



INNOVATIVE: Journal Of Social Science Research

Volume 4 Nomor 5 Tahun 2024 Page 7242-7257

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Penentuan Penerima Bantuan Program Keluarga Harapan menggunakan Algoritma Random Forest di Desa Kebonrejo

Krisna Giana Putra

Universitas Islam Balitar

Email: krisnagiana@gmail.com

Abstrak

Penelitian ini bertujuan untuk meningkatkan efektivitas penyaluran bantuan Program Keluarga Harapan (PKH) di Desa Mronjo dengan menerapkan algoritma Random Forest. Penelitian ini menggunakan metode kuantitatif dengan pendekatan deskriptif. Data dikumpulkan melalui observasi, wawancara, dan studi literatur, dengan sampel 99 orang dari populasi 838 warga Dusun Kebunrejo, yang dipilih menggunakan teknik purposive sampling. Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan penerima PKH dengan baik. Faktor-faktor penting yang mempengaruhi klasifikasi meliputi pendapatan, pekerjaan, kondisi rumah tangga, penyakit kronis, dan status kesesuaian. Evaluasi model menggunakan Confusion Matrix menunjukkan bahwa model memiliki akurasi, presisi, recall, dan nilai F1 yang tinggi, masing-masing mencapai 90%. Meskipun sebagian besar prediksi sesuai dengan data asli, masih terdapat beberapa kesalahan klasifikasi yang perlu diperbaiki. Secara keseluruhan, model Random Forest terbukti efektif dalam penentuan penerima bantuan PKH dan dapat menjadi acuan untuk peningkatan kebijakan penyaluran bantuan sosial di masa mendatang.

Kata Kunci: Kemiskinan, Program Keluarga Harapan, Algoritma Random Forest

Abstract

This study aims to improve the effectiveness of the distribution of Family Hope Program (PKH) assistance in Mronjo Village by applying the Random Forest algorithm. The manual method used previously was often ineffective and prone to errors. This study uses a quantitative method with a descriptive approach. Data were collected through observation, interviews, and literature studies, with a sample of 99 people from a population of 838 residents of Kebunrejo Hamlet, who were selected using a purposive sampling technique. The results showed that the Random Forest algorithm was able to classify PKH recipients well. Important factors that influence classification include income, employment, household conditions, chronic diseases, and suitability status. Model evaluation using the Confusion Matrix showed that the model had high accuracy, precision, recall, and F1 value, each reaching 90%. Although most of the predictions were in accordance with the original data, there were still some classification errors that needed to be corrected. Overall, the Random Forest model has proven effective in determining PKH assistance recipients and can be a reference for improving social assistance distribution policies in the future.

Keywords: *Poverty, Family Hope Program, Random Forest Algorithm*

PENDAHULUAN

Pemerintah Indonesia meluncurkan Program Keluarga Harapan (PKH) untuk membantu Rumah Tangga Sangat Miskin (RTSM). Optimalisasi teknologi, khususnya *machine learning*, menjadi penting dalam membantu pengguna bantuan Program Keluarga Harapan (PKH) dalam memprediksi permasalahan dalam konteks globalisasi dan kemajuan teknologi informasi untuk meminimalisir ketidakakuratan dan ketidaktepatan dalam menjangkau kelompok sasaran yang sebenarnya membutuhkan bantuan.

Pada tahun 2022 terdapat 252 penerima bantuan PKH, dan pada tahun 2023 terdapat 243 dari 838 penduduk yang penerima bantuan ini hampir 25% penduduk desa mendapatkan bantuan ini, dimana itu merupakan bantuan maksimal yang dapat diberikan oleh pemerintah setiap desanya, ketepatan sasaran menjadi penting. Berdasarkan cara penentuan kriteria penerima bantuan PKH ini adalah antara lain Lansia, penyandang disabilitas, ibu hamil, anak usia dini umur 0-6 tahun.

Persyaratan untuk menjadi penerima PKH antara lain berstatus sebagai warga negara Indonesia, memiliki Kartu Keluarga Sejahtera (KKS) dan Terdaftar dalam Data terpadu (DTKS), berada dalam kondisi sosial ekonomi miskin dan rentan, serta memiliki anggota keluarga seperti yang disebutkan di atas.

Karena banyaknya bantuan dan banyaknya warga yang membutuhkan bantuan yang ada saat ini keefisienan dan ketepatan sasaran menjadikan hal yang penting dalam menyalurkan bantuan apapun pada hal ini adalah PKH, implementasi teknologi klasifikasi

menjadi sangat penting karena dapat membantu mengatasi tantangan administratif dan penyaluran sumber daya yang tidak efisien. Dengan menggunakan pendekatan ini, pemerintah dan lembaga yang terlibat dalam distribusi bantuan dapat memaksimalkan dampak positifnya, mengurangi pemborosan, dan memastikan bahwa bantuan tersebut benar-benar mencapai mereka yang membutuhkannya. Karena adanya masalah tersebut, maka diperlukan cara untuk mempermudah dalam penentuan layak atau tidaknya seorang warga dalam menerima bantuan. Dengan menggunakan algoritma random forest maka dapat membantu meningkatkan keefisienan dan ketepatan sasaran bantuan tersebut.

