

Reduksi False Positive Pada Klasifikasi Job Placement dengan Hybrid Random Forest dan Auto Encoder

M. Riza Pahlevi¹, Errissya Rasywir^{1*}, Yovi Pratama¹, Fachruddin², Marrylinteri Istoningtyas¹, Muhammad Yaasin²

¹ Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

² Fakultas Ilmu Komputer, Program Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹rizapahlevi@unama.ac.id, ²errissya.rasywir@gmail.com, ³yovi.pratama@gmail.com, ⁴fachruddin.stikom@gmail.com,

⁴marrylinteri.jurnal@gmail.com, ⁵m.yaasin@gmail.com

Email Penulis Korespondensi: errissya.rasywir@gmail.com

Submitted: 23/01/2024; Accepted: 30/03/2024; Published: 30/03/2024

Abstrak— Interpretasi False Positive (FP) menunjukkan hasil prediksi negatif dan merupakan jawaban kesalahan tipe 1 dengan hasil prediksi positif yang salah. Berdasarkan ini, kami mencoba melakukan reduksi untuk kesalahan tipe 1 untuk meningkatkan nilai akurasi dari hasil klasifikasi. Tingkat FP yang rendah sangat penting untuk penggunaan sistem Computer Aided Detection (CAD). Dalam usulan penelitian ini untuk mereduksi FP kami menggunakan rancangan hasil evaluasi Random Forest (RF) yang akan diinterpretasikan ulang oleh algoritma Auto Encoder (AE). Pemilihan algoritma RF dipilih karena sebagai jenis ensemble learning yang dapat mengoptimalkan akurasi dengan paralel. RF dipilih karena melakukan bagging terhadap seluruh output Decision Tree (DT) yang digunakan. Untuk menekan reduksi TP lebih kuat kami menggunakan algoritma Auto Encoder (AE) untuk memproses ulang kelas hasil bagging dari RF menjadi input dalam layer AE. AE menggunakan kesalahan rekonstruksi, yang dalam hal ini adalah klasifikasi Job Placement. Dari hasil pengujian, diperoleh bahwa mengkombinasikan penggunaan random forest yang menggunakan C4.5 sebagai decision tree dengan Autoencoder dapat meningkatkan akurasi pada task Klasifikasi Job Placement dengan selisih 0.004652 lebih baik dibanding tanpa dikombinasikan dengan auto encoder. Selain itu, pada pengujian menggunakan kombinasi RF dan AE dihasilkan lebih sedikit nilai False Positive (FP) yakni sebesar 11 item pada Pengujian Cross Validation-5 (CV-5), selanjutnya 13 item pada pengujian Cross Validation-10 (CV-10) serta pada pengujian split data training sebesar 60% dihasilkan FP sebesar 12 saja. Nilai tersebut lebih sedikit dibandingkan dengan False Positive yang dihasilkan pengujian tanpa Autoencoder yakni 12 item pada CV-5, sebesar 15 item pada CV-10, dan sebesar 13 pada split data training.

Kata Kunci: Klasifikasi; Job Placement; Random Forest; Autoencoder; Hybrid

Abstract— The False Positive (FP) interpretation shows a negative prediction result and is a type 1 error answer with an incorrect positive prediction result. Based on this, we try to reduce type 1 errors to increase the accuracy value of the classification results. A low FP rate is critical for the use of Computer Aided Detection (CAD) systems. In this research proposal, to reduce FP, we use a Random Forest (RF) evaluation result design which will be reinterpreted by the Auto Encoder (AE) algorithm. The RF algorithm was chosen because it is a type of ensemble learning that can optimize accuracy in parallel. RF was chosen because it performs bagging on all Decision Tree (DT) outputs used. To suppress TP reduction more strongly, we use the Auto Encoder (AE) algorithm to reprocess the class bagging results from RF into input in the AE layer. AE uses reconstruction errors, which in this case is Job Placement classification. From the test results, it was found that combining the use of a random forest using C4.5 as a decision tree with an Autoencoder can increase accuracy in the Job Placement Classification task by a difference of 0.004652 better than without combining it with an autoencoder. Apart from that, in testing using a combination of RF and AE, fewer False Positive (FP) values were produced, namely 11 items in the Cross Validation-5 (CV-5) Test, then 13 items in the Cross Validation-10 (CV-10) test and in testing split training data of 60%, the FP was only 12. This value is less than the false positives produced by testing without Autoencoder, namely 12 items on CV-5, 15 items on CV-10, and 13 on split training data.

Keywords: Classification; Job Placement; Random Forest; Autoencoder; Hybrid

1. PENDAHULUAN

Machine Learning membangun proses komputer untuk belajar dari data [1], [2]. Komputer butuh data untuk bisa belajar. Salah satu teknik belajar komputer (machine learning) adalah supervised learning. Klasifikasi salah satu model prediksi yang bersifat diskrit. Pengukuran kinerja dapat dilakukan dengan membandingkan nilai aktual dengan nilai prediksi [3]–[5]. Confusion Matrix mengukur kinerja klasifikasi yang memiliki 4 kombinasi berbeda dari nilai prediksi dan nilai actual antara lain True Positif (TP), True Negatif (TN), False Positive (FP), dan False Negatif (FN) [3], [6], [7]. TP memiliki interpretasi bahwa hasil prediksi positif yang benar. TN dapat diinterpretasi dengan hasil prediksi negatif yang benar. FP diklasifikasikan sebagai kesalahan tipe 1 dengan hasil prediksi positif yang salah. Sedangkan, FN dikenal sebagai kesalahan tipe 2, yang merupakan tipe yang berbahaya dalam mengevaluasi hasil prediksi. Interpretasi FP adalah bahwa hasil prediksi negatif dan merupakan jawaban yang salah [8]–[11]. Berdasarkan ini, kami mencoba melakukan reduksi untuk kesalahan tipe 1 untuk meningkatkan nilai akurasi dari hasil klasifikasi.

