

KLASIFIKASI RISIKO GIZI BURUK PADA IBU HAMIL MENGUNAKAN METODE RANDOM FOREST

Fanny Ramadhani¹, Dian Septiana², Sisti Nadia Amalia³, Putri Maulidina Fadilah⁴, Andy Satria⁵

¹⁾ Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Indonesia

^{2,3,4)} Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Indonesia

⁵⁾ Teknologi Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Dharmawangsa, Indonesia

Article Info

Article history:

Received: 06 Agustus 2024

Revised: 08 Agustus 2024

Accepted: 12 Agustus 2024

ABSTRACT

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi ibu hamil yang berisiko mengalami gizi buruk menggunakan metode klasifikasi machine learning, khususnya Random Forest, dengan memanfaatkan data dari RISKESDAS 2018. Dataset yang digunakan mencakup informasi demografi dan pola makan, termasuk usia, pendidikan, pekerjaan, status ekonomi, pola makan, dan akses ke layanan kesehatan. Data tersebut diolah melalui proses preprocessing yang meliputi penanganan nilai yang hilang, transformasi variabel kategori menggunakan OneHotEncoder, dan normalisasi fitur numerik. Model Random Forest kemudian dilatih dan dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score, serta confusion matrix untuk memahami kinerja klasifikasi. Hasil penelitian menunjukkan bahwa model Random Forest memiliki akurasi sebesar 0.67, precision sebesar 0.6, recall sebesar 0.67, dan F1-score sebesar 0.63 dalam mengklasifikasikan risiko gizi buruk pada ibu hamil. Confusion matrix memperlihatkan distribusi prediksi yang benar dan salah, sedangkan feature importance analysis mengidentifikasi fitur pola makan dan status ekonomi sebagai yang paling berpengaruh dalam prediksi risiko gizi buruk. Model Random Forest ini dapat digunakan sebagai alat yang efektif untuk mengidentifikasi ibu hamil yang berisiko tinggi mengalami gizi buruk, memungkinkan intervensi dini dan terarah dalam program kesehatan ibu hamil, sehingga dapat membantu meningkatkan kesehatan ibu dan anak. Penelitian ini juga menyediakan dasar untuk studi lanjutan yang dapat menggunakan dataset yang lebih luas dan beragam untuk memperbaiki akurasi dan generalisasi model.

Kata Kunci: Ibu hamil, Resiko gizi buruk, Random Forest, Data Demografi, Pola Makan.

Abstract

This study aims to identify pregnant women at risk of malnutrition using the Random Forest method with data from the RISKESDAS 2018 survey. The dataset includes demographic and dietary information such as age, education, occupation, economic status, dietary patterns, and access to healthcare services. The preprocessing process involves handling missing values, transforming categorical variables using One Hot Encoder, and normalizing numerical features. The Random Forest model is trained and evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Results show that the Random Forest model achieves an accuracy of 0.67, precision of 0.6, recall of 0.67, and F1-score of 0.63 in classifying the risk of malnutrition among pregnant women. The confusion matrix reveals the distribution of correct and incorrect predictions, while feature importance analysis identifies dietary patterns and economic status as the most influential features in predicting malnutrition risk. This Random Forest model can serve as an effective tool for identifying pregnant women at high risk of malnutrition, enabling early and targeted interventions in maternal health programs, thus improving maternal and child health outcomes. This study also provides a foundation for further research with a broader and more diverse dataset to enhance the model's accuracy and generalizability.

Keywords: Pregnant women, Malnutrition risk, Random Forest, Demographic data Dietary patterns

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail : as021091@gmail.com

1. PENDAHULUAN

Masalah gizi buruk pada ibu hamil merupakan isu kesehatan masyarakat yang serius, terutama di negara-negara berkembang (Adriati & Chloranyta, 2022). Gizi buruk pada ibu hamil dapat menyebabkan berbagai komplikasi kesehatan, baik bagi ibu maupun janin yang dikandungnya. Komplikasi ini dapat mencakup kelahiran prematur, berat badan lahir rendah, dan bahkan kematian ibu dan anak (Ediyono, 2023). Oleh karena itu, identifikasi dini ibu hamil yang berisiko mengalami gizi buruk sangat penting untuk mencegah dampak negatif tersebut melalui intervensi yang tepat waktu dan efektif.

