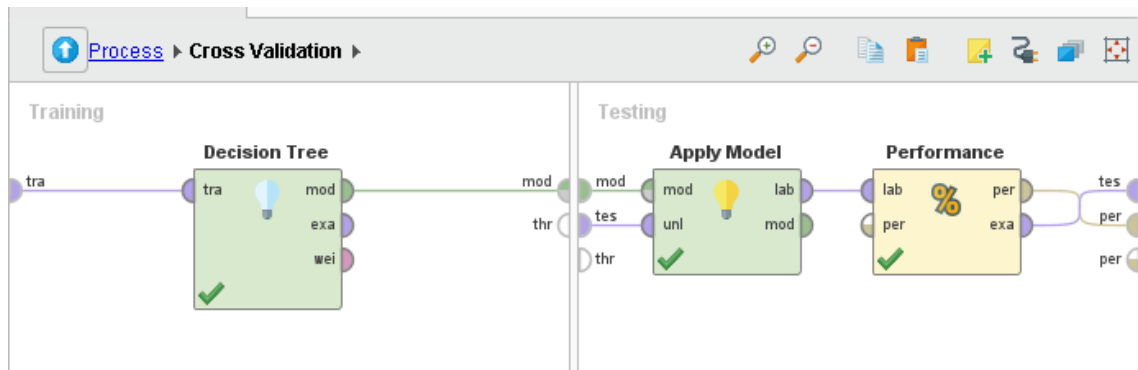
Error! Reference source not found. menunjukkan proses komputasi dan operator yang digunakan pada *Rapid Miner* dalam membangun model klasifikasi. Proses komputasi terdiri dari data asli psikogram dan data hasil augmentasi yang akan digabungkan menggunakan operator *append*, operator *cross validation* yang akan melakukan validasi silang untuk melakukan estimasi performa

statistik dari *learning model*. Lalu ada operator *apply model* yang digunakan untuk mengaplikasikan model yang sudah dilatih oleh algoritma kedalam *example set* lainnya atau dalam hal ini merupakan *data testing* dengan tujuan untuk memperoleh kelas prediksi model. Adapun pada operator *pre-processing* berisi gabungan operator yang digunakan dalam *pre-processing* data. Adapun penjelasan terkait operator yang digunakan dalam tahap ini dapat dilihat pada Gambar 4.5.2.



Gambar 4.5.2 Tahap *Pre-processing* Data

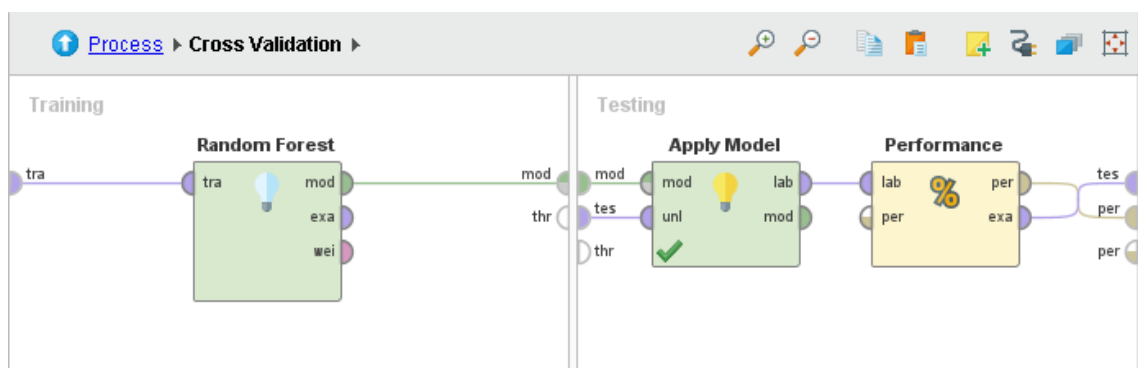
Operator yang digunakan dalam tahap *pre-processing* yaitu *shuffle* untuk mengacak data, *remove useless attributes* yang digunakan untuk menghapus atribut yang tidak digunakan atau tidak terpakai dan dapat dilihat pada **Error! Reference source not found.** Selanjutnya yaitu operator *nominal to numerical* yang digunakan untuk mengubah tipe data dan *instance* dari data atau dengan kata lain proses transformasi data non-numerikal menjadi data numerikal. Adapun atribut yang melalui tahap ini dapat dilihat pada **Error! Reference source not found.** Operator selanjutnya yaitu *remap binominals* yang digunakan untuk memodifikasi *value mapping* dari atribut binominal dalam hal ini adalah atribut label kelas dengan mengubah label kelas positif dan negatif. Tahap yang ditunjukkan pada **Error! Reference source not found.** dan Gambar 4.5.2 untuk semua algoritma adalah sama dan tidak ada perbedaan terkecuali pada operator *cross validation* yang akan dijelaskan pada Gambar 4.5.3.



Gambar 4.5.3 Pembangunan Model *Decision Tree* Menggunakan *Cross Validation*

Gambar 4.5.3 menunjukkan pembangunan model *decision tree* dengan *cross validation*. Operator ini memiliki dua sub proses yang pertama yaitu *training subprocess* yang digunakan untuk training model dan berisi operator algoritma yang akan kita gunakan, dalam hal ini *decision tree* dan *testing subprocess* yang mana model yang sudah dilatih pada *training subprocess* akan diaplikasikan kedalam *testing subprocess* menggunakan operator *apply model* dan akan diukur nilai performa model dengan bantuan operator *performance*. Adapun nilai performa yang akan dianalisis yaitu akurasi, AUC, *recall*, *precision*, F1 score, dan *classification error*.

Pembangunan model selanjutnya yaitu menggunakan algoritma *Random Forest* menggunakan *cross validation* yang ditunjukkan pada Gambar 4.5.4.



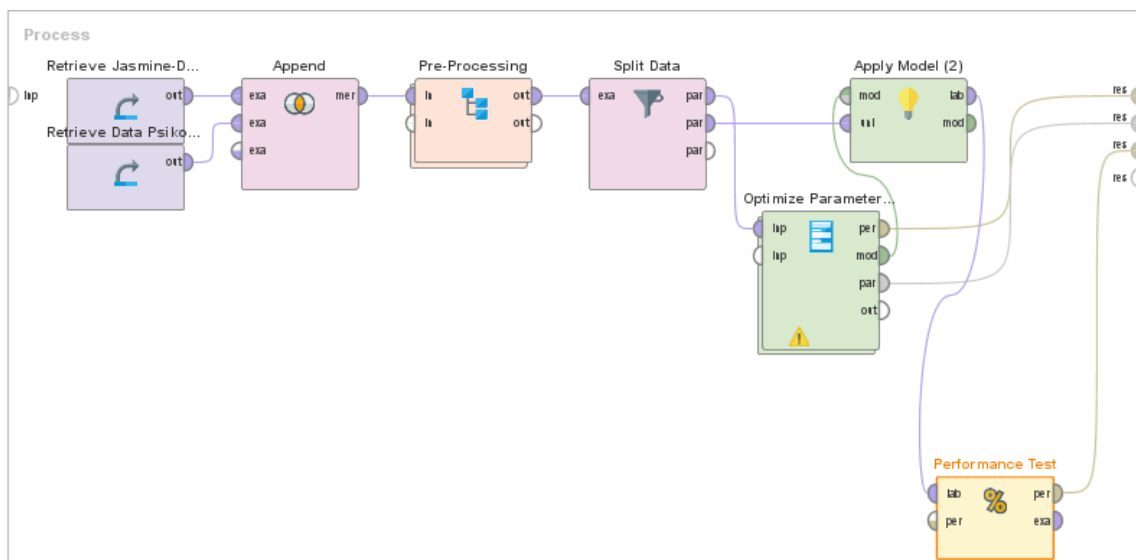
Gambar 4.5.4 Pembangunan Model *Random Forest* Menggunakan *Cross Validation*

Pembangunan model menggunakan *random forest* tidak jauh berbeda dengan *decision tree*. Proses mengganti model dapat dengan mudah dilakukan dengan mengganti operator algoritma yang akan digunakan dalam membangun model klasifikasi pada *rapid miner*.

Pada Gambar 4.5.4 diperlihatkan bahwa operator yang digunakan sebagian besar sama hanya berbeda pada algoritma yang digunakan yaitu *random forest*.

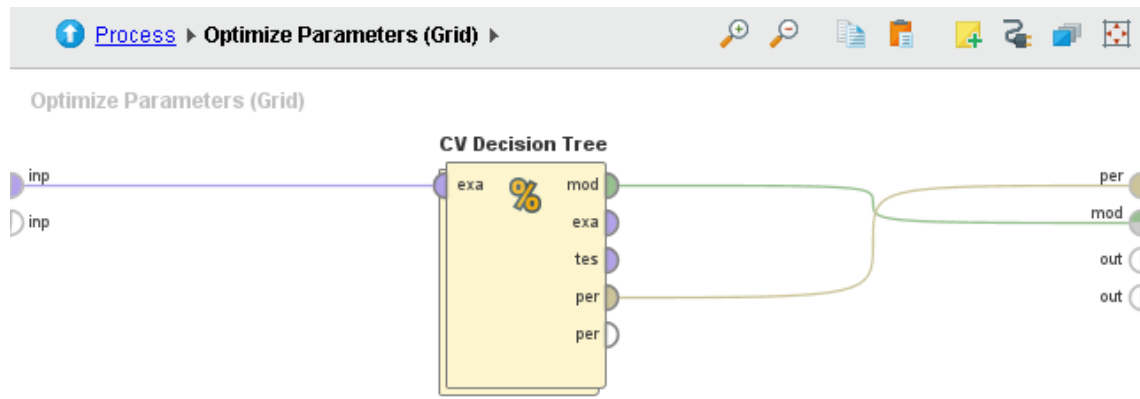
4.6 Pengolahan Model Klasifikasi dengan *Parameter Tuning*

Pembangunan model dengan melakukan parameter tuning secara garis besar memiliki kesamaan dengan pembangunan model pada umumnya, yang membedakan proses ini adalah adanya penambahan operator baru berupa *Optimize Parameters (Grid)* yang akan mencari nilai optimal dari parameter yang dipilih untuk operator dalam subprosesnya. Proses komputasi menggunakan *Optimize Parameters (Grid)* dapat dilihat pada Gambar 4.6.1.



Gambar 4.6.1 Pembangunan Model Menggunakan *Parameter Tuning*

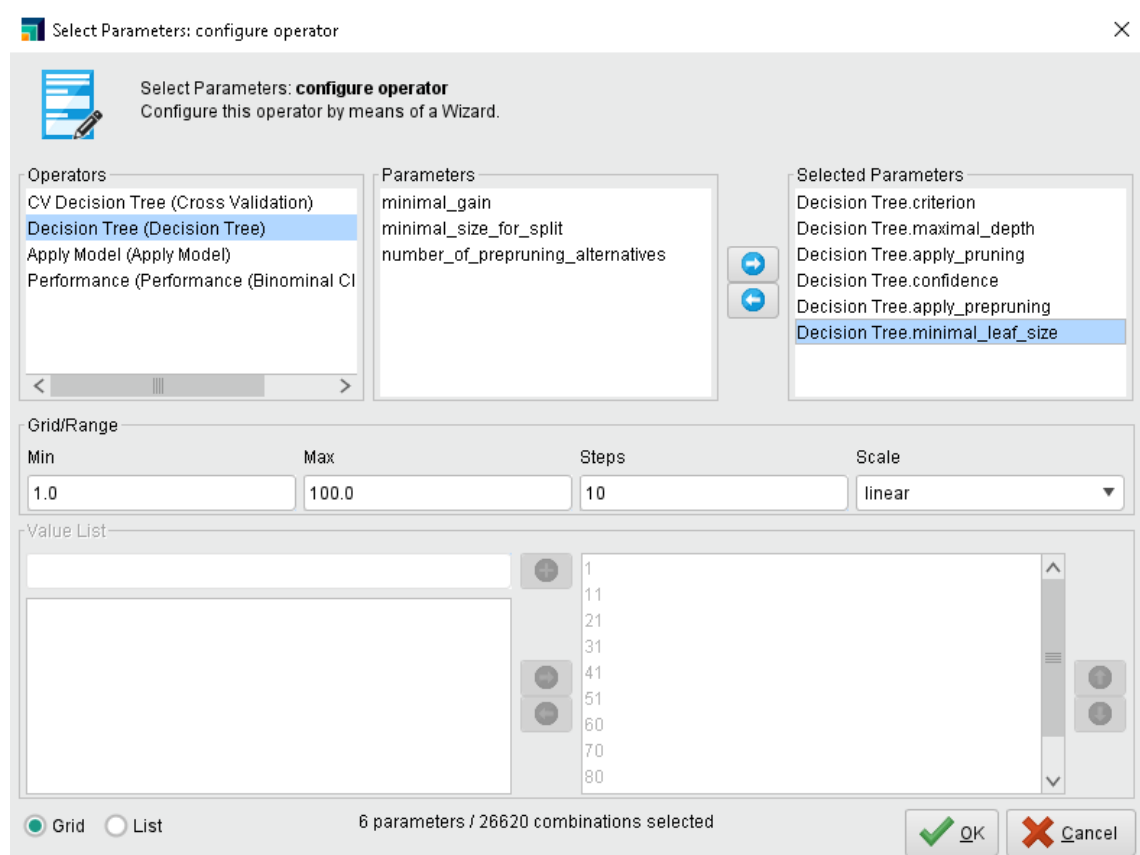
Gambar 4.6.1 menunjukkan penggunaan operator *Optimize Parameters (Grid)* dalam membangun model klasifikasi. Operator pada proses ini tidak jauh berbeda pada proses sebelumnya dan untuk lebih jelas proses *parameter tuning* dapat dilihat pada Gambar 4.6.2.



Gambar 4.6.2 *Parameter Tuning* pada Algoritma *Decision Tree*

Pada *parameter tuning*, operator yang akan dimasukkan yaitu *cross validation* yang sudah berisi operator *decision tree*, *apply model*, dan *performance*. Untuk bagian operator yang digunakan dalam *cross validation* juga berisi sama dengan tahap sebelumnya. Yang perlu diganti dari bagian *cross validation* adalah algoritma yang akan digunakan.

Untuk mulai melakukan *parameter tuning*, pada bagian *Parameters*, klik *Edit Parameter Settings*. Lalu akan muncul *pop-up* pada Gambar 4.6.3.