Tingkat FP yang rendah sangat penting untuk penggunaan sistem Computer Aided Detection (CAD). Dalam sebuah sistem yang diperlukan bukan hanya akurasi yang tinggi, tapi juga dibutuhkan akurasi tanpa tingkat kesalahan dalam prediksi [12]–[14]. Pada zaman berbasis teknologi ini hampir semua kegiatan berbasis Komputer, salah satunya Job Placement. Sebuah Manajemen Sumber Daya Manusia (MSDM) memiliki proses pencapaian tujuan yang organisasi [15]–[17]. MSDM merupakan sistem pendayagunaan, penilaian, dan pengelolaan secara efektif dan efisien untuk mencapai tujuan [18]–[20]. Diperlukan proses perencanaan, pengorganisasian, dan pengawasan terhadap pengadaan, pemberian kompensasi serta pemeliharaan untuk mencapai tujuan. Meningkatnya kebutuhan akan



individu yang berpendidikan dan berbakat, terutama di negara berkembang, merekrut lulusan baru merupakan praktik rutin bagi organisasi [21]–[23]. Metode rekrutmen dan proses seleksi konvensional dapat rentan terhadap kesalahan dan untuk mengoptimalkan keseluruhan proses, diperlukan beberapa metode inovatif [19], [23]–[25]. Salah satunya adalah dengan menerapkan machine learning dalam MSDM salah satunya adalah dengan klasifikasi Job Placement .

Penelitian ini dilakukan dengan menguji suatu metode usulan yang kami rancang untuk mengurangi tingkat FP dalam melakukan Job Placement . Penelitian terkait reduksi FP telah dilakukan antara lain oleh Zhenghao Shi et al (2019) yang melakukan penelitian dengan CNN yang disetel untuk melatih pengklasifikasi SVM. Output dari SVM yang dilatih digunakan untuk klasifikasi akhir [10]. Selanjutnya, Zhan Wu et al (2020) melakukan penelitian dengan hasil percobaan menunjukkan klasifikasi dengan skor CPM 0,9008 artinya deteksi nodul paru efisien, akurat, dan andal untuk diagnosis klinis [26]. Dandan Zhao et al (2022) melakukan penelitian dengan hasil eksperimen menunjukkan bahwa kerangka kerja CNN multiskala yang diusulkan dapat mengurangi waktu pelatihan dari 36 jam menjadi 8 jam dan memastikan kinerja nodul positif sejati dengan informasi domain spasial dan frekuensi terintegrasi [6]. Selain itu, kerangka kerja CNN multiskala yang diusulkan mencapai sensitivitas 95,2% dan 98,1% [6]. Juezhao Yu, et al (2020) melakukan penelitian dengan dataset yang terdiri dari 1600 pemeriksaan CT dengan Arsitektur CNN multiskala yang dikembangkan memiliki kinerja yang lebih baik daripada arsitektur vanilla [7]. SA El-Regaily et al (2020) melakukan penelitian dengan algoritma CNN yang diusulkan untuk deteksi Node pada paru yang mencapai sensitivitas deteksi tinggi 85,256%, spesifisitas 90,658%, dan akurasi 89,895% [6]. Hasil percobaan menunjukkan bahwa algoritma yang diusulkan mengungguli sebagian besar algoritma lain dalam hal akurasi dan sensitivitas. Solusi yang diusulkan mencapai pertukaran yang baik antara efisiensi dan efektivitas dan menghemat banyak waktu perhitungan.

Dalam usulan penelitian ini untuk mereduksi FP kami menggunakan rancangan hasil evaluasi Random Forest (RF) yang akan diinterpretasikan ulang oleh algoritma Auto Encoder (AE). Pemilihan algoritma RF dipilih karena sebagai jenis ensemble learning yang dapat mengoptimalkan akurasi dengan parallel. RF dipilih karena melakukan bagging terhadap seluruh output Decision Tree (DT) yang digunakan. Secara teoritis proses bagging dalam RF merupakan salah satu Tindakan untuk mereduksi TP. Random forest untuk klasifikasi data yang besar. RF merupakan algoritma pembelajaran mesin yang menggabungkan keluaran dari beberapa pohon keputusan untuk sampai pada satu hasil. Kemudahan penggunaan dan fleksibilitasnya telah memicu pengadopsiannya, karena menangani masalah klasifikasi dan regresi. Kelebihan RF mampu mengatasi noise dan missing value dalam jumlah yang besar. Dan kekurangan RF yaitu interpretasi yang sulit serta optimalisasi tuning model yang tepat.

Untuk menekan reduksi TP lebih kuat kami menggunakan algoritma Auto Encoder (AE) untuk memproses ulang kelas hasil bagging dari RF menjadi input dalam layer AE. AE menggunakan kesalahan rekonstruksi, yang dalam hal ini adalah klasifikasi Job Placement pada dataset Kaggle. Hipotesis dari proses Encoder dalam AE yakni di mana satu sampel dengan skor lebih tinggi dapat diklasifikasi sebagai kelas dari klasifikasi Job Placement pada dataset Kaggle. Adapun kelas yang akan di bagging dalam RF dan di encode ulang oleh AE adalah **Placed** dan **Not Placed**. Berdasarkan masalah di atas maka penelitian ini akan menguji sebuah proposed metode untuk mereduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

2. METODOLOGI PENELITIAN

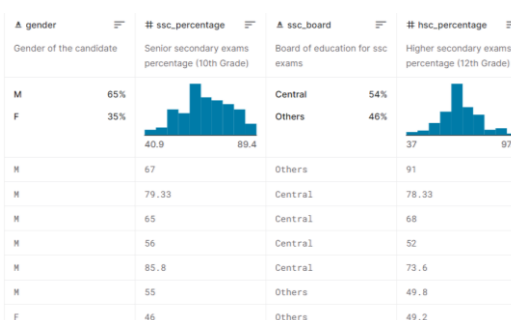
2.1 Kerangka Penelitian

Kerangka kerja penelitian Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder dapat dilihat pada sub bagian dibawah.