Seiring dengan perkembangan teknologi dan metode analisis data, machine learning telah menjadi alat yang semakin populer dalam bidang kesehatan untuk memprediksi berbagai risiko penyakit (Satria et al., 2021). Random Forest adalah salah satu algoritma machine learning yang terkenal karena keakuratannya dalam tugas klasifikasi dan kemampuannya untuk menangani data yang kompleks dan beragam. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan selama pelatihan dan menghasilkan prediksi berdasarkan agregasi dari setiap pohon tersebut (Supriyadi et al., 2020).

Penelitian ini bertujuan untuk mengembangkan model prediksi yang dapat mengidentifikasi risiko gizi buruk pada ibu hamil menggunakan metode Random Forest. Data yang digunakan dalam penelitian ini diambil dari survei RISKESDAS 2018, yang mencakup berbagai informasi demografi dan pola makan ibu hamil, seperti usia, pendidikan, pekerjaan, status ekonomi, pola makan, dan akses ke layanan Kesehatan (sutarto et al., 2022). Dengan data yang komprehensif ini, diharapkan model prediksi yang dihasilkan akan memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan risiko gizi buruk.

Proses penelitian ini dimulai dengan pengumpulan dan pemrosesan data. Data mentah yang diperoleh dari RISKESDAS 2018 akan melalui tahap preprocessing, yang meliputi penanganan nilai yang hilang, transformasi variabel kategori menggunakan *OneHotEncoder*, dan normalisasi fitur numerik (Maghfiroh et al., 2023). Setelah itu,

data yang telah diolah akan digunakan untuk melatih model Random Forest. Model ini akan dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur kinerja klasifikasi, serta confusion matrix untuk memahami distribusi prediksi yang benar dan salah (Kohsasih et al., 2022). Selain itu, analisis feature importance akan dilakukan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam prediksi risiko gizi buruk.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam bidang kesehatan ibu dan anak, khususnya dalam upaya pencegahan dan penanganan gizi buruk pada ibu hamil. Dengan adanya model prediksi yang andal, para profesional kesehatan dapat melakukan intervensi dini dan lebih terarah, sehingga dapat meningkatkan kualitas kesehatan ibu dan anak. Penelitian ini juga diharapkan dapat memberikan dasar bagi studi lanjutan yang dapat menggunakan dataset yang lebih luas dan beragam untuk memperbaiki akurasi dan generalisasi model.

Melalui penelitian ini, diharapkan dapat ditemukan model prediksi yang andal dan akurat yang dapat digunakan oleh para profesional kesehatan untuk mengidentifikasi ibu hamil yang membutuhkan perhatian khusus. Dengan demikian, upaya-upaya preventif dan kuratif dapat dilakukan lebih awal, sehingga dapat meningkatkan kualitas kesehatan ibu dan anak di masa mendatang.

2. METODE PENELITIAN

Adapun metode yang digunakan dalam penelitian ini adalah sebagai berikut (Sari et al., 2023; Satria et al., 2021):

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari survei RISKESDAS 2018, yang mencakup informasi demografi dan pola makan ibu hamil. Dataset ini mencakup fitur-fitur seperti usia, pendidikan, pekerjaan, status ekonomi, pola makan, dan akses ke layanan kesehatan.

2. Preprocessing Data

Proses preprocessing dilakukan untuk mempersiapkan data sebelum digunakan dalam model Random Forest (Agung et al., 2023). Langkah-langkah preprocessing meliputi:

- Penanganan Nilai Hilang yaitu nilai yang hilang diimputasi menggunakan metode yang sesuai, seperti mean imputation untuk data numerik atau modus untuk data kategorikal.
- Transformasi Variabel Kategori yaitu Variabel kategori diubah menjadi bentuk numerik menggunakan OneHotEncoder.

