Volume 14, Nomor 1, Maret 2024

P-ISSN 2088-1770
E-ISSN 2503-3247

JURNAL SAINTEKOM

SAINS, TEKNOLOGI, KOMPUTER DAN MANAJEMEN

1995

STMIK PALANGKARAYA

JURNAL SAINTEKOM

Sains, Teknologi, Komputer, dan Manajemen

Penerbit:
Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya

Penanggung Jawab:
Ketua Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya

Alamat Redaksi:
Kampus STMIK Palangkaraya Jl.George Obos No.114
Telp. (0536) 3225515 Hp. 081349775351, 085651387908
Palangka Raya, Kalimantan Tengah, Indonesia
e-mail: jurnalsaintekom@stmikplk.ac.id

Jurnal Saintekom merupakan jurnal ilmiah yang berfungsi sebagai media mengkomunikasikan ide, gagasan, dan pemikiran seputar kajian actual tentang sains, teknologi, komputer, dan manajemen antar akademisi dan peneliti, yang diterbitkan oleh Sekolah Tinggi Manajemen Informatika dan komputer (STMIK) Palangkaraya. Terbit pertama kali tahun 2011 berdasarkan SK Ketua STMIK Palangkaraya Nomor 09 Tahun 2010, tanggal 14 Juni 2010. Jurnal Saintekom terbit setiap 6 bulan (2 kali setahun: Maret & September). Jurnal Saintekom mengundang semua pihak untuk berpartisipasi mengirimkan artikel sesuai dengan misi jurnal.

JURNAL SAINTEKOM

Sains, Teknologi, Komputer, dan Manajemen

Terbit secara Berkala Dua Kali Setahun
Pada Bulan Maret dan September

Penerbit:
Sekolah Tinggi Manajemen Informatika dan Komputer
(STMIK) Palangka Raya

Volume 14, Nomor 1, September 2024 **P-ISSN : 2088-1770**
E-ISSN : 2503-3247

JURNAL SAINTEKOM

Sains, Teknologi, Komputer, dan Manajemen

DAFTAR ISI

JURNAL SAINTEKOM

Sains, Teknologi, Komputer Dan Manajemen

Volume 14, Nomor 1, Maret 2024

9 772088 177004

JURNAL SAINTEKOM

Volume 14, Nomor 1, Maret 2024

SAIN TEKOM
ISSN 2088-1770 (Print)
ISSN 2503-3247(Online)
Vol. 14 No.1 (2024) 13-25

JURNAL
SAINTEKOM
Sains, Teknologi, Komputer, dan Manajemen

Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir

*Hendri Mahmud Nawawi[1], Agung Baitul Hikmah[2], Ali Mustopa[3], Ganda Wijaya[4]
[1)]Informatika, Fakulas Teknologi Informasi, Universitas Nusa Mandiri
[2)]Sistem Informasi Kampus Kota Tasikmalaya, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
[3)]Teknik Informatika Kampus Kota Pontianak, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
[4)]Sistem Informasi, Fakulas Teknologi Informasi, Universitas Nusa Mandiri
[1,4] Jl. Raya Jatiwaringin No.2, Cipinang Melayu, Kec. Makasar, Kota Jakarta Timur, DKI Jakarta
[2,3] Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta
Email: [1]hendri.hiw@nusamandiri.ac.id, [2]agung.abl@bsi.ac.id, [3]ali.aop@bsi.ac.id, [4]ganda.gws@nusamandiri.ac.id

ABSTRACT

The complexity of the job market requires individuals and organizations to understand the trends and needs of the world of work. One of the main challenges is the right career placement. That is becoming increasingly popular is the use of Machine Learning (ML) algorithms in the decision-making process. ML classification models such as Random Forest, Decision Tree, Naïve Bayes, KNN, and SVM have demonstrated their potential in uncovering hidden patterns from data, including a person's educational history, work experience and interests. In this research, the application of the ML classification model is aimed at predicting career placement. From the data sample used of 215, this research evaluates the effectiveness of various ML models in the context of career placement. As a result, the Random Forest Model is superior to other proposed models with an accuracy value of 87% and an AUC/ROC value of 0.93 which indicates a very good classification value. Meanwhile, the SVM model with Linear Kernel shows the lowest performance with an accuracy value of 67%. Apart from getting information on the best accuracy and AUC/ROC values, the results of this research found that the 'ssc_presentage' attribute (high school exam percentage) is an important factor in career placement decisions.

