

PREDIKSI DIABETES MELLITUS TIPE 2 DENGAN ALGORITMA LOGISTIC REGRESSION UNTUK PENDETEKSIAN DINI

Sepriyan Gunawan^{1*}, Rini Astuti², Willy Priharto³, Ryan Hamonangan⁴

^{1,3,4} STMIK IKMI Cirebon; Jl. Perjuangan No. 10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Telp: 0231-490480

² STMIK LIKMI Bandung; Jl. Ir. H. Juanda No.96, Lebakgede, Kecamatan Coblong, Kota Bandung, Jawa Barat, Telp: 022-2502121

Received: 18 Desember 2024
Accepted: 14 Januari 2025
Published: 20 Januari 2025

Keywords:

Logistic Regression,
Prediksi, Diabetes,
Data Mining,
Health Data.

Correspondent Email:

sepriyangunawan16@gmail.com

Diabetes Mellitus Tipe 2 merupakan penyakit kronis yang prevalensinya terus meningkat di dunia, termasuk Indonesia. Deteksi dini risiko diabetes penting untuk mencegah komplikasi serius. Penelitian ini bertujuan untuk menerapkan algoritma Logistic Regression dalam memprediksi risiko Diabetes Mellitus Tipe 2 menggunakan data kesehatan pasien dari Puskesmas Jatibarang. Data yang digunakan meliputi variabel umur, jenis kelamin, kebiasaan merokok, konsumsi alkohol, hasil IMT, tekanan darah sistole dan diastole, serta riwayat penyakit. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan AUC, serta diukur tingkat kesalahannya dengan RMSE. Hasil penelitian menunjukkan bahwa Logistic Regression memiliki performa baik dengan akurasi 82,51%, precision 88,04%, recall 89,76%, F1-score 88,89%, AUC 83,12%, dan RMSE sebesar 0.4182. Fitur utama yang berpengaruh terhadap prediksi adalah umur, pola makan (kurang sayur/buah), dan tekanan darah diastolik. Model ini efektif untuk deteksi dini risiko Diabetes Mellitus Tipe 2 dan dapat digunakan sebagai dasar pengambilan keputusan di fasilitas kesehatan primer.

Type 2 Diabetes Mellitus is a chronic disease with increasing prevalence worldwide, including in Indonesia. Early risk detection is crucial to prevent severe complications. This study aims to implement the Logistic Regression algorithm to predict Type 2 Diabetes Mellitus risk using patient health data from Puskesmas Jatibarang. The dataset includes variables such as age, gender, smoking habits, alcohol consumption, BMI, systolic and diastolic blood pressure, and medical history. Model evaluation was conducted using accuracy, precision, recall, F1-score, and AUC metrics, with error measured by RMSE. Results indicate that Logistic Regression performed well with an accuracy of 82.51%, precision of 88.04%, recall of 89.76%, F1-score of 88.89%, AUC of 83.12%, and RMSE of 0.4182. Key contributing features include age, diet (low vegetable/fruit intake), and diastolic blood pressure. The model is effective for early detection of Type 2 Diabetes Mellitus risk and can serve as a basis for decision-making in primary healthcare facilities.

1. PENDAHULUAN

Diabetes Mellitus Tipe 2 (DMT2) adalah penyakit kronis dengan prevalensi yang terus meningkat secara global [1]. Di Indonesia,

angka kejadian DMT2 menunjukkan tren yang mengkhawatirkan, terkait dengan pola makan yang buruk, obesitas, dan gaya hidup tidak sehat [2]. DMT2 sering terdeteksi terlambat, yang

dapat menyebabkan komplikasi serius seperti penyakit jantung dan kebutaan [3]. Oleh karena itu, deteksi dini sangat penting untuk mengidentifikasi individu berisiko tinggi dan memberikan intervensi lebih awal [4].

Beberapa penelitian telah menguji penerapan *Logistic Regression* (LR) untuk prediksi Diabetes Melitus Tipe 2 (DMT2) menggunakan berbagai dataset, baik yang berbasis data global maupun lokal. Sebagai contoh, penelitian oleh [5] membandingkan LR dengan *K-Nearest Neighbor* (KNN) dalam klasifikasi penyakit kanker serviks, dan LR menunjukkan hasil yang lebih stabil dengan akurasi 85,5%.