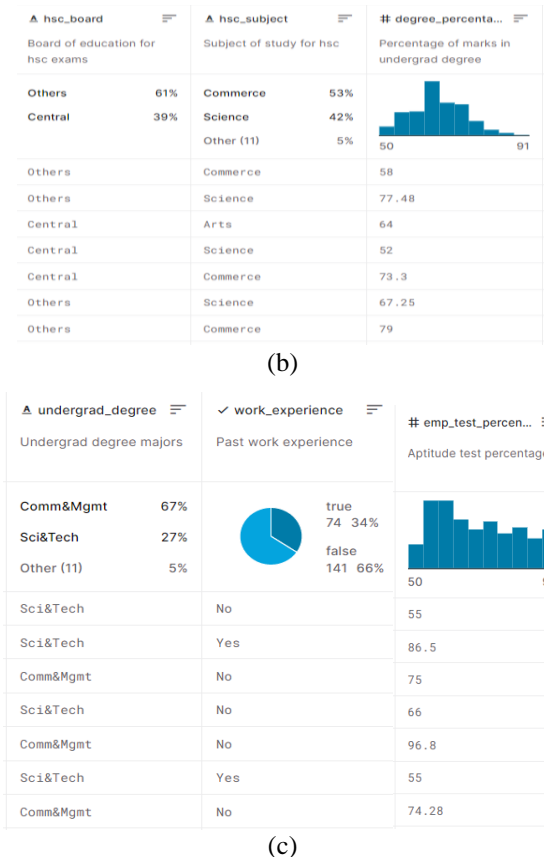
a. Penelitian Awal

Masalah inisial adalah bagaimana hasil Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder. Teori yang diambil tentang Random Forest, Auto Encoder dan Evaluasi dengan Confussion Matrix.

b. Pengumpulan Data Job Placement.



(a)

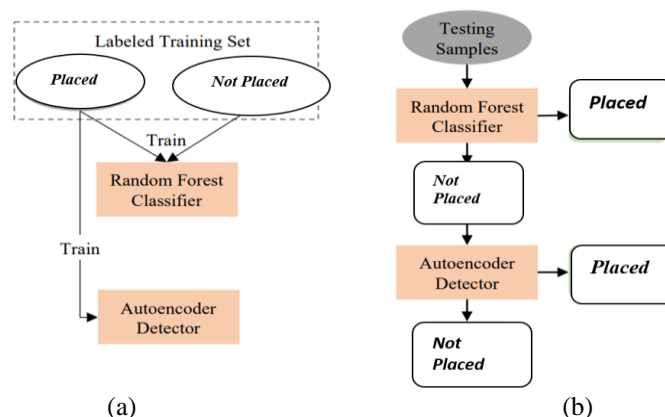


Gambar 1. Dataset Klasifikasi Job Placement pada dataset Kaggle (a) Atribut Gender, SSC Percentage, SSC board, HSC Percentage (b) Atribut Other Central, Commerce Science (c) Atribut Undergrad degree, Work Experience & Emp Test Percentage

Gambar 1 di atas adalah visualisasi atribut dari Dataset Klasifikasi Job Placement pada dataset Kaggle. Atribut-atribut pada dataset Job Placement antara lain adalah Atribut Gender yang mendeskripsikan jenis kelamin pekerja, SS (senior & secondary exam) mendeskripsikan nilai sekolah pekerja, Percentage yang menggambarkan mendeskripsikan persentasi nilai sekolah pekerja, SSC board mendeskripsikan nilai board education pekerja, HSC Percentage (higher secondary exam) yang mendeskripsikan nilai sekolah pendidikan tinggi pekerja, Other Central, Commerce Science, Undergrad degree, Work Experience & Emp Test (hasil nilai ujian aptitude) Percentage.

c. Merancang & Implementasi Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

Dalam penelitian ini, ditargetkan skenario di mana telah dikumpulkan kelas Job Placement pada dataset Kaggle, tetapi mungkin ada beberapa klasifikasi Job Placement yang tidak diketahui sebagai variasi klasifikasi Job Placement pada dataset Kaggle atau baru jenis lain. Ketika ada klasifikasi Job Placement pada dataset Kaggle yang tidak diketahui, pengklasifikasi RF tradisional akan salah mengklasifikasikannya sebagai Placed, jadi digunakan pendekatan probabilistik untuk membuatnya.



Gambar 2. Rancangan Proposed Methode Random Forest dan Auto Encoder untuk Reduksi False Positive (a) Tahapan Training (b) Tahapan Testing



