

OPTIMASI MODEL KLASIFIKASI KANKER PAYUDARA MELALUI PENDEKATAN DECISION TREE, ARTIFICIAL NEURAL NETWORK, DAN SUPPORT VECTOR MACHINE

OPTIMIZATION OF BREAST CANCER CLASSIFICATION MODEL THROUGH DECISION TREE, ARTIFICIAL NEURAL NETWORK, AND SUPPORT VECTOR MACHINE APPROACHES

Fahrurrozi¹, Muhammad Aminullah², Endi Febriyanto³, Handika Adityo⁴, M. Farid Setiawan⁵
^{1,2,3,4,5}IIB Darmajaya Lampung

Email : fahrurrozyghalib@gmail.com

Abstract: Optimization Of Breast Cancer Classification Model Through Decision Tree, Artificial Neural Network, And Support Vector Machine Approaches. This research aims to optimize the breast cancer classification model using the Decision Tree, Artificial Neural Network and Support Vector Machine approaches. Breast cancer is a dangerous disease that seriously threatens women's health. Data from this research shows that the classification model using Artificial Neural Network and Support Vector Machine produces high accuracy, with better results after hyperparameter tuning. Experimental results show that the Support Vector Machine model has the best performance, followed by Artificial Neural Network and Decision Tree. The Hyperparameter-tuned Decision Tree algorithm shows improved performance compared to the regular Decision Tree algorithm. It is hoped that the results of this research can make a significant contribution to the development of more sophisticated and accurate breast cancer diagnosis tools.

Keywords: Classification, Breast Cancer, Decision Tree, ANN, SVM, Hyperparameter Tuning

Abstrak: Optimasi Model Klasifikasi Kanker Payudara Melalui Pendekatan Decision Tree, Artificial Neural Network, Dan Support Vector Machine.

Penelitian ini bertujuan untuk mengoptimalkan model klasifikasi kanker payudara menggