

LAPORAN PRAKTIKUM PM

Penerapan Teori Model Linear Untuk Klasifikasi Pada Dataset Health Care

Dede Masita ¹⁾, Ayu Erlinawati ²⁾, Sella Dianka Fitri ³⁾, Rizki Adrian Bennovry ⁴⁾, Saiful Haris Muhammad ⁵⁾, Muhammad Kaisar Firdaus ⁶⁾

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

Email : dede.120450007@student.itera.ac.id ¹⁾ ayu.1204500025@student.itera.ac.id

²⁾ sella.1204500043@student.itera.ac.id ³⁾ rizki.120450073@student.itera.ac.id ⁴⁾

saiful.120450115@student.itera.ac.id ⁵⁾ muhammad.120450135@student.itera.ac.id ⁶⁾

Abstrak

Pentingnya perlindungan privasi dalam data perawatan kesehatan yang sensitif dan rahasia. Dengan adanya regulasi privasi yang ketat, data perawatan kesehatan sintetis dibuat untuk tujuan pendidikan dan penelitian. Metode analisis data, seperti Exploratory Data Analysis (EDA) dan pemodelan menggunakan Decision Tree, Random Forest, Neural Network, dan Support Vector Machine (SVM), digunakan untuk memahami dan menghasilkan wawasan yang berguna dari dataset ini, penggunaan metode analisis data dan pemodelan memberikan pemahaman yang mendalam terhadap dataset perawatan kesehatan sintetis untuk keperluan pendidikan dan penelitian.

Kata Kunci: Perlindungan Privasi, Data Perawatan Kesehatan, EDA, Decision Tree, Random Forest, Neural Network, Support Vector Machine.

1. Pendahuluan

Pembelajaran mesin (ML) semakin mendapat tempat penting dalam masyarakat kita. Digunakan di banyak bidang, meskipun hal ini dapat berdampak langsung pada warga negara seperti dalam perekrutan proses [1]. Data perawatan kesehatan seringkali sensitif karena berisi informasi pribadi dan medis yang sangat rahasia. Informasi ini mencakup riwayat medis, kondisi kesehatan, obat-obatan yang dikonsumsi dan informasi pribadi lainnya. Karena informasi ini sangat pribadi, maka perlindungan privasi sangat penting untuk mencegah penyalahgunaan informasi dan diskriminasi. Oleh karena itu data perawatan kesehatan patuh ada peraturan privasi yang ketat dan sulit diakses untuk pembelajaran dan eksperimen. Kepercayaan dapat ditingkatkan setelah kita memahami bagaimana dan mengapa model mendalam membentuk batasan keputusan tertentu dan dengan demikian membuat keputusan tertentu[2].

Karena sensitifnya data perawatan dibuatlah dataset perawatan kesehatan sintetis yang dirancang untuk memberikan data perawatan kesehatan yang praktis dan beragam

untuk tujuan pendidikan dan penelitian. Kami akan menggunakan beberapa metode seperti EDA dan modeling Decision Tree, Random Forest dan Neural Network. Metode-metode ini akan membantu dalam memahami data dengan lebih baik dan menghasilkan wawasan yang berguna. Sehingga dalam laporan ini akan dibahas hasil analisis dan interpretasi data menggunakan metode-metode ini.

2. Metode

2.1 Data

Setiap kolom menyediakan informasi spesifik tentang pasien, penerimaan mereka, dan layanan perawatan kesehatan yang diberikan, sehingga dataset ini cocok untuk berbagai tugas analisis data dan pemodelan di domain perawatan kesehatan. Berikut adalah penjelasan singkat tentang setiap kolom dalam dataset:

- Nama: Kolom ini mewakili nama pasien yang terkait dengan catatan kesehatan.
- Usia: Usia pasien pada saat masuk, diungkapkan dalam tahun.
- Jenis Kelamin: Menunjukkan jenis kelamin pasien, baik “Laki-laki” atau “Perempuan”.
- Golongan Darah: Golongan darah pasien, yang dapat menjadi salah satu golongan darah umum (misalnya, “A +”, “O-”, dll.).
- Kondisi Medis: Kolom ini menentukan kondisi medis atau diagnosis utama yang terkait dengan pasien, seperti “Diabetes,” “Hipertensi,” “Asma,” dan lainnya.
- Tanggal Masuk: Tanggal ketika pasien masuk ke fasilitas kesehatan.
- Dokter: Nama dokter yang bertanggung jawab atas perawatan pasien selama masa perawatan.
- Rumah Sakit: Mengidentifikasi fasilitas kesehatan atau rumah sakit tempat pasien dirawat.
- Penyedia Asuransi: Kolom ini menunjukkan penyedia asuransi pasien, yang dapat menjadi salah satu dari beberapa opsi, termasuk “Aetna,” “Blue Cross,” “Cigna,” “UnitedHealthcare,” dan “Medicare”.
- Jumlah Tagihan: Jumlah uang yang dibebankan untuk layanan kesehatan pasien selama masa perawatan mereka. Ini dinyatakan sebagai bilangan pecahan.
- Nomor Kamar: Nomor kamar tempat pasien diakomodasi selama masa perawatan mereka.
- Jenis Penerimaan: Menentukan jenis penerimaan, yang dapat berupa “Darurat,” “Elektif,” atau “Mendesak,” mencerminkan keadaan penerimaan.
- Tanggal Keluar: Tanggal ketika pasien keluar dari fasilitas kesehatan, berdasarkan tanggal masuk dan sejumlah hari acak dalam rentang yang realistis.
- Obat: Mengidentifikasi obat yang diresepkan atau diberikan kepada pasien selama masa perawatan mereka. Contohnya termasuk “Aspirin,” “Ibuprofen,” “Penicillin,” “Paracetamol,” dan “Lipitor.”

- Hasil Tes: Menggambarkan hasil tes medis yang dilakukan selama masa perawatan pasien. Nilai yang mungkin termasuk “Normal,” “Abnormal,” atau “Inconclusive,” yang menunjukkan hasil tes.