Gambar 4.6.3 *Select parameters: Configure operator* Algoritma *Decision Tree*

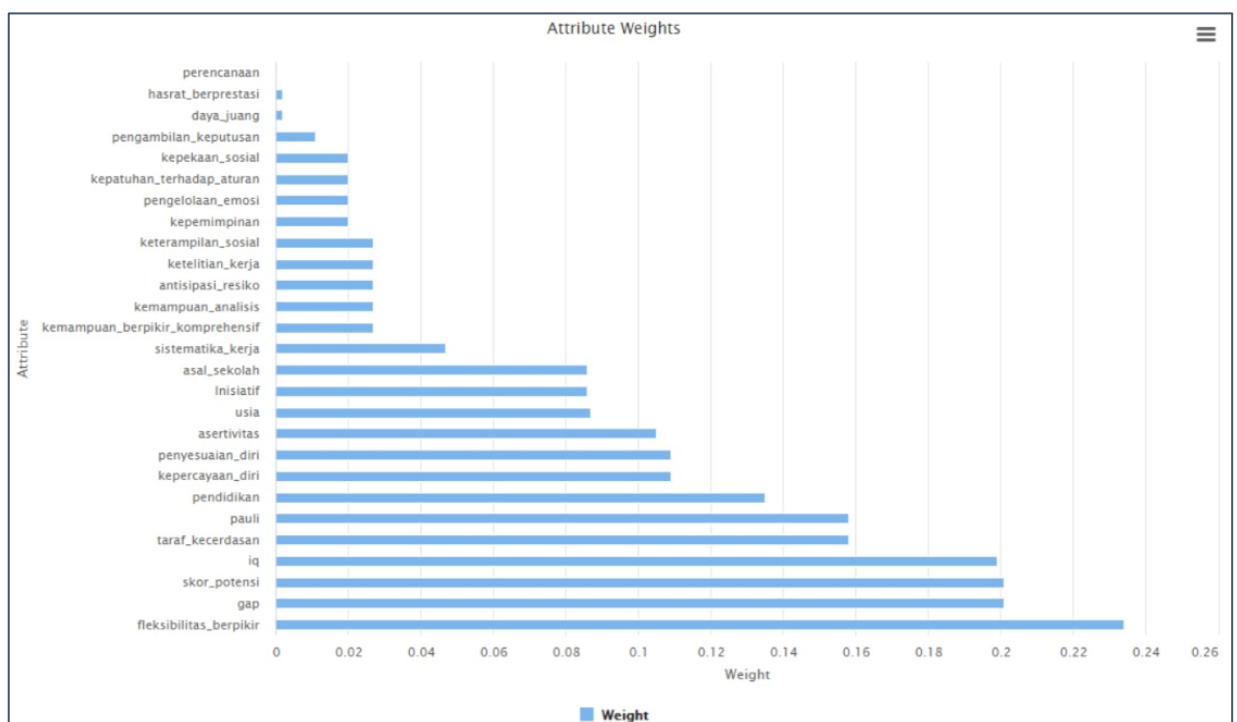
Pada Gambar 4.6.3, akan dilakukan *parameter tuning* dari algoritma *decision tree*. Sebelumnya, tentukan parameter apa saja yang akan di *tuning* atau dicari nilai optimalnya dengan cara pilih *Decision Tree (Decision Tree)* pada *operators* dan pindahkan parameter yang akan di *tuning* dari kolom *Parameters* ke kolom *Selected Parameters*. Lalu klik *Ok* dan jalankan proses komputasi. Lamanya proses parameter *tuning* akan bergantung pada banyaknya parameter yang akan dipilih untuk dicari nilai optimalnya. Semakin banyak parameter yang akan di *tuning*, akan semakin lama proses komputasi parameter *tuning*. Selain itu, kemampuan *hardware* perangkat yang digunakan juga akan sangat memengaruhi lamanya proses *tuning*.

4.7 Features Importance

Features importance pada penelitian ini dilakukan menggunakan operator *weight by information gain* untuk mengetahui kepentingan setiap fitur dalam model klasifikasi *Person-Job Fit*. Penggunaan *information gain* dalam tahap ini adalah karena kemampuan dari *information gain* dalam menangani variasi dalam data (Shu, Yan, & Qian, 2023) dan dapat digunakan dalam menentukan nilai entropy maksimum untuk penilaian bobot dari fitur atau atribut (Prasetyowati, Maulidevi, & Surendro, 2021). Dimana nilai entropi maksimum ini dapat membantu dalam melakukan ekstraksi fitur atau atribut yang paling informatif dan relevan dari sebuah data (Chen, Zhang, & Pan, 2023; Schnöoll, Ferner, & Wegenkittl, 2019). Adapun data yang dipakai untuk *features importance* adalah data yang sudah melalui tahap *cleaning data*. Dikarenakan atribut dengan *instance* yang homogen tidak memiliki variasi dalam data dan tidak mampu memberikan informasi yang berguna pada model sehingga dapat diabaikan dalam proses *feature importance* (Saarela & Jauhiainen, 2021).

Gambar 4.7.1, menunjukkan urutan kepentingan setiap fitur dalam membangun model klasifikasi yang dapat membantu dalam mengidentifikasi fitur yang memiliki kepentingan terbesar dalam model klasifikasi dan. Fitur fleksibilitas berpikir memiliki tingkat kepentingan terbesar dengan bobot mendekati 0,24 yang menunjukkan bahwa kemampuan seorang calon karyawan untuk berpikir secara fleksibel dan kreatif dalam memecahkan masalah menjadi poin penting yang menentukan kecocokan calon karyawan tersebut dengan pekerjaannya. Selanjutnya yaitu fitur *gap* dan skor potensi di tempat yang sama dengan bobot 0,2 kemudian diikuti oleh *Iq* dengan bobot yang mendekati 0,2. Lalu

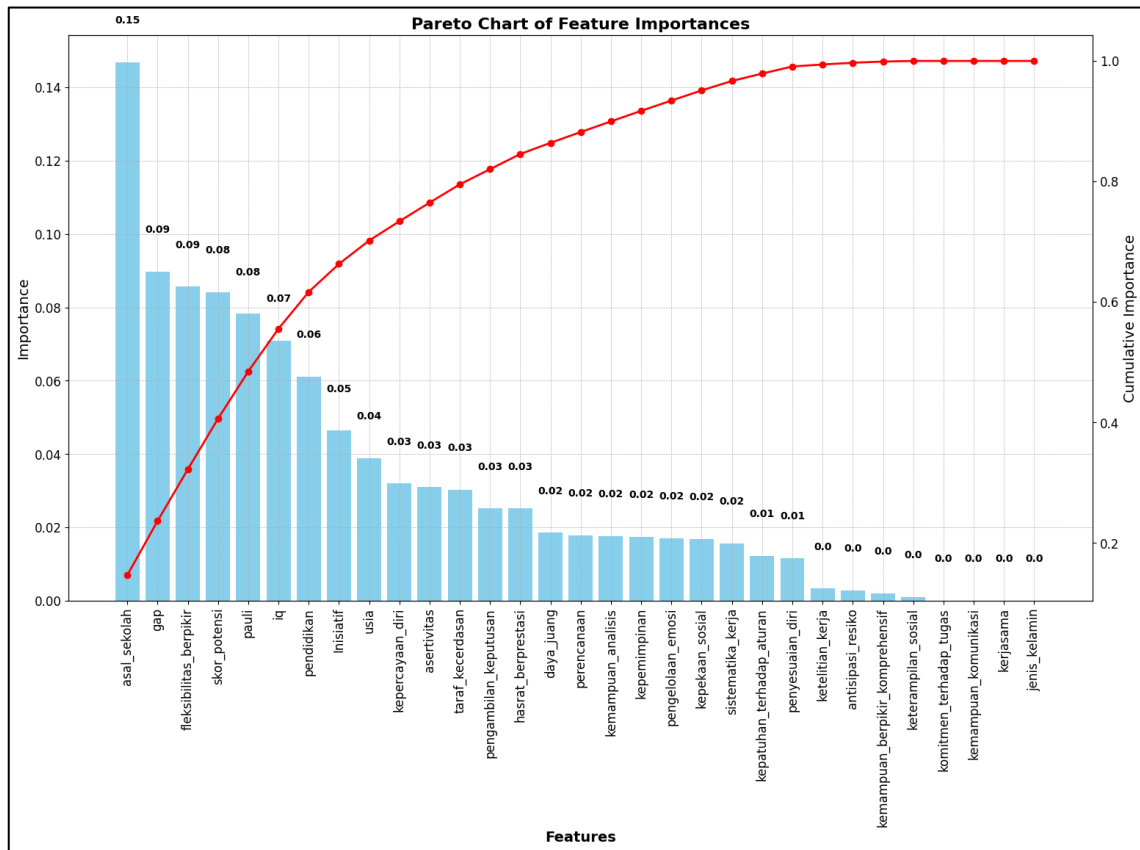
fitur taraf kecerdasan dengan bobot yang mendekati 0,16. Penggunaan *information gain* digunakan untuk mengukur seberapa banyak informasi yang dapat diberikan dari setiap atribut atau fitur terhadap target atau kelas data yang akan diprediksi. Sehingga atribut dengan bobot *information gain* yang lebih tinggi dianggap lebih penting. Oleh karena itu, fitur fleksibilitas berpikir merupakan atribut yang lebih penting dalam memprediksi *person-job fit*.



Gambar 4.7.1 *Features Importance*

Untuk mengukur urutan kepentingan fitur yang digunakan pada penelitian ini adalah diagram pareto. *Pareto chart* atau diagram pareto merupakan sebuah alat visualisasi yang umumnya digunakan dalam pengendalian mutu. Tujuannya adalah untuk mengidentifikasi dan memprioritaskan masalah yang paling signifikan. Diagram pareto bekerja dengan melakukan identifikasi fitur atau variable yang paling signifikan atau yang memiliki dampak terbesar, sehingga organisasi dapat lebih fokus pada variabel yang paling produktif atau yang menyumbangkan hasil dan pengaruh yang paling signifikan (Wardhani, 2022). Prinsip dari diagram pareto dikenal juga sebagai prinsip 80/20 yang dapat diartikan bahwa 80% konsekuensi datang dari 20% penyebab. Diagram pareto

mengilustrasikan prinsip ini dengan menunjukkan bahwa sebagian besar konsekuensi atau masalah diakibatkan oleh sejumlah kecil penyebab (Sudirman, 2021).



Gambar 4.7.2 Pareto Chart Fitur terhadap *Person-Job Fit*

Berdasarkan Gambar 4.7.2 Pareto Chart Fitur terhadap *Person-Job Fit*, dapat diketahui bahwa fitur asal sekolah menyumbang 15% konsekuensi atau memiliki signifikansi sebesar 15% dari total kepentingan yang dapat mempengaruhi hasil memprediksi kelas *person-job fit*. Penggunaan kedua metode dalam mencari *feature importance* pada penelitian ini digunakan secara komplementer. Dimana pencarian *feature importance* menggunakan bobot *information gain* ditunjukkan untuk mengetahui fitur yang paling informatif atau memberikan informasi paling banyak terhadap *person-job fit*, sedangkan pareto digunakan untuk mengetahui distribusi pentingnya fitur dalam model klasifikasi *person-job fit*. Sehingga nantinya perusahaan dapat lebih efektif dalam mengidentifikasi dan memilih prioritas fitur yang paling signifikan dalam model klasifikasi *person-job fit*.

BAB V

PEMBAHASAN

5.1 Performa Model Klasifikasi *Decision Tree* dan *Random Forest*

Perbandingan performa model klasifikasi menggunakan algoritma *Decision Tree* dan *Random Forest*:

Tabel 5.1.1 Perbandingan Performa model Klasifikasi Algoritma *Decision Tree*

Matriks Evaluasi	<i>Decision Tree</i>	<i>Decision Tree Tuning</i>	<i>Performance Data Testing</i>
Akurasi	85%	89,29%	91,67%
AUC	45%	21,4%	50%
<i>Precision</i>	93,33%	93,75%	87,5%
<i>Recall</i>	80%	88,24%	100%
<i>F1 Score</i>	87,5%	90,91%	93,33%
<i>Class Precision (Highly Fit)</i>	93,33%	93,75%	87,5%
<i>Class Precision (Moderately Fit)</i>	76,92%	83,33%	100%
<i>Class Recall (Highly Fit)</i>	82,35%	88,24%	100%
<i>Class Recall (Moderately Fit)</i>	90,91%	90,91%	80%
<i>Classification Error</i>	15%	10,71%	8,33%

Hasil performa model menggunakan algoritma *Decision Tree* dapat dilihat pada Tabel 5.1.1. Semua metrik evaluasi menunjukkan peningkatan saat dilakukan poses *parameter tuning*. Sedangkan performa saat diuji pada *data testing* ada penurunan pada beberapa metrik. Adapun penjelasan dan interpretasi lebih lanjut dari hasil diatas adalah sebagai berikut:

1. Akurasi menunjukkan peningkatan performa menjadi 89,29% dan semakin meningkat saat diterapkan ke data uji menjadi sebesar 91,67%.
2. Skor AUC mengalami penurunan yang cukup signifikan dari 45% menjadi 21,4% setelah *parameter tuning* dan kembali meningkat dengan skor 50% pada saat diaplikasikan ke *data testing*.
3. Selanjutnya yaitu *Precision* mengalami sedikit peningkatan performa menjadi 93,75%. Namun penurunan terjadi saat model diaplikasikan pada *data testing* dengan skor hanya sebesar 87,5%.

4. Metrik *Recall* menunjukkan peningkatan performa dari 80% menjadi 88,24% dan semakin meningkat menjadi 100% saat diaplikasikan ke *data testing*.
5. Berikutnya yaitu metrik *F1 score* menunjukkan peningkatan nilai dari 87,5% menjadi 90,91% setelah dilakukan proses *parameter tuning*, dan semakin meningkat saat diterapkan pada data uji 93,33%.
6. Selanjutnya yaitu presisi dari kelas positif (*Highly Fit*) menunjukkan peningkatan performa dari 93,33% menjadi 93,75% setelah *parameter tuning* dan menurun saat model diaplikasikan kedalam *data testing* menjadi sebesar 87,5%.
7. Pada presisi dari kelas negatif (*Moderately Fit*) terlihat model mengalami peningkatan yang sangat signifikan dari 76,92% menjadi 83,33% setelah dilakukan *parameter tuning* dan semakin meningkat menjadi 100% saat diaplikasikan.
8. Metrik *recall* dari kelas positif (*Highly Fit*) menunjukkan peningkatan dari 82,35% menjadi 88,24% setelah dilakukan *parameter tuning* dan meningkat menjadi 100% saat diaplikasikan ke *data testing*.
9. Metrik *recall* dari kelas negatif (*Moderately Fit*) yang menunjukkan stabilitas performa sebesar 90,91% baik sebelum dan sesudah *parameter tuning* namun mengalami penurunan menjadi 80% diaplikasikan.
10. Terakhir metrik *classification error* menurun dari 15% menjadi 10,71% dan semakin menurun menjadi 8,33%.