Dalam penelitian lainnya, [4] menggunakan LR untuk memprediksi risiko DMT2 dan mencapai hasil yang baik, dengan akurasi 84%, precision 82,5%, recall 85,3%, dan F1-score 83,8%. Penelitian oleh [1] juga melaporkan bahwa LR memiliki kemampuan untuk memprediksi DMT2 dengan akurasi 82% pada dataset yang berfokus pada faktor-faktor risiko kesehatan.

Namun, sebagian besar penelitian tersebut menggunakan dataset yang lebih besar dan lebih terstruktur, sedangkan penelitian ini berfokus pada penggunaan data lokal dengan karakteristik yang lebih bervariasi dan keterbatasan data yang sering kali tidak lengkap.

Penelitian ini bertujuan untuk menguji kinerja LR dalam memprediksi DMT2 dengan menggunakan data lokal dari Puskesmas Jatibarang. Meskipun penelitian sebelumnya banyak menggunakan dataset besar dan komprehensif, penelitian ini fokus pada aplikasi LR dalam konteks data lokal untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap prediksi DMT2.

Tujuan utama dari penelitian ini adalah untuk menerapkan algoritma LR dalam memprediksi DMT2 menggunakan data lokal dari Puskesmas Jatibarang dan mengevaluasi kinerjanya dalam mendeteksi individu berisiko tinggi terkena DMT2. Hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem deteksi dini diabetes yang lebih efektif di layanan kesehatan primer di Indonesia.

2. TINJAUAN PUSTAKA

2.1. *Diabetes Mellitus Tipe 2 (DMT2)*

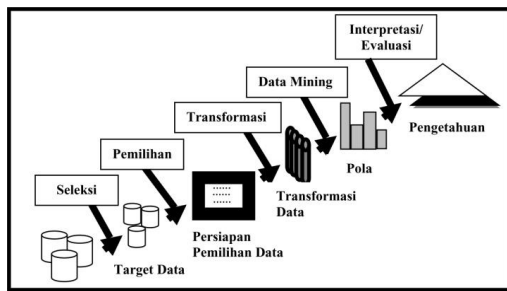
Diabetes Mellitus Tipe 2 (DMT2) adalah penyakit kronis yang ditandai dengan peningkatan kadar gula darah akibat resistensi insulin [6]. Beberapa Faktor risiko utama DMT2 meliputi usia, obesitas, hipertensi, pola makan tidak sehat, kebiasaan merokok, serta riwayat keluarga [2], [7]. Deteksi dini DMT2 sangat penting untuk mencegah komplikasi berat, seperti gangguan jantung dan kerusakan ginjal [3], [8]. Oleh karena itu, penelitian ini menggunakan data kesehatan untuk mengidentifikasi individu berisiko tinggi terkena DMT2 secara dini.

2.2. *Logistic Regression (LR)*

Logistic Regression (LR) adalah metode statistik yang digunakan untuk memodelkan probabilitas suatu variabel biner berdasarkan hubungan antara variabel independen dan variabel dependen [9]. LR sering digunakan dalam aplikasi kesehatan untuk memprediksi kejadian seperti risiko penyakit [4]. Model ini memanfaatkan fungsi logit untuk mengubah hubungan linier menjadi probabilitas, yang memudahkan dalam klasifikasi biner [10]. LR dapat mengidentifikasi faktor risiko utama, seperti umur, tekanan darah, dan indeks massa tubuh, yang berperan dalam prediksi DMT2 [2], [7].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis dan memprediksi risiko Diabetes Mellitus Tipe 2 (DMT2) dengan algoritma *Logistic Regression*. Metode yang digunakan mencakup pengumpulan, pengolahan, analisis, dan penyajian data secara objektif untuk mengidentifikasi pola risiko yang dapat mendeteksi DMT2 lebih awal [8]. Penelitian ini menggunakan metodologi *Knowledge Discovery in Databases* (KDD), yang terdiri dari lima tahap utama: pemilihan data, pra-pemrosesan, transformasi, penambangan data, dan evaluasi [11]. Proses ini bertujuan untuk menemukan pola-pola penting yang dapat mendukung pengambilan keputusan dalam sistem deteksi dini DMT2.