2.2 Rancangan Metode dan Model

1. Support Vector Machine

Algoritma Machine Learning yang digunakan untuk klasifikasi dan regresi. Berfokus ada menemukan hyperplane yang memisahkan dua kelas data dengan margin maksimum. Hyperplane dapat berupa garis lurus atau non-linier. Dengan Model umum ;

$$y_i \in \{-1, 1\}$$

$$w^T x_i + b \geq 1$$

2. Decision Tree

Merupakan model yang menggambarkan keputusan berdasarkan serangkaian aturan yang berhierarki. Model prediktif yang menggunakan struktur pohon untuk memetakan input ke output. Fokus utamanya adalah memahami bagaimana pemisahan berdasarkan fitur dapat mengklasifikasikan data dengan efisien dan berusaha meminimalkan ketidakpastian dengan memilih fitur terbaik untuk memisahkan data. **Entropy** mengukur ketidakpastian dalam kumpulan data. **Information Gain** mengukur perbedaan antara entropi sebelum dan sesudah, **Gini index** mengukur ketidakmurnian dalam kumpulan data.

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

3. Neural Network

Model Prediktif yang terinspirasi oleh cara kerja otak manusia. Berfokus pada pembelajaran dari data melalui jaringan *neural* buatan dengan Tujuan mengenali pola kompleks dan memodelkan hubungan non-linier antara input dan output. Pada *Neural Network* terdapat **Activation Function** yang mengkonversi input menjadi output, **Backpropagation** algoritma untuk memperbarui bobot jaringan berdasarkan kesalahan prediksi dan **Gradient Descent** metode untuk menemukan minimum lokal dalam fungsi dengan menggunakan gradien.

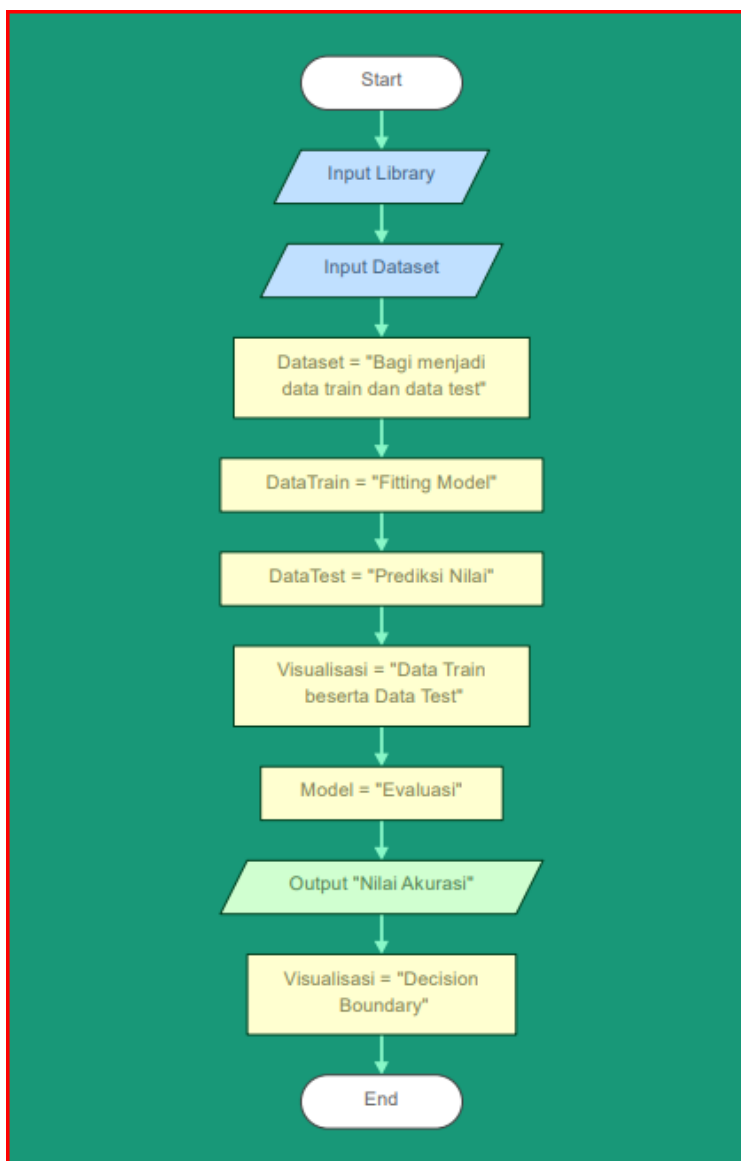
$$\Delta w_{ij} = - \eta dE/w_{ij}$$

$$f(x) = 1 : (1 + e^{-x})$$

4. Random Forest

Algoritma *ensemble* yang menggabungkan hasil dari beberapa *decision tree* untuk menghasilkan prediksi yang lebih akurat. Kita dapat memahami prinsip kerjanya dengan menggabungkan konsep-konsep dari *decision tree* dan *bagging*. Pada laporan ini kami menggunakan perbandingan Random Forest dengan *Bernoulli*, Regresi Logistic dan SVC untuk menemukan model yang cocok untuk dataset yang kami gunakan.

2.3 Flowchart



2.4 Pseudocode

1. Input Library: Mengimpor library yang diperlukan (seperti scikit-learn, matplotlib, dll.)
2. Input Dataset: Membaca dataset dari file atau sumber data lainnya.
3. Bagi Dataset: Memisahkan dataset menjadi data pelatihan (train data) dan data pengujian (test data).
4. Fitting Model: Memilih model yang sesuai (misalnya, Decision Tree, SVM, atau model lainnya) dan melakukan fitting model menggunakan data pelatihan.
5. Prediksi Nilai: Menggunakan model yang telah difitting untuk membuat prediksi menggunakan data uji.
6. Visualisasi Data Train dan Data Test: Membuat visualisasi untuk data pelatihan dan data pengujian (scatter plot, plot decision boundary, dll.).
7. Evaluasi Model: Mengukur kinerja model, biasanya menggunakan metrik seperti akurasi, presisi, recall, dll.
8. Output Nilai Akurasi: Menampilkan atau menyimpan nilai akurasi model.
9. Visualisasikan Decision Boundary: Jika model adalah model yang dapat divisualisasikan (seperti Decision Tree atau SVM), gambarkan decision boundary.
10. Selesai.