Dalam membangun model menggunakan algoritma *decision tree*, model menunjukkan nilai performa yang dapat dikatakan cukup baik bahkan sebelum dilakukan proses *parameter tuning* pada sebagian besar metrik evaluasi. Namun proses *parameter tuning* memaksimalkan performa model klasifikasi menggunakan *decision tree*. Hal ini dapat dilihat bahwa melalui metrik Akurasi yang terjadi peningkatan menjadi 89,29%. Peningkatan ini menunjukkan bahwa adanya peningkatan efektivitas model dalam melakukan prediksi saat dilakukan tuning parameter dibandingkan saat menggunakan nilai parameter *default*. Pada metrik AUC terjadi penurunan signifikan menjadi 21,4% dari 45%. Metrik *Precision* juga mengalami peningkatan meskipun tidak secara signifikan menjadi 93,75% yang menunjukkan adanya sedikit peningkatan pada kemampuan model untuk mengklasifikasikan data dengan benar (Andi, 2023). Peningkatan performa selanjutnya yaitu pada metrik *Recall* menjadi 88,24% yang menunjukkan dimana model menjadi lebih baik dalam mengingat atau melakukan

identifikasi label positif yang ada pada data. Peningkatan performa juga terjadi pada metrik *F1 Score* menjadi 90,91%. Nilai *F1 score* yang lebih tinggi menunjukkan bahwa model klasifikasi telah mengalami perbaikan dalam menjaga keseimbangan yang baik antara metrik *precision* dan *recall* dalam mengidentifikasi kelas positif (Udit, 2022). Selanjutnya *Class Precision (Highly Fit)* menunjukkan peningkatan performa meskipun hanya sedikit menjadi 93,75% yang berarti bahwa terjadi peningkatan model dalam melakukan identifikasi *instance* yang termasuk kedalam kelas *Highly Fit*. *Class Precision (Moderately Fit)* juga menunjukkan peningkatan menjadi 88,33% yang artinya terjadi peningkatan pada model dalam melakukan identifikasi *instance* yang termasuk kedalam kelas *Moderately Fit*. Pada *Class Recall (Highly Fit)* juga terjadi peningkatan performa menjadi 88,24% yang menunjukkan bahwa model lebih baik dalam mengidentifikasi *instance* yang sebenarnya masuk kedalam kelas *Highly Fit*. *Class Recall (Moderately Fit)* menunjukkan stabilitas skor pada 90,91%. Pada metrik *classification error* terjadi penurunan *error* model dalam melakukan klasifikasi menjadi 10,71% atau dapat dikatakan bahwa model menjadi semakin akurat dalam melakukan proses klasifikasi. Hal ini sesuai dengan penelitian yang dilakukan oleh Rizky dkk (2024) dan Mantovani dkk (2024), yang mengatakan bahwa *hyperparameter tuning* akan meningkatkan performa model klasifikasi pada algoritma *decision tree*.

Meskipun demikian, diketahui pula bahwa saat model diterapkan pada data uji terdapat penurunan skor pada beberapa metrik. Untuk penjelasan lebih lanjut yaitu pada metrik akurasi mengalami peningkatan menjadi 91,67% artinya model menjadi sangat baik dalam melakukan klasifikasi pada data. Selanjutnya ada metrik AUC, meskipun terdapat peningkatan skor saat diterapkan ke data uji menjadi 50% yang berarti adanya perbaikan yang berarti dalam model klasifikasi, skor ini dalam AUC termasuk kedalam kategori buruk atau artinya dengan model memiliki kinerja yang hanya setara atau sebatas sebaik pemilihan acak berdasarkan probabilitas kelas (Ahad, Antar, & Ahmed, 2021). Dapat dikatakan pula bahwa model sama saja dengan memilih secara acak antara karyawan yang *Highly Fit* dan *Moderately Fit*. Selanjutnya yaitu pada metrik *F1 score* yang terdapat peningkatan menjadi 93,33%. Lalu peningkatan skor juga terjadi pada *classification error* menjadi 8,33%. Adapun pada metrik *recall*, *class precision (Moderately fit)*, *class recall (Highly Fit)* peningkatan skor saat diterapkan pada data uji menjadi 100%. Penurunan performa pada model menggunakan *decision tree* dapat dilihat pada metrik *precision* menjadi 87,5%, *class precision (Highly Fit)* menjadi 87,5% dan

class recall (Moderately Fit) menjadi 80%. Hal tersebut menunjukkan adanya *trade-off* antara *recall* dan *precision* dimana jika terjadi peningkatan pada *recall*, biasanya akan mengakibatkan penurunan *precision* dan berlaku pula sebaliknya. Hal ini umum dan sangat mungkin terjadi jika ingin melakukan peningkatan salah satu metrik dengan mengorbankan metrik lainnya, meskipun sebenarnya model yang baik akan memiliki peningkatan atau skor yang baik pada kedua metrik tersebut (Hilis & Hoormann, 2016).

Tabel 5.1.2 Perbandingan Performa model Klasifikasi Algoritma *Random Forest*

Matriks Evaluasi	<i>Random Forest</i>	<i>Random Forest Tuning</i>	<i>Performance Data Testing</i>
Akurasi	85%	96,67%	91,67%
AUC	90%	100%	85,7%
<i>Precision</i>	93,33%	100%	87,5%
<i>Recall</i>	80%	94,44%	100%
<i>F1 Score</i>	87,5%	96,67%	93,33%
<i>Class Precision (Highly Fit)</i>	93,33%	100%	87,5%
<i>Class Precision (Moderately Fit)</i>	76,92%	91,67%	100%
<i>Class Recall (Highly Fit)</i>	82,35%	94,12%	100%
<i>Class Recall (Moderately Fit)</i>	90,91%	100%	80%
<i>Classification Error</i>	15%	3,33%	8,33%

Selanjutnya, perbandingan hasil performa model klasifikasi menggunakan algoritma *Random Forest* dapat dilihat pada Tabel 5.1.2. Dapat diketahui bahwa performa dari model klasifikasi menggunakan *Random Forest* sebelum *parameter tuning* sudah sangat baik. Meskipun demikian, setelah dilakukan *parameter tuning*, terdapat kenaikan skor performa model meskipun tidak secara signifikan. Adapun penjelasan mengenai hasil performa model yaitu:

1. Terdapat peningkatan performa metrik akurasi dari 85% menjadi 96,67% dan menurun saat model diterapkan pada *data testing* dengan akurasi hanya sebesar 91,67%.
2. Skor AUC mengalami peningkatan performa dari 90% menjadi 100% dan menurun dengan skor 85,7% pada saat diaplikasikan ke *data testing*.
3. Selanjutnya yaitu *Precision* menunjukkan peningkatan skor sesudah *parameter tuning* menjadi sebesar 100%, namun mengalami penurunan menjadi 87,5% saat diaplikasikan ke *data testing*.
4. Matriks *Recall* memiliki peningkatan skor menjadi sebesar 94,44% sesudah *parameter tuning* dan kembali meningkat menjadi 100% saat diaplikasikan.

5. Berikutnya yaitu metrik *F1 score* terdapat peningkatan menjadi 96,67%, namun terjadi penurunan performa pada *data testing* menjadi 93,33%.
6. Selanjutnya yaitu presisi dari kelas positif (*Highly Fit*) menunjukkan peningkatan performa dari 93,33% menjadi 100% setelah *parameter tuning* dan menurun saat model diaplikasikan kedalam *data testing* menjadi sebesar 87,5%.
7. Pada presisi dari kelas negatif (*Moderately Fit*) terlihat model mengalami peningkatan menjadi 91,67% setelah dilakukan *parameter tuning* dan kembali meningkat menjadi 100% saat diaplikasikan.
8. Metrik *recall* dari kelas positif (*Highly Fit*) mengalami peningkatan menjadi 94,12% setelah dilakukan *parameter tuning* dan kembali meningkat menjadi 100% saat diaplikasikan.
9. Metrik *recall* dari kelas negatif (*Moderately Fit*) yang menunjukkan peningkatan performa menjadi 100% dan menurun menjadi 80% saat diaplikasikan.
10. Terakhir metrik *classification error* memiliki penurunan *error* menjadi 3,33% namun *error* kembali meningkat menjadi 8,33% saat diaplikasikan.

Dalam membangun model menggunakan algoritma *random forest*, model menunjukkan nilai performa yang dapat dikatakan cukup baik bahkan sebelum dilakukan proses *parameter tuning* pada sebagian besar metrik evaluasi. Namun proses *parameter tuning* memaksimalkan performa model klasifikasi menggunakan *random forest*. Hal ini dapat dilihat bahwa melalui metrik Akurasi yang terjadi peningkatan menjadi 96,67%. Peningkatan ini menunjukkan bahwa adanya peningkatan efektivitas model dalam melakukan prediksi saat dilakukan tuning parameter dibandingkan saat menggunakan nilai parameter *default*. Pada metrik AUC terjadi peningkatan signifikan menjadi 100%. Metrik *Precision* juga mengalami peningkatan meskipun tidak secara signifikan menjadi 100% yang menunjukkan adanya kesempurnaan pada kemampuan model untuk mengklasifikasikan data dengan benar (Andi, 2023). Peningkatan performa selanjutnya yaitu pada metrik *Recall* menjadi 94,44% yang menunjukkan dimana model menjadi sangat baik dalam mengingat atau melakukan identifikasi label positif yang ada pada data. Peningkatan performa juga terjadi pada metrik *F1 Score* menjadi 96,67%. Nilai *F1 score* yang lebih tinggi menunjukkan bahwa model klasifikasi telah mengalami perbaikan dalam menjaga keseimbangan yang baik antara metrik *precision* dan *recall* dalam mengidentifikasi kelas positif (Udit, 2022). Selanjutnya *Class Precision (Highly Fit)*

menunjukkan peningkatan performa menjadi 100% yang berarti bahwa terjadi peningkatan model dalam melakukan identifikasi *instance* yang termasuk kedalam kelas *Highly Fit*. *Class Precision (Moderately Fit)* juga menunjukkan peningkatan menjadi 91,67% yang artinya terjadi peningkatan pada model dalam melakukan identifikasi *instance* yang termasuk kedalam kelas *Moderately Fit*. Pada *Class Recall (Highly Fit)* juga terjadi peningkatan performa menjadi 94,12% yang menunjukkan bahwa model lebih baik dalam mengidentifikasi *instance* yang sebenarnya masuk kedalam kelas *Highly Fit*. *Class Recall (Moderately Fit)* menunjukkan adanya peningkatan menjadi 100%. Pada metrik *classification error* terjadi penurunan *error* model dalam melakukan klasifikasi menjadi 3,33% atau dapat dikatakan bahwa model menjadi semakin akurat dalam melakukan proses klasifikasi. Secara garis besar, dari hasil penelitian ini sesuai dengan penelitian yang dilakukan oleh Mantovani dkk (2024) dan Lötsch & Mayer (2022) yang mengatakan bahwa *hyperparameter tuning* akan meningkatkan performa model klasifikasi pada algoritma *random forest*.

Meskipun model memiliki peningkatan performa setelah melalui *parameter tuning*, berdasarkan **Error! Reference source not found.** dapat diketahui bahwa performa model saat diaplikasikan pada *data testing* tidak memiliki peningkatan skor yang berarti pada sebagian besar metrik evaluasi. Metrik akurasi, AUC, *precision*, F1 score, *class precision (highly fit)*, *class recall (moderately fit)* dan pada metrik *classification error* mengalami penurunan performa. Akurasi menurun menjadi 91,67%, *precision* menjadi 87,5%, F1 score menjadi 93,33%, *class precision (highly fit)* 87,5%, *class recall (moderately fit)* menjadi 80% dan pada metrik *classification error* performa menurun dengan tingkat *error* yang semakin meningkat menjadi 8,33%. Metrik AUC pada metode *decision tree* mengalami kenaikan meskipun hanya menjadi 50%. Dengan skor 50% model memiliki kinerja yang hanya setara atau sebatas sebaik pemilihan acak berdasarkan probabilitas kelas (Ahad, Antar, & Ahmed, 2021). Sedangkan pada metode *random forest* penurunan skor AUC menjadi 85,7% masih dapat dikatakan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kelas positif dan negatif. Hal ini sering juga disebut sebagai *model performance mismatch*.

Model performance mismatch merupakan sebuah situasi dimana performa model yang diobservasi pada data pelatihan tidak sejalan dengan performanya pada data uji. Hal ini mencerminkan tantangan generalisasi model, di mana model yang menunjukkan hasil memuaskan saat dilatih, gagal untuk menunjukkan performa yang sama ketika

dihadapkan pada data yang berbeda. Ketidakcocokan antara model dan data yang diuji akan mengakibatkan hasil yang kurang akurat yang berujung pada performa yang tidak optimal (Barbier, Chen, Panchenko, & Sáenz, 2021). Meskipun demikian, metrik *recall*, *class precision (moderately fit)*, dan *class recall (highly fit)* justru mengalami peningkatan performa menjadi 100% saat diterapkan pada data uji. Hal tersebut menunjukkan adanya *trade-off* antara *recall* dan *precision* dimana jika terjadi peningkatan pada *recall*, biasanya akan mengakibatkan penurunan *precision* dan berlaku pula sebaliknya. Hal ini umum dan sangat mungkin terjadi jika ingin melakukan peningkatan salah satu metrik dengan mengorbankan metrik lainnya, meskipun sebenarnya model yang baik akan memiliki peningkatan atau skor yang baik pada kedua metrik tersebut (Hilis & Hoormann, 2016). Dalam model klasifikasi ini, baik *recall* maupun *precision* tergolong sangat baik meskipun terjadi kasus *trade-off* dengan perbandingan skor yang tidak terlalu signifikan dan berada di atas 80%. Sehingga masih dapat dikatakan bahwa model klasifikasi menggunakan *random forest* ini sudah memiliki performa yang sangat baik dalam melakukan klasifikasi.