Menurut Breiman L. (2001) Kelebihan dari algoritma Random Forest antara lain : 1) Berjalan efisien untuk database yang besar. 2) Dapat mengolah ribuan variabel input tanpa harus menghapus variabel lainnya. 3) Dapat memberikan perkiraan variabel yang dianggap penting dalam klasifikasi. 4) Menghasilkan hasil perkiraan yang tidak bias dalam hasil foresting (pengelompokan). 5) Mampu menjaga akurasi pengolahan data meskipun terdapat data yang hilang. 6) Mampu mendistribusikan dan menyeimbangkan nilai bias pada kelas dengan dataset yang memiliki populasi yang tidak seimbang. 7) Memungkinkan dilakukannya deteksi interaksi antar variabel. Penulis tertarik untuk mengimplementasikan algoritma Random Forest dan pengujian menggunakan *Confusion Matrix*.

Dalam praktiknya, pemberian bantuan PKH yang ada di desa tersebut masih dilakukan secara manual oleh para fasilitator. Karena belum berbasis sistem teknologi informasi maka sering kali terjadi permasalahan pada efektivitas program bantuan sosial di desa tersebut. Hasil penelitian diharapkan dapat digunakan untuk memperbaiki implementasi PKH, memastikan bahwa bantuan tepat sasaran, dan meningkatkan kesejahteraan keluarga miskin dan rentan di Desa Mronjo. Selain itu, penelitian ini juga dapat memberikan rekomendasi kebijakan yang relevan untuk pemerintah daerah dan nasional dalam mengelola program bantuan sosial serupa. Oleh karena itu sesuai dengan upaya untuk meningkatkan ketepatan sasaran bantuan Program Keluarga Harapan maka penelitian skripsi ini dilakukan dengan judul "Penentuan Penerima Bantuan Program Keluarga Harapan Menggunakan Algoritma Random Forest (Studi Kasus: Dusun Kebonrejo Kelurahan Mronjo)".

Berdasarkan latar belakang di atas, penelitian ini bertujuan untuk mengetahui implementasi Algoritma klasifikasi Random Forest di Dusun Kebonrejo dan untuk mengetahui hasil akurasi pengujian *Confusion Matrix* pada klasifikasi penerima bantuan Program Keluarga Harapan dengan data yang ada menggunakan algoritma random forest.

METODE PENELITIAN

Penelitian ini dilakukan dari November 2023 hingga Juli 2024. Pengambilan data pada penelitian ini dilakukan di Kantor Desa Mronjo dan pengolahan data dilakukan di laboratorium komputer fakultas teknologi dan informasi Universitas Islam Balitar Blitar.

Jenis penelitian yang akan dilakukan merupakan jenis penelitian deskriptif. Penelitian deskriptif merupakan suatu penelitian yang menggambarkan atau menganalisis suatu hasil penelitian tetapi tidak dapat digunakan untuk menarik kesimpulan secara luas (Sugiyono, 2005). Namun, penelitian deskriptif disini menggunakan pendekatan kuantitatif.

Dalam proses pengumpulan data, membutuhkan teknik dan instrument pengumpulan data yang sesuai. Peneliti menggunakan teknik wawancara dan studi literatur. Jenis data yang digunakan dalam penelitian adalah data sekunder. Data ini didapat dari arsip Kelurahan Mronjo.

Populasi dalam penelitian ini adalah warga Dusun Kebunrejo, kelurahan Mronjo yang berjumlah 838 orang. Untuk menentukan sampel yang akan digunakan dalam penelitian, terdapat berbagai teknik sampling yang digunakan. Penelitian ini menggunakan *Nonprobability sampling*. Teknik pengambilan sampel pada penelitian ini menggunakan *purposive sampling* yaitu teknik penarikan sampel yang dilakukan dengan memilih subyek berdasarkan kriteria spesifik yang ditetapkan peneliti. Sesuai dengan Purposive sampling pada penelitian ini, kriteria yang digunakan yaitu: Penerima bantuan Program Keluarga Harapan (PKH) pada tahun 2023-2024 di Dusun Kebonrejo Kelurahan Mronjo. Sampel yang dibutuhkan dan diambil dalam penelitian ini adalah paling sedikit adalah 99 orang.

Setelah data terkumpul, dilakukan seleksi data karena data yang diperoleh perlu diolah untuk digunakan sesuai dengan atribut yang dibutuhkan dengan melakukan seleksi untuk dijadikan Dataset. Sebagai contoh pada penelitian ini atribut yang dipilih adalah pendapatan, pekerjaan, rumah tangga tunggal lanjut usia, anggota keluarga miskin yang sakit menahun/kronis. pada tahap ini akan menghilangkan data yang null.