3. Hasil dan Analisis

3.1 EDA

a.) df.info()

Melakukan proses df.info() yang digunakan untuk menampilkan informasi nilai null dan tipe data dari setiap kolomnya. Dari hasil yang didapat, ada 14 kolom dengan data type berbeda di setiap kolomnya. Terdapat 3 jenis data type pada data yang digunakan yaitu; object, int64, float64.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 15 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   Name                  10000 non-null object  
 1   Age                   10000 non-null int64   
 2   Gender                10000 non-null object  
 3   Blood Type            10000 non-null object  
 4   Medical Condition     10000 non-null object  
 5   Date of Admission     10000 non-null object  
 6   Doctor                10000 non-null object  
 7   Hospital              10000 non-null object  
 8   Insurance Provider    10000 non-null object  
 9   Billing Amount         10000 non-null float64  
10   Room Number           10000 non-null int64   
11   Admission Type        10000 non-null object  
12   Discharge Date        10000 non-null object  
13   Medication            10000 non-null object  
14   Test Results          10000 non-null object  
dtypes: float64(1), int64(2), object(12)
memory usage: 1.1+ MB
```

b.) df.isna().sum()

Melakukan proses df.isna().sum() yang merupakan bagian dari data reprocessing, untuk melihat apakah pada data terdapat nilai null dari setiap atributnya, Berdasarkan hasil yang didapat dapat dilihat tidak adanya nilai null dalam setiap atribut pada data ini..