Gambar 2 di atas adalah gambaran umum dari metode yang diusulkan. (a) proses pelatihan; (b) proses pengujian. Karena AE hanya melatih data Placed, sampel Placed akan memiliki data yang lebih rendah MSE daripada yang klasifikasi dengan kedua kelas pada dataset Job Placement . Dari sudut pandang ini, kita dapat menentukan ambang threshold yang lebih rendah, dan sampel yang lebih rendah dari ambang ini memiliki tingkat akurasi yang lebih tinggi bahwa mereka termasuk dalam kelas Placed. Berdasarkan asumsi ini, diintegrasikan kedua proses keputusan ini. Setelah mendapatkan model yang dilatih, selama fase pengujian, diterapkan dua langkah strategi klasifikasi. Dalam usulan dicantumkan prosedur klasifikasi dalam algoritma. Ada dua hyperparameter dipertimbangkan untuk keputusan tersebut. Yang pertama adalah T1, digunakan untuk probabilitas RF, dan yang kedua satu adalah T2, digunakan untuk MSE. Pada awalnya, sampel xi diklasifikasikan menggunakan RF classifier. Ketika probabilitas sampel lebih besar dari T1, itu akan diklasifikasikan sebagai klasifikasi Job Placement pada dataset Kaggle. Setelah itu, digunakan AE untuk memeriksa kembali sampel klasifikasi Job Placement pada dataset Kaggle yang diprediksi oleh RF. Di bagian ini, ketika kesalahan rekonstruksi lebih kecil dari T2, sampel direklasifikasi menjadi normal. Dalam pendekatan dua langkah ini, sampel dalam tahap pengujian dapat diklasifikasikan dengan benar, terutama sampel klasifikasi Job Placement pada dataset Kaggle yang salah klasifikasi.

d. Melakukan Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder

Proses ini adalah eksekusi dari alur penelitian yang telah dibangun dari gambar 6 dengan menggunakan Bahasa pemrograman Pyhton.

e. Hasil awal Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

Kegiatan ini pelaporan hasil Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

f. Progress Report eksperimen.

Kegiatan ini dilakukan untuk melaporkan langkah kegiatan apa saja yang telah terhadap Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

g. Evaluasi Akhir Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

Pada kegiatan ini dilakukan evaluasi akhir terhadap Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder.

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisi hasil dan pembahasan dari topik penelitian, yang bisa di buat terlebih penerapan metode yang digunakan, baik secara sederhana dengan mengemukakan data yang ada pada penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, tabel dan lainnya mengenai Hasil Evaluasi klasifikasi Job Placement Dengan Model Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

3.1 Hasil Pengujian Klasifikasi Job Placement dengan Model Hybrid Random Forest dan Autoencoder

Berikut ditampilkan Hasil Evaluasi klasifikasi Job Placement Dengan Model Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning) . Parameter yang digunakan adalah Recall Sensitivity True Positive Rate (TPR), False Positive Rate (FPR) False Alarm rate, Specificity True Negative Rate (TNR), Precision, False Negative Rate (FNR) dan Accuracy.

Tabel 1. Hasil Evaluasi klasifikasi Job Placement Dengan Model Random Forest (Origin)

Parameter	Formula	Split 60%	CV -10	CV-5
Recall Sensitivity True Positive Rate (TPR)	$TP / (FN+TP)$	0.916256	0.944162	0.920792
False Positive Rate (FPR) False Alarm rate	$FP / (TN +FP)$	0.928571	0.833333	0.923077
Specificity True Negative Rate (TNR)	$TN/(TN +FP)$	0.071429	0.166667	0.076923
Precision	$TP/(TP +FP)$	0.934673	0.925373	0.939394
False Negative Rate (FNR)	$FN / (FN +TP)$	0.083744	0.055838	0.079208
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	0.861751	0.87907	0.869767

Tabel 1 di atas adalah Hasil Evaluasi klasifikasi Job Placement Dengan Model Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Pada hasil tersebut nilai TPR tertinggi dihasilkan oleh klasifikasi dengan model pengujian Cross-Validation sebesar 10 (CV-10) yakni sebesar 0.944162, selanjutnya nilai FPR tertinggi



dihasilkan oleh model pengujian dengan Split data test sebesar 60% yakni sebesar 0.928571. Untuk akurasi keseluruhan dihasilkan oleh pengujian (CV-10) sebesar 0.87907.

Tabel 2. Hasil Evaluasi klasifikasi Job Placement Dengan Model Hybrid Random Forest dan Autoencoder

Parameter	Formula	Split 60%	CV 10	CV 5
Recall Sensitivity True Positive Rate (TPR)	$TP / (FN+TP)$	0.916256	0.934673	0.920792
False Positive Rate (FPR)	$FP / (TN +FP)$	0.857143	0.8125	0.846154
False Alarm rate				
Specificity True Negative Rate (TNR)	$TN/(TN +FP)$	0.142857	0.1875	0.153846
Precision	$TP/(TP +FP)$	0.939394	0.934673	0.944162
False Negative Rate (FNR)	$FN / (FN +TP)$	0.083744	0.065327	0.079208
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	0.866359	0.87907	0.874419

Tabel 2 di atas adalah Hasil Evaluasi klasifikasi Job Placement Dengan Model Hybrid Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Pada hasil tersebut nilai TPR tertinggi dihasilkan oleh klasifikasi dengan model pengujian Cross-Validation sebesar 10 (CV-10) yakni sebesar 0.934673, selanjutnya nilai FPR tertinggi dihasilkan oleh model pengujian dengan Split data test sebesar 60% yakni sebesar 0.857143. Untuk akurasi keseluruhan dihasilkan oleh pengujian (CV-10) sebesar 0.87907.

Tabel 3. Hasil Convolution Matrix (Test Mode : Split 60%) Dengan Random Forest (Origin)

Dataset: 215 Job Placement			
Predicted			
		Negatif	Positif
Actual	Negatif	1	13
	Positif	17	186

Tabel 3 di atas adalah Hasil Convolution Matrix (Test Mode : Split 60%) Dengan Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 4. Hasil Convolution Matrix (Test Mode : CV: 10) Dengan Random Forest (Origin)

Dataset: 215 Job Placement			
Predicted			
		Negatif	Positif
Actual	Negatif	3	15
	Positif	11	186

Tabel 4 di atas adalah Hasil Convolution Matrix (Test Mode : CV: 10) Dengan Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 5. Hasil Convolution Matrix (Test Mode : CV: 5) Dengan Random Forest (Origin)

Dataset: 215 Job Placement			
Predicted			
		Negatif	Positif
Actual	Negatif	1	12
	Positif	16	186