Tahap pengolahan data yang sudah bersih yang menjadi Dataset akan diolah dengan menggunakan Algoritma Random Forest, yaitu dengan membagi data latih dan data uji kemudian hasil data latih akan menghasilkan data prediksi yang akan dibandingkan dengan data uji. Dari data yang sudah didapat perlu dikonversi menjadi representasi numerik agar bisa digunakan dalam perhitungan dengan algoritma Random Forest. Sebelumnya, variabel kategorikal harus diubah menjadi representasi numerik.

Tahap pengujian hasil dari data prediksi akan dibandingkan dengan data uji untuk menentukan nilai akurasi, presisi, dan recall dengan menggunakan *Confusion Matrix*. Metode ini membantu untuk memproses hasil analisis *classifier*. Setelah *Confusion Matrix* maka hasil akhirnya bisa dilihat berapa hasil nilai *accuracy*, *recall*, *precision*, dan *f1-score*.

HASIL DAN PEMBAHASAN

Implementasi Algoritma Random Forest untuk Penentuan Klasifikasi Penerima Bantuan Program Keluarga Harapan

Pada tahap *Modeling*, algoritma *Random Forest* diimplementasikan untuk melakukan klasifikasi penerima bantuan Program Keluarga Harapan. Dengan menggunakan fitur seperti pendapatan, pekerjaan, rumah tangga (tunggal/lanjut usia), penyakit kronis, status kelayakan, model telah dilatih untuk membedakan antara berbagai kategori. Berikut merupakan sebagian dari *output* DataFrame yang menunjukkan hasil klasifikasi dari penentuan penerima PKH menggunakan algoritma *Random Forest*.

	Usia	Penghasilan	Jumlah Anggota	Jumlah Tanggungan	Pekerjaan	S
2	-1.457644	-0.638158	-0.765981	-0.917539	0	
3	-1.372101	-1.436842	-0.290216	-0.404947	0	
4	-1.457644	-0.538816	0.185548	-0.917539	0	
5	-1.286559	-1.137500	-1.241745	0.107644	0	
6	-1.201016	-0.538816	-0.290216	-0.404947	0	

Dari potongan DataFrame tersebut, terdapat beberapa baris yang menunjukkan bahwa label prediksi cocok dengan label asli. Hal tersebut memiliki pengertian bahwa model telah berhasil mengklasifikasikan jenis penerima tersebut dengan benar. Seperti contoh pada indeks ke-862320 yang menunjukkan bahwa nilai pada 'Label_Prediksi' bernilai 3 dan 'Label_Prediksi' bernilai 3 sesuai dengan klasifikasi yang sebenarnya. Namun, masih terdapat prediksi model yang tidak sesuai dengan label yang sebenarnya yaitu seperti pada baris dengan indeks 4288986 di mana label prediksi tidak cocok dengan label asli, menunjukkan bahwa model tersebut membuat kesalahan dalam klasifikasi pada contoh tersebut.

Pengukuran nilai parameter terbaik dilakukan menggunakan model *Random Forest* yang berfungsi untuk mengevaluasi pentingnya setiap fitur dalam membuat prediksi yang akurat. Model ini telah dilatih dan mengidentifikasi atribut pada dataset pelatihan yang paling signifikan dalam memprediksi jenis penerima yang layak. Adapun hasil yang didapatkan sebagai berikut. Grafik tersebut ditunjukkan menggunakan horizontal bar chart yang menggambarkan pentingnya fitur-fitur dalam sebuah model *Random Forest*. Setiap bar mewakili fitur yang berbeda dari dataset pelatihan dengan panjang yang sebanding dengan skor pentingnya. Terdapat 4 fitur yang digunakan sebagai pengukuran dan berkontribusi dalam melakukan prediksi menggunakan model *Random Forest*, antara lain pendapatan, pekerjaan, rumah tangga (tunggal/lanjut usia), dan penyakit kronis.

Dari grafik tersebut, dapat dilihat bahwa fitur 'pendapatan' adalah fitur yang paling penting dalam model, diikuti oleh fitur 'pekerjaan'. Fitur-fitur 'kondisi rumah tangga' dan fitur 'penyakit kronis' memiliki tingkat pentingnya yang rendah. Hasil tersebut menunjukkan bahwa model lebih banyak mengandalkan pendapatan dan pekerjaan untuk membuat prediksi daripada faktor-faktor lainnya. Pentingnya fitur, atau "feature importances" tersebut memberikan *insight* tentang kontribusi relatif dari setiap fitur terhadap keputusan yang dibuat oleh model.