Gambar 1 Tahapan KDD

3.1. Seleksi Data (Data Selection)

Pada tahap ini, data yang relevan untuk penelitian ini dipilih, yaitu data rekam medis dari Puskesmas Jatibarang terdiri dari 890 record dengan 13 atribut. Variabel yang dipilih meliputi faktor risiko yang berhubungan langsung dengan DMT2, seperti jenis kelamin, umur, kebiasaan merokok, konsumsi alkohol, kurangnya konsumsi sayur/buah, hasil IMT, tekanan darah sistole dan diastole, hipertensi, riwayat penyakit jantung, hiperlipidemia, dan riwayat keluarga. Pemilihan variabel ini didasarkan pada literatur yang menunjukkan bahwa faktor-faktor ini berperan penting dalam perkembangan DMT2 [2], [3], [4].

Tabel 1 Atribut Dataset

Atribut	Keterangan
Jenis Kelamin	Menunjukkan jenis kelamin pasien (Laki-laki/Perempuan)
Umur	Usia pasien dalam tahun
Merokok	Status kebiasaan merokok pasien (Ya/Tidak).
Konsumsi Alkohol	Status konsumsi alkohol pasien (Ya/Tidak).
Kurang Sayur/Buah	Indikator kurangnya konsumsi sayur/buah (Ya/Tidak).
Hasil IMT	Kategori Indeks Massa Tubuh (IDEAL, LEBIH, GEMUK, SANGAT GEMUK).
Sistole	Tekanan darah sistolik pasien dalam mmHg.
Diastole	Tekanan darah diastolik pasien dalam mmHg.
Hipertensi	Indikator adanya hipertensi pada pasien (Ya/Tidak).
Riwayat Penyakit Jantung	Indikator adanya riwayat penyakit jantung pada pasien (Ya/Tidak).
Hiperlipidemia	Status kadar lemak darah tinggi (Ya/Tidak).
Riwayat Keluarga	Indikator riwayat keluarga dengan diabetes (Ya/Tidak).
Diabetes (target)	Status diagnosis Diabetes Mellitus Tipe 2 (Positif/Negatif).

3.2. Pemilihan Data (Data Pre-processing)

Data yang dipilih akan dibersihkan dengan menghapus data yang tidak lengkap atau tidak relevan. Proses ini melibatkan penghapusan baris yang memiliki nilai hilang. Langkah ini dilakukan untuk menjaga kualitas dataset sehingga dapat meningkatkan akurasi model yang akan dibangun [10].

3.3. Transformasi Data

Pada tahap ini, data kategorikal diubah menjadi numerik, dan variabel numerik seperti tekanan darah dinormalisasi menggunakan *StandardScaler* untuk memastikan skala yang seragam. Setelah normalisasi, data dibagi menjadi dua subset: training data (70%) dan testing data (30%), untuk mempersiapkan proses pelatihan dan evaluasi model dengan data yang belum pernah dilihat sebelumnya.

3.4. Data Mining (Implementasi Model)

Pada tahap ini, algoritma *Logistic Regression* digunakan untuk membangun model prediksi risiko DMT2. *Logistic Regression* dipilih untuk memodelkan probabilitas kejadian berdasarkan variabel risiko. Teknik *k-fold cross-validation* diterapkan untuk menghindari overfitting dan memastikan model dapat digeneralisasi dengan baik [12]. Data dibagi menjadi *k folds*, dan proses pelatihan dilakukan sebanyak *k* iterasi untuk menghasilkan estimasi akurasi yang lebih stabil.

3.5. Evaluasi (Interpretation/Evaluation)

Model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, RMSE (*Root Mean Squared Error*), dan AUC untuk menilai efektivitas dalam memprediksi risiko DMT2. Selain itu, *confusion matrix* digunakan untuk menganalisis performa model dengan melihat distribusi *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* [8], [11]. Metrik ini memberikan gambaran menyeluruh mengenai kemampuan model dalam mendeteksi kasus positif dan negatif. Hasil evaluasi ini memberikan wawasan tentang kesalahan prediksi dan faktor risiko utama yang memengaruhi hasil, sehingga model dapat dioptimalkan untuk deteksi dini DMT2 yang lebih akurat [13].