```
Name          0
Age            0
Gender         0
Blood Type     0
Medical Condition 0
Date of Admission 0
Doctor         0
Hospital       0
Insurance Provider 0
Billing Amount 0
Room Number    0
Admission Type 0
Discharge Date 0
Medication     0
Test Results   0
dtype: int64
```

c.) df.describe()

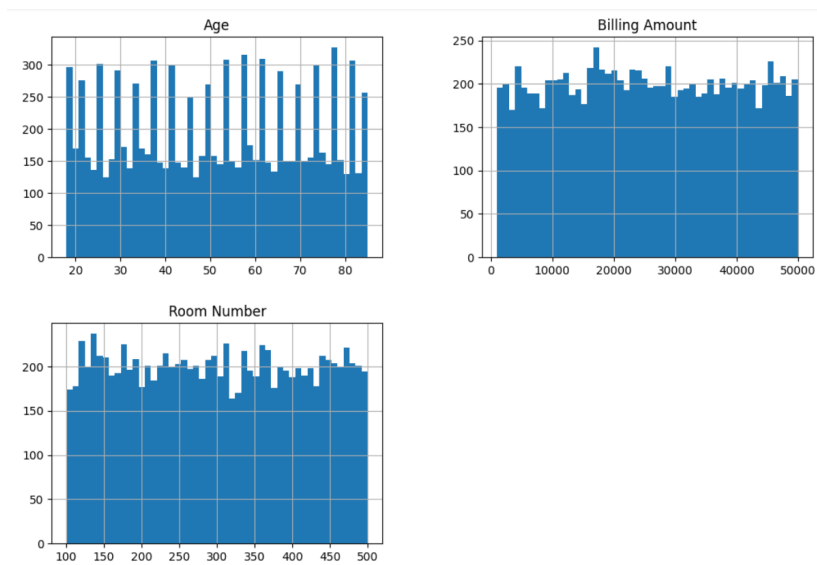
Melakukan fungsi df.describe(), fungsi ini akan menunjukkan deskripsi statistik data dari kolom bertipe numerik, dapat dilihat ada 3 kolom yang bertipe numerik, yang setiap

kolomnya menampilkan nilai count, mean, std, min, nilai 25%(Q1), 50%(Q2), 75%(Q3) dan nilai max.

	Age	Billing Amount	Room Number
count	10000.000000	10000.000000	10000.000000
mean	51.452200	25516.806778	300.082000
std	19.588974	14067.292709	115.806027
min	18.000000	1000.180837	101.000000
25%	35.000000	13506.523967	199.000000
50%	52.000000	25258.112566	299.000000
75%	68.000000	37733.913727	400.000000
max	85.000000	49995.902283	500.000000

d.) Histogram

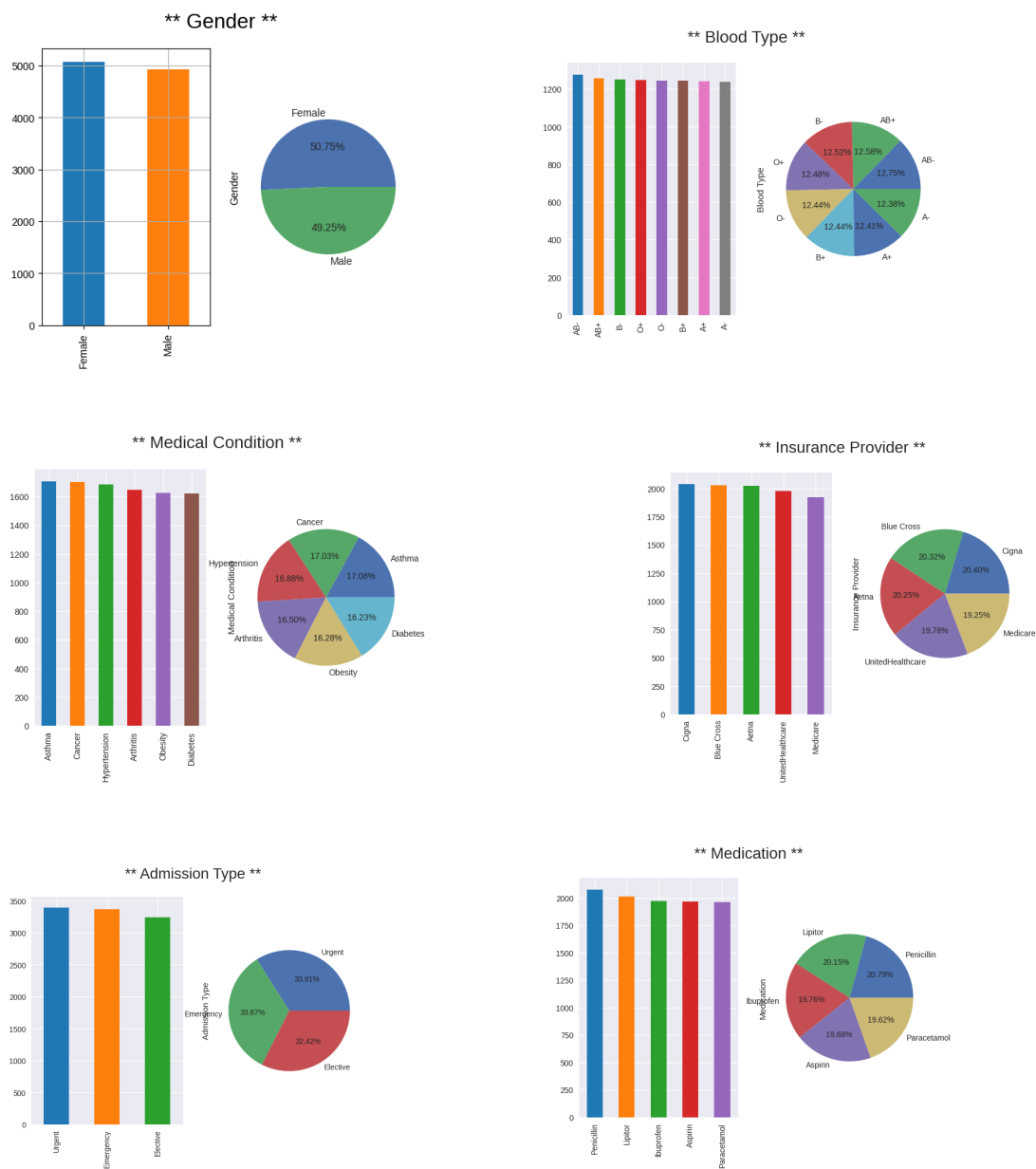
Histogram digunakan untuk menampilkan visualisasi dari setiap atribut numerik. Dari hasil visualisasi diatas, dapat diketahui bahwa kolom "Age" memiliki interval yang cukup tinggi sedangkan "Billing Amount" dan "Room Number" memiliki Interval yang cukup rendah.



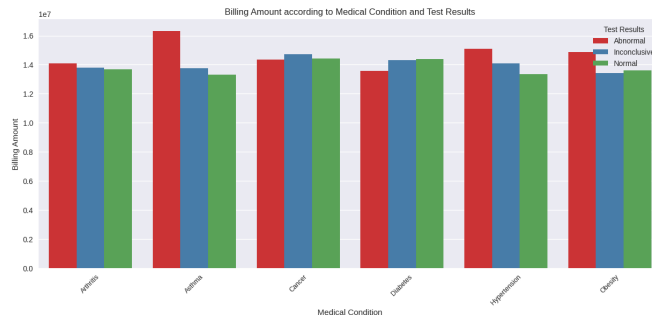
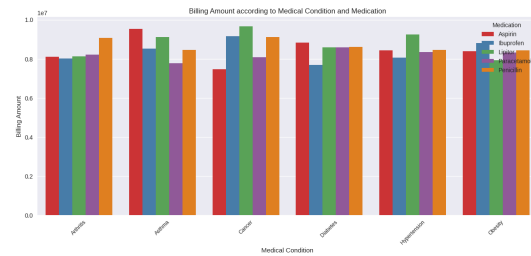
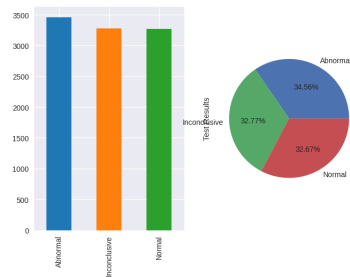
e.) Barchart dan Piechart

Barchart dan Piechart digunakan untuk menampilkan visualisasi dari setiap atribut kategori, Berdasarkan hasil yang didapat adanya perbedaan setiap kategori pada


masing-masing kolom/attribut. Barchart berwarna menunjukkan jumlah dari setiap kategori pada kelas sedangkan piechart menampilkan persentase dari setiap kategori.



**** Test Results ****



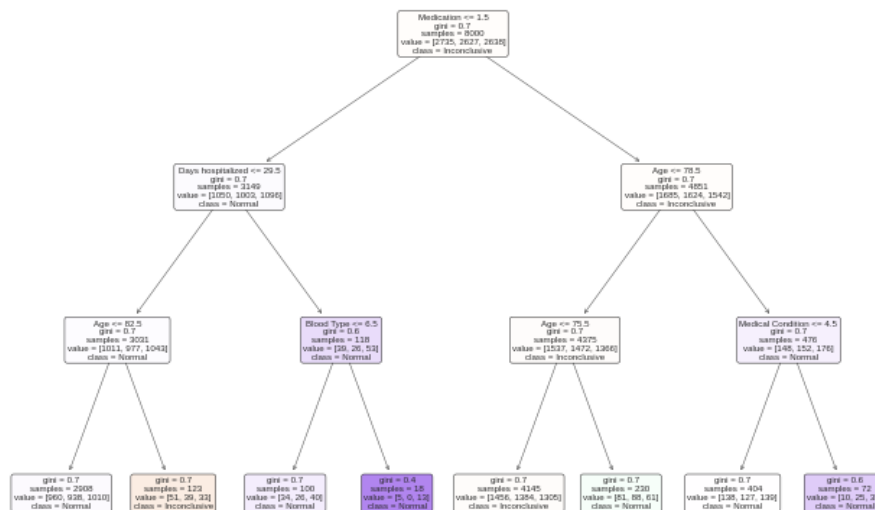
3.1 Decision Tree

 Score Of Model = 0.3415
Accuracy = 0.3415

Dari hasil metode Decision Tree yang ditampilkan pada kasus ini, dapat kita lihat bahwa nilai score of model atau nilai akurasi dari model Decision Tree sebesar 0.3415 atau 34,15 % yang artinya model ini memberikan prediksi yang kurang baik dengan kinerja rendah dalam memprediksi label pada dataset namun nilai akurasi memberikan indikasi tingkat prediksi yang benar. Dalam hal ini kita dapat mempertimbangan model model yang lain dalam hal keakurasian.

Visualisasi Decision Tree

Plotted the decision tree