Setelah didapatkan parameter terbaik dari pemodelan *Random Forest* tersebut. Nilai parameter akan diimplementasikan berupa visualisasi dari sebuah pohon keputusan individu yang merupakan komponen dari model *Random Forest* yang telah di latih. Dalam visualisasi model tersebut, simpul atas (*root node*) merepresentasikan fitur dengan kepentingan terbesar dalam melakukan prediksi. Setiap simpul internal di bawahnya merepresentasikan fitur dengan pertanyaan biner yang membagi data lebih lanjut berdasarkan jawaban ya atau tidak, yang mengarah pada simpul atau daun selanjutnya berdasarkan hasil prediksi terbaik saat itu. Nilai dalam simpul menunjukkan distribusi kelas pada titik pembagian tersebut, dan proses ini terus berlanjut hingga mencapai simpul daun yang memberikan prediksi akhir.

Nilai Hasil Akurasi Pengujian Menggunakan *Confusion Matrix* pada Penentuan Klasifikasi Penerima Bantuan Program Keluarga Harapan Menggunakan Algoritma Random Forest

Hasil dari proses pemodelan mengindikasikan bahwa model *Random Forest* telah berhasil mencapai tujuan klasifikasi penerima bantuan Program Keluarga Harapan dengan memberikan gambaran yang informatif mengenai penentuan calon penerima bantuan Program Keluarga Harapan. Evaluasi kinerja model dilakukan mencakup metrik-metrik seperti akurasi, presisi, *recall*, dan *F1-score*, untuk memberikan pemahaman tentang sejauh mana model mampu mengklasifikasikan dengan tepat penentuan penerima bantuan Program Keluarga Harapan berdasarkan data yang telah diberikan dengan menggunakan pengukuran dan penilaian kinerja.

Untuk mengukur keberhasilan model dalam menghasilkan prediksi yang tepat, perhitungan metrik evaluasi dilakukan yang meliputi perhitungan akurasi (*accuracy*), presisi (*precision*), *recall*, dan nilai F1 (*f1-score*). Kinerja model tercermin dalam hasil dari metrik evaluasi tersebut, yang memberikan gambaran tentang kemampuannya dalam mengklasifikasikan setiap kelas data dengan tepat, beserta kemampuan model untuk menyeimbangkan antara presisi dan *recall* secara efektif. Berikut merupakan nilai dari masing-masing pengukuran metrik evaluasi.

Accuracy: 0.9115258283033455
Precision: 0.9131120656144661
Recall: 0.9115258283033455
F1 Score: 0.9091279112981441

```
classification_report:
      precision    recall  f1-score   support

0         0.92      0.70      0.80     252490
1         0.89      0.96      0.93     2576659
2         0.96      0.72      0.82     313490
3         0.87      0.90      0.89     1268575
4         0.84      0.60      0.70     160550

 accuracy                   0.91     6422315
 macro avg                 0.92      0.80      0.85     6422315
 weighted avg              0.91      0.91      0.91     6422315
```

Berdasarkan hasil tersebut, nilai akurasi ditunjukkan sebesar 91% yang mengidentifikasi Tingkat keberhasilan model sama mengklasifikasikan secara benar. Hasil tersebut menandakan bahwa model mampu mengidentifikasi calon penerima yang benar sesuai dengan jenis yang telah ditetapkan. Dalam aspek presisi, nilai yang dihasilkan adalah 91%, menerangkan bahwa model sangat tepat dalam prediksinya, dengan probabilitas tinggi bahwa calon penerima yang diprediksi oleh model memang sesuai dengan klasifikasi yang akurat. Nilai *recall* juga berada pada tingkat yang sama, yaitu 91%, menunjukkan bahwa model tersebut mampu secara konsisten mengidentifikasi dan tidak melewatkan calon penerima yang relevan. Terakhir, skor F1 yang diperoleh adalah 90%, yang mencerminkan keseimbangan yang harmonis antara presisi dan *recall*.

Berdasarkan hasil tersebut, model algoritma *Random Forest* menunjukkan performa yang sangat baik dalam mengklasifikasikan calon penerima bantuan Program Keluarga Harapan dengan memadukan nilai akurasi, presisi, dan *recall* yang tinggi. Berdasarkan hasil evaluasi tersebut pula, jika dibandingkan dengan penelitian terdahulu yang relevan pada bab II. Hasil menunjukkan bahwa terdapat penelitian-penelitian yang menghasilkan nilai akurasi lebih tinggi dari hasil yang peneliti dapatkan.