Keywords : machine learning; random forest; job placement; classification; prediction

ABSTRAK

Kompleksitas pasar kerja mengharuskan individu dan organisasi memahami tren serta kebutuhan dunia kerja. Salah satu tantangan utama adalah penempatan karir yang tepat. Pendekatan yang semakin populer adalah penggunaan algoritma *Machine Learning* (ML) dalam proses pengambilan keputusan. Model klasifikasi ML seperti *Random Forest, Decision Tree, Naïve Bayes*, KNN, dan SVM telah menunjukkan potensinya dalam mengungkap pola tersembunyi dari data, termasuk riwayat pendidikan, pengalaman kerja dan minat seseorang. Dalam penelitian ini, penerapan model klasifikasi ML ditujukan untuk memprediksi penempatan karir. Dari sampel data yang digunakan sebanyak 215, penelitian ini mengevaluasi efektivitas berbagai model ML dalam konteks penempatan karir. Hasilnya, Model *Random Forest* lebih unggul dibandingkan model lain yang diusulkan dengan nilai akurasi mencapai 87% dan nilai AUC/ROC 0,93 yang menunjukkan nilai klasifikasi sangat baik. Sedangkan, model SVM dengan Kernel Linier menunjukkan performa terendah dengan nilai akurasi sebesar 67%. Selain mendapatkan informasi nilai akurasi dan AUC/ROC terbaik hasil penelitian ini menemukan bahwa atribut *'ssc_presentage'* (persentase ujian sekolah menengah atas) merupakan faktor penting dalam keputusan penempatan karir.

Kata kunci : *machine learning; random forest*; penempatan karir; klasifikasi; prediksi

Diterima Redaksi: 9-10-2023 | Selesai Revisi: 27-12-2023 | Diterbitkan Online: 31-3-2024

1. PENDAHULUAN

Era modern telah dipenuhi oleh kompleksitas pasar kerja yang terus berubah, maka dari itu penting bagi individu maupun organisasi untuk memahami dan mengantisipasi tren serta kebutuhan di dunia kerja (Umar et al., 2018). Salah satu aspek krusial dalam mengelola sumber daya manusia adalah penempatan karir yang tepat (Danuri & Jaroji, 2019). Hal ini menyebabkan munculnya bidang-bidang pengetahuan baru seperti analisis sumber daya manusia (SDM) (yang mencakup topik terkait lainnya, seperti analisis tenaga kerja dan analisis sumber daya manusia (Pessach et al., 2020). Keputusan yang akurat dan tepat dalam menempatkan individu pada peran tertentu tidak hanya mempengaruhi produktivitas dan kepuasan karyawan, tetapi juga berdampak pada kesuksesan organisasi secara keseluruhan (Metalita et al., 2021).

Dalam upaya untuk mengoptimalkan proses penempatan karir, muncul paradigma baru yang mengandalkan teknologi dan analisis data (Sukmawati et al., 2022). Di sinilah model klasifikasi berbasis *machine learning* hadir sebagai alat yang efektif untuk mengatasi kompleksitas dan variabilitas dalam pengambilan keputusan penempatan karir.

Prinsip dalam algoritma *Machine Learning* (ML) telah menjadi populer saat ini untuk memecahkan berbagai masalah dan mendukung dalam banyak proses pengambilan keputusan (Aravind et al., 2019). Model-model ini dapat memanfaatkan berbagai data seperti riwayat pendidikan, pengalaman kerja, keterampilan, minat, serta faktor-faktor lain yang berpotensi mempengaruhi kecocokan individu dengan peran tertentu (Bao et al., 2023).

Penelitian tentang klasifikasi penempatan kerja terhadap pekerja imigran menggunakan dua model algoritma C45 dan naïve bayes, akurasi tertinggi oleh model C45 dengan presentasi 84,84% sedangkan naïve bayes sebesar 58,29% (Sukmawati et al., 2022). Model klasifikasi *machine learning* juga diterapkan pada kasus klasifikasi kejadian berat badan lahir di Indonesia, hasil penelitian menunjukan metode *machine learning* seperti *Random Forest, CART, Naïve Bayes*, dan *SVM* untuk klasifikasi risiko Berat Badan Lahir Rendah (BBLR). Dengan teknik *resample*, ketepatan klasifikasi

DOI: https://doi.org/10.33020/saintekom.v14i1.512

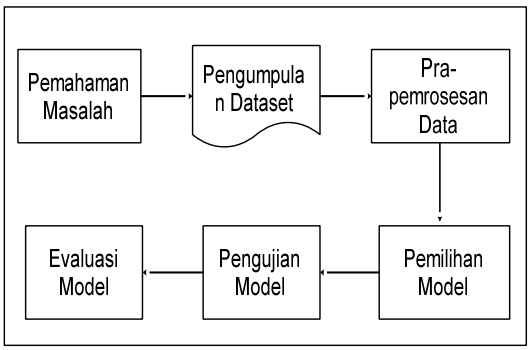
meningkat, terutama pada data *imbalanced* (tidak seimbang). *Random Forest* menunjukkan hasil terbaik, dengan beberapa faktor kunci mempengaruhi risiko Berat Badan Lahir Rendah (BBLR) seperti jarak kelahiran, pemeriksaan kehamilan, dan umur ibu (Sihombing & Yuliati, 2021).