Dari visualisasi pohon keputusan tersebut dapat dilihat pohon keputusan tersebut memiliki kedalaman maksimum sebesar 3, kita dapat melihat bagaimana model membuat keputusan berdasarkan fitur-fitur yang ada. Dengan membatasi kedalaman pohon keputusan, kita dapat dengan lebih mudah melihat bagaimana model membuat pemisahan antara kelas-kelas yang berbeda.

3.2 Random Forest

0.315

Skor akurasi yang didapat adalah 0.315

Dari hasil metode Random Forest yang ditampilkan pada kasus ini, dapat kita lihat bahwa akurasi dari model Random Forest sebesar 0.315 atau 31,5 % yang artinya model ini memiliki kinerja rendah dalam memprediksi label pada dataset. Model ini menggunakan 100 pohon keputusan ($n_estimators=100$).

Metode SVC

0.3385

Pada metode SVC, nilai akurasinya 0.3385

Dari hasil metode SVC (Support Vector Classification) yang ditampilkan pada kasus ini, dapat kita lihat bahwa akurasi dari model SVC sebesar 0.3385 atau 33,85 % yang artinya model ini memiliki kinerja rendah dalam mengklasifikasi label pada dataset.

Metode Bernoulli

0.3545

Pada metode Bernoulli, nilai akurasinya 0.3545

Dari hasil metode SVC (Support Vector Classification) yang ditampilkan pada kasus ini, dapat kita lihat bahwa akurasi dari model SVC sebesar 0.3385 atau 33,85 % yang artinya model ini memiliki kinerja rendah dalam memprediksi label pada dataset. Pada model ini menggunakan kelas BernoulliNB dari Naive Bayes untuk memprediksi probabilitas setiap kelas dimana data tersebut diubah menjadi data biner (0 dan 1) untuk mempermudah prediksi.

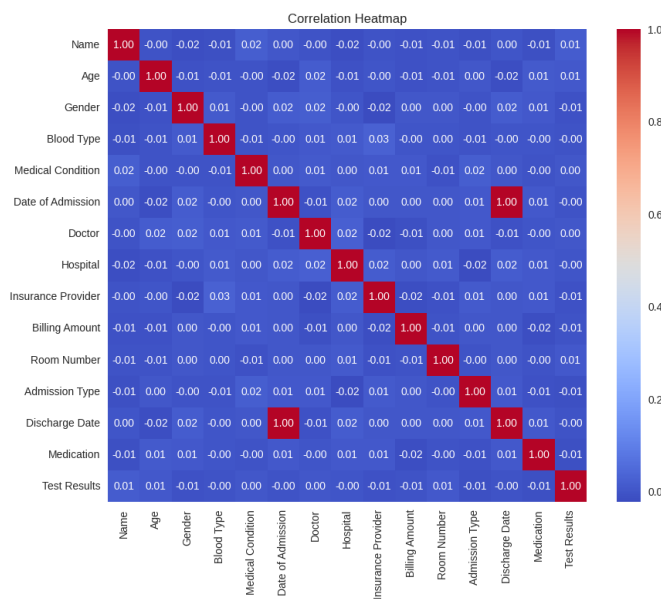
Metode Regresi Logistik

0.3355

Pada metode Regresi Logistik, nilai akurasinya 0.3355

Dari hasil metode Regresi Logistik yang ditampilkan pada kasus ini, dapat kita lihat bahwa akurasi dari modelnya sebesar 0.3355 atau 33,55 % yang artinya model ini memiliki kinerja rendah dalam memprediksi label pada dataset.

Correlation Heatmap



Visualisasi diatas merupakan Correlation Heatmap, dimana kita dapat melihat tampilan tingkat korelasi antara setiap atributnya. Warna merah memiliki arti bahwa korelasinya bernilai 1 yang menunjukkan hubungan linear positive sempurna. Dari gambar diatas dapat kita lihat bahwa kolom Discharge Date dan kolom Date of Admission memiliki hubungan linear positif sempurna yang artinya kedua kolom tersebut saling berkorelasi.

Melihat Akurasi dari Model XGBoost