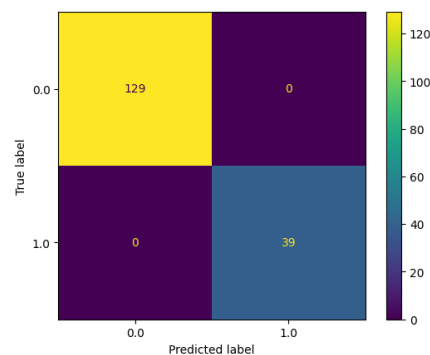
Hal tersebut dapat disebabkan karena beberapa hal, diantaranya adalah karena bentuk data dalam hal ini data statistik kependudukan yang digunakan pada setiap penelitian berbeda, melakukan optimasi algoritma yang tepat (Malik dkk, 2021), penambahan efektivitas algoritma khusus yaitu RNN-LSTM yang digunakan dalam memodelkan seri waktu (Upadhayula dkk, 2019) melakukan pemilihan dan penghindaran fitur khusus yang memiliki hubungan langsung dengan variabel target untuk mencegah *underfitting* sehingga meningkatkan kinerja model machine learning (Aljabri dkk, 2022; Obaid dkk, 2021) serta penambahan proses *soft-classification* untuk deteksi anomali (Yeng dkk, 2020).

Pada pengimplementasian model *Random Forest*, ditampilkan pula *Confusion Matrix* untuk menunjukkan efektivitas algoritma *Random Forest* dalam tugas mengklasifikasikan

penentuan klasifikasi penerima bantuan Program Keluarga Harapan, berdasarkan data dari data kependudukan di kelurahan Mronjo. *Confusion Matrix* tersebut menggambarkan kinerja model dalam mengklasifikasikan data penerima ke dalam berbagai kategori atau jenis

Dari hasil heatmap, nilai-nilai pada diagonal menunjukkan jumlah *true positives*, yang mencerminkan jumlah prediksi yang benar karena bernilai lebih dari 90% oleh model untuk masing-masing kelas. Seperti contoh pada kelas 1 memiliki sejumlah besar prediksi yang tepat. Warna gelap pada heatmap menandakan frekuensi yang lebih tinggi dalam prediksi atau kelas sebenarnya, dengan skala di sisi kanan yang mengindikasikan kisaran jumlah prediksi tersebut. Dari informasi tersebut, dapat dilihat bahwa model memiliki kecenderungan untuk mengklasifikasikan dengan benar untuk beberapa kelas, sementara untuk kelas lainnya terdapat ruang untuk peningkatan.

Berikut adalah heatmap dari pengujian confusion Matrix.



Gambar 1. Heatmap Confusion Matrix

Program Aplikasi

Tahap selanjutnya untuk melakukan proses penyiapan data agar siap dan mudah diolah dalam program, penulis melakukan proses pengkodean menggunakan paperspace dengan tahapan sebagai berikut.

1. Install *Library*

Tahap pertama adalah melakukan proses install dari beberapa *library* yang akan digunakan dalam python. *Library* tersebut antara lain adalah "tldextract" untuk mengekstrak informasi data dari arsip kependudukan di desa, "pandas" untuk mempermudah manipulasi dan analisis data, dan "scikit-learn" untuk pengembangan dan implementasi model *machine learning*.

```
!pip install tldextract pandas scikit-learn
```

2. Pengunduhan Dataset yang Sudah Didapatkan

Tahap ini bertujuan untuk mengunduh dan mengekstrak dataset pada Paperspace

menggunakan 2 modul yaitu modul 'urllib.request' dan 'zipfile'. Modul 'urllib.request' digunakan untuk mengambil dataset dari URL yaitu lokasi dataset berada dan menyimpannya sebagai file "dataset.zip" dengan fungsi 'urlretrieve'.

```
import urllib.request

urllib.request.urlretrieve("https://nopladi-
index.vercel.app/api/raw/?path=/dataset/krisna/dataset.z
```

Selanjutnya, modul 'zipfile' digunakan untuk mengekstrak isi dataset tersebut. Proses ini dilakukan untuk mempersiapkan dataset agar siap digunakan dalam melakukan pemodelan.

```
import zipfile

with zipfile.ZipFile("dataset.zip", 'r') as zip_ref:
```

3. Memasukkan Dataset Teks Ke Dalam List Array

Pada tahap ini bertujuan untuk memasukkan isi dataset ke dalam sebuah list array. Direktori dataset ditentukan dalam variabel 'folder_dataset' dan di inialisasi dalam sebuah list kosong bernama 'data'. List tersebut kemudian diubah menjadi satu dimensi dengan nama flat_data dan dilakukan konversi huruf kecil pada setiap elemen, hasilnya disimpan dalam list flat_data_lower. Proses ini menyiapkan data dalam bentuk list array untuk analisis atau pemrosesan selanjutnya

```
import os
folder_dataset = "dataset"
data = []
path = os.path.join(os.getcwd(), folder_dataset)
for filename in os.listdir(folder_dataset):
    with open(os.path.join(path, filename), 'r') as f:
        data2 = f.readlines()
        data.append(data2)

flat_data = [item for sublist in data for item in
sublist]

flat_data_lower = [string.lower() for string in
flat_data]
```