Penelitian ini bertujuan untuk menjelaskan dan menggambarkan penerapan model klasifikasi *machine learning* dalam prediksi ketepatan penempatan karir. Model klasifikasi yang digunakan pada penelitian ini yaitu *Random Forest, Decision Tree, Naïve Bayes, KNN* dan *SVM* untuk mempelajari pola-pola tersembunyi dari dataset penelitian. Model terpilih akan dijadikan sebagai usulan untuk pengembangan penelitian dan membuat keputusan penempatan karir yang lebih tepat dan akurat, penelitian ini menghasilkan nilai akurasi jauh lebih baik dari penelitian sebelumnya dengan nilai akurasi sebesar 87,16%. Perbedaan dari penelitian sebelumnya dengan yang diusulkan jika pada penelitian (Sukmawati et al., 2022) jenis data yang diteliti memiliki kesamaan dari objek yang diteliti tentang klasifikasi penempatan kerja sedangkan pada penelitian (Sihombing & Yuliati, 2021) karakteristik dataset yang diteliti memiliki kesamaan dengan inputan multiple attribut terhadap satu variabel target mengusulkan banyak model untuk mencari model terbaik.

Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi terhadap pengembangan strategi pengelolaan sumber daya manusia yang lebih efisien dan berdampak positif bagi individu untuk memprediksi posisi karir sesuai dengan pendidikan dan kemampuan maupun organisasi dalam hal rekrutmen terhadap kandidat karyawan.

2. METODE

Agar penelitian yang dilakukan dapat terarah berdasarkan sesuai urutan maka diperlukan metode penelitian. Pada penelitian ini tahapan atau metode penelitian digambarkan pada Gambar 1.



Gambar 1. Metode Penelitian

Metode penelitian dalam *machine learning* (ML) biasanya

DOI: https://doi.org/10.33020/saintekom.v14i1.512

melibatkan serangkaian langkah yang sistematis untuk mengembangkan, melatih, dan mengevaluasi model ML.

2.1. Pemahaman Masalah

Sebelum memulai penelitian, penting untuk memahami masalah yang ingin diselesaikan terutama pada kasus data yang kompleks. Apakah itu klasifikasi, regresi, pengelompokan, atau jenis tugas *Machine Learning* lainnya (Pusporani et al., 2019). Pada penelitian ini jenis dataset yang digunakan cenderung memiliki karakteristik klasifikasi dikarenakan terdiri dari dua kelas utama yaitu ya dan tidak.

2.2. Pengumpulan Data

Data yang digunakan bersumber dari dataset publik yang dapat diakses secara terbuka dari repository *kaggle*. Penelitian ini menggunakan dataset *Job Placement* (Penempatan Karir) yang terdiri dari 13 atribut utama dimana 12 atribut sebagai label dan satu atribut sebagai variabel target.

2.3. Pra-pemprosesan Data

Tahap pra pemrosesan data atau *preprocessing* merupakan tahapan yang sangat penting (Afiasari et al., 2023). Tujuannya agar data yang diproses tidak terdapat informasi yang kosong atau *null* sehingga dapat mempengaruhi hasil yang diharapkan (Gata et al., 2023). Data mentah seringkali perlu dibersihkan dan diproses sebelum digunakan. Ini bisa melibatkan penghapusan data yang hilang, normalisasi, dan transformasi fitur (Anggraini et al., 2021).

2.4. Pemilihan Model

Dalam analisis ini, berbagai model klasifikasi telah diuji untuk memprediksi kesuksesan penempatan posisi pekerjaan berdasarkan dataset penelitian. Model-model yang dianalisis termasuk *Random Forest, Naive Bayes, SVM, KNN* dan *Decision Tree*. Dasar pemilihan model yang diajukan pada penelitian ini melihat kecenderungan atau karakteristik dataset yang diteliti seperti jenis fitur data, ukuran dataset, dan kompleksitas komputasi.

a. *Random Forest*

Random forest adalah teknik yang membangun sejumlah pohon keputusan dengan menggabungkan *root* node, node internal, dan node daun. Pada proses pembentukannya, atribut dan data dipilih secara random berdasarkan aturan tertentu (Sihombing & Yuliati, 2021). Rumus Persamaan 1 *random forest*:

DOI: https://doi.org/10.33020/saintekom.v14i1.512

$Entropy\ (Y) = \sum i\ p\ (c|Y) log2p(C|Y)$(1)

Y adalah himpunan kasus, dan $(c|Y)$ adalah proporsi nilai Y terhadap kelas c.

b. *Naïve Bayes*

Metode Naïve Bayes merupakan salah satu teknik klasifikasi yang terkenal dan termasuk dalam daftar sepuluh algoritma utama di bidang data mining (Pusporani et al., 2019). Rumus naïve bayes pada Persamaan 2.

$$P\ (Y|X) = P\ (X|Y) * \frac{P(Y)}{P(X)} \quad (2)$$

Y adalah data class (belum diketahui), P(Y|X) adalah probabilitas Y berdasar kondisi X, P(Y|X) adalah probabilitas X berdasar kondisi Y, P(Y) adalah probabilitas dari Y, dan P(X) adalah probabilitas dari X.