Tabel 5 di atas adalah Hasil Convolution Matrix (Test Mode : CV: 5) Dengan Random Forest (Origin) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 6. Hasil Convolution Matrix (Test Mode : Split 60%) Dengan Hybrid Random Forest dan Autoencoder

Dataset: 215 Job Placement			
Predicted			
		Negatif	Positif
Actual	Negatif	2	12
	Positif	17	186

Tabel 6 di atas adalah Hasil Convolution Matrix (Test Mode : Split 60) Dengan Hybrid Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 7. Hasil Convolution Matrix (Test Mode: CV-10) Dengan Hybrid Random Forest dan Autoencoder

Dataset: 215 Job Placement			
		Predicted	
		Negatif	Positif
Actual	Negatif	3	13
	Positif	13	186

Tabel 7 di atas adalah Hasil Convolution Matrix (Test Mode : CV 10) Dengan Hybrid Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 8. Hasil Convolution Matrix (Test Mode : CV 5) Dengan Hybrid Random Forest dan Autoencoder

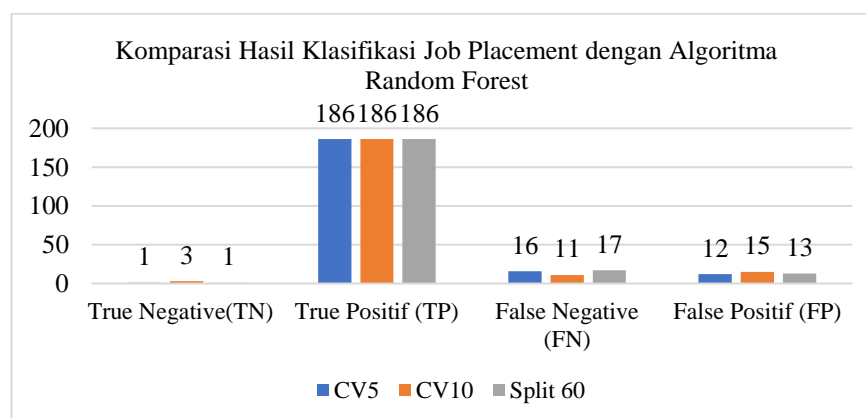
Dataset:215 Job Placement			
		Predicted	
		Negatif	Positif
Actual	Negatif	2	11
	Positif	16	186

Tabel 8 di atas adalah Hasil Convolution Matrix (Test Mode : CV-5) Dengan Hybrid Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 9. Komparasi Hasil Klasifikasi Job Placement dengan Algoritma Random Forest

Hasil Klasifikasi	CV5	CV10	Split 60
True Negative(TN)	1	3	1
True Positif (TP)	186	186	186
False Negative (FN)	16	11	17
False Positive (FP)	12	15	13

Tabel 9 di atas adalah Komparasi Hasil Klasifikasi Job Placement dengan Algoritma Random Forest pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Pada pengujian menggunakan Kombinasi Random Forest dan Autoencoder dihasilkan lebih sedikit nilai False Positive (FP) yakni sebesar 12 item pada Pengujian Cross Validation sebesar 5 %, selanjutnya 15 item pada pengujian Cross Validation sebesar 10% serta pada pengujian split data training sebesar 60% dihasilkan FP sebesar 13 saja.



Gambar 3. Komparasi Hasil Klasifikasi Job Placement dengan Algoritma Random Forest

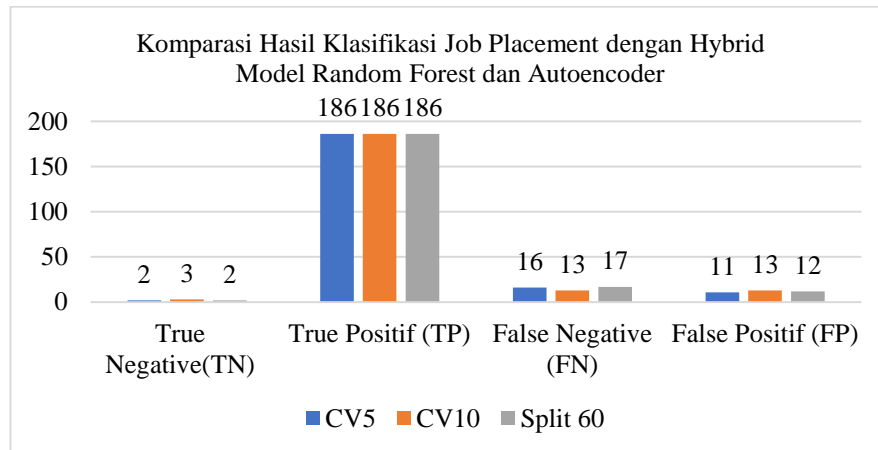
Gambar 3 di atas adalah Komparasi Hasil Klasifikasi Job Placement dengan Algoritma Random Forest pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 10. Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Algoritma Random Forest dan Autoencoder

Hasil Klasifikasi	CV5	CV10	Split 60
True Negative(TN)	2	3	2

True Positif (TP)	186	186	186
False Negative (FN)	16	13	17
False Positif (FP)	11	13	12

Tabel 10 di atas adalah Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Algoritma Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Pada pengujian menggunakan Kombinasi Random Forest dan Autoencoder dihasilkan lebih sedikit nilai False Positive (FP) yakni sebesar 11 item pada Pengujian Cross Validation sebesar 5 %, selanjutnya 13 item pada pengujian Cross Validation sebesar 10% serta pada pengujian split data training sebesar 60% dihasilkan FP sebesar 12 saja.