4. Ekstrak Data dengan Regular Expression

Pada tahap ini dilakukan ekstraksi data dari setiap string dalam list dan disusun ke dalam Data Frame yang dinamai 'dataset_separated_by_regex' setelah pencocokan pola menggunakan 're.findall'. Di dalam DataFrame tersebut disajikan kolom-kolom seperti no., nama, pendapatan, pekerjaan, rumah tangga (tunggal/lanjut usia), penyakit kronis, status kelayakan yang diekstrak dari data asli sehingga mempermudah pengolahan data menjadi lebih terstruktur.

```

no., nama, pendapatan, pekerjaan, kondisi, penyakit,
status kelayakan = [], [], [], [], [], [], []

for data in flat_data_lower:

    match = re.findall(pattern, data)
    for m in match:

        no.append(m[0])
        nama.append(m[1])
        pendapatan.append(m[2])
        pekerjaan.append(m[5])
        kondisi.append(m[3])
        penyakit.append(m[4])
        kelayakan.append(m[6])

dataset_separated_by_regex = pd.DataFrame({'No.':no.,
'Nama': nama, 'Pendapatan': pendapatan, 'Pekerjaan':
pekerjaan, 'Kondisi': kondisi, 'Penyakit': penyakit,
'status kelayakan': kelayakan})

```

Selanjutnya data tersebut akan disimpan dalam format CSV sebagai berikut.

```

dataset_separated_by_regex.to_csv('dataset_separated_by
_regex.csv', index=False)

```

Transformation

Dalam tahap ini, terjadi proses transformasi data menjadi format yang cocok untuk keperluan data mining. Tujuannya adalah untuk mempermudah koordinasi data yang diolah menggunakan Google Colaboratory dalam konteks klasifikasi dengan algoritma decision tree dan random forest.

	KELUARGA	ANAK	USIA	KEPEMILIKAN RUMAH	MOTOR	MOBIL	PENGHASILAN \
0	4	0.285714	40	0	2	0	0.473684
1	3	0.142857	31	1	1	0	0.210526
2	4	0.285714	29	0	1	0	0.298246
3	3	0.142857	31	1	1	0	0.263158
4	3	0.142857	32	1	1	0	0.263158

	STATUS
0	1.0
1	0.0
2	0.0
3	0.0
4	0.0

Gambar 4. Transformasi Data

Nilai dalam dataset dengan tujuan untuk meningkatkan kualitas atau model *machine learning*. Dalam penelitian ini, *data transformation* digunakan untuk mengubah tipe data pada kolom "waktu" dari format string ke format integer.

```
#sudah dioptimalkan dengan pendekatan berbeda
# Konversi kolom 'Waktu' menjadi objek datetime
df_encoded['Waktu'] = pd.to_datetime(df_encoded['Waktu'],
```

Konversi kolom “waktu” dari format string ke objek datetime dilakukan dengan menggunakan fungsi ‘pd.to_datetime’. Selanjutnya, nilai waktu dihitung sebagai selisih waktu dari Tengah malam (00:00:00.000) dalam detik, dan hasilnya dinormalisasikan ke rentang 0-1 dengan membaginya dengan jumlah detik dalam sehari (24*3600).

Fitur Seleksi

Keberhasilan Fitur Seleksi terbukti dalam upaya meningkatkan hasil evaluasi pada tahap data mining. Dalam penelitian ini, algoritma Decision Tree dan Random Forest digunakan untuk menjalankan seleksi fitur. Proses seleksi fitur dilakukan setelah mengidentifikasi skenario dengan tingkat akurasi tertinggi pada tahap data mining. Akurasi tersebut selanjutnya dimanfaatkan sebagai nilai fungsi kecocokan (*fitness function*) dalam algoritma Decision Tree dan Random Forest.

Data Mining

Dalam penelitian ini, terdapat penerapan tahap Data Mining yang berfokus pada klasifikasi menggunakan algoritma Decision Tree dan Random Forest. Perhitungan klasifikasi dilakukan dengan melibatkan 4 variabel diantaranya pendapatan, pekerjaan, kondisi rumah tangga (tunggal/lanjut usia), penyakit kronis, dan status kelayakan. Dengan harapan dapat mencapai tingkat akurasi yang optimal. Proses pengujian menggunakan data uji sebesar 0.3 (30%) dari seluruh dataset.

```
[31] from sklearn.model_selection import train_test_split

X = df.drop('STATUS', axis=1)
y = df.STATUS

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Gambar 1. Kode Presentase Data Testing

Decision Tree

Dari hasil akurasi decision tree, menunjukkan bahwa pada dataset latihan, model berhasil mencapai tingkat akurasi yang optimal sebesar 100%, dan seluruh metrik lainnya juga mencapai nilai maksimum (1.0). Kemungkinan terdapat overfitting dapat diindikasikan, di mana model dapat terlalu terampil dalam menyesuaikan diri dengan data latihan tetapi kurang dapat melakukan generalisasi terhadap data baru. Pada dataset pengujian, kinerja model tetap sangat baik dengan tingkat akurasi sebesar 97.86%. Meskipun terdapat sedikit perbedaan antara nilai precision dan recall untuk kelas 1, secara keseluruhan, model ini

memberikan hasil yang memuaskan pada data yang belum pernah dilihat sebelumnya.