c. *Support Vector Machine*

SVM termasuk dalam kategori *supervised learning*, teknik SVM dalam pembelajaran mesin dirancang untuk mengklasifikasikan data, baik yang bersifat linear maupun non-linear. Konsep utama dari SVM adalah meningkatkan jarak *hyperplane* (batas keputusan) terhadap data yang di klasifikasikannya (Sihombing & Yuliati, 2021). Kernel Linear digunakan pada penelitian ini dikarenakan karakteristik dataset yang menghasilkan klasifikasi biner 0 dan 1 sehingga lebih sederhana dalam mengambil sebuah keputusan. Rumus SVM dengan kernel linear pada persamaan 3:

$$K(x, y) = x^T y \quad (3)$$

K(x,y) adalah nilai kernel linier antara dua vektor x dan y, x^T adalah transpos dari vektor x, dan y adalah vektor kedua.

d. K-*Nearest Neighbor*

K-*Nearest Neighbor* (K-NN) adalah metode berbasis supervisi yang membutuhkan data latihan untuk mengkategorikan objek berdasarkan kedekatan jaraknya. Esensi dari K-NN adalah menentukan seberapa dekat suatu informasi yang akan dinilai dengan k tetangga terdekatnya dalam dataset pelatihan (Hozairi et al., 2021). Rumus K-NN pada Persamaan 4

$$di = \sqrt{\sum_{i=1}^{n}(xij - pj)^2} \quad (4)$$

di adalah jarak sampel, xij adalah data sampel pengetahuan, pj adalah data input var ke-j, dan n adalah jumlah sampel.

DOI: https://doi.org/10.33020/saintekom.v14i1.512

e. *Decision Tree*

Struktur model *Decision Tree* mirip dengan diagram alir, setiap node internal menggambarkan variabel prediktor yang memisahkan data, sementara setiap node daun menandakan kelas hasil dari proses klasifikasi (Pusporani et al., 2019). Rumus *Decision Tree* pada Persamaan 5.

$$Entropy\ (S) = \sum_{i=1}^{n} pi * log2pi \quad (5)$$

S = himpunan kasus,

n = jumlah partisi S, dan

Pi = proporsi dari Si terhadap S.

2.5. Pelatihan Model

Setelah memilih model klasifikasi machine learning langkah selanjutnya adalah melakukan pelatihan terhadap dataset penelitian dimana data akan dibagi menjadi dua bagian yaitu data pelatihan (*training*) dan data pengujian (*testing*) pada penelitian ini pembagian data dibagi menjadi 80:20 dimana sebanyak 80% dijadikan data *training* dan 20% sebagai data *testing*.

Dalam melakukan pelatihan model teknik yang biasa digunakan adalah dengan menggunakan *cross validation*. *Cross validation* adalah pendekatan untuk mengeksploitasi sampel pelatihan dan penilaian akurasi beberapa kali sehingga berpotensi meningkatkan keandalan hasil (Ramezan et al., 2019). Teknik ini sering digunakan dalam pembelajaran mesin untuk memeriksa seberapa baik suatu model akan generalisasi ke kumpulan data baru (Firdaus, 2022).

2.6. Evaluasi Model

Dalam mengukur kinerja performa model klasifikasi yang digunakan dalam klasifikasi binari *Confusion Matriks* merupakan matriks yang biasanya digunakan untuk memperoleh informasi mengenai prediksi aktual dan prediksi yang dibuat oleh model (Visa et al., n.d.). *Confusion matriks* terdiri dari empat bagian:

1. ***True Positive* (TP):** Kasus dimana hasil prediksi model adalah positif dan sesuai dengan kondisi aktual yang juga positif.
2. ***False Positive* (FP):** Kasus dimana model memprediksi positif, tapi kenyataannya negatif.
3. ***True Negative* (TN):** Kasus dimana model secara akurat memprediksi hasil negatif yang sesuai dengan realitas yang memang negatif.
4. ***False Negative* (FN):** Kasus dimana model memprediksi negatif, tapi kenyataannya positif.

DOI: https://doi.org/10.33020/saintekom.v14i1.512

Kriteria yang diusulkan dalam memberikan rekomendasi terhadap model pilihan dilihat dari seberapa besarnya nilai akurasi yang ditunjukan oleh setiap model. Beberapa evaluasi model yang digunakan pada penelitian ini diantaranya:

a. Akurasi. Seberapa sering model menghasilkan prediksi yang benar (Sihombing & Yuliati, 2021). Rumus akurasi pada Persamaan 6.

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \quad (6)$$

b. Recall: Dari semua kasus positif, seberapa banyak yang benar-benar dideteksi oleh model (Firdaus, 2022). Rumus Recall pada Persamaan 7.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

c. Precision: Dari semua prediksi positif, seberapa banyak yang benar-benar positif (Kharisma et al., 2023). Rumus Precision pada Persamaan 8.

$$Precison = \frac{TP}{TP+FP} \quad (8)$$

d. F1-Score: Sebuah ukuran yang menggabungkan recall dan precision (Firdaus, 2022). Seperti pada Persamaan 9.

$$FScore = 2 * \frac{Precision+recall}{Precision+Recall} \quad (9)$$

e. AUC/ROC: Area di bawah kurva karakteristik operasi penerima, yang menggambarkan kinerja model dalam mengklasifikasikan data dalam kategori positif atau negatif (Siahaan et al., 2021). Aturan pengambilan keputusan dalam menentukan nilai AUC/ROC pada Tabel 1.