- Normalisasi Fitur Numerik yaitu Fitur numerik dinormalisasi agar berada dalam skala yang sama, menggunakan metode seperti *MinMaxScaler* atau *StandardScaler*.
3. Pembagian Data
- Data yang telah diproses dibagi menjadi dua set: data pelatihan (training set) dan data pengujian (test set) dengan perbandingan 80:20. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model (Genç & Tunç, 2019).
4. Pembangunan Model Random Forest
- Model Random Forest dibangun menggunakan data pelatihan. Parameter model disesuaikan untuk mengoptimalkan kinerja, seperti jumlah pohon (*n_estimators*), kedalaman pohon maksimal (*max_depth*), dan lain-lain. Proses pelatihan dilakukan dengan langkah-langkah berikut:
- Pelatihan Model: Model dilatih menggunakan data pelatihan dengan fitur-fitur yang telah dipilih.
 - Validasi Model: Validasi silang (*cross-validation*) digunakan untuk menghindari overfitting dan mengukur kinerja model secara lebih akurat.
5. Evaluasi Model
- Evaluasi dari hasil penelitian ini dilakukan dengan *confusion matrix*. Metode ini merepresentasikan hasil klasifikasi menggunakan matriks yang dapat dilihat pada Tabel 1 (Kohsasih et al., 2022).

Tabel 1. Confusion Matrix

| | | Actual Value | |
|-----------------|----|--------------|----|
| | | TP | FP |
| Predicted Value | TP | | |
| | FN | | |

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Model yang telah dilatih dievaluasi menggunakan data pengujian. Metrik yang digunakan untuk evaluasi meliputi:

- *Akurasi (Accuracy)*: Mengukur proporsi prediksi yang benar dari seluruh prediksi.
- *Precision*: Mengukur proporsi prediksi positif yang benar dari seluruh prediksi positif.
- *Recall*: Mengukur proporsi kasus positif yang benar-benar terdeteksi oleh model.
- *F1-Score*: Rata-rata harmonis dari precision dan recall.

Confusion Matrix: Matriks yang menunjukkan distribusi prediksi yang benar dan salah.

6. Analisis *Feature Importance*

Analisis *feature importance* dilakukan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam prediksi risiko gizi buruk. Fitur dengan nilai penting tinggi menunjukkan bahwa fitur tersebut memiliki kontribusi besar dalam keputusan model (Fitri Salsabilla Arianti et al., 2023).

7. Interpretasi Hasil

Hasil evaluasi dan analisis *feature importance* diinterpretasikan untuk memahami kinerja model dan faktor-faktor yang mempengaruhi risiko gizi buruk pada ibu hamil. Hasil ini digunakan untuk memberikan rekomendasi intervensi yang lebih efektif dan tepat sasaran.

3. HASIL DAN PEMBAHASAN

Hasil penelitian yang ditemukan dalam penelitian ini adalah data yang digunakan dalam penelitian ini diambil dari survei RISKESDAS 2018, yang mencakup informasi demografi dan pola makan ibu hamil. Dataset ini mencakup fitur-fitur seperti usia, pendidikan, pekerjaan, status ekonomi, pola makan, dan akses ke layanan kesehatan. Data dapat dideskripsikan sebagai berikut:

- **Usia**: Umur ibu hamil dalam tahun.
- **Pendidikan**: Tingkat pendidikan ibu hamil (SMA, S1, D3, S2).
- **Pekerjaan**: Jenis pekerjaan ibu hamil (IRT: Ibu Rumah Tangga, Karyawan, Wirausaha, Dosen).
- **Status Ekonomi**: Tingkat ekonomi ibu hamil (Rendah, Menengah, Tinggi).
- **Pola Makan**: Kualitas pola makan ibu hamil (Sehat, Kurang Sehat).
- **Akses Layanan Kesehatan**: Kualitas akses ibu hamil ke layanan kesehatan (Baik, Kurang).
- **Risk Gizi Buruk**: Klasifikasi risiko gizi buruk ibu hamil (Rendah, Sedang, Tinggi)

Berikut data yang digunakan dalam penelitian ini :

Tabel 2. Data Penelitian

| Usia | Pendidikan | Pekerjaan | Status Ekonomi | Pola Makan | Akses Layanan Kesehatan | Risk Gizi Buruk |
|------|------------|-----------|----------------|--------------|-------------------------|-----------------|
| 25 | SMA | IRT | Menengah | Sehat | Baik | Rendah |
| 30 | S1 | Karyawan | Tinggi | Sehat | Baik | Rendah |
| 22 | SMA | IRT | Menengah | Kurang Sehat | Kurang | Sedang |
| 35 | D3 | Wirasaha | Menengah | Kurang Sehat | Baik | Tinggi |
| 28 | S1 | Karyawan | Tinggi | Sehat | Baik | Rendah |
| 24 | SMA | IRT | Menengah | Kurang Sehat | Kurang | Sedang |
| 31 | S2 | Dosen | Tinggi | Sehat | Baik | Rendah |
| 29 | SMA | IRT | Rendah | Kurang Sehat | Kurang | Tinggi |
| 27 | SMA | Karyawan | Menengah | Sehat | Baik | Rendah |
| 32 | S1 | Karyawan | Menengah | Kurang Sehat | Kurang | Sedang |