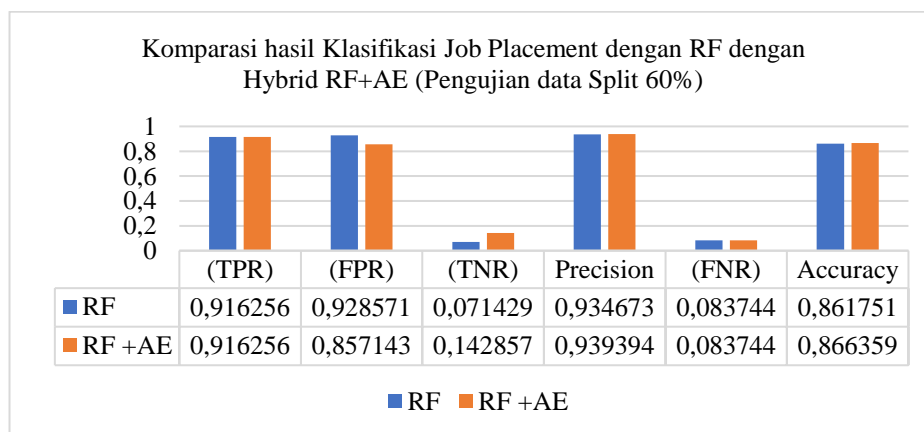
Gambar 4. Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Autoencoder

Gambar 4 adalah Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 11. Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Autoencoder

Evaluasi	Split 60%		
	RF	RF +AE	Delta
Parameter (TPR)	0.916256	0.916256	0
(FPR)	0.928571	0.857143	-0.07143
(TNR)	0.071429	0.142857	0.071428
Precision	0.934673	0.939394	0.004721
(FNR)	0.083744	0.083744	0
Accuracy	0.861751	0.866359	0.004608

Tabel 11 di atas adalah Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).



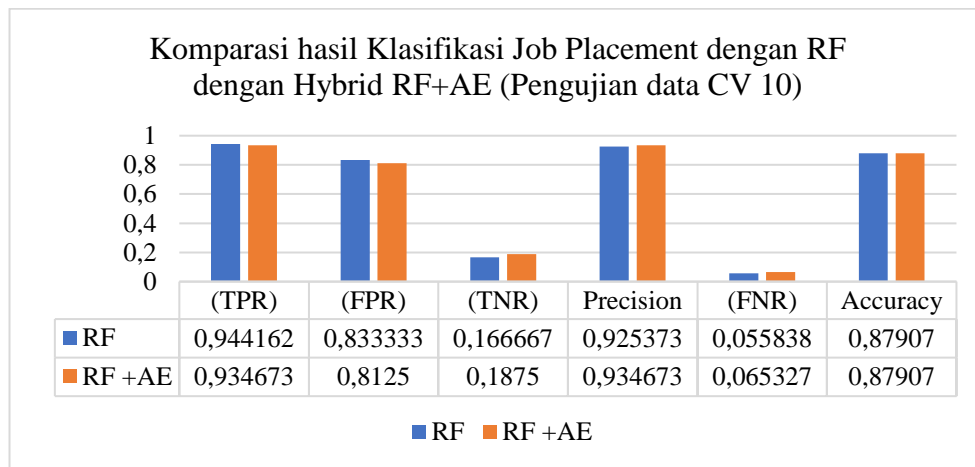
Gambar 5. Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%)

Gambar 5 di atas adalah Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 12. Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%)

Evaluasi	CV- 10		
	Parameter	RF	RF +AE
(TPR)	0.944162	0.934673	-0.00949
(FPR)	0.833333	0.8125	-0.02083
(TNR)	0.166667	0.1875	0.020833
Precision	0.925373	0.934673	0.0093
(FNR)	0.055838	0.065327	0.009489
Accuracy	0.87907	0.87907	0

Tabel 12 di atas adalah Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).



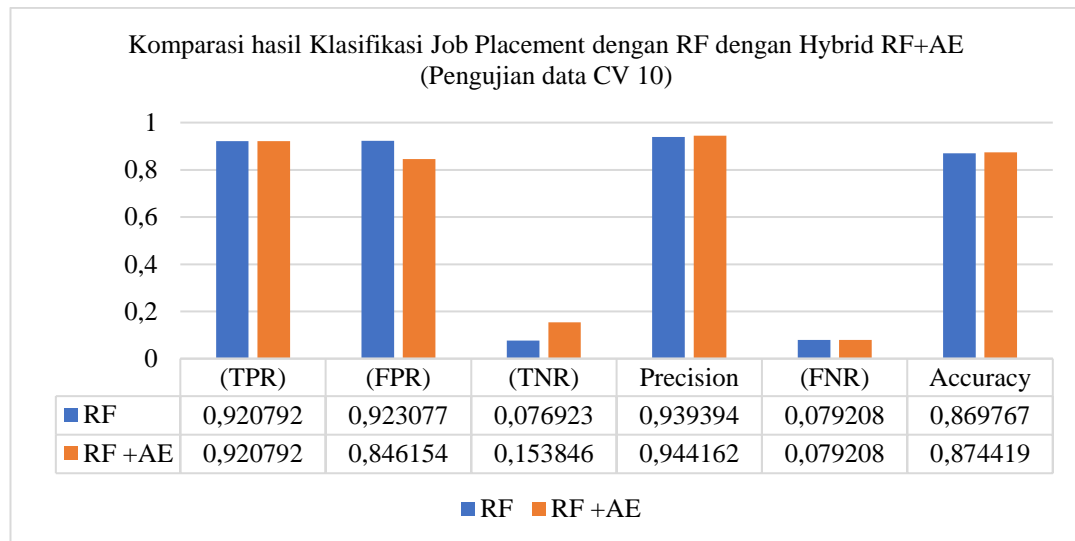
Gambar 6. Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%)

Gambar 6 di atas adalah Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).