Tabel 1. Nilai AUC/ROC

Nilai	Klasifikasi
0,90-1,00	Sangat baik
0.80-0.90	Baik
0,70-0,80	Sedang
0,60-0,70	Buruk
0,50-0,60	Salah

Sumber: (Siahaan et al., 2021)

3. HASIL DAN PEMBAHASAN

Hasil yang ditampilkan pada bagian ini, menyajikan dan membahas temuan utama yang diperoleh dari analisis data penempatan karir. Analisis ini bertujuan untuk memahami faktor-faktor yang mempengaruhi penempatan pekerjaan serta untuk membandingkan kinerja berbagai model klasifikasi dalam memprediksi kesuksesan penempatan. Dengan memahami hasil ini, perusahaan dapat merumuskan strategi yang lebih tepat dalam proses rekrutmen serta meningkatkan efisiensi program pelatihan mereka.

Pertama-tama, pada bagian ini hasil yang ditampilkan adalah eksplorasi data awal yang memberikan gambaran tentang distribusi dan

DOI: https://doi.org/10.33020/saintekom.v14i1.512

karakteristik dari data yang ada. Selanjutnya, membahas kinerja dari masing-masing model klasifikasi yang telah diterapkan, termasuk *Random Forest, Naive Bayes, SVM, KNN* dan *Decision Tree* (C45).

3.1. Dataset Penelitian

Pada penelitian ini dataset yang digunakan bersumber dari kaggel.com dimana jumlah total data sebanyak 215 dan terdiri dari 12 atribut label dan 1 atribut target. Secara detail atribut-atribut yang digunakan pada penelitian ini dijelaskan pada Tabel 2.

Tabel 2. Atribut Dataset

Atribut	Keterangan
gender	Jenis kelamin kandidat
ssc_percentage	Persentase ujian sekolah menengah atas (Kelas 10)
ssc_board	Dewan pendidikan untuk ujian ssc
hsc_percentage	Persentase ujian sekolah menengah atas (Kelas 12)
hsc_borad	Dewan pendidikan untuk ujian hsc
hsc_subject	Mata pelajaran untuk hsc
degree_percentage	Persentase nilai dalam gelar sarjana
undergrad_degree	Jurusan gelar sarjana
work_experience	Pengalaman kerja sebelumnya
emp_test_percentage	Persentase tes bakat
spesialisasi	Jurusan pasca sarjana - (spesialisasi MBA)
mba_percent	Persentase nilai dalam gelar MBA
status (target)	Ditempatkan / Tidak Ditempatkan

Sumber: (Raza, 2023)

Pada Tabel 2 menjelaskan masing-masing keterangan dari setiap atribut penelitian dimana berdasarkan dari 12 atribut inputan akan menghasilkan klasifikasi atau target apakah kandidat ditempatkan atau tidak.

Detail dari dataset yang akan diteliti dijelaskan pada Tabel 3.

Tabel 3. Dataset Penelitian

gender	ssc percentage	ssc board	hsc percentage	hsc board	hsc subject	degree percentage	undergrad degree	work experience	Emp test percentage	specialization	Mba percent	status
M	67	Others	91	Others	Commerce	58	Sci & Tech	No	55	Mkt&HR	58.8	Placed
M	79.33	Central	78.33	Others	Science	77.48	Sci & Tech	Yes	86.5	Mkt&Fin	66.28	Placed
F	77	Central	87	Central	Commerce	59	Comm& Mgmt	No	68	Mkt&Fin	68.63	Placed
..	..	..	..	..	..	..	..	..	...	…	…	…

DOI: https://doi.org/10.33020/saintekom.v14i1.512

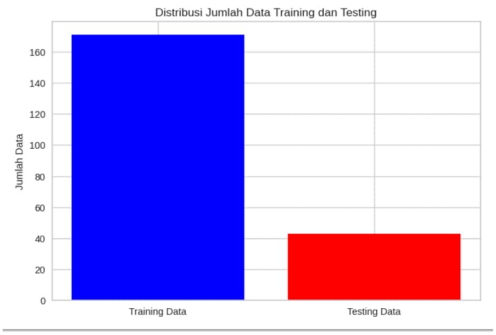
F	46	Others	49.2	Others	Commerce	79	Comm & Mgmt	No	74.28	Mkt&Fin	53.29	Not Placed
M	82	Central	64	Central	Science	66	Sci & Tech	Yes	67	Mkt&Fin	62.14	Placed

Sumber : (Raza, 2023)

Pada Tabel 3 *sample* dataset penelitian dijelaskan secara rinci masing-masing atribut masukan (*input*) dan menghasilkan satu atribut target pada kolom status (*placed/not placed*) sebagai *output*.