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Data yang digunakan dalam penelitian ini adalah rekam medis pasien dari Puskesmas Jatibarang selama periode Januari hingga September 2024. Dataset terdiri dari 890 baris dan 13 atribut. Seleksi data dilakukan untuk memastikan bahwa hanya atribut yang relevan dengan risiko diabetes yang digunakan, sementara data pribadi seperti nama pasien dan NIK dihilangkan.

	Jenis Kelamin	Umur	Merokok	Konsumsi Alkohol	Kurang Sayur/Buah	Hasil IMT	Sistolik	Diastolik	hipertensi	Riwayat Penyakit Jantung	hyperlipidaemia	Riwayat Keturunan	Diabetes
0	Pemempuan	50.0	Tidak	Tidak	Tidak	LEBIH	100	70	Tidak	Tidak	Tidak	Tidak	Tidak
1	Laki-laki	51.0	Ya	Tidak	Tidak	LEBIH	140	80	Tidak	Tidak	Tidak	Tidak	Tidak
2	Pemempuan	53.0	Tidak	Tidak	Tidak	IDEAL	140	80	Tidak	Tidak	Tidak	Tidak	Positif
3	Pemempuan	52.0	Tidak	Tidak	Tidak	IDEAL	150	90	Ya	Ya	Tidak	Tidak	Positif
4	Pemempuan	58.0	Tidak	Tidak	Tidak	IDEAL	100	70	Tidak	Ya	Tidak	Tidak	Positif
...													
885	Pemempuan	53.0	Tidak	Tidak	Tidak	IDEAL	140	80	Ya	Tidak	Tidak	Tidak	Negatif
886	Pemempuan	33.0	Tidak	Tidak	Tidak	IDEAL	120	80	Tidak	Tidak	Tidak	Tidak	Negatif
887	Pemempuan	30.0	Tidak	Tidak	Tidak	IDEAL	120	80	Tidak	Tidak	Tidak	Tidak	Negatif
888	Pemempuan	31.0	Tidak	Tidak	Tidak	IDEAL	120	80	Tidak	Tidak	Tidak	Tidak	Negatif
889	Pemempuan	55.0	Tidak	Tidak	Tidak	GEMUK	130	90	Tidak	Tidak	Tidak	Tidak	Negatif

Gambar 2 Dataset Diabetes Mellitus tipe 2

4.2. Data Pre-processing

Tahap pre-processing dilakukan untuk memastikan kualitas dataset sebelum digunakan. Langkah pertama adalah identifikasi nilai kosong (*missing values*) dengan menggunakan Python.

```
print(df.isnull().sum())
```

```
Jenis Kelamin      0
Umur                9
Merokok            1
Konsumsi Alkohol   1
Kurang Sayur/Buah  1
Hasil IMT          4
Sistolik           0
Diastolik          0
hipertensi         1
Riwayat Penyakit Jantung 0
hyperlipidaemia    0
Riwayat Keturunan  0
Diabetes           0
dtype: int64
```

Gambar 3 Missing Values

Gambar 3 menunjukkan hasil pemeriksaan data kosong pada dataset dengan fungsi `df.isnull().sum()`. Kolom Umur memiliki 9 data kosong, Hasil IMT 4, dan kolom lain seperti kurang buah/sayur, hipertensi, Merokok dan Konsumsi Alkohol masing-masing 1. Baris dengan data kosong akan dihapus menggunakan `dropna()` untuk menjaga validitas analisis.

4.3. Transformasi Data

Data yang telah dibersihkan ditransformasikan untuk analisis lebih lanjut. Variabel kategorik, seperti jenis kelamin dan kebiasaan merokok, dikonversi menjadi numerik menggunakan *Label Encoding*, sementara variabel numerik, seperti IMT dan

tekanan darah, dinormalisasi dengan *StandardScaler* untuk menyamakan skala menjadi distribusi dengan rata-rata 0 dan standar deviasi 1.

```
label_encoders = {}
for column in df.columns:
    if df[column].dtype == 'object':
        le = LabelEncoder()
        df[column] = le.fit_transform(df[column])
        label_encoders[column] = le
```

```
scaler = StandardScaler()
X = df.drop('Diabetes', axis=1)
X = scaler.fit_transform(X)
y = df['Diabetes']
```