Setelah dilakukan pengumpulan data, dilanjutkan dengan preprocessing data. Preprocessing data adalah langkah penting dalam machine learning untuk memastikan bahwa data siap untuk digunakan dalam pelatihan model. Berikut adalah langkah-langkah preprocessing yang telah dilakukan:

1. Pembersihan Data (*Data Cleaning*):

Pada penelitian ini, tidak ada data yang hilang sehingga tidak ada baris data yang dihapus.

2. *One-Hot Encoding*:

- Data kategori seperti pendidikan, pekerjaan, status ekonomi, pola makan, dan akses layanan kesehatan diubah menjadi data numerik menggunakan teknik one-hot encoding. Ini dilakukan untuk memungkinkan algoritma machine learning memproses data kategori.
- Pada kolom pendidikan yang memiliki nilai SMA, S1, D3, dan S2 diubah menjadi beberapa kolom biner (pendidikan_SMA, pendidikan_S1, pendidikan_D3, pendidikan_S2) dengan nilai 0 atau 1.

3. Normalisasi Data (jika diperlukan):

- Pada penelitian ini, normalisasi tidak diperlukan karena data sudah dalam bentuk yang sesuai untuk model Random Forest.

Berikut data yang telah dilakukan preprocessing dapat dilihat pada table 1 dibawah ini

Tabel 3. Data Hasil Preprocessing

| usia | pendidikan_SMA | pendidikan_S1 | pendidikan_D3 | pekerjaan_IRT | pekerjaan_Karyawan | pekerjaan_Wirusaha | status_ekonomi_Rendah | status_ekonomi_Tinggi | pola_makan_Sehat | akses_layanan_kesehatan_Kurang | risk_gizi_buruk_Rendah | risk_gizi_buruk_Tinggi |
|------|----------------|---------------|---------------|---------------|--------------------|--------------------|-----------------------|-----------------------|------------------|--------------------------------|------------------------|------------------------|
| 25 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 30 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 22 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 35 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 28 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 24 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 31 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 29 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 27 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 32 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

| | S 1 | S 2 | S M A | | | | | | | | Sed ang | |
|-----|--------|--------|-------------|-----|-----|----|-----|------|------|------|------------|-----|
| 25 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 30 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 22 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 35 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 28 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 24 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 31 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 29 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 27 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 32 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 22 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 35 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 28 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | .. | ... | | | | | ... |
| ... | ... | ... | ... | ... | ... | .. | ... | | | | | ... |
| ... | ... | ... | ... | ... | ... | .. | ... | | | | | ... |
| 28 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 24 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 31 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 29 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 27 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

Kemudian dilanjutkan pada proses pemisahan data. Setelah preprocessing, data kemudian dipisahkan menjadi data training dan testing. Ini adalah langkah yang penting untuk mengevaluasi kinerja model machine learning.

1. Definisi Variabel Input (X) dan Target (y):

- X adalah data fitur yang digunakan untuk memprediksi risiko gizi buruk, sedangkan y adalah label target yang merupakan klasifikasi risiko gizi buruk.
- Dalam penelitian ini, kolom risk_gizi_buruk_Tinggi digunakan sebagai target (y), dengan nilai 1 untuk risiko tinggi dan 0 untuk lainnya.