```

Accuracy: 0.35
      precision    recall  f1-score   support

     0       0.36      0.38      0.37        706
     1       0.34      0.35      0.34        636
     2       0.34      0.31      0.32        658

 accuracy          0.35        2000
 macro avg       0.34      0.34      0.34        2000
 weighted avg    0.35      0.35      0.34        2000

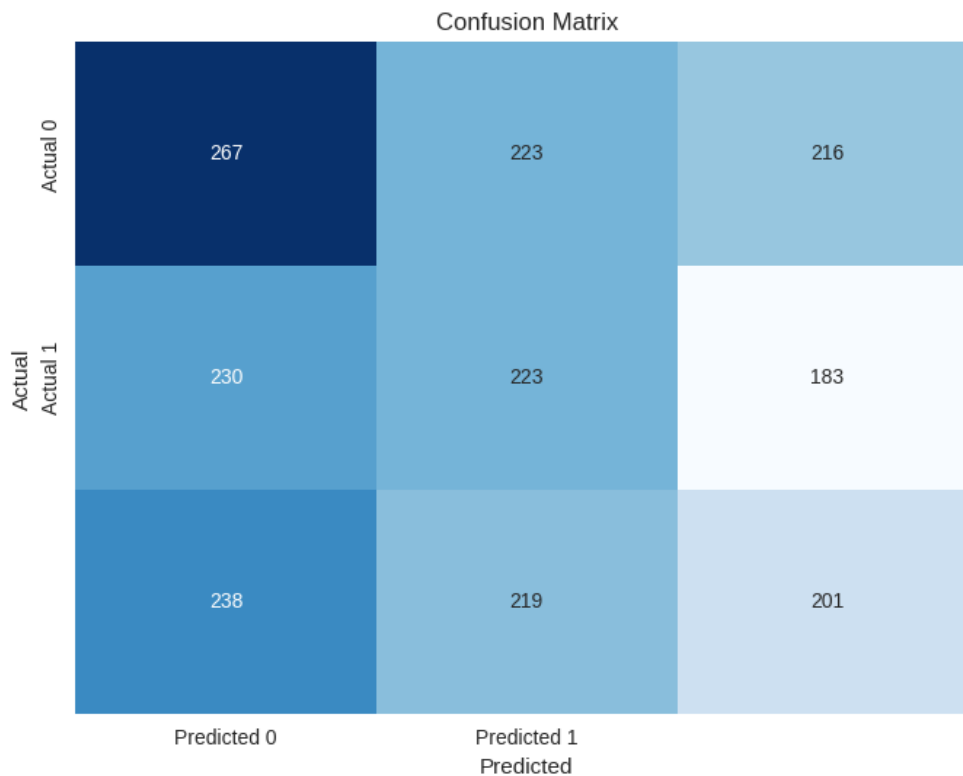
[[267 223 216]
 [230 223 183]
 [238 219 201]]

```

Dari hasil yang ditampilkan diatas, dapat kita lihat bahwa akurasi yang didapatkan sebesar 0.35 atau 35 % yang artinya model XGBoost ini memiliki kinerja sedang dalam memprediksi data.

Dapat kita ketahui juga terdapat output yang berisikan Classification Report yang menjelaskan seberapa baik model XGBoost, dapat kita lihat pada kelas 0 lebih baik dibandingkan kelas 1 dan 2 dimana jumlah instance (Support) lebih tinggi dibandingkan kelas 1 dan 2. Walaupun begitu baik kelas 0,1 dan 2 pada uji model XGBoost ini memiliki kinerja yang seragam pada setiap kelasnya, dan hasilnya tampak cukup seimbang.

Confusion Matrix



Visualisasi diatas merupakan Confusion Matrix, dimana kita dapat melihat tampilan matriks kebingungan dengan membandingkan jumlah instans yang diklasifikasikan dengan benar dan yang salah untuk setiap kelas. Dalam gambar tersebut menjelaskan bahwa kolom True Negative for class 0 memiliki nilai yang besar menunjukkan bahwa model dengan baik memprediksi dan mempertahankan klasifikasi yang benar untuk instans yang sebenarnya bukan kelas 0. Sebaliknya dapat dilihat juga bahwa False Negative (FN) untuk Kelas 2 memiliki nilai yang rendah dalam confusion matrix, itu menunjukkan bahwa model cenderung baik dalam mengidentifikasi dan memprediksi instans yang sebenarnya adalah Kelas 2. Dalam konteks ini, model secara relatif baik dalam menghindari kesalahan "False Negative" untuk Kelas 2.

3.3 Metode Neural Network



Neural Network Accuracy: 1.0

Dari hasil metode Neural Network yang ditampilkan diatas, sebelum mendapatkan hasil yang diminta kita mendeklarasikan nilai fitur yaitu Age, Billing

Amount dan Room Number untuk kita tinjau terhadap nilai target yaitu Medical Condition. Dalam kasus ini, dapat kita lihat bahwa akurasi dari model Neural Network sebesar 1 atau 100 % yang artinya model ini sempurna dalam memprediksi semua sampel pada data.

3.3 Metode SVM



SVM Accuracy: 0.168

Classification Report:

	precision	recall	f1-score	support
Arthritis	0.00	0.00	0.00	319
Asthma	0.17	0.53	0.25	332
Cancer	0.00	0.00	0.00	363
Diabetes	0.00	0.00	0.00	303
Hypertension	0.17	0.47	0.25	336
Obesity	0.00	0.00	0.00	347