```
Train Result:
=====
Accuracy Score: 100.00%
-----
CLASSIFICATION REPORT:
      0.0  1.0  accuracy  macro avg  weighted avg
precision  1.0  1.0      1.0      1.0      1.0
recall    1.0  1.0      1.0      1.0      1.0
f1-score   1.0  1.0      1.0      1.0      1.0
support   581.0  74.0      1.0     655.0     655.0
-----
Confusion Matrix:
[[581  0]
 [ 0  74]]
-----
Test Result:
=====
Accuracy Score: 97.86%
-----
CLASSIFICATION REPORT:
      0.0  1.0  accuracy  macro avg  weighted avg
precision  0.984314  0.923077  0.978648  0.953695  0.978212
recall    0.992095  0.857143  0.978648  0.924619  0.978648
f1-score   0.988189  0.888889  0.978648  0.938539  0.978294
support   253.000000  28.000000  0.978648  281.000000  281.000000
-----
Confusion Matrix:
[[251  2]
 [ 4  24]]
```

Gambar 2. Hasil Akurasi Decision Tree

Random Forest

Dari hasil penerapan model random forest, diketahui bahwa model menunjukkan kinerja yang sangat baik pada data pelatihan dengan mencapai tingkat akurasi 100%. Perlu diperhatikan bahwa terdapat potensi overfitting, karena performa optimal pada data pelatihan tidak selalu mencerminkan kemampuan model untuk menghasilkan hasil yang baik pada data baru. Pada tahap pengujian, model mencapai tingkat akurasi yang tinggi, yakni 97.86%. Walaupun begitu, terdapat beberapa kesalahan dalam mengidentifikasi instance positif untuk kelas 1. Ini merupakan fokus area yang dapat diperbaiki untuk meningkatkan recall kelas 1 tanpa mengorbankan precision secara signifikan.

```
Train Result:
=====
Accuracy Score: 100.00%
-----
CLASSIFICATION REPORT:
      0.0  1.0  accuracy  macro avg  weighted avg
precision  1.0  1.0      1.0      1.0      1.0
recall    1.0  1.0      1.0      1.0      1.0
f1-score   1.0  1.0      1.0      1.0      1.0
support   581.0  74.0      1.0     655.0     655.0
-----
Confusion Matrix:
[[581  0]
 [ 0  74]]
-----
Test Result:
=====
Accuracy Score: 97.86%
-----
CLASSIFICATION REPORT:
      0.0  1.0  accuracy  macro avg  weighted avg
precision  0.984314  0.923077  0.978648  0.953695  0.978212
recall    0.992095  0.857143  0.978648  0.924619  0.978648
f1-score   0.988189  0.888889  0.978648  0.938539  0.978294
support   253.000000  28.000000  0.978648  281.000000  281.000000
-----
Confusion Matrix:
[[251  2]
 [ 4  24]]
```

Gambar 3. Algoritma Random Forest

Pengujian

Tahap terakhir yaitu evaluasi dilakukan untuk mengukur kinerja model yang telah dilatih dan mengetahui apakah model yang telah dibuat menghasilkan hasil yang akurat. Dalam evaluasi, model akan diuji terhadap data yang belum pernah dilihat sebelumnya, dalam hal ini menggunakan data pengujian. Proses evaluasi dilakukan dengan menggunakan metrik evaluasi untuk melakukan pengukuran dan penilaian kinerja suatu

model berdasarkan kriteria tertentu. Adapun evaluasi yang dilakukan sebagai berikut.

1. Pengukuran terhadap tingkat *accuracy*, *precision*, *recall*, dan *f1-score* untuk melihat sejauh mana model memberikan prediksi yang akurat dan sesuai dengan tujuan. Adapun implementasi sebagai berikut.

```
from sklearn.metrics import
classification_report,

accuracy_score, precision_score,
```

Metrik akurasi dihitung untuk mengevaluasi sejauh mana model mampu mengklasifikasikan data secara benar secara keseluruhan. Berikut adalah

```
# Evaluate Accuracy

accuracy = accuracy_score(y_test, y_pred)
```

implementasinya.

Dalam mengukur sejauh mana model dapat memberikan prediksi yang tepat untuk setiap kelas dalam klasifikasi, digunakan metrik evaluasi *precision*. Presisi dihitung dengan mempertimbangkan bobot kelas yang seimbang, dan implementasinya dapat dilihat pada kode berikut.

Metrik evaluasi *recall* digunakan mengukur sejauh mana model dapat mengidentifikasi dan mengklasifikasikan dengan benar suatu kelas dari total

```
# Evaluate Precision

precision = precision_score(y_test,
```

kemunculan kelas tersebut dalam data aktual. Implementasinya dapat dilihat pada kode berikut.