3.2. *Preprocessing*

Pada tahap *preprocessing* dataset dibagi menjadi dua bagian yaitu data *training* dan *testing* pembagiannya pada Gambar 2.



Gambar 2. Data *Training* dan *Testing*
(Sumber : Penelitian, 2023)

Pada Gambar 2 data *training* dan data *testing* dibagi berdasarkan skala 80:20 artinya 80% data atau sebanyak 172 dijadikan data pelatihan (*training)* ditunjukan pada bar berwarna biru dan 43 data sebagai data pengujian (*testing*) ditunjukan pada bar dengan warna merah teknik *sampling* yang digunakan dengan *cross validation* artinya sample dipilih secara acak. Penelitian ini menggunakan semua *feature* atribut untuk mendapatkan *feature* paling berpengaruh dari dataset yang diteliti.

3.3. *Modeling*

Pada tahapan ini merupakan tahapan paling penting dalam klasifikasi dengan memodelkan metode klasifikasi pada penelitian ini yaitu *Random Forest, Decision Tree, Naïve Bayes, KNN* dan *SVM* untuk dilihat seberapa besar nilai yang yang dihasilkan dengan menggunakan kelima metode yang dibandingkan nilai akurasi terbaik adalah yang menjadi rekomendasi model terbaik untuk hasil penelitian. Secara rinci Tabel 4 menjelaskan hasil dari performa model.

Tabel 4. Performa Model Klasifikasi

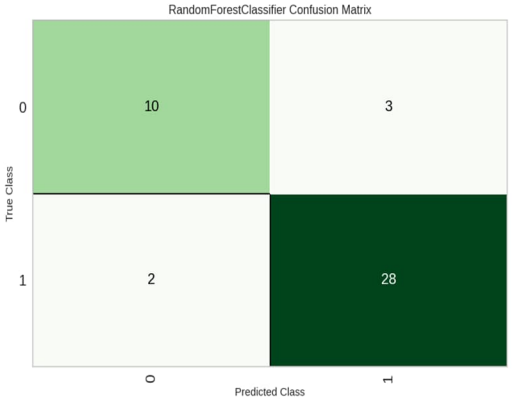
Model	Akurasi	AUC	Recall	Precision	F1-Score
Random Forest Classifier	**0.8716**	**0.9304**	**0.9492**	**0.8783**	**0.9109**
K Neighbors Classifier	0.8425	0.8815	0.9485	0.8450	0.8920
Decision Tree Classifier	0.8144	0.7580	0.9061	0.8397	0.8693
Naive Bayes	0.7958	0.8524	0.8712	0.8443	0.8543
SVM - Linear Kernel	0.6752	0.0000	0.6659	0.7858	0.6428

DOI: https://doi.org/10.33020/saintekom.v14i1.512

Berdasarkan hasil penelitian pada Tabel 4, Model *Random Forest* menunjukkan kinerja terbaik dalam semua metrik yang diberikan dibandingkan model lainnya yang diuji pada penelitian ini. Nilai akurasi *Random Forest* sebesar 0,87 atau 87%, diikuti oleh KNN dengan nilai akurasi sebesar 0,84 (84%), *Decision Tree* 0,81 (81%), Naïve Bayes 0,79 (79%) sementara SVM dengan Kernel Linier memiliki kinerja yang paling rendah nilai akurasinya 0.67 atau 67%.

3.4. Evaluasi Model

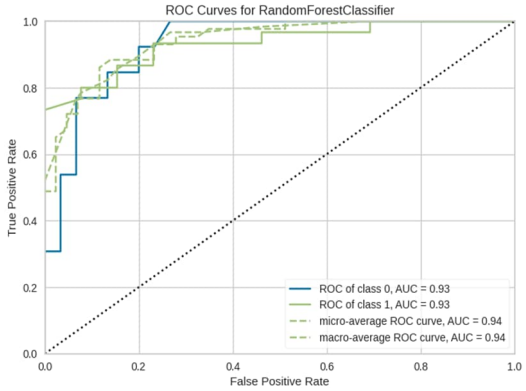
Setelah melihat hasil pada Tabel 4, performa model *Random Forest* memiliki nilai paling baik diantara model klasifikasi lainnya yang diuji pada penelitian ini. Berikut Gambar 3 nilai *confusion matriks.*



Gambar 3. *Confusion Matrix*

Pada Gambar 3 merupakan hasil *confusion matrix* dari model *random forest* dimana dari 43 data *testing* atau 20% dari total jumlah dataset berdasarkan hasil klasifikasi dapat diambil kesimpulan 10 data yang diprediksi benar hasilnya juga benar (sesuai), 3 data yang diprediksi benar tetapi hasilnya salah (tidak sesuai), 2 data yang diprediksi salah tetapi hasilnya benar (tidak sesuai) dan 28 data yang diprediksi salah hasilnya juga salah (sesuai).