Gambar 4 Proses Transformasi Data

```
array([[ 0.63903062, -0.65076439, -0.37301708, ..., -0.17139779,
        -0.21873932, -0.20094787],
       [-1.56487025, -0.12543531,  2.68084238, ..., -0.17139779,
        -0.21873932, -0.20094787],
       [ 0.63903062,  0.92522284, -0.37301708, ..., -0.17139779,
        -0.21873932, -0.20094787],
       ...,
       [ 0.63903062, -1.96408708, -0.37301708, ..., -0.17139779,
        -0.21873932, -0.20094787],
       [ 0.63903062, -1.61386769, -0.37301708, ..., -0.17139779,
        -0.21873932, -0.20094787],
       [ 0.63903062,  0.22478407, -0.37301708, ..., -0.17139779,
        -0.21873932, -0.20094787]])
```

Gambar 5 hasil Transformasi Data

Transformasi ini memastikan data kompatibel dengan algoritma *machine learning* dan meningkatkan akurasi model.

Setelah transformasi, dataset dibagi menjadi data latih dan data uji dengan perbandingan 70:30 menggunakan fungsi `train_test_split` dari *Scikit-learn*. Data latih (`X_train` dan `y_train`) terdiri atas 613 sampel, sedangkan data uji (`X_test` dan `y_test`) terdiri atas 263 sampel.

4.4. Data Mining

Pada tahap ini, algoritma *Logistic Regression* diterapkan untuk membangun model prediksi Diabetes Mellitus Tipe 2 menggunakan pustaka *Scikit-learn*. Model dilatih menggunakan data latih (`X_train` dan `y_train`) dengan metode `fit()`.

```
from sklearn.model_selection import train_test_split, cross_val_score, KFold
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression

logreg = LogisticRegression()
logreg.fit(X_train, y_train)
```

```
LogisticRegression
```

Gambar 6 Model Logistic Regression

```
kfold = KFold(n_splits=5, shuffle=True, random_state=42)
cv_scores_logreg = cross_val_score(logreg, X, y, cv=kfold, scoring='accuracy')
```

Gambar 7 Teknik K-Fold Cross Validation

Teknik *K-Fold Cross Validation* diterapkan dengan 5 fold ($n_splits=5$), di mana data diacak ($shuffle=True$) untuk memastikan pembagian yang merata, dan parameter $random_state=42$ digunakan untuk menghasilkan hasil yang konsisten. Teknik ini memastikan model bekerja stabil pada data baru dan mengurangi risiko overfitting.

4.5. Evaluasi (Interpretation/Evaluation)

4.5.1. Evaluasi Model Logistic Regression

Evaluasi model Logistic Regression dilakukan menggunakan beberapa metrik, yaitu akurasi, *precision*, *recall*, *F1-score*, dan *AUC-ROC Curve*.

```
accuracy = accuracy_score(y_test, y_pred_logreg)
precision = precision_score(y_test, y_pred_logreg)
recall = recall_score(y_test, y_pred_logreg)
f1 = f1_score(y_test, y_pred_logreg)
auc_logreg = roc_auc_score(y_test, logreg.predict_proba(X_test)[:,1])

print(f'Logistic Regression - Akurasi: {accuracy}')
print(f' Precision: {precision}')
print(f' Recall: {recall}')
print(f' F1 Score: {f1}')
print(f' AUC score :{auc_logreg}')
```

Logistic Regression - Akurasi: 0.8250950570342205
Precision: 0.8803827751196173
Recall: 0.8975609756097561
F1 Score: 0.8888888888888888
AUC score :0.8312447434819176

Gambar 8 Evaluasi Model Logistic Regression

Hasil evaluasi menunjukkan bahwa model memiliki akurasi 82,51%, *precision* 88,04%, *recall* 89,76%, *F1-score* 88,89%, dan AUC 83,12%. Nilai-nilai ini menunjukkan performa yang baik dalam mendeteksi risiko Diabetes Mellitus Tipe 2, dengan kemampuan model yang cukup andal dalam mengidentifikasi kasus positif dan negatif.