Masalah *model performance mismatch* yang paling terlihat terjadi pada algoritma *random forest*. Hal ini dapat disebabkan karena beberapa hal yaitu *imbalanced dataset* yang disebabkan oleh kurangnya proses belajar model pada kelas minoritas (Kaope & Pristyanto, 2023) dan *overfitting* (Rice, Wong, & Kolter, 2020). Mengacu pada Gambar 4.4.1 Persentase Kelas Data *Person-Job Fit*, kelas *Moderately Fit* memiliki rasio sebesar 40% yang masuk kedalam kategori *mild* atau ringan. Hal ini dilihat berdasarkan Tabel 5.1.3.

Tabel 5.1.3 Derajat Ketidakseimbangan

<i>Degree of imbalance</i>	<i>Proportion of Minority Class</i>
<i>Mild</i>	20 - 40% dari total data
<i>Moderate</i>	1 - 20% dari total data
<i>Extreme</i>	< 1% dari total data

Sumber: (Google for Developers, 2023)

Namun, meskipun derajat ketidakseimbangan pada penelitian ini masih tergolong ringan, kecilnya ukuran data yang digunakan tampaknya memengaruhi kinerja model klasifikasi pada penelitian ini (Johnson & Khoshgoftaar, 2019). Hal tersebut dapat terlihat karena

terdapat kasus *model performance mismatch*. Kecilnya data ini juga memungkinkan terjadinya kasus *overfitting*. *Overfitting* akan terjadi ketika model mempelajari data dengan terlalu baik yang mengakibatkan model tidak dapat digeneralisasikan dengan data baru (Trivedi, Bhatt, & Srivastava, 2021). *Overfitting* dapat disebabkan oleh banyak hal, beberapa diantaranya yang terlihat dari penelitian ini yaitu kecilnya data *training* yang digunakan untuk membangun model karena sedikitnya data yang dapat digunakan oleh model untuk “belajar” pola yang ada didalam data. Selain itu, data dengan tingkat korelasi yang tinggi juga menjadi salah satu penyebab munculnya masalah *overfitting* pada penelitian ini karena model akan ikut mempelajari hubungan yang spesifik atau *correlated noise* dalam data yang menyebabkan data tidak akan dapat digeneralisasikan pada data baru (Zhu, Lévy-Leduc, & Ternès, 2021).

Selain itu, data yang digunakan pada penelitian ini merupakan hasil kombinasi data asli dengan data augmentasi yang merupakan sebuah proses untuk membuat data baru dari data yang sudah ada dan secara artifisial menambahkan jumlah data dengan tujuan untuk membantu model dalam mempelajari pola dalam data (Nanthini, et al., 2023). Data yang diperoleh secara artifisial oleh algoritma mungkin tidak dapat menangkap hubungan atau pola yang tersembunyi dari data asli, terlebih data asli yang digunakan pada penelitian berasal dari hasil tes psikolog yang cenderung muncul keterkaitan antar aspek-aspeknya (Fenn, Tan, & George, 2020). Mengingat semua hal yang terkait dengan manusia merupakan yang hal yang kompleks dan tidak ada manusia yang sama satu sama lainnya, hasil tes psikolog juga membawa sifat karakteristik yang kuat dari setiap orang yang berisi pola-pola yang rumit, unik dan tersembunyi yang tidak dapat ditangkap oleh mesin (Sanbonmatsu, Cooley, & Butner, 2021). Sehingga sangat besar kemungkinan model klasifikasi tidak dapat bekerja secara maksimal karena pada saat proses augmentasi data tidak dapat menangkap kerumitan pada data asli. Melalui penelitian ini, dapat terlihat bahwa algoritma *decision tree* sudah sangat mumpuni dalam membangun model klasifikasi *Person-Job Fit* dikarenakan memiliki hasil performa pada *data testing* yang hampir sama baiknya saat diterapkan ke data uji. Sedangkan pada algoritma *random forest* performa menggunakan parameter standar terlihat sudah sangat baik, dilakukannya *parameter tuning* mungkin menyebabkan model menjadi terlalu banyak belajar dan terlalu menyesuaikan diri pada data tersebut sehingga muncul kasus *overfitting* pada algoritma *Random Forest*.

5.2 Perbandingan Performa Model Klasifikasi

Adapun perbandingan performa model klasifikasi dengan *Decision Tree* dan *Random Forest* dapat dilihat pada

Tabel 5.2.1 Perbandingan Performa Model

Matriks Evaluasi			<i>Parameter Tuning</i>		<i>Data Testing</i>	
	<i>Decision Tree</i>	<i>Random Forest</i>	<i>Decision Tree</i>	<i>Random Forest</i>	<i>Decision Tree</i>	<i>Random Forest</i>
Akurasi	85%	85%	89,29%	96,67%	91,67%	91,67%
AUC	45%	90%	21,4%	100%	50%	85,7%
<i>Precision</i>	93,33%	93,33%	93,75%	100%	87,5%	87,5%
<i>Recall</i>	80%	80%	88,24%	94,44%	100%	100%
<i>F1 Score</i>	87,5%	87,5%	90,91%	96,67%	93,33%	93,33%
<i>Class Precision (Highly Fit)</i>	93,33%	93,33%	93,75%	100%	87,5%	87,5%
<i>Class Precision (Moderately Fit)</i>	76,92%	76,92%	83,33%	91,67%	100%	100%
<i>Class Recall (Highly Fit)</i>	82,35%	82,35%	88,24%	94,12%	100%	100%
<i>Class Recall (Moderately Fit)</i>	90,91%	90,91%	90,91%	100%	80%	80%
<i>Classification Error</i>	15%	15%	10,71%	3,33%	8,33%	8,33%

Berdasarkan perbandingan yang tunjukkan pada Tabel 5.2.1, dapat diketahui bahwa kedua model memiliki akurasi yang relatif serupa yaitu sebesar 85%. Namun, setelah dilakukan *parameter tuning*, model dengan *random forest* mengalami peningkatan yang cukup signifikan menjadi 96,67%, sedangkan kenaikan model dengan *decision tree* hanya mencapai 91,67% yang berarti *random forest* lebih efektif dalam melakukan klasifikasi. Pada metrik AUC, *random forest* memiliki skor yang jauh lebih baik sebesar 90% dibandingkan *decision tree* yang hanya sebesar 45%. Setelah dilakukan *parameter tuning* pun *random forest* memiliki peningkatan menjadi 100% dibandingkan *decision tree* yang justru semakin menurun dengan 21,4%. Hal ini menunjukkan *random forest* lebih baik dalam membedakan kelas yang positif dan negatif. Pada metrik *precision* kedua model memiliki skor yang hampir sama, namun setelah *parameter tuning*, *random forest* mencapai 100% sedangkan *decision tree* hanya sampai 87,5% yang berarti bahwa *random forest* memiliki kinerja yang lebih baik dalam mengklasifikasikan data dengan benar. Pada metrik *recall*, setelah dilakukan *parameter tuning*, *random forest* memiliki skor yang lebih baik sebesar 94,44% dibandingkan dengan *decision tree* sebesar 80% yang berarti

random forest lebih baik dalam mengidentifikasi semua *instance* atau label yang positif. *Random forest* juga memiliki *F1 score* yang lebih tinggi setelah dilakukan *parameter tuning* sebesar 96,67% dibandingkan dengan *decision tree* yang hanya sebesar 87,5% yang berarti *random forest* lebih baik menjaga keseimbangan antara *precision* dan *recall*. Pada metrik klasifikasi pada kelas *highly fit* dan *moderately fit*, dapat terlihat bahwa *random forest* memiliki performa yang lebih baik dalam membedakan kelas data. Pada metrik *classification error*, *random forest* juga memiliki tingkat kesalahan yang lebih rendah yaitu sebesar 3,33% dibandingkan dengan *decision tree* dengan tingkat kesalahan sebesar 8,33%. Disisi lain, pengujian model menggunakan *data testing* memiliki performa yang serupa pada Sebagian besar metrik evaluasi. Namun, *random forest* unggul dalam performa AUC sebesar 85,7% dibandingkan *decision tree* yang hanya sebesar 50%. Dari penjelasan tersebut, untuk kasus *person-job fit* dengan 40 data dan 27 atribut pada penelitian ini, *random forest* merupakan metode terbaik yang dapat dipilih dikarenakan performanya yang lebih baik secara keseluruhan.

5.3 Analisis Perbaikan

Permasalahan yang dialami PT PNR adalah belum adanya pemanfaatan dengan maksimal data hasil psikotes dan gejala yang menunjukkan bahwa adanya tantangan dalam memilih karyawan yang memiliki potensi yang sesuai dengan kebutuhan perusahaan. Oleh karena itu, dilakukanlah penelitian ini dengan salah satu fungsi dan algoritma dari *machine learning* yaitu klasifikasi dengan bantuan algoritma *decision tree* dan *random forest*. Penelitian ini menggunakan data historis hasil tes psikologi dan nilai performa karyawan Divisi PL PT PNR dengan tahun masuk perusahaan antara 2015 hingga 2019 yang berjumlah 18 data dengan total 33 atribut. Dari 33 atribut tersebut, 1 merupakan target dan 32 sisanya merupakan prediktor. Teknik augmentasi juga dilakukan karena kecilnya ukuran data sehingga diperoleh total data keseluruhan sebanyak 40 data, hal ini dapat juga dilihat dengan lebih rinci pada **Error! Reference source not found.** Sedangkan untuk data yang akan digunakan dalam pembangunan model klasifikasi dapat dilihat pada Tabel 4.3.3 Data Hasil *Cleaning* dan diketahui bahwa variable akhir yang akan digunakan yaitu sebanyak 28 variabel dengan 1 label atau target dan 27 atribut regular. Hasil yang dapat diperoleh dari pembangunan model klasifikasi dapat terlihat pada Tabel 5.2.1 Perbandingan Performa Model yang menunjukkan bahwa dalam kasus dalam penelitian

ini, *random forest* merupakan algoritma yang memiliki performa terbaik pada sebagian besar metrik evaluasi.

Penelitian ini dilakukan sebagai langkah awal untuk memaksimalkan pemanfaatan data hasil psikotes dan sebagai langkah preventif serta antisipatif dalam upaya mengurangi jumlah kasus ketidakcocokan karyawan dalam proses rekrutmen. Hasil dari pemanfaatan hasil psikotes ini yaitu diketahuinya variabel yang penting dalam seleksi rekrutmen. Pada Tabel 4.3.2 *Cleaning Data*, terdapat 3 variabel yang tergolong sebagai dasar kemampuan yang sangat penting dalam dunia pekerjaan. Dalam penelitian ini berdasarkan hasil data psikotes, diketahui bahwa variabel komitmen terhadap tugas, kerjasama dan kemampuan komunikasi memiliki nilai yang sama pada 18 data yang ada. Hal ini berarti 3 variabel tersebut merupakan kemampuan dasar atau standar yang wajib dimiliki oleh semua calon karyawan untuk keberhasilan di tempat kerja (Lang, Dony, & Roberts, 2022; Friedman, 2024; Jain, 2023). Namun, variabel tersebut tidak digunakan dalam membangun model klasifikasi *person-job fit* dikarenakan kurangnya variasi pada *instance*. Meskipun demikian, variabel tersebut tetap akan dipertimbangkan sebagai dasar kemampuan yang wajib dimiliki oleh seluruh calon karyawan.

Penelitian ini juga membantu untuk mengetahui variabel penting selain ketiga variabel tersebut yang dapat dilihat dari hasil Gambar 4.7.1 *Features Importance* yang menunjukkan bahwa variabel fleksibilitas berpikir merupakan salah satu variabel penting yang paling informatif yang patut untuk mengetahui kecocokan karyawan dengan pekerjaannya. Selain itu, Gambar 4.7.2 *Pareto Chart* Fitur terhadap *Person-Job Fit* juga menunjukkan variabel asal sekolah yang memiliki konsekuensi terbesar dalam mempengaruhi hasil memprediksi kelas *person-job fit*. Dari pemaparan tersebut, dapat diketahui pentingnya memanfaatkan variabel potensi yang ada dalam hasil psikotes. Tanpa psikotes, perusahaan akan menghadapi resiko memilih kandidat yang tidak memiliki spesifikasi yang sesuai dengan kebutuhan pekerjaan, yang nantinya berdampak negatif pada kinerja dan produktivitas dalam tim (Appsensi, 2022). Kandidat yang tidak sesuai dengan spesifikasi yang ada juga lebih mungkin untuk meninggalkan perusahaan dalam waktu yang singkat dan akan meningkatkan *turnover* perusahaan (Idris, 2020) yang juga akan berimbas pada meningkatnya pengeluaran pada proses rekrutmen dan pelatihan karyawan baru (Appsensi, 2022).

Kedepannya jika tersedia jumlah data dengan jumlah yang lebih besar untuk membangun model prediksi *Person-Job Fit*, akan sangat disarankan menggunakan

algoritma yang lebih kompleks dan lebih sesuai seperti *XGBoost* karena memiliki parameter yang mampu mengatasi data yang rentan akan masalah *overfitting* seperti data hasil psikotes karyawan atau PCA (*Principal Component Analysis*) yang mampu menangani data berkorelasi tinggi. Penelitian dan analisis lebih mendalam juga diperlukan jika ingin melakukan penelitian menggunakan data hasil psikotes ini dikarenakan data tes psikologi merupakan data yang rumit dengan tingkat individualisme yang tinggi yang melambangkan perbedaan pada setiap orang, penting untuk mencari solusi dari masalah yang mungkin muncul dalam data seperti korelasi yang tinggi antar fitur.

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan hasil dan analisis yang sudah dipaparkan, berikut kesimpulan yang diperoleh dari penelitian:

1. Adapun untuk hasil performa model klasifikasi dengan *tree algorithm* yaitu sebagai berikut:

Algoritma Decision Tree: Akurasi 91,67%, AUC 50%, *Precision* 87,5%, *Recall* 100%, *F1 Score* 93,33%, *Class Precision (Highly Fit)* 87,5%, *Class Precision (Moderately Fit)* 100%, *Class Recall (Highly Fit)* 100%, dan *Class Recall (Moderately Fit)* 80%, dan *Classification Error* sebesar 8,33%.