accuracy			0.17	2000
macro avg	0.06	0.17	0.08	2000
weighted avg	0.06	0.17	0.08	2000

Dari hasil metode SVM yang ditampilkan diatas, sebelum mendapatkan hasil yang diminta kita mendeklarasikan nilai fitur yaitu Age, Billing Amount dan Room Number untuk kita tinjau terhadap nilai target yaitu Medical Condition. Dalam kasus ini, dapat kita lihat bahwa akurasi dari model SVM sebesar 0.168 atau 16,8 % yang artinya model ini memiliki kinerja rendah dalam mengklasifikasikan data.

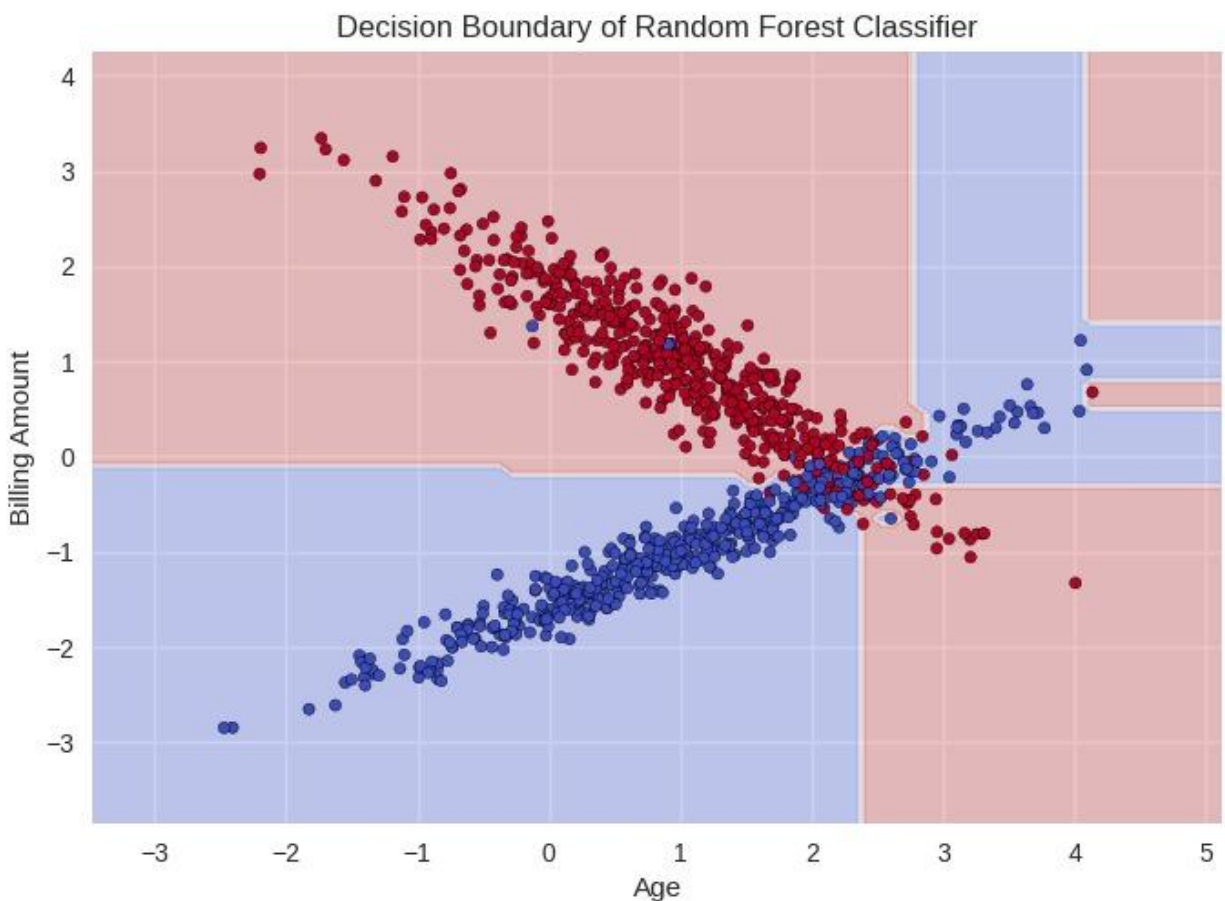
Dapat kita ketahui juga bahwa output tersebut menjelaskan classification report dimana classification report ini bertujuan untuk menjelaskan model dalam kelas pada dataset apakah model tersebut efektif atau tidak terhadap kelas yang dituju. Berikut analisis tiap kelas yang ditampilkan pada output:

- Arthritis: Precision dan Recall keduanya rendah, menunjukkan bahwa model tidak efektif dalam mengidentifikasi kasus arthritis.
- Asthma: Precision dan Recall cukup rendah, namun Recall relatif lebih tinggi, menunjukkan bahwa model mungkin lebih baik dalam menangkap kasus asthma, tetapi masih tidak memadai.
- Cancer: Precision dan Recall keduanya rendah, menunjukkan bahwa model tidak efektif dalam mengidentifikasi kasus kanker.
- Diabetes: Precision dan Recall keduanya rendah, menunjukkan bahwa model tidak efektif dalam mengidentifikasi kasus diabetes.

- Hypertension: Precision dan Recall keduanya cukup rendah, menunjukkan bahwa model tidak efektif dalam mengidentifikasi kasus hipertensi.
- Obesity: Precision dan Recall keduanya rendah, menunjukkan bahwa model tidak efektif dalam mengidentifikasi kasus obesitas.

Dapat kita simpulkan bahwa di beberapa kelas model SVM ini tidak efektif dalam mengklasifikasikan kelas tersebut, dari analisis diatas kita melihat kelas seperti Arthritis, Cancer, Diabetes, Obesity tidak efektif terhadap model ini karena memiliki nilai precision dan recall sama dengan nol.

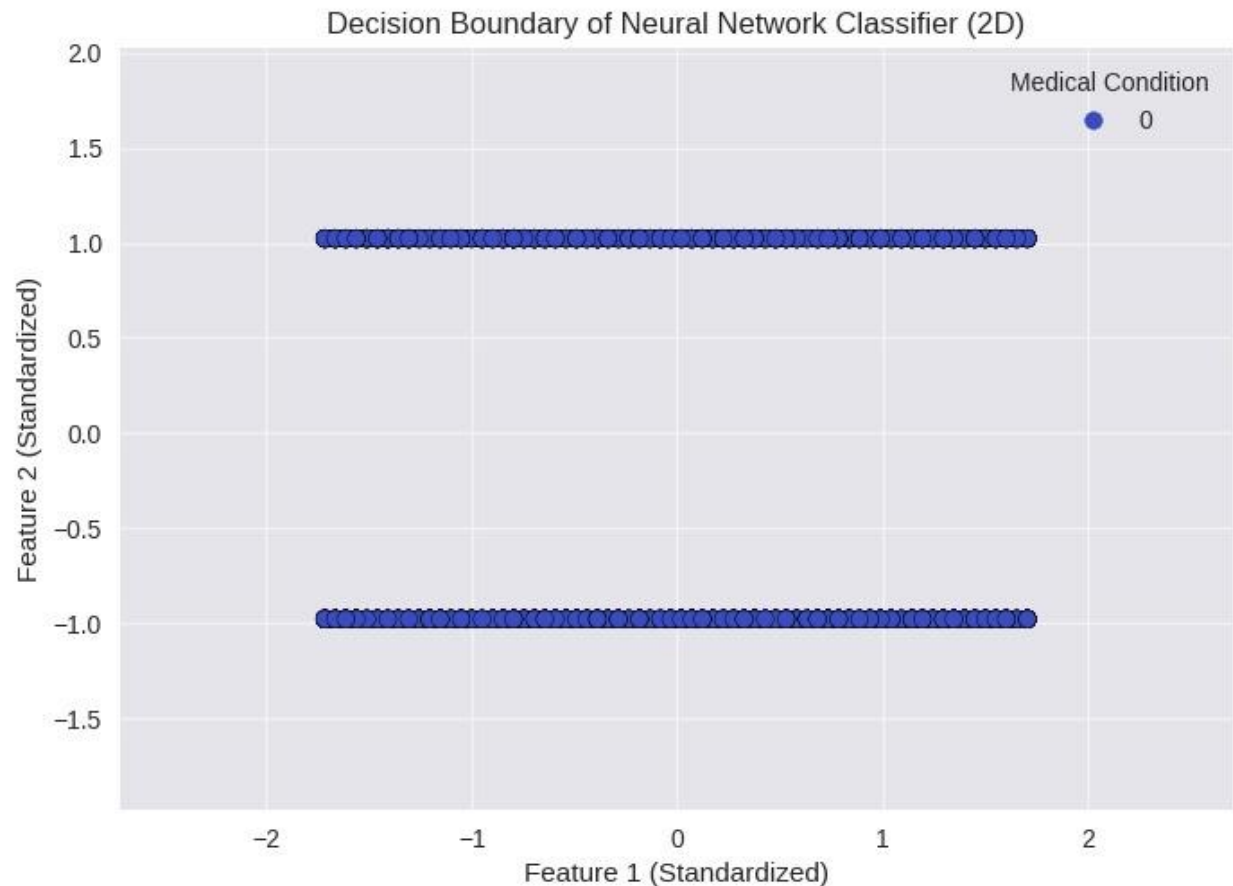
3.4 Decision Random forest



Klasifikasi menggunakan Random Forest melalui decision boundary (batas keputusan) memberikan pemahaman tentang kemampuan model dalam memisahkan kelas-kelas yang berbeda berdasarkan fitur-fitur yang ada. Dalam random forest, decision boundary dapat terbentuk melalui kombinasi dari beberapa pohon keputusan yang sangat varied. Setiap pohon dalam random forest menghasilkan decision boundary sendiri, dan saat digabungkan, mereka dapat membentuk batas yang lebih kompleks dan mudah beradaptasi dengan data yang rumit. Berdasarkan output diatas Sumbu X dan Y: Sumbu X diberi label "Age" dan sumbu Y diberi