4.5.2. Evaluasi dengan Metrik Error

```
from sklearn.metrics import mean_squared_error, mean_absolute_error

rmse_logreg = mean_squared_error(y_test, y_pred_logreg, squared=False)
print(f'Logistic Regression ')
print(f'RMSE: {rmse_logreg}')
```

Logistic Regression
RMSE: 0.418216382947607

Gambar 9 Metrik Error

Gambar 9 menunjukkan Model *Logistic Regression* dievaluasi menggunakan metrik RMSE (*Root Mean Squared Error*) yang menghasilkan nilai sebesar 0.4182. Nilai ini menunjukkan rata-rata kesalahan prediksi model dalam satuan asli variabel target. Dengan nilai error yang relatif kecil, model ini menunjukkan kemampuan yang baik dalam

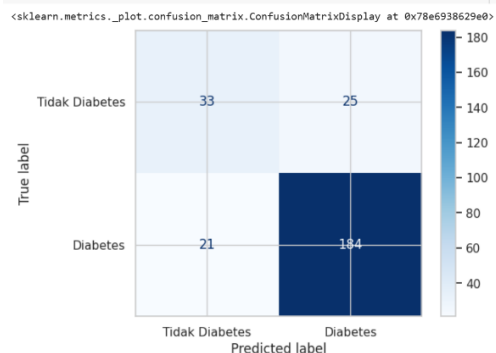
memberikan prediksi yang akurat untuk risiko Diabetes Mellitus Tipe 2.

4.5.3. Analisis Confusion Matrix

```
# Confusion Matrix logistic Regression
import matplotlib.pyplot as plt
import seaborn as sns

cm_logreg = confusion_matrix(y_test, y_pred_logreg)
disp_logreg = ConfusionMatrixDisplay(cm_logreg, display_labels=['Tidak Diabetes', 'Diabetes'])
disp_logreg.plot(cmap='Blues')
```

Gambar 10 Proses Confusion Matrix



Gambar 11 Confusion Matrix

Confusion Matrix memberikan gambaran rinci tentang prediksi model, dengan 184 *True Positives* (TP), 33 *True Negatives* (TN), 25 *False Positives* (FP), dan 21 *False Negatives* (FN). Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data dan lebih efektif dalam mendeteksi kasus diabetes.

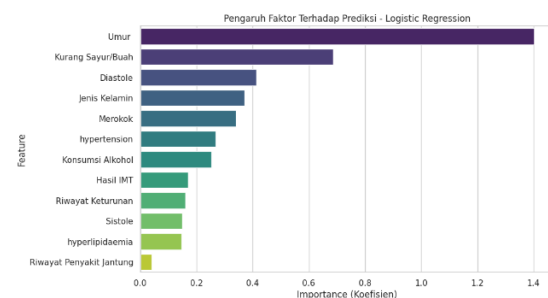
4.5.4. Analisis Koefisien Fitur

```
# Mendapatkan koefisien dari model Logistic Regression
feature_importance_logreg = pd.DataFrame({
    'Feature': df.drop('Diabetes', axis=1).columns,
    'Importance': np.abs(logreg.coef_[0])
}).sort_values(by='Importance', ascending=False)

# Menampilkan hasil
print("Faktor yang Paling Mempengaruhi Berdasarkan Logistic Regression:")
print(feature_importance_logreg)

plt.figure(figsize=(10, 6))
sns.barplot(data=feature_importance_logreg, x='Importance', y='Feature', palette='viridis')
plt.title("Pengaruh Faktor Terhadap Prediksi - Logistic Regression")
plt.xlabel("Importance (Koefisien)")
plt.ylabel("Feature")
plt.show()
```

Gambar 12 Proses Koefisien Fitur



Gambar 13 Koefisien Fitur

Koefisien Logistic Regression digunakan untuk mengidentifikasi fitur yang paling memengaruhi prediksi. Umur merupakan fitur dengan pengaruh terbesar, diikuti oleh Kurang Sayur/Buah dan Diastole, yang menyoroti pentingnya usia dan pola makan dalam memprediksi risiko diabetes. Faktor lain seperti Jenis Kelamin, Merokok, dan *Hypertension* juga memberikan kontribusi signifikan.