Selanjutnya menampilkan nilai AUC/ROC yang dihasilkan oleh model. Hasilnya pada Gambar 4.

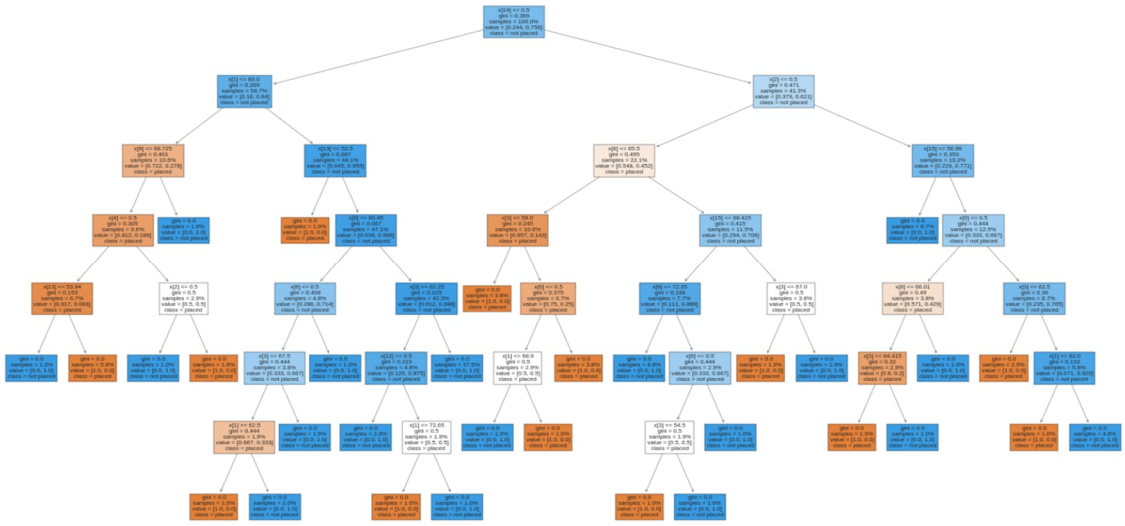


Gambar 4. *Curva AUC/ROC*

Gambar 4 menampilkan kurva AUC/ROC dari model *random forest* garis warna biru pada kurva mendekati angka 1 atau 100% karena nilai kinerja modelnya mencapai 93%, maka dari itu berdasarkan ketentuan nilai AUC/ROC kinerja model *random forest* dalam mengklasifikasikan data ketepatan penempatan karir bernilai sangat baik.

Visualisasi berdasarkan model terpilih yaitu *random forest* ditampilkan pada Gambar 5.

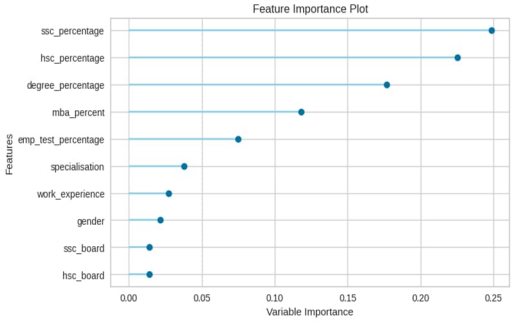
DOI: https://doi.org/10.33020/saintekom.v14i1.512



Gambar 5. Visualisasi Pohon Keputusan

Gambar 5 menunjukan prediksi yang kemungkinan terjadi pada kasus penempatan karir, warna biru tua menunjukan kelas *placed* (ditempatkan) dan warna orange tua menunjukan kelas *not placed* (tidak ditempatkan).

Selanjutnya melihat *Feature Importance yang* digunakan untuk menentukan atribut atau fitur mana dalam dataset yang paling berkontribusi terhadap kinerja model prediktif dengan *Random Forest* hasilnya pada Gambar 6.



Gambar 6. Visualisasi Fitur *Importance*

Gambar 6 menunjukan bahwa atribut *ssc_presentage* atau Persentase ujian sekolah menengah atas (Kelas 10) merupakan fitur paling penting dari penempatan karir berdasarkan dataset penelitian. Hasil ini menampilkan atribut paling penting yang ditemukan oleh model penelitian terpilih yang tidak disampaikan pada penelitian sebelumnya. Penelitian sebelumnya hanya berfokus pada hasil nilai akurasi dan AUC terbaik saja.

Pada hasil yang didapatkan berdasarkan pengujian model *Random forest* merupakan model terbaik dibandingkan dengan model lainnya, maka dari itu dapat disimpulkan bahwa karakteristik dataset penelitian ini memiliki klasifikasi dengan berbagai kriteria keputusan yang kompleks.