label “Billing Amount”. Grafik Tersebut memiliki dua warna, merah dan biru, yang mewakili dua kelas yang berbeda. Batas keputusannya adalah garis diagonal yang memisahkan dua kelas tersebut. Distribusi Data tersebut berdasarkan kelas merah sebagian besar berada di sudut kiri atas grafik, sedangkan kelas biru sebagian besar berada di sudut kanan bawah.

3.4 Decision Neural Network



Hasil output dari model Neural Network dalam klasifikasi menggunakan Decision Neural Network (DNN) dapat memberikan wawasan tentang kinerja dan kemampuan model dalam memprediksi kelas target. Berdasarkan hasil diatas Sumbu X diberi label “Feature 1 (Standardized)” dan sumbu Y diberi label “Feature 2 (Standardized)”. Grafik ini memiliki dua kelas yang berbeda, yang mewakili dua kondisi medis yang berbeda. Batas keputusan adalah garis biru yang memisahkan dua kelas tersebut. Distribusi Data tersebut dapat dilihat pada garis biru diatas terdapat salah satu kelas sebagian besar berada di atas garis biru, sedangkan kelas lainnya sebagian besar berada di bawah garis biru.

4. Kesimpulan

Setelah melakukan berbagai metode analisis, akurasi hasil pengujian menunjukkan bahwa Decision Tree mencapai 0.3415, Random Forest sebesar 0.315, Neural Network memperoleh nilai akurasi maksimum 1.0, dan SVM dengan nilai akurasi 0.16. Oleh karena itu, Neural Network sebagai metode terbaik berdasarkan tingkat akurasi yang tinggi.

Selain itu, ketika melihat Decision Boundary, Neural Network juga menunjukkan hasil yang superior, memperlihatkan klasifikasi yang lebih baik di antara metode-metode lainnya. Terdapat kejelasan dalam pemisahan setiap kelas, kelas tidak saling bercampur. Hal tersebut menunjukkan ketepatan dan ketelitian yang tinggi dalam prediksi.

Sehingga dapat disimpulkan bahwa Neural Network adalah metode yang paling terbaik, baik dari segi akurasi maupun dalam pemahaman batas keputusan yang lebih baik dibandingkan metode-metode lain yang dievaluasi.

Referensi

Nanfack, Geraldin, Valentin Delchevalerie, and Benoît Frénay. *"Boundary-based fairness constraints in decision trees and random forests."* ESANN 2021 Proceedings-29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2021.

Karimi, Hamid, and Jiliang Tang. "Decision boundary of deep neural networks: Challenges and opportunities." *Proceedings of the 13th International Conference on Web Search and Data Mining*. 2020.

Lampiran

Berikut kami lampirkan kode yang telah dibuat pada percobaan kali ini :

 PM3_KELOMPOK.ipynb

link video presentasi:

[Presentasi Kelompok 2 Modul 3](#)