Algoritma Random Forest: Akurasi 91,67%, AUC 85,7%, *Precision* 87,5%, *Recall* 100%, *F1 Score* 93,33%, *Class Precision (Highly Fit)* 87,5%, *Class Precision (Moderately Fit)* 100%, *Class Recall (Highly Fit)* 100%, *Class Recall (Moderately Fit)* 80%, dan *Classification Error* sebesar 8,33%.

2. Berdasarkan hasil *features importance*, adapun fitur yang paling informatif dalam model klasifikasi *Person-Job Fit* yaitu fleksibilitas berpikir. Selain itu, variable asal sekolah memiliki konsekuensi terbesar dalam mempengaruhi hasil memprediksi kelas *person-job fit*.

6.2 Saran

Saran yang dapat diberikan untuk pengembangan dimasa mendatang yaitu:

1. Permasalahan yang muncul pada penelitian ini adalah kecilnya jumlah data bagi model untuk “berlatih”, maka diperlukan sebuah langkah untuk meningkatkan kemampuan model dalam melakukan prediksi pada data baru dengan menambah input data dari karyawan yang akan digunakan sebagai *data training* dengan memperluas jangkauan rentang masa kerja.
2. Diperlukan tahap yang lebih memperhatikan kasus *imbalanced* pada kelas data seperti menggunakan teknik SMOTE (*Synthetic Minority Oversampling Technique*), *Random Over Sampling*, *Borderline SMOTE*, *Kmeans SMOTE* dan metode serupa lainnya.

3. Untuk menangani data berkorelasi tinggi seperti hasil tes psikologi, diperlukan analisis yang lebih mendalam dan sebuah metode yang lebih tepat untuk digunakan seperti PCA (*Principal Component Analysis*).

DAFTAR PUSTAKA

- Abbasi, S., Tahir, M., Abbas, M., & Shabbir, M. (2020). Examining the relationship between recruitment & selection practices and business growth: An exploratory study. *Public Affairs*, 22(2). doi:<https://doi.org/10.1002/pa.2438>
- Abenov, T., Grabbert, T., Franklin-Hensler, M., & Larrat, T. (2023, Februari 14). Has mining lost its luster? Why talent is moving elsewhere and how to bring them back. *Metals & Mining*. Retrieved from <https://www.mckinsey.com/industries/metals-and-mining/our-insights/has-mining-lost-its-luster-why-talent-is-moving-elsewhere-and-how-to-bring-them-back>
- Ahad, M., Antar, A., & Ahmed, M. (2021). Performance Evaluation in Activity Classification: Factors to Consider. Dalam *IoT Sensor-Based Activity Recognition* (Vol. 173, hal. 133-147). Springer, Cham. doi:https://doi.org/10.1007/978-3-030-51379-5_8
- Ali, B., & Anwar, G. (2021). Employee Turnover Intention and Job Satisfaction. *International Journal of Advanced Engineering, Management and Science*, 7(6), 22–30. doi:<https://doi.org/10.22161/ijaems.76.3>
- Andi. (2023, Agustus 12). *Memahami dan Menerapkan Matriks Evaluasi ROC-AUC dalam Machine Learning*. Retrieved from Medium: <https://medium.com/@andimrinaldisaputraa/memahami-dan-menerapkan-matriks-evaluasi-roc-auc-dalam-machine-learning-4468e5fcb9a>
- Anonim. (2015). *Principles of Management*. University of Minnesota Libraries Publishing. doi:<https://doi.org/10.24926/8668.1801>
- Ardhani, H., & Indriati, F. (2023). Employee Perceptions on the Implementation of Talent Management at PT Pegadaian Office Region VIII Jakarta. *Journal of Social Research*, 2(11). doi:<https://doi.org/10.55324/josr.v2i11.1495>
- Astuti W., & Amalia, L. (2021). The Relationship between Work Motivation, Job Satisfaction, and Employee Performance: The Moderating Role of Psychology Capital and The Mediating Role of Organizational Commitment. *Jurnal*

- Manajemen Teori dan Terapan*, 14(2), 103-128.
doi:<https://doi.org/10.20473/jmtt.v14i2.26192>
- Barbier, J., Chen, W., Panchenko, D., & Sáenz, M. (2021). PERFORMANCE OF BAYESIAN LINEAR REGRESSION IN A MODEL WITH MISMATCH. *Mathematics*. doi:<https://doi.org/10.48550/arXiv.2107.06936>
- Beaumont-Oates, W. (2023, Juni 23). Characteristics of high potential employees. *Succession Planning*. Retrieved from <https://www.thomas.co/resources/type/hr-blog/characteristics-high-potential-employees>
- Biswas, S., Fole, A., Khare, N., & Agrawal, P. (2023). Enhancing correlated big data privacy using differential privacy and machine learning. *Journal of Big Data*. doi:<https://doi.org/10.1186/s40537-023-00705-8>
- Blockeel, H., Devos, L., Frénay, B., Nanack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. *Front. Artif. Intell.*, 6. doi:<https://doi.org/10.3389/frai.2023.1124553>
- Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends (JASTT)*, 2(1), 20 - 28. doi:<https://doi.org/10.38094/jastt20165>
- Chen, Q., Zhang, A., & Pan, G. (2023). A maximum-entropy-attention-based convolutional neural network for image perception. *Neural Computing and Applications*, 35, 8647–8655. Retrieved from <https://link.springer.com/article/10.1007/s00521-022-07564-z>
- Chen, X., Ran, L., Zhang, Y., Yang, J., Yao, H., Zhu, S., & Tan, X. (2019). Moderating role of job satisfaction on turnover intention and burnout among workers in primary care institutions: a cross-sectional study. *BMC Public Health*, 19, 1526. doi:<https://doi.org/10.1186/s12889-019-7894-7>
- Chi, N. (2014). Person-Job Fit, Transformational Leadership, and Job Performance. In A. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (1 ed., pp. 4778–4781). Springer Dordrecht. doi:https://doi.org/10.1007/978-94-007-0753-5_4051
- Chuang, A., Shen, C., & Judge, T. (2015). Development of a Multidimensional Instrument of Person–Environment Fit: The Perceived Person–Environment Fit Scale (PPEFS). *Applied Psychology*, 65(1), 66-98. doi:<https://doi.org/10.1111/apps.12036>

- Dachi, J., & Sitompul, P. (2023, Oktober). Analisis Perbandingan Algoritma XGBoost dan Algoritma. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam (JURRIMIPA)*, 2(2), 87-103. doi:<https://doi.org/10.55606/jurrimipa.v2i2.1470>
- De Cooman, R., & Vleugels, W. (2022). Person–Environment Fit: Theoretical Perspectives, Conceptualizations, and Outcomes. *Business and Management*. doi:<https://doi.org/10.1093/acrefore/9780190224851.013.377>
- Edwards, J. (2008). Person–Environment Fit in Organizations: An Assessment of Theoretical Progress. *Academy of Management Annals*, 2(1), 167-230. doi:<https://doi.org/10.5465/19416520802211503>
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., . . . Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164, 102–111. doi:<https://doi.org/10.1016/j.enconman.2018.02.087>
- Fenn, J., Tan, C., & George, S. (2020). Development, validation and translation of psychological tests. *BJPsych Advances*, 26(5), 306 - 315. doi:<https://doi.org/10.1192/bja.2020.33>
- Fernandes França , T., São Mamede, H., Manuel Pereira Barroso, J., & Manuel Pereira Duarte dos Santos, V. (2023). Artificial intelligence applied to potential assessment and talent identification in an organisational context. *Heliyon*, 9(4). doi:<https://doi.org/10.1016/j.heliyon.2023.e14694>
- Fernández-Aráoz, C., Groyberg, B., & Nohria, N. (2011, Oktober). How to Hang On to Your High Potentials. Retrieved from <https://hbr.org/2011/10/how-to-hang-on-to-your-high-potentials>
- Fife, D., & D’Onofrio, J. (2023). Common, uncommon, and novel applications of random forest in psychological research. *Behavior Research Methods*, 55, 2447–2466. doi:<https://doi.org/10.3758/s13428-022-01901-9>
- Flatau-Harrison, H., Wilson, M., & Vleugels, W. (2023). The reciprocal relationship between safety behaviour and person-job fit: A self-regulation perspective. *Safety Science*, 168. doi:<https://doi-org.proxy.undip.ac.id/10.1016/j.ssci.2023.106289>
- Goetz, N., & Wals, A. (2022). Similar but different? The influence of job satisfaction, organizational commitment and person-job fit on individual performance in the continuum between permanent and temporary organizations. *International*

- Journal of Project Management*, 40(3), 251-261.
doi:<https://doi.org/10.1016/j.ijproman.2022.03.001>
- Google for Developers. (2023, 06 09). *Foundational Course - Imbalanced Data*. Retrieved from Google for Developers: <https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data>
- Grailey, K., Murray, E., Reader, T., & Brett, S. (2021). The presence and potential impact of psychological safety in the healthcare setting: an evidence synthesis. *BMC Health Services Research*. doi:<https://doi.org/10.1186/s12913-021-06740-6>
- Green, D. (2022, February 24). Talent at a turning point: How people analytics can help. *People & Organizational Performance Practice*. (B. Hancock, & B. Schaninger, Interviewers) McKinsey & Company. Retrieved November 9, 2023, from <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/talent-at-a-turning-point-how-people-analytics-can-help#/>
- Guo, Y., & Hou, X. (2022). The effects of job crafting on tour leaders' work engagement: the mediating role of person-job fit and meaningfulness of work. *International Journal of Contemporary Hospitality Management*, 34(5), 1649-1667. doi:<https://doi.org/10.1108/IJCHM-09-2021-1082>
- Hilis, S., & Hoormann, S. (2016, Januari 20). *Precision and Recall: Understanding the Trade-Off*. Retrieved from Medium: <https://medium.com/opex-analytics/why-you-need-to-understand-the-trade-off-between-precision-and-recall-525a33919942>
- Iqbal, N., Khan, M., Mohmand, Y., & Mujtaba, B. (2020). The Impact of in-Service Training and Motivation on Job Performance of Technical & Vocational Education Teachers: Role of Person-Job Fit. *Public Organization Review*, 20, 529-548. doi:<https://doi.org/10.1007/s11115-019-00455-3>
- Johnson, J., & Khoshgoftaar, T. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6, article number 27. Retrieved from <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0192-5>
- Ju, D., Yao, J., & Ma, I. (2021). Person–Job Fit And Job Involvement: The Curvilinear Effect And The Moderating Role Of Goal Orientation. *Journal of Managerial Psychology*, 36(5), 433-446. doi:<https://doi.org/10.1108/JMP-02-2020-0095>

- Kabalina, V., & Osipova, A. (2022). Identifying and assessing talent potential for future needs of a company. *Journal of Management Development*, 3, 147-162. doi:<https://doi.org/10.1108/JMD-11-2021-0319>
- Kaope, C., & Pristyanto, Y. (2023). The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 22(2), 227-238. doi:<https://doi.org/10.30812/matrik.v22i2.2515>
- Kaur, H., & Kaur, R. (2023). Longitudinal effects of high-performance work practices on job performance via person–job fit. *The Bottom Line*, 36(2), 161-180. doi:<https://doi.org/10.1108/BL-02-2022-0030>
- Kemnaker. (2019). *SKKNI Kemnaker*. Retrieved from Kemnaker: <https://skkni.kemnaker.go.id/tentang-skkni/dokumen?page=16&limit=20>
- Khan, F., Kanwal, S., Alamri, S., & Mumtaz, B. (2020). Hyper-Parameter Optimization of Classifiers, Using an Artificial Immune Network and Its Application to Software Bug Prediction. *IEEE Access*, 8, 20954 - 20964. doi:<https://doi.org/10.1109/ACCESS.2020.2968362>
- Kim, T., Schuh, S., & Cai, Y. (2019, Agustus 20). Change in Person-Job Fit and Its Impact on Employee Work Attitudes over Time. *CEIBS Knowledge*. Retrieved from <https://www.ceibs.edu/new-papers-columns/change-person-job-fit-and-its-impact-employee-work-attitudes-over-time>
- Kim, T., Schuh, S., & Cal, Y. (2019). Person or job? Change in person-job fit and its impact on employee work attitudes over time. *Journal of Management Studies*, 57(2), 287-313. doi:<https://doi.org/10.1111/joms.12433>
- Kristof-Brown, A., & Guay, R. (2011). Person–environment fit. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 3, pp. 3-50). American Psychological Association. doi:<https://psycnet.apa.org/doi/10.1037/12171-001>
- Lötsch, J., & Mayer, B. (2022). A Biomedical Case Study Showing That Tuning Random Forests Can Fundamentally Change the Interpretation of Supervised Data Structure Exploration Aimed at Knowledge Discovery. *BioMedInformatics*, 2(4), 544-552. doi:<https://doi.org/10.3390/biomedinformatics2040034>
- Mantovani, R., Horváth, T., Rossi, A., Cerri, R., Barbon Junior, S., Vanschoren, J., & de Carvalho, A. (2024). Better trees: an empirical study on hyperparameter tuning

- of classification decision tree induction algorithms. *Data Mining and Knowledge Discovery*. doi:<https://doi.org/10.1007/s10618-024-01002-5>
- Mona, E., Syukri, A., & Risnita. (2022). The Relationship Of Employee Potential, Employee Competence And Work Motivation With Performance Of Employees In The Public Works And People's Housing Department Jambi Province. *International Journal of Economics and Management Research*, 1(2), 194-208. Retrieved from <https://ijemr.politeknikpratama.ac.id/index.php/ijemr/article/view/49>
- Montero Guerra, J., Danvila-del-Valle, I., & Méndez-Suárez, M. (2023). The impact of digital transformation on talent management. *Technological Forecasting and Social Change*, 188. doi:<https://doi.org/10.1016/j.techfore.2022.122291>
- Motaghifard, A., Omidvari, M., & Kazemi, A. (2023). Forecasting of safe-green buildings using decision tree algorithm: data mining approach. *Environment, Development and Sustainability*, 25, 10323–10350. doi:<https://doi.org/10.1007/s10668-022-02491-4>
- Nanthini, K., Sivabalaselvamani, D., Chitra, K., Gokul, P., KavinKumar, S., & Kishore, S. (2023). A Survey on Data Augmentation Techniques. *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, 913-920. doi:10.1109/ICCMC56507.2023.10084010
- Nurtjahjono, G., Nimran, U., Al Musadieg, M., & Utami, H. (2020). The Effect of Job Characteristic, Person-Job Fit, Organizational Commitment on Employee Performance (Study of East Java BPJS Employees). *Journal of Public Administration Studies*, 5(1). doi:<https://doi.org/10.21776/ub.jpas.2020.005.01.2>
- Pranto, B. (2020, Januari 3). *Entropy Calculation, Information Gain & Decision Tree Learning*. Retrieved Januari 2024, from Medium: <https://medium.com/analytics-vidhya/entropy-calculation-information-gain-decision-tree-learning-771325d16f>
- Prasetyowati, M., Maulidevi, N., & Surendro, K. (2021). Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *Journal of Big Data*, 8, article number 84. Retrieved from <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00472-4>
- Qin, C., Zhu, H., Xu, T., Zhu, C., Ma, C., Chen, E., & Xiong, H. (2020). An Enhanced Neural Network Approach to Person-Job Fit in Talent Recruitment. *ACM*