Hasil ini memberikan wawasan mendalam tentang faktor risiko utama yang dapat menjadi fokus dalam upaya deteksi dini dan pencegahan diabetes, serta menunjukkan efektivitas *Logistic Regression* dalam prediksi berbasis data kesehatan.

5. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan sebagai berikut:

1. Tahapan prediksi risiko Diabetes Mellitus Tipe 2 menggunakan algoritma *Logistic Regression* meliputi Seleksi data, Pemihilan data, *Transformation*, Data Mining, dan *Evaluation*.
2. Hasil evaluasi model *Logistic Regression* menunjukkan performa yang baik dengan akurasi 82,51%, *precision* 88,04%, *recall* 89,76%, *F1-score* 88,89%, AUC 83,12%, dan RMSE sebesar 0.4182, yang mengindikasikan tingkat kesalahan prediksi rendah.
3. Model menunjukkan kemampuan mendeteksi kasus positif diabetes dengan 184 *True Positives* (TP) dan hanya 21 *False Negatives* (FN).
4. Faktor-faktor utama yang memengaruhi prediksi risiko diabetes adalah Umur, Kurang Sayur/Buah, dan Tekanan Diastolik, yang berperan signifikan dalam hasil prediksi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," *IEEE Access*, vol. 9, no. February, pp. 43711–43735, 2021, doi: 10.1109/ACCESS.2021.3059343.
- [2] P. Vas, D. Hopkins, M. Feher, F. Rubino, and M. B. Whyte, "Diabetes, obesity and COVID-19: A complex interplay," *Diabetes, Obes. Metab.*, vol. 22, no. 10, pp. 1892–1896, Oct. 2020, doi: 10.1111/dom.14134.
- [3] J. B. Buse *et al.*, "2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)," *Diabetes Care*, vol. 43, no. 2, pp. 487–493, Feb. 2020, doi: 10.2337/dci19-0066.
- [4] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Appl. Comput. Informatics*, vol. 18, no. 1/2, pp. 90–100, Mar. 2022, doi: 10.1016/j.aci.2018.12.004.
- [5] N. D. Azzahra, A. Ambarwati, A. Desiani, S. I. Maiyanti, and I. Ramayanti, "Perbandingan Algoritma K-Nearest Neighbor Dan Logistic Regression Dalam Klasifikasi Penyakit Kanker Serviks," *Energy J. Ilm. Ilmu-Ilmu Tek.*, vol. 14, no. 1, pp. 1–8, 2024, doi: 10.51747/energy.v14i1.1843.
- [6] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 2, no. January, pp. 40–46, 2021, doi: 10.1016/j.ijcce.2021.01.001.
- [7] Y. Denggos, "Penyakit Diabetes Mellitus Umur 40-60 Tahun di Desa Bara Batu Kecamatan Pangkep," *Heal. J. Ilm. Kesehat.*, vol. 2, no. 1, pp. 55–61, Mar. 2023, doi: 10.47709/healthcaring.v2i1.2177.
- [8] C. A. Rahayu, "PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, pp. 261–266, Aug. 2023, doi: 10.23960/jitet.v11i3.3055.
- [9] D. Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, Feb. 2024, doi: 10.58794/jekin.v4i1.640.
- [10] I. N. Laela and W. M. Baihaqi, "Analisis Klasifikasi Penyakit Multiple Sclerosis Menggunakan Algoritma Logistic Regression dan SVM," *Gener. J.*, vol. 8, no. 1, pp. 27–33, Jan. 2024, doi: 10.29407/gj.v8i1.20646.
- [11] N. H. Alfajr and S. Defiyanti, "PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE RANDOM FOREST DAN PENERAPAN PRINCIPAL COMPONENT

- ANALYSIS (PCA),” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, pp. 3457–3464, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5055.
- [12] B. M. Karomah, “PENERAPAN METODE STACKING DALAM MENGLASIFIKASIKAN PENDERITA PENYAKIT DIABETES,” *J. Publ. Ilmu Komput. dan Multimed.*, vol. 1, no. 3, pp. 188–194, Jan. 1970, doi: 10.55606/jupikom.v1i3.522.
- [13] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, “IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION,” *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 493, Dec. 2022, doi: 10.37600/tekinkom.v5i2.698.