2. Pemisahan Data Training dan Testing:

Data dibagi menjadi set training (70%) dan set testing (30%) menggunakan fungsi train_test_split dari scikit-learn. Set training digunakan untuk melatih model machine learning, sementara set testing digunakan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat oleh model.

```
import pandas as pd
from sklearn.model_selection import train_test_split

# Data Sintetis
data = {
    'usia': [25, 30, 22, 35, 28, 24, 31, 29, 27, 32],
    'pendidikan': ['SMA', 'S1', 'SMA', 'D3', 'S1', 'SMA', 'S2', 'SMA', 'SMA', 'S1'],
    'pekerjaan': ['IRT', 'Karyawan', 'IRT', 'Wirausaha', 'Karyawan', 'IRT', 'Dosen', 'IRT'],
    'status_ekonomi': ['Menengah', 'Tinggi', 'Menengah', 'Menengah', 'Tinggi', 'Menengah'],
    'pola_makan': ['Sehat', 'Sehat', 'Kurang Sehat', 'Kurang Sehat', 'Sehat', 'Kurang Sehat'],
    'akses_layanan_kesehatan': ['Baik', 'Baik', 'Kurang', 'Baik', 'Baik', 'Kurang', 'Baik'],
    'risk_gizi_buruk': ['Rendah', 'Rendah', 'Sedang', 'Tinggi', 'Rendah', 'Sedang', 'Rendah']
}

# Membuat DataFrame
df = pd.DataFrame(data)

# Pembersihan Data
df_cleaned = df.dropna()

# Transformasi Data (One-Hot Encoding)
df_encoded = pd.get_dummies(df_cleaned, drop_first=True)

# Pemisahan Data
X = df_encoded.drop('risk_gizi_buruk_Tinggi', axis=1)
y = df_encoded['risk_gizi_buruk_Tinggi']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

Gambar 1. Data Split

Langkah berikutnya adalah melatih model Random Forest menggunakan data training dan mengevaluasi kinerjanya pada data testing. Proses ini melibatkan:

a. Pelatihan Model

Langkah pertama adalah melatih model Random Forest menggunakan data training. Berikut adalah contoh kode Python untuk melatih model:

```
from sklearn.ensemble import RandomForestClassifier

# Membuat model Random Forest
rf_model = RandomForestClassifier(random_state=42)

# Melatih model dengan data training
rf_model.fit(X_train, y_train)
```

Gambar 2. Training Data dengan Python

b. Hyperparameter Tuning

Untuk mendapatkan performa terbaik dari model Random Forest, kita dapat melakukan pencarian hyperparameter menggunakan Grid Search. Berikut adalah contoh kode untuk melakukan hyperparameter tuning:

```
from sklearn.model_selection import GridSearchCV

# Mendefinisikan grid hyperparameter yang akan dicari
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [10, 20, 30],
    'max_features': ['auto', 'sqrt', 'log2']
}

# Menggunakan GridSearchCV untuk menemukan hyperparameter terbaik
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=5, scoring='accuracy')
grid_search.fit(X_train, y_train)

# Mengambil model terbaik dari hasil pencarian
best_rf_model = grid_search.best_estimator_
print("Best parameters found: ", grid_search.best_params_)
```

Gambar 3. Hyperparameter Tuning dengan Python

c. Evaluasi Model

Setelah model terbaik ditemukan, langkah berikutnya adalah mengevaluasi kinerjanya menggunakan data testing. Evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, F1-score, dan confusion matrix.

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Memprediksi label untuk data testing
y_pred = best_rf_model.predict(X_test)

# Menghitung metrik evaluasi
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')

print(f'Accuracy: {accuracy}')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1 Score: {f1}')

# Confusion Matrix
conf_matrix = confusion_matrix(y_test, y_pred)
sns.heatmap(conf_matrix, annot=True, fmt='d')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

Gambar 4. Proses Evaluasi Model

d. Interpretasi Model

Menggunakan feature importance untuk mengidentifikasi fitur-fitur yang paling berkontribusi terhadap prediksi risiko gizi buruk.

```
# Mengambil feature importance dari model
feature_importance = best_rf_model.feature_importances_
features = X.columns
importances = pd.DataFrame({'Feature': features, 'Importance': feature_importance})
importances = importances.sort_values(by='Importance', ascending=False)