Tabel 13. Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%)

Evaluasi	CV 5		
	Parameter	RF	RF +AE
(TPR)	0.920792	0.920792	0
(FPR)	0.923077	0.846154	-0.07692
(TNR)	0.076923	0.153846	0.076923
Precision	0.939394	0.944162	0.004768
(FNR)	0.079208	0.079208	0
Accuracy	0.869767	0.874419	0.004652

Tabel 13 di atas adalah Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Pengujian data Split 60%) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning).



Gambar 7. Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Penguujian data CV 10)

Gambar 7 di atas adalah Komparasi hasil Klasifikasi Job Placement dengan RF dengan Hybrid RF+AE (Penguujian data CV 10) pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Dari Tabel 14 dan gambar 11 dapat disimpulkan bahwa mengkombinasikan penggunaan random forest yang menggunakan C4.5 sebagai decision tree dengan Autoencoder dapat meningkatkan akurasi pada task Klasifikasi Job Placement dengan selisih 0.004652 lebih baik disbanding tanpa dikombinasikan dengan autoencoder.

4. KESIMPULAN

Pemilihan algoritma Random Forest (RF) sebagai jenis ensemble learning dapat mengoptimalkan akurasi dengan paralel. RF dipilih karena melakukan bagging terhadap seluruh output Decision Tree (DT) yang digunakan. Untuk menekan reduksi False Positive (FP) lebih kuat, penelitian ini menggunakan algoritma Auto Encoder (AE) untuk memproses ulang kelas hasil bagging dari RF menjadi input dalam layer AE. AE menggunakan kesalahan rekonstruksi, yang dalam hal ini adalah klasifikasi Job Placement. Dari hasil pengujian, diperoleh bahwa mengkombinasikan penggunaan random forest yang menggunakan C4.5 sebagai decision tree dengan Autoencoder dapat meningkatkan akurasi pada task Klasifikasi Job Placement dengan selisih 0.004652 lebih baik dibanding tanpa dikombinasikan dengan autoencoder. Selain itu, pada pengujian menggunakan kombinasi RF dan AE dihasilkan lebih sedikit nilai False Positive (FP) yakni sebesar 11 item pada Pengujian Cross Validation-5 (CV-5), selanjutnya 13 item pada pengujian Cross Validation-10(CV-10) serta pada pengujian split data training sebesar 60% dihasilkan FP sebesar 12 saja. Nilai tersebut lebih sedikit dibandingkan dengan False Positive yang dihasilkan pengujian tanpa Autoencoder yakni 12 item pada CV-5, sebesar 15 item pada CV-10, dan sebesar 13 pada split data training. Berdasarkan Komparasi Hasil Klasifikasi Job Placement dengan Hybrid Algoritma Random Forest dan Autoencoder pada Reduksi Nilai False Positive Pada Klasifikasi Job Placement dengan Hybrid Model Random Forest dan Auto Encoder (Ensemble of multimodal Deep Learning). Pada pengujian menggunakan Kombinasi Random Forest dan Autoencoder dihasilkan lebih sedikit nilai False Positive (FP) yakni sebesar 11 item pada Pengujian Cross Validation sebesar 5%, selanjutnya 13 item pada pengujian Cross Validation sebesar 10% serta pada pengujian split data training sebesar 60% dihasilkan FP sebesar 12 saja. Nilai tersebut lebih sedikit dibandingkan dengan False Positive yang dihasilkan pengujian tanpa Autoencoder yakni 12 item pada CV-5, sebesar 15 item pada CV-10, dan sebesar 13 pada Split Datatest sebesar 60%.

REFERENCES

- [1] D. Z. Abidin, S. Nurmaini, Erwin, E. Rasywir, and Y. Pratama, "Indoor Positioning System in Learning Approach Experiments," *J. Electr. Comput. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6592562.
- [2] C. C. Lee, V. C. Koo, T. S. Lim, Y. P. Lee, and H. Abidin, "A multi-layer perceptron-based approach for early detection of BSR disease in oil palm trees using hyperspectral images," *Heliyon*, vol. 8, no. 4, p. e09252, 2022, doi: 10.1016/j.heliyon.2022.e09252.
- [3] E. Rasywir, R. Sinaga, and Y. Pratama, "Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network (CNN)," *J. Paradig. UBSI*, vol. 22, no. 2, pp. 117-123, 2020.
- [4] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, no. July 2018, pp. 70-90, 2018, doi: 10.1016/j.compag.2018.02.016.
- [5] H. Bunyamin, Heriyanto, S. Novianti, and L. Sulistiani, "Topic clustering and classification on final project reports: A comparison of traditional and modern approaches," *IAENG Int. J. Comput. Sci.*, vol. 46, no. 3, pp. 1-6, 2019.