Transactions on Information Systems, 38(2), 1 - 33.
doi:<https://doi.org/10.1145/3376927>

- Rajper, Z., Ghumro, I., & Mangi, R. (2020). Rajper, Z.A., Ghumro, I.A., & Mangi, R.A. (2020). The impact of person job fit and person organization fit on employee job performance: A study among employees of services sector. *Abasyn Journal of Social Sciences*. Rajper, Z.A., Ghumro, I.A., & Mangi, R.A. (2020). *The impact of person job fit and person organization fit on employee job performance: A study among employees of services sector. Abasyn Journal of Social Sciences*. doi:<https://doi.org/10.34091/jass.13.1.05>
- Rebele, R. (2019). Can We Really Test People for Potential? *People Analytics*. Retrieved November 15, 2023, from <https://sloanreview.mit.edu/article/can-we-really-test-people-for-potential/>
- Rice, L., Wong, E., & Kolter, Z. (2020). Overfitting in adversarially robust deep learning. *Proceedings of the 37th International Conference on Machine Learning*, 8093-8104. Retrieved from <https://proceedings.mlr.press/v119/rice20a>
- Rizky, M., Faisal, M., Budiman, I., Kartini, D., & Abadi, F. (2024). Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithm for Software Defect Prediction. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 18(1), 95-106. doi:<https://doi.org/10.22146/ijccs.90437>
- Robledo-Ardila, C., & Roman-Calderon, J. (2020). Individual potential and its relationship with past and future performance. *Journal of Education for Business*, 95(1), 45-52. doi:<https://doi.org/10.1080/08832323.2019.1596872>
- Rusconi, G. (2023, Mei 25). Skill vs Capability vs Competency: Understanding the Differences. Retrieved from <https://cloudassess.com/blog/skill-vs-capability-vs-competency/>
- Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3, article number 272. doi:10.1007/s42452-021-04148-9
- Sanbonmatsu, D., Cooley, E., & Butner, J. (2021). The Impact of Complexity on Methods and Findings in Psychological Science. *Frontiers in Psychology*, 11. doi:<https://doi.org/10.3389/fpsyg.2020.580111>

- Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2. doi:<https://doi.org/10.1007/s42979-021-00592-x>
- Sartori, R., Costantini, A., & Ceschi, A. (2020). Psychological assessment in human resource management: discrepancies between theory and practice and two examples of integration. *Personnel Review*, 51(1), 284-298. doi:<https://doi.org/10.1108/PR-05-2019-0281>
- Schnöoll, M., Ferner, C., & Wegenkittl, S. (2019). The Effectiveness of the Max Entropy Classifier for Feature Selection. *Data Science – Analytics and Applications*, 37–39. Retrieved from https://link.springer.com/chapter/10.1007/978-3-658-27495-5_4
- Schweimer, C., & Scher, S. (2022). Quantifying probabilistic robustness of tree-based classifiers against natural distortions. doi:<https://doi.org/10.48550/arXiv.2208.10354>
- Shekar, B., & Dagneu, G. (2019). Grid search-based hyperparameter tuning and classification of microarray cancer data. In *Proceedings of the Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. doi:<https://doi.org.ezproxy.ugm.ac.id/10.1109/ICACCP.2019.8882943>
- Shu, W., Yan, Z., & Qian, W. (2023). Information gain-based semi-supervised feature selection for hybrid data. *Applied Intelligence*, 7310–7325. Retrieved from <https://link.springer.com/article/10.1007/s10489-022-03770-353>
- Solorio-Ramírez, J.-L., Jiménez-Cruz, R., Villuendas-Rey, Y., & Yáñez-Márquez, C. (2023). Random forest Algorithm for the Classification of Spectral Data. *Algorithms*, 16(6), 293. doi:<https://doi.org/10.3390/a16060293>
- Sørli, H., Hetland, J., Bakker, A., Espevik, R., & Olsen, O. (2022). Daily autonomy and job performance: Does person-organization fit act as a key resource? *Journal of Vocational Behavior*, 133. doi:<https://doi.org/10.1016/j.jvb.2022.103691>
- Stephanie, C., & Sarno, R. (2019). Classification Talent of Employee Using C4.5, KNN, SVM. *2019 International Conference on Information and Communications Technology (ICOIACT)*. doi:<https://doi.org/10.1109/ICOIACT46704.2019.8938508>

- Sugiyono. (2006). *Metode Penelitian Kuantitatif, Kualitatif dan R & D*. Bandung: Alfabeta.
- Susanto, P., Parmenas, N., Suryawan, R., & Apriyani, I. (2023). Determinant Attitude and Employee Recruitment: Analysis Psikotest, Assessment, Behavioral Event Interview and Experience (Study Literature). *I*. doi:<https://doi.org/10.38035/ijphs.v1i1>
- Tantithamthavorn, C., McIntosh, S., Hassan, A., & Matsumoto, K. (2019). The impact of automated parameter optimization on defect prediction. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, 45(7), 683–771. doi:10.1109/TSE.2018.2794977
- Trivedi, U., Bhatt, M., & Srivastava, P. (2021). Prevent Overfitting Problem in Machine Learning: A Case Focus on Linear Regression and Logistics Regression. *Innovations in Information and Communication Technologies (IICT-2020)*, 345-349. doi:https://doi.org/10.1007/978-3-030-66218-9_40
- Tsai, P., Kao, Y., & Kuo, S. (2023). Exploring the critical factors influencing the outlying island talent recruitment and selection evaluation model: Empirical evidence from Penghu, Taiwan. *Evaluation and Program Planning*, 99. doi:<https://doi.org/10.1016/j.evalprogplan.2023.102320>
- Udit. (2022, Desember 31). *Navigating the Precision-Recall Tradeoff: Understanding F1 Score*. Retrieved from Medium: <https://itsudit.medium.com/navigating-the-precision-recall-tradeoff-understanding-f1-score-544d481f1393>
- Vanneschi, L., & Silva, S. (2023). Decision Tree Learning. *Lectures on Intelligent Systems*, 149-159. doi:https://doi.org/10.1007/978-3-031-17922-8_6
- Wang, Z., Wei, W., Xu, C., Xu, J., & Mao, X. (2022). Person-job fit estimation from candidate profile and related recruitment history with co-attention neural network. *Neurocomputing*, 501, 14-24. doi:<https://doi.org/10.1016/j.neucom.2022.06.012>
- Wenehenubun, S. (2023). Analisis Peran Person Job-Fit Terhadap Human Resources Development Practices Dalam Meningkatkan Employee Job Performance. *Jurnal Humanipreneur*, 2(1), 1-7. doi:<https://doi.org/10.53091/hum.v2i1.26>
- Wulandari, W. (2021). Linking Person Job Fit, Person Organization Fit and Organizational Culture to Employee Performance in Islamic Banks: the Mediating Role of Job Motivation. *Journal of Islamic Economic and Business Research*, 1(2), 125-139. doi:<https://doi.org/10.18196/jiebr.v1i2.17>

- Xiao, Y., Dong, M., Shi, C., Zheng, W., Shao, Z., Xie, H., & Li, G. (2021). Person–environment fit and medical professionals’ job satisfaction, turnover intention, and professional efficacy: A cross-sectional study in Shanghai. *PLoS ONE*, *16*(4). doi:<https://doi.org/10.1371/journal.pone.0250693>
- Yan, X., Chen, L., & Guo, G. (2022). Kernel-based data transformation model for nonlinear classification of symbolic data. *Soft Comput*, *26*, 1249–1259. doi:<https://doi.org/10.1007/s00500-021-06600-9>
- Ye, Y., Zhu, H., Xu, T., Zhuang, F., Yu, R., & Xiong, H. (2020). Identifying High Potential Talent: A Neural Network Based Dynamic Social Profiling Approach. *2019 IEEE International Conference on Data Mining (ICDM)*. doi:<https://doi.org/10.1109/ICDM.2019.00082>
- Zaky, M. (2021). Penerapan teknologi informasi dalam proses rekrutmen dan seleksi sumber daya manusia. *KOMITMEN: Jurnal Ilmiah Manajemen*, *2*(2), 112-123. Retrieved from <https://digilib.uinsgd.ac.id/67542/>
- Zhu, W., Lévy-Leduc, C., & Ternès, N. (2021). A variable selection approach for highly correlated predictors in high-dimensional genomic data. *Bioinformatics*, *37*(16), 2238–2244. doi:<https://doi.org/10.1093/bioinformatics/btab114>

LAMPIRAN

1.1. Proses Augmentasi Data

```

import warnings
import numpy as np
import pandas as pd

# Memanggil Dataset
data = pd.read_csv('/content/Data Psikogram.csv')

# Untuk menampilkan jumlah baris dan kolom yang ada pada data
print("There are {:,} rows and {} columns in the
data.".format(data.shape[0], data.shape[1]))
print("There are {} missing values in the
data.".format(data.isnull().sum().sum()))

# Untuk menampilkan 5 data teratas pada database
data.head()

# Remove unnecessary features
df = data.drop(['nama'], axis=1)

df.head()

#Making a copy of original data
df_original = df.copy()

#Encode Categorical Features
#Import necessary libraries
from sklearn.preprocessing import LabelEncoder

# Create a label encoder object
le = LabelEncoder()

# Identify the features to be encoded
columns_to_encode =
['pendidikan', 'asal_sekolah', 'pj_fit', 'jenis_kelamin']

# Dictionary to store mapping {original_value: encoded_value}
label_mapping = {}

# Loop over columns to encode
for column in columns_to_encode:
    # Convert the column to string type
    df[column] = df[column].astype(str)

```

```

# Fit and transform the column
df[column] = le.fit_transform(df[column])

# Get mapping of original values to encoded values
label_mapping[column] = dict(zip(le.classes_,
le.transform(le.classes_)))

# Print the mappings
for column, mapping in label_mapping.items():
    print(f'Column: {column}')
    for original_value, encoded_value in mapping.items():
        print(f'{original_value}: {encoded_value}')
    print('\n')

# Show data tabel and summary of dataframe
df

# Define your features and target variable
X = df.drop('pj_fit', axis=1)
y = df['pj_fit']

from sklearn.model_selection import StratifiedKFold, cross_val_score
from sklearn.ensemble import RandomForestClassifier
from imblearn.over_sampling import RandomOverSampler
import numpy as np

# Membuat objek ROS
ros = RandomOverSampler(random_state=42)

# Membuat objek classifier
clf = RandomForestClassifier(random_state=42)

# List untuk menyimpan skor rata-rata validasi silang untuk setiap
nilai n_splits
avg_cv_scores = []

# Range nilai n_splits yang akan dicoba
n_splits_range = list(range(2, 11))

for n_splits in n_splits_range:
    # Membuat objek StratifiedKFold dengan n_splits saat ini
    skf = StratifiedKFold(n_splits=n_splits, random_state=42,
shuffle=True)

    # List untuk menyimpan skor validasi silang untuk setiap lipatan
    cv_scores = []

```

```

for train_index, test_index in skf.split(X, y):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    # Menggunakan ROS pada data X_train dan y_train
    X_train_res, y_train_res = ros.fit_resample(X_train, y_train)

    # Melatih classifier dan menghitung skor validasi silang
    clf.fit(X_train_res, y_train_res)
    score = clf.score(X_test, y_test)
    cv_scores.append(score)

    # Menghitung rata-rata skor validasi silang untuk n_splits saat
    ini dan menyimpannya
    avg_cv_scores.append(np.mean(cv_scores))

# Mencetak nilai n_splits yang memberikan skor rata-rata validasi
silang tertinggi
best_n_splits = n_splits_range[np.argmax(avg_cv_scores)]
print(f"Nilai terbaik untuk n_splits adalah: {best_n_splits}")

# Sekarang Anda dapat menggunakan nilai best_n_splits untuk membagi
data Anda menggunakan StratifiedKFold
skf_best = StratifiedKFold(n_splits=best_n_splits, random_state=42,
shuffle=True)

from sklearn.model_selection import StratifiedKFold
from imblearn.over_sampling import RandomOverSampler
from sklearn.ensemble import RandomForestClassifier

# Membuat objek ROS
ros = RandomOverSampler(random_state=42)

# Membuat objek StratifiedKFold dengan n_splits terbaik
skf_best = StratifiedKFold(n_splits=6, random_state=42,
shuffle=True)

for train_index, test_index in skf_best.split(X, y):
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    # Menggunakan ROS pada data X_train dan y_train
    X_train_res, y_train_res = ros.fit_resample(X_train, y_train)

# Mengubah X_train_res menjadi DataFrame
df_X = pd.DataFrame(X_train_res, columns=X.columns)

# Mengubah y_train_res menjadi DataFrame

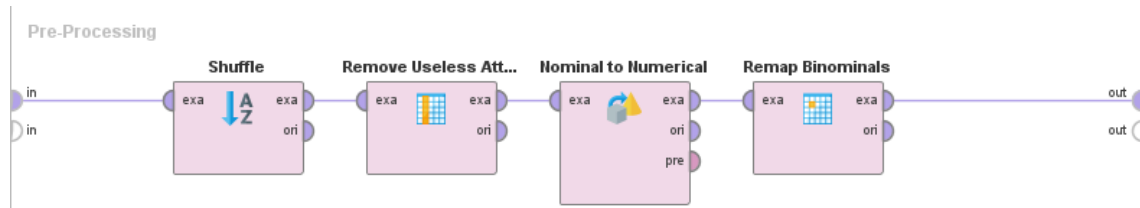
```

```
df_y = pd.DataFrame(y_train_res, columns=['pj_fit'])

# Menggabungkan df_X dan df_y menjadi satu DataFrame
df = pd.concat([df_X, df_y], axis=1)
df

df.to_csv('df.csv', index=False)
```