```
# Evaluate Recall

recall = recall_score(y_test, y_pred,
```

Metrik evaluasi *F1-score* digunakan memberikan informasi tentang sejauh mana model dapat mencapai keseimbangan antara *precision* dan *recall* dalam mengklasifikasikan suatu kelas. Implementasinya dapat dilihat pada kode berikut.

Metrik evaluasi dihitung dan disajikan untuk memberikan gambaran singkat tentang performa model. Berikut adalah implementasinya.

2. Berdasarkan hasil prediksi dari model tersebut, *Confusion Matrix* dibuat untuk

```
# Evaluate F1-Score
```

```
f1 = f1_score(y_test, y_pred,  
             average='weighted')
```

as positif dan kelas
negatif (TN), false

```
classification_rep =  
classification_report(y_test, y_pred)  
  
import matplotlib.pyplot as plt  
# Print the metrics  
print("Accuracy:", accuracy)  
print("Precision:", precision)  
print("Recall:", recall) print("F1-  
score:", f1)  
  
print("\nclassification_report:\n",  
      classification_report(y_test, y_pred))  
  
# Plot confusion matrix  
plt.figure(figsize=(10, 8))  
  
sns.heatmap(cm, annot=True, fmt='g',  
            cmap='Blues', xticklabels=np.unique(y),  
            yticklabels=np.unique(y))  
plt.xlabel('Predicted')  
  
plt.ylabel('Actual')  
  
plt.title('Confusion Matrix for Random  
Forest  
Classifier')  
  
plt.show()
```

SIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan, maka terdapat beberapa kesimpulan sebagai berikut.

Implementasi algoritma *random forest* untuk klasifikasi penentuan penerima bantuan program keluarga harapan di Desa Kebonrejo pada hasil pemodelan mencakup kolom-kolom penting seperti pendapatan, pekerjaan, rumah tangga (tunggal/lanjut usia), penyakit kronis, serta status kesesuaian. Evaluasi model menunjukkan sebagian besar prediksi sesuai dengan label asli, menandakan keberhasilan model, namun terdapat beberapa kasus prediksi yang tidak cocok, mengindikasikan potensi kesalahan klasifikasi.

Nilai hasil akurasi pengujian menggunakan *Confusion Matrix* pada klasifikasi penentuan penerima bantuan program keluarga harapan di Desa Kebonrejo menggunakan algoritma Random Forest dapat diukur melalui akurasi, presisi, recall, dan nilai F1. Dengan nilai akurasi, presisi, recall, dan F1 mencapai 90%, maka model ini menunjukkan kemampuan yang tinggi dalam mengidentifikasi dan mengklasifikasikan penerima dengan tepat. Berdasarkan hal tersebut, model Random Forest memberikan hasil yang sangat baik dalam tugas klasifikasi penentuan penerima bantuan Program Keluarga Harapan.

DAFTAR PUSTAKA

- Aurelien Geron. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.
- Fauzan, M., Gusti, S. K., Jasril, J., & Pizaini, P. (2023). *Penerapan Seleksi Fitur Untuk Klasifikasi Penerima Bantuan Sosial Pangkalan Sesai Menggunakan Metode K-Nearest Neighbor*. *Jurnal Sistem Komputer dan Informatika (JSON)*, 5(1), 1-10.
- Han, J. and Kamber, M. (2006) *Data Mining: Concepts and Techniques*. 2nd Edition, Morgan Kaufmann Publishers, San Francisco.
- Kothari, C. R. (2004). *Research methodology: Methods and techniques*. *New Age International*.
- Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). *Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin*. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(2), 421-428.

- Madia, N., Septiarini, A., Hatta, H. R., Hamdani, H., & Wati, M. *Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode Naïve Bayes (Studi Kasus: Kabupaten Penajam Paser Utara)*. JISKA (Jurnal Informatika Sunan Kalijaga), 8(1), 36-49.
- Patasik, E. S., & Yulianto, S. (2023). *Classification of Regional Languages Using Methods Gradient Boots and Random Forest*. Jurnal Teknik Informatika (JUTIF), 4(5), 1249-1255.
- Putri, H., Purnamasari, A. I., Dikananda, A. R., Nurdiawan, O., & Anwar, S. (2021). *Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN*. Building of Informatics, Technology and Science (BITS), 3(3), 331-337.
- Sitanggang, B. (2014). *Implementasi kebijakan penyaluran hibah dan bantuan sosial kemasyarakatan di Kabupaten Kubu Raya* (Doctoral dissertation, Tanjungpura University).
- Sugianto, C. A., & Maulana, F. R. (2019). *Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)*. Techno. com, 18(4), 321-331.
- Tin Kam Ho. (1995). *Random Decision Forests*. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1, 278-282.