DOI: https://doi.org/10.33020/saintekom.v14i1.512

4. KESIMPULAN

Model klasifikasi *machine learning* diterapkan pada klasifikasi penempatan karir dengan jumlah sampel 215 data, Berdasarkan hasil penelitian menunjukan model *random forest* memiliki nilai akurasi paling unggul dengan nilai akurasi sebesar 0,87 atau 87% dan nilai AUC/ROC sebesar 0,93 dengan tingkat klasifikasi sangat baik, hal ini menunjukkan bahwa dataset dalam penelitian ini mengandung klasifikasi yang kompleks dengan beragam kriteria keputusan dan atribut *ssc_presentage* atau persentase ujian sekolah menengah atas kelas 10 di identifikasi sebagai fitur terpenting dalam penempatan karir berdasarkan hasil penelitian.

DAFTAR PUSTAKA

Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. *Jurnal SAINTEKOM*, *13*(1), 100–110. https://doi.org/10.33020/saintekom.v13i1.402

Anggraini, S., Akbar, M., Wijaya, A., Syaputra, H., & Sobri, M. (2021). Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Machine Learning. *Journal of Software Engineering Ampera*, *2*(1), 57–68. https://doi.org/10.51519/journalsea.v2i1.105

Aravind, T., Reddy, B. S., Avinash, S., & Jeyakumar, G. (2019). Information of Under Graduate Students. *Proceedings of the Third International Conference on I-SMAC*, 542–546.

Bao, Y., Peng, Y., & Wu, C. (2023). Deep Learning-Based Job Placement in Distributed Machine Learning Clusters With Heterogeneous Workloads. *IEEE/ACM Transactions on Networking*, *31*(2), 634–647. https://doi.org/10.1109/TNET.2022.3202529

Danuri, D., & Jaroji, J. (2019). Elisitasi Kebutuhan Perangkat Lunak Rekrutmen Pegawai Dengan Pendekatan Soft System Methodology. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, *10*(2), 42. https://doi.org/10.46964/justti.v10i2.110

Firdaus, I. A. (2022). Deteksi Infeksi Mycoplasma Pneumoniae Pneumonia Menggunakan Komparasi Algoritma Klasifikasi Machine Learning. *JOINTECS (Journal of Information Technology and Computer Science)*, *7*(1), 35. https://doi.org/10.31328/jointecs.v7i1.3242

Gata, W., Surohman, S., & Nawawi, H. M. (2023). *Twitter in analysis of policy sentiments of the omnibus law work creative design*. 020011. https://doi.org/10.1063/5.0128546

Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor,

DOI: https://doi.org/10.33020/saintekom.v14i1.512

Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. https://doi.org/10.21107/nero.v6i2.237

Kharisma, I. L., Septiani, D. A., Fergina, A., & Kamdan, K. (2023). Penerapan Algoritma Decision Tree untuk Ulasan Aplikasi Vidio di Google Play. *Jurnal Nasional Teknologi dan Sistem Informasi*, *9*(2), 218–226. https://doi.org/10.25077/TEKNOSI.v9i2.2023.218-226

Metalita, E., Handiyani, H., Afriani, T., & Rayatin, L. (2021). Analisis Jenjang Karir dan Minat Menjadi Perawat Intensif. *Jurnal Keperawatan Silampari*, *5*(1), 156–167. https://doi.org/10.31539/jks.v5i1.2907

Pessach, D., Singer, G., Avrahami, D., Chalutz Ben-Gal, H., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems*, *134*, 113290. https://doi.org/10.1016/j.dss.2020.113290

Pusporani, E., Qomariyah, S., & Irhamah, I. (2019). Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning. *Inferensi*, *2*(1), 25. https://doi.org/10.12962/j27213862.v2i1.6810

Ramezan, C. A., Warner, T. A., & Maxwell, A. E. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, *11*(2). https://doi.org/10.3390/rs11020185

Raza, A. (2023). *Job Placement Dataset* [dataset]. https://www.kaggle.com/datasets/ahsan81/job-placement-dataset?resource=download

Siahaan, R. A., Nasution, M., & Hasibuan, M. N. S. (2021). Model Data Mining untuk Perancangan Aplikasi Diagnostik Inflammatory Liver Disease. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 145–153. https://doi.org/10.54367/jtiust.v6i1.1277

Sihombing, P. R., & Yuliati, I. F. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *20*(2), 417–426. https://doi.org/10.30812/matrik.v20i2.1174

Sukmawati, S., Sulastri, S., Februariyanti, H., & Jananto, A. (2022). Perbandingan Algoritma C4.5 Dan Algoritma Naive Bayes Untuk Klasifikasi Pekerja Migran Indonesia. *I N F O R M A T I K A*, *14*(1), 7. https://doi.org/10.36723/juri.v14i1.280

Umar, R., Fadlil, A., & Yuminah, Y. (2018). Sistem Pendukung Keputusan dengan Metode AHP untuk Penilaian Kompetensi Soft Skill Karyawan. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, *4*(1), 27–34. https://doi.org/10.23917/khif.v4i1.5978

Visa, S., Ramsay, B., & Ralescu, A. (n.d.). *Confusion Matrix-based Feature Selection*.

DOI: https://doi.org/10.33020/saintekom.v14i1.512