# Visualisasi feature importance
sns.barplot(x='Importance', y='Feature', data=importances)
plt.title('Feature Importance')
plt.show()
```

Gambar 5. Feature Importance

Model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Confusion matrix juga digunakan untuk mengevaluasi kinerja model. Feature importance dihitung untuk menentukan kontribusi masing-masing fitur.

Model Random Forest menunjukkan kinerja yang baik dalam mengklasifikasikan risiko gizi buruk pada ibu hamil. Berikut adalah hasil evaluasi:

Tabel 4. Hasil Evaluasi

| | |
|---------|------|
| Akurasi | 0.85 |
| Presisi | 0.84 |
| Recall | 0.85 |

| | |
|----------|------|
| F1 Score | 0.84 |
|----------|------|

Feature Importance

Hasil evaluasi menunjukkan bahwa model Random Forest memiliki kinerja yang baik dalam mengklasifikasikan risiko gizi buruk pada ibu hamil. Confusion matrix menunjukkan bahwa model mampu mengklasifikasikan risiko gizi buruk dengan baik, meskipun ada beberapa kesalahan klasifikasi. Feature importance menunjukkan bahwa fitur usia dan akses layanan kesehatan memiliki kontribusi terbesar dalam prediksi risiko gizi buruk.

4. SIMPULAN

Penelitian ini menunjukkan bahwa metode Random Forest efektif dalam mengklasifikasikan risiko gizi buruk pada ibu hamil berdasarkan data demografi dan pola makan. Model ini dapat digunakan untuk mengidentifikasi ibu hamil yang berisiko tinggi dan memerlukan intervensi lebih lanjut. Penelitian di masa depan dapat menggunakan data yang lebih luas dan beragam untuk meningkatkan akurasi dan generalisasi model.

UCAPAN TERIMAKASIH

Terimakasih kepada LPPM UNIMED yang telah memberikan dana penelitian sehingga penelitian ini dapat diselesaikan dengan baik.

PUSTAKA

- Adriati, F., & Chloranyta, S. (2022). Status Gizi Ibu Hamil Berdasarkan Pengukuran Lingkar Lengan Atas (LILA). *Jurnal Kesehatan Panca Bhakti Lampung*, 10(2), 127. <https://doi.org/10.47218/jkpbl.v10i2.194>
- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05.
- Ediyono, S. (2023). Dampak Kurangnya Nutrisi Pada Ibu Hamil Terhadap Risiko Stunting Pada Bayi Yang Dilahirkan. In *Jurnal Ilmu Keperawatan dan Kebidanan* (Vol. 14, Issue 1).
- Fitri Salsabilla Arianti, D., Arum, L., Burhanudin, M., Yani Yogyakarta, A., & Siliwangi Ringroad Barat, J. (2023). Analisis Prediksi Kematian Pasien Covid-19 di Meksiko Menggunakan Algoritma Random Forest. In *Indonesian Journal on Data Science* (Vol. 1, Issue 2). <https://ejournal.unjaya.ac.id/index.php/ijds>
- Genç, B., & Tunç, H. (2019). Optimal training and test sets design for machine learning. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(2), 1534–1545. <https://doi.org/10.3906/elk-1807-212>
- Kohsasih, K. L., Situmorang, Z., & Artikel, I. (2022). Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular. *JURNAL INFORMATIKA*, 9(1), 13–17. <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Maghfiroh, A., Findawati, Y., & Indahyanti, U. (2023). Klasifikasi Penipuan pada Rekening Bank menggunakan Pendekatan Ensemble Learning. *Building of Informatics, Technology and Science (BITS)*, 4(4). <https://doi.org/10.47065/bits.v4i4.3212>

-
- Sari, I. P., Al-Khowarizmi, Ramadhani, F., Satria, A., & Sulaiman, O. K. (2023). Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. *6th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2023 - Proceeding*, 486–491. <https://doi.org/10.1109/ISRITI60336.2023.10467242>
- Satria, A., Sitompul, O. S., & Mawengkang, H. (2021). 5-Fold Cross Validation on Supporting K-Nearest Neighbour Accuration of Making Consimilar Symptoms Disease Classification. *Proceedings - 2nd International Conference on Computer Science and Engineering: The Effects of the Digital World After Pandemic (EDWAP), IC2SE 2021*. <https://doi.org/10.1109/IC2SE52832.2021.9792094>
- Supriyadi, R., Gata, W., Maulidah, N., Fauzi, A., Komputer, I., & Nusa Mandiri Jalan Margonda Raya No, S. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *13(2)*, 67–75. <http://journal.stekom.ac.id/index.php/E-Bisnis> page67
- sutarto, sutarto, Oktarlina, R. Z., Adila, N. T. H., & Indriyani, R. (2022). Status Kesehatan Anak Stunting Indonesia (Analisa Data Riskesdas 2018). *Jurnal Dunia Kesmas*, *11(2)*. <https://doi.org/10.33024/JDK.V11I2.6930>