1.2. Tahap *Pre-Processing* pada Data

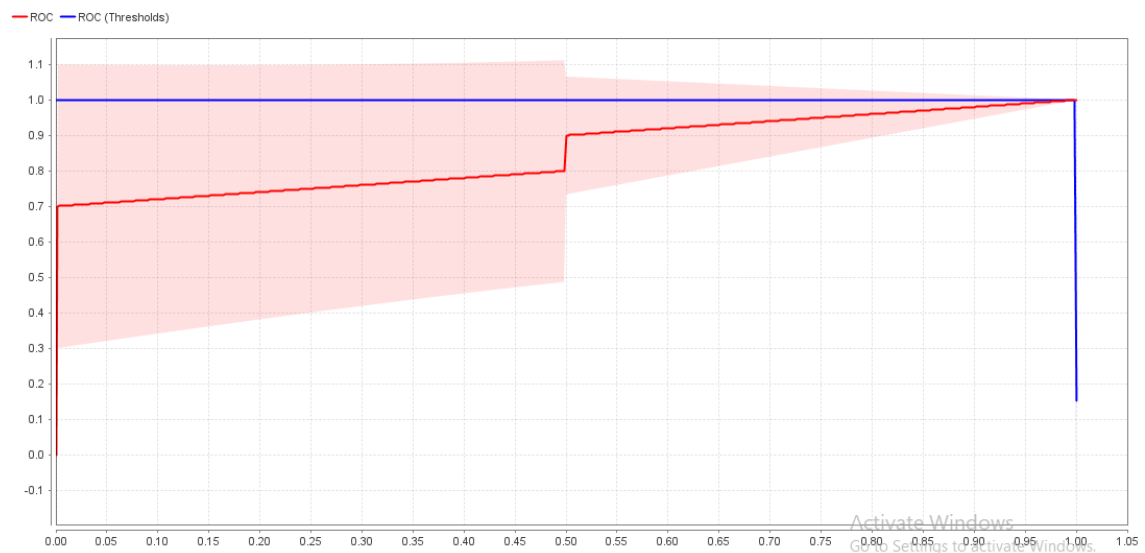


1.3. Hasil Performa Data Pelatihan *Decision Tree*

accuracy: 85.00% +/- 19.95% (micro average: 85.71%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

AUC: 0.450 +/- 0.158 (micro average: 0.450) (positive class: Highly Fit)



precision: 93.33% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

recall: 80.00% +/- 34.96% (micro average: 82.35%) (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

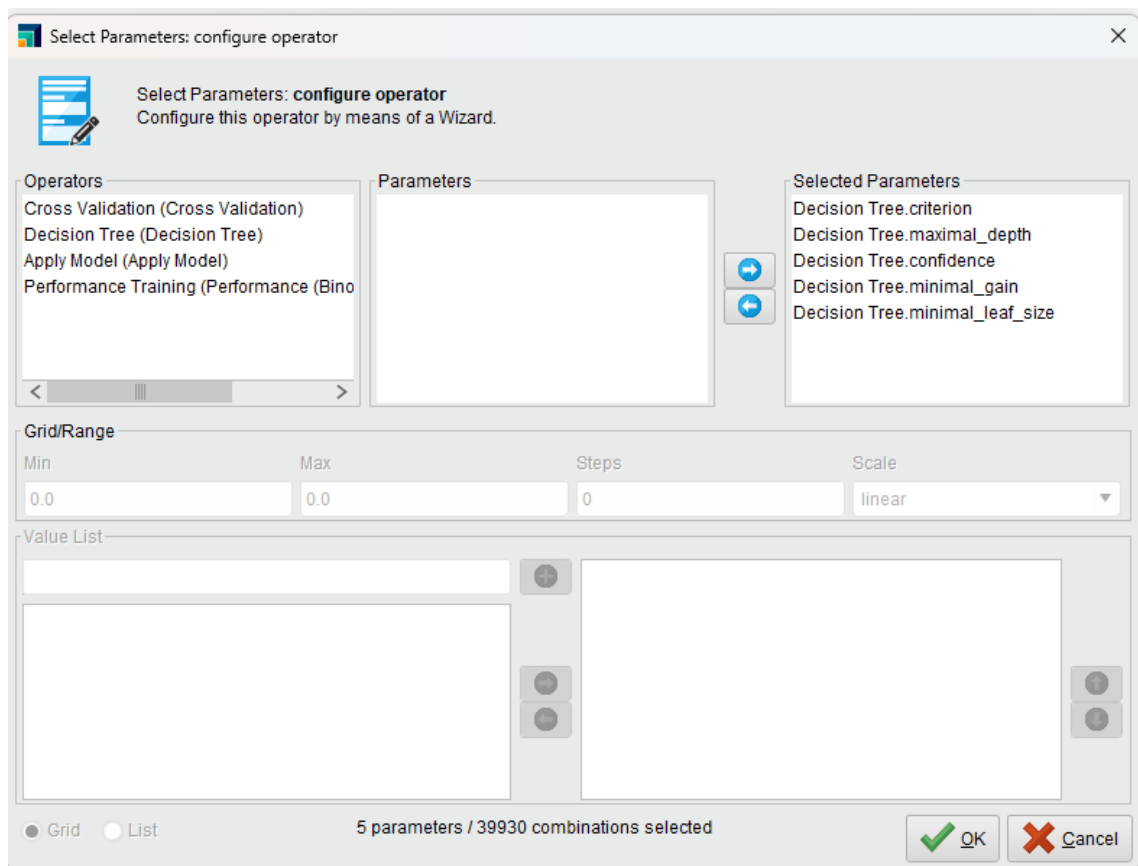
f_measure: 87.50% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

classification_error: 15.00% +/- 19.95% (micro average: 14.29%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

1.4. Process *Select ParametersConfigure Operator* Algorithm *Decision Tree*



1.5. Hasil *Parameter Set Parameter Tuning* Algorithm *Decision Tree*

Parameter set:

Performance:

PerformanceVector [

```

-----accuracy: 89.29% +/- 31.50% (micro average: 89.29%)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
-----classification_error: 10.71% +/- 31.50% (micro average: 10.71%)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
-----AUC: 0.214 +/- 0.252 (micro average: 0.214) (positive class:
Highly Fit)
-----precision: 93.75% (positive class: Highly Fit)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
-----recall: 88.24% (positive class: Highly Fit)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
-----lift: 154.41% (positive class: Highly Fit)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
-----f_measure: 90.91% (positive class: Highly Fit)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:10      2
Highly Fit:      1      15
]
Decision Tree.criterion      = gain_ratio
Decision Tree.maximal_depth  = 5
Decision Tree.confidence     = 0.185
Decision Tree.minimal_gain    = 0.1
Decision Tree.minimal_leaf_size = 1

```

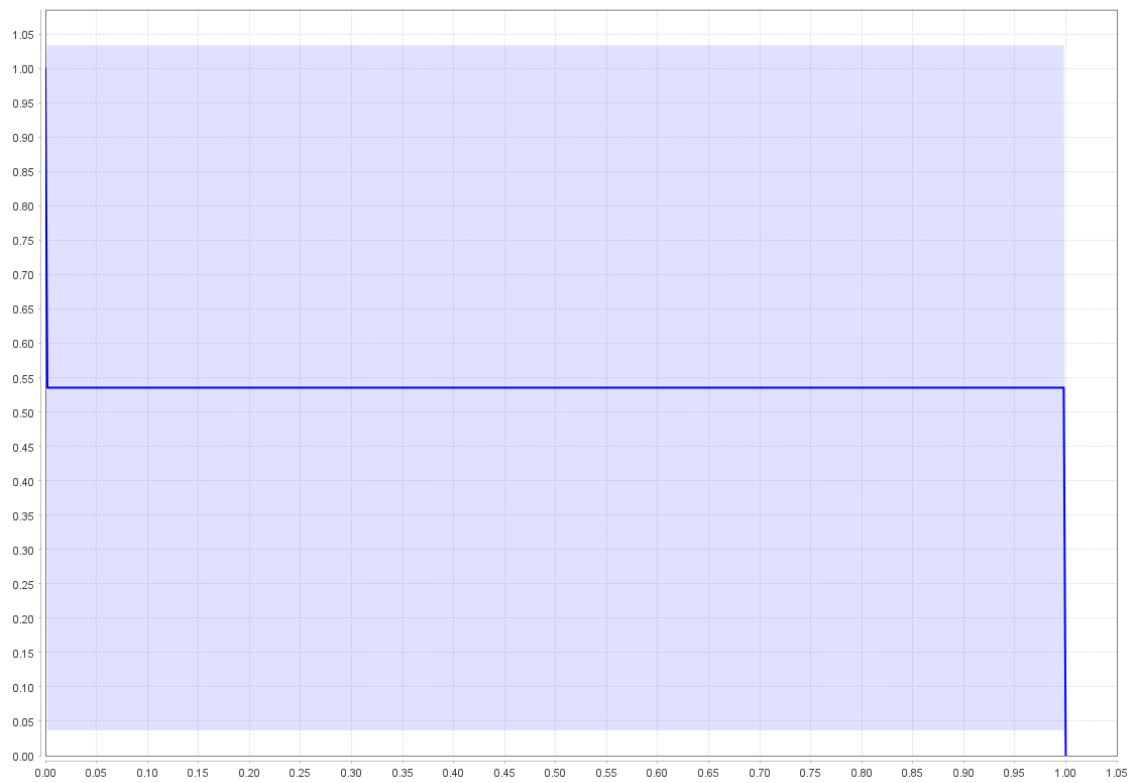
1.6. Hasil Performa Data Pelatihan *Decision Tree* (Setelah *Parameter Tuning*)

accuracy: 89.29% +/- 31.50% (micro average: 89.29%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	2	83.33%
pred. Highly Fit	1	15	93.75%
class recall	90.91%	88.24%	

AUC: 0.214 +/- 0.252 (micro average: 0.214) (positive class: Highly Fit)

— ROC — ROC (Thresholds)



precision: 93.75% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	2	83.33%
pred. Highly Fit	1	15	93.75%
class recall	90.91%	88.24%	

recall: 88.24% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	2	83.33%
pred. Highly Fit	1	15	93.75%
class recall	90.91%	88.24%	

f_measure: 90.91% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	2	83.33%
pred. Highly Fit	1	15	93.75%
class recall	90.91%	88.24%	

classification_error: 10.71% +/- 31.50% (micro average: 10.71%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	2	83.33%
pred. Highly Fit	1	15	93.75%
class recall	90.91%	88.24%	

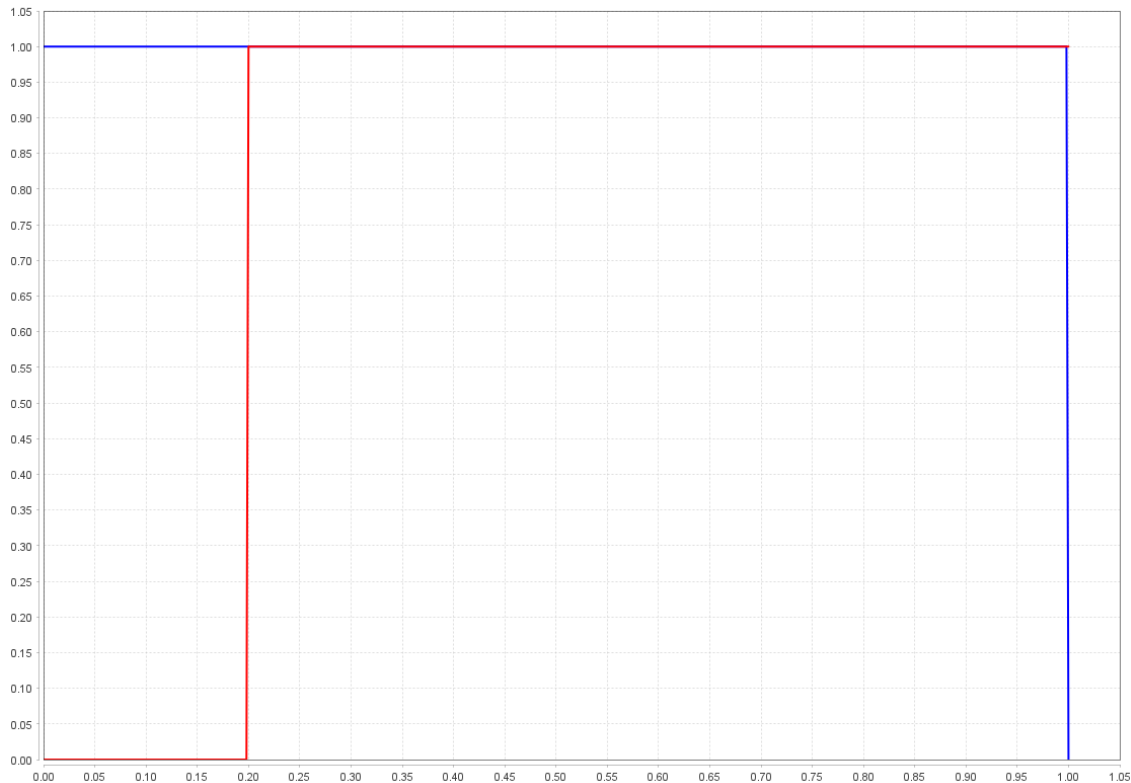
1.7. Hasil Performa Data Uji *Decision Tree* (Setelah *Parameter Tuning*)

accuracy: 91.67%

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

AUC: 0.500 (positive class: Highly Fit)