- [6] D. Zhao, Y. Liu, H. Yin, and Z. Wang, “A novel multi-scale CNNs for false positive reduction in pulmonary nodule detection,” *Expert Syst. Appl.*, vol. 207, p. 117652, 2022, doi: <https://doi.org/10.1016/j.eswa.2022.117652>.
- [7] J. Yu, B. Yang, J. Wang, J. K. Leader, D. O. Wilson, and J. Pu, “2D CNN versus 3D CNN for false-positive reduction in lung cancer screening,” *J. Med. Imaging*, vol. 7, no. 5, p. 51202, 2020, doi: 10.1117/1.JMI.7.5.051202.
- [8] F. Wulandari, P. A. Jusia, and J. Jasmir, “Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan,” *J. Manaj. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 214–227, 2020.
- [9] I. K. S. Buana, “DETEKSI GERAKAN KEPALA DAN KEDIPAN MATA DENGAN HAAR CASCADE CLASSIFIER CONTOUR DAN MORFOLOGI DALAM PENGOPERASIAN KOMPUTER UNTUK KAUM DIFABLE,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 1, pp. 29–36, 2019, doi: 10.33330/jurteksi.v5i1.273.
- [10] Z. Wu *et al.*, “MD-NDNet: a multi-dimensional convolutional neural network for false-positive reduction in pulmonary nodule detection,” *Phys. Med. Biol.*, vol. 65, no. 23, p. 235053, Dec. 2020, doi: 10.1088/1361-6560/aba87c.
- [11] J. Zeng, K. Ben, X. Li, and X. Zhang, “Fast code clone detection based on weighted recursive autoencoders,” *IEEE Access*, vol. 7, pp. 125062–125078, 2019, doi: 10.1109/ACCESS.2019.2938825.
- [12] F. Qin, T. Zuo, and X. Wang, “Ccpso: Wifi fingerprint indoor positioning system based on cdae-cnn,” *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–17, 2021, doi: 10.3390/s21041114.
- [13] A. Labrada and B. D. Barkana, “A Comprehensive Review of Computer-Aided Models for Breast Cancer Diagnosis Using Histopathology Images,” *Bioengineering*, vol. 10, no. 11, p. 1289, 2023, doi: 10.3390/bioengineering10111289.
- [14] P. Jiang *et al.*, *A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis*, vol. 56, no. s2. Springer Netherlands, 2023.
- [15] M. A. C. Perdana, N. W. Sulistyowati, A. Ninasari, Jainudin, and S. Mokodenseho, “Analisis Pengaruh Pembiayaan, Skala Usaha, dan Ketersediaan Sumber Daya Manusia terhadap Profitabilitas UMKM,” *Sanskara Ekon. dan Kewirausahaan*, vol. 1, no. 03, pp. 135–148, 2023, doi: 10.58812/sek.v1i03.120.
- [16] Fachrudin, M. R. Pahlevi, M. Ismail, E. Rasywir, and Y. Pratama, “Analisis Usability Pada Implementasi Sistem Pengelolaan Keuangan Masjid Menggunakan USE Questionnaire,” *J. Media Inform. Budidarma*, vol. 4, pp. 1216–1224, 2020, doi: 10.30865/mib.v4i4.2518.
- [17] G. A. Gulo, N. W. Nurhidayati, D. Aprillia, and M. Maruloh, “Sistem Penunjang Keputusan Pemilihan Karyawan Terbaik Di Restoran Soto Pak J Menggunakan Metode Ahp,” *J. Larik Ldng. Artik. Ilmu Komput.*, vol. 1, no. 1, pp. 11–20, 2021, doi: 10.31294/larik.v1i1.504.
- [18] M. D. Rahmandika, “PENGALAMAN PELANGGAN, ULASAN PELANGGAN SECARA DARING, DAN VARIASI PRODUK TERHADAP KEPUTUSAN PEMBELIAN ULANG,” *J. Manaj. Pemasar. dan Perilaku Konsum.*, vol. 01, no. 1, pp. 1–12, 2022, doi: <http://dx.doi.org/10.21776/jmppk.2022.01.1.01>.
- [19] A. Muhammad Azizi, P. Korespondensi, and F. Ariany, “Sistem Informasi Pengajuan Cuti Pegawai Menggunakan Metode Pengujian Iso 25010 (Study Kasus : Pt Mutiara Ferindo Internusa),” vol. 4, no. 3, pp. 326–334, 2023, [Online]. Available: <https://doi.org/10.33365/jatika.v4i3.3721>.
- [20] C. R. Widestyanto and S. Samsinar, “Sistem Pendukung Keputusan Pemilihan Karyawan Terbaik Pada PT. Menara Depok Asri Menggunakan Metode Profile Matching,” *Pros. Semin. Nas.*, no. September, pp. 1893–1903, 2022.
- [21] A. Bahrudin, Jupriyadi, and Permata, “Optimasi Arsip Penyimpanan Dokumen Foto Menggunakan Algoritma Kompresi Deflate (Studi Kasus :Studio Muezzart),” *J. Ilm. Infrastruktur Teknol. Inf.*, vol. 1, no. 2, pp. 14–18, 2020, doi: 10.33365/jiiti.v1i2.582.
- [22] F. Utami, S. Suhendri, and M. Abdul Mujib, “Implementasi Algoritma Haar Cascade pada Aplikasi Pengenalan Wajah,” *J. Inf. Technol.*, vol. 3, no. 1, pp. 33–38, 2021, doi: 10.47292/joint.v3i1.45.
- [23] M. Adenuddin Alwy, “Manajemen Sumber Daya Manusia Di Era Digital Melalui Lensa Manajer Sumber Daya Manusia Generasi Berikutnya,” *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 1, no. 10, pp. 2265–2276, 2022, doi: 10.54443/sibatik.v1i10.334.
- [24] B. R. Putra and A. Diana, “Rancang Bangun Sistem Pendukung Keputusan Pemilihan Karyawan Terbaik Dengan Metode Analytical Hierarchy Process (AHP) Pada Rumah Makan Ciganea Pusat,” *J. Perad. Sains, Rekayasa Dan Teknol.*, vol. 9, no. 2, pp. 250–264, 2021, doi: 10.47065/tin.v2i8.1261.
- [25] L. G. N. Rismawati, Manurung Harlia, Ardila Jonner and R. Silaeen, “ANALISIS MANAJEMEN SUMBER DAYA MANUSIA BERBASIS KOMPETENSI DALAM KINERJA KARYAWAN DI PDAM TIRTANADI MEDAN,” *J. Glob. Manaj.*, vol. 11, no. 2, pp. 22–33, 2022.
- [26] Z. Wu *et al.*, “MD-NDNet: a multi-dimensional convolutional neural network for false-positive reduction in pulmonary nodule detection,” *Phys. Med. Biol.*, vol. 65, no. 23, p. 235053, 2020, doi: 10.1088/1361-6560/aba87c.