— ROC — ROC (Thresholds)



precision: 87.50% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

recall: 100.00% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

f_measure: 93.33% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

classification_error: 8.33%

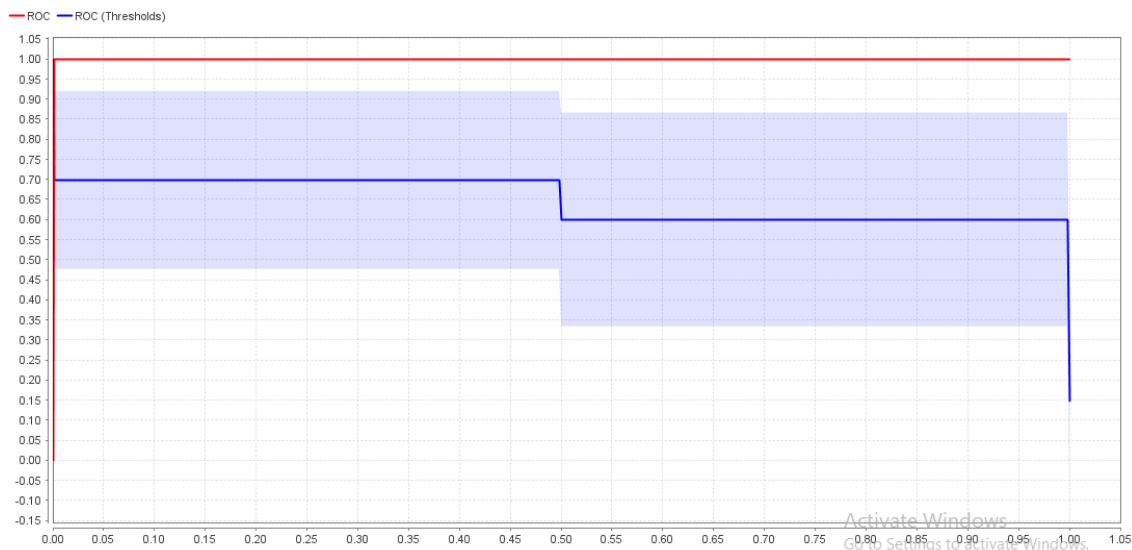
	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

1.8. Hasil Performa Data Pelatihan *Random Forest*

accuracy: 85.00% +/- 19.95% (micro average: 85.71%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

AUC: 0.900 +/- 0.316 (micro average: 0.900) (positive class: Highly Fit)



precision: 93.33% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

recall: 80.00% +/- 34.96% (micro average: 82.35%) (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

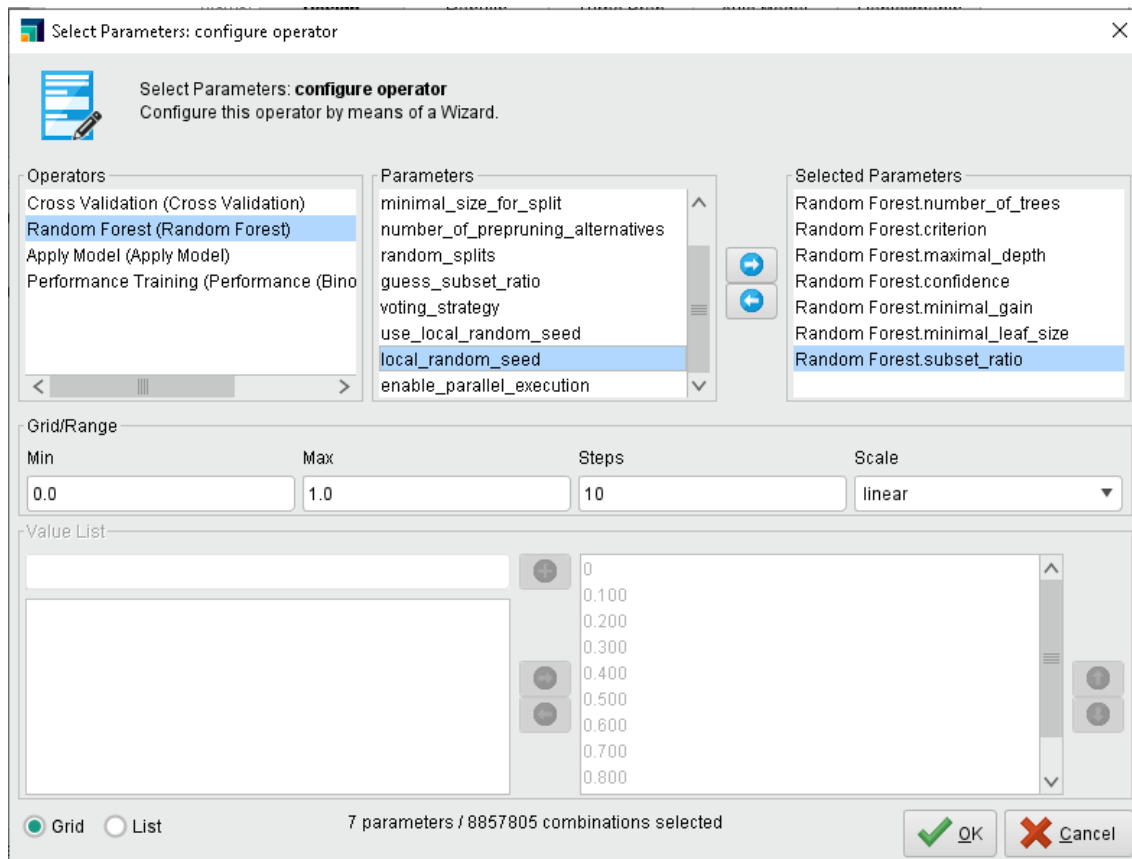
f_measure: 87.50% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

classification_error: 15.00% +/- 19.95% (micro average: 14.29%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	10	3	76.92%
pred. Highly Fit	1	14	93.33%
class recall	90.91%	82.35%	

1.9. Process *Select ParametersConfigure Operator* Algoritma *Random Forest*



1.10. Hasil Parameter Set Parameter Tuning Algoritma Random Forest

Parameter set:

Performance:

PerformanceVector [

-----accuracy: 96.67% +/- 8.16% (micro average: 96.43%)

ConfusionMatrix:

True: Moderately Fit Highly Fit

Moderately Fit:11 1

Highly Fit: 0 16

-----AUC: 1.000 +/- 0.000 (micro average: 1.000) (positive class:

Highly Fit)

-----precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Highly Fit)

ConfusionMatrix:

True: Moderately Fit Highly Fit

Moderately Fit:11 1

Highly Fit: 0 16

-----recall: 94.44% +/- 13.61% (micro average: 94.12%) (positive class: Highly Fit)

ConfusionMatrix:

True: Moderately Fit Highly Fit

Moderately Fit:11 1

Highly Fit: 0 16

-----lift: 169.44% +/- 42.71% (micro average: 164.71%) (positive class: Highly Fit)

ConfusionMatrix:

True: Moderately Fit Highly Fit

```

Moderately Fit:11      1
Highly Fit:      0      16
-----f_measure: 96.67% +/- 8.16% (micro average: 96.97%) (positive
class: Highly Fit)
ConfusionMatrix:
True:  Moderately Fit Highly Fit
Moderately Fit:11      1
Highly Fit:      0      16
]
Random Forest.number_of_trees = 41
Random Forest.maximal_depth   = 11
Random Forest.confidence      = 0.3875716431433165
Random Forest.subset_ratio    = 0.017544087464573033
Random Forest.minimal_gain    = 0.6368727539324707
Random Forest.minimal_leaf_size = 22
Cross Validation.number_of_folds = 6

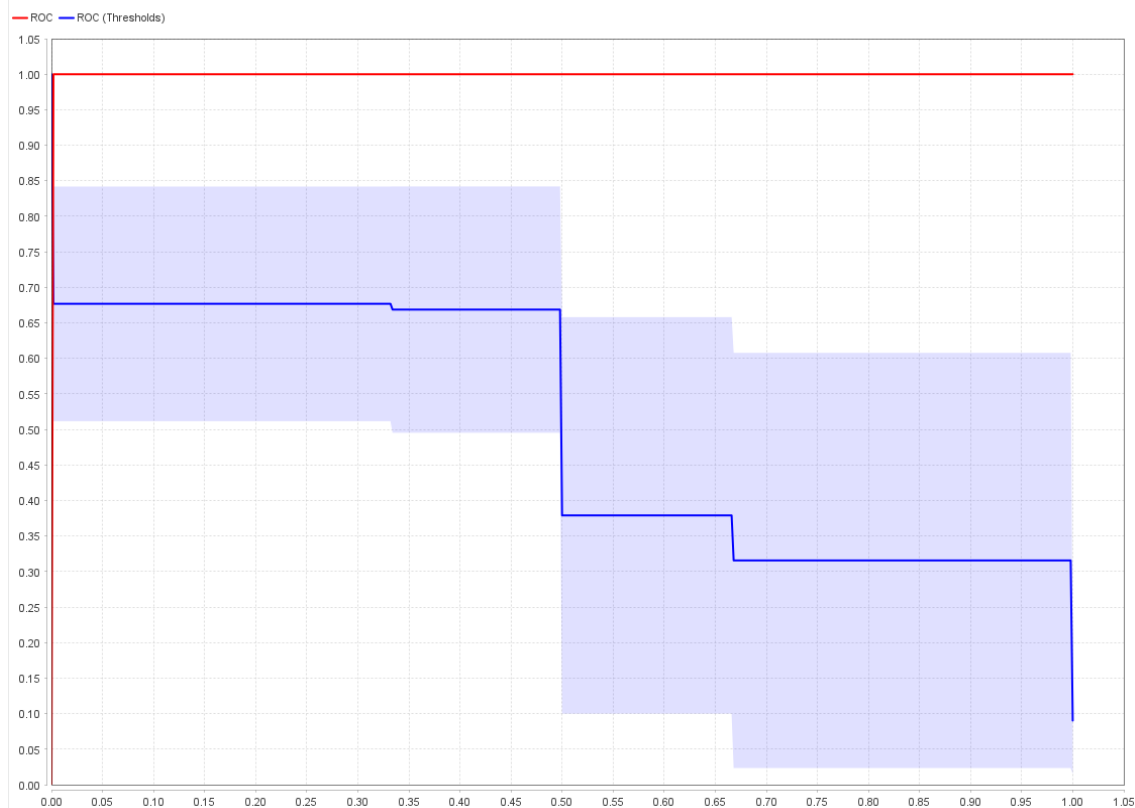
```

1.11. Hasil Performa Data Pelatihan *Random Forest* (Setelah *Parameter Tuning*)

accuracy: 96.67% +/- 8.16% (micro average: 96.43%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	11	1	91.67%
pred. Highly Fit	0	16	100.00%
class recall	100.00%	94.12%	

AUC: 1.000 +/- 0.000 (micro average: 1.000) (positive class: Highly Fit)



precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	11	1	91.67%
pred. Highly Fit	0	16	100.00%
class recall	100.00%	94.12%	

recall: 94.44% +/- 13.61% (micro average: 94.12%) (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	11	1	91.67%
pred. Highly Fit	0	16	100.00%
class recall	100.00%	94.12%	

f_measure: 96.67% +/- 8.16% (micro average: 96.97%) (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	11	1	91.67%
pred. Highly Fit	0	16	100.00%
class recall	100.00%	94.12%	

classification_error: 3.33% +/- 8.16% (micro average: 3.57%)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	11	1	91.67%
pred. Highly Fit	0	16	100.00%
class recall	100.00%	94.12%	

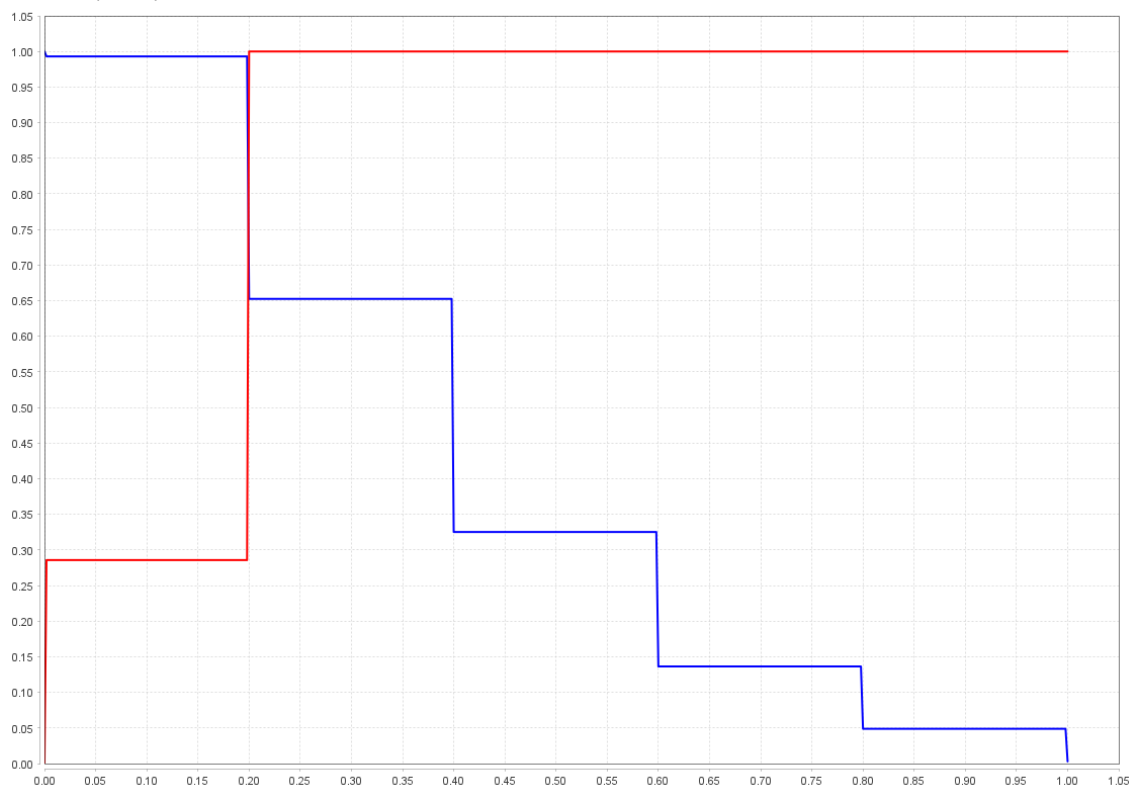
1.12. Hasil Performa Data Uji *Random Forest* (Setelah *Parameter Tuning*)

accuracy: 91.67%

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

AUC: 0.857 (positive class: Highly Fit)

— ROC — ROC (Thresholds)



precision: 87.50% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

recall: 100.00% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

f_measure: 93.33% (positive class: Highly Fit)

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

classification_error: 8.33%

	true Moderately Fit	true Highly Fit	class precision
pred. Moderately Fit	4	0	100.00%
pred. Highly Fit	1	7	87.50%
class recall	80.00%	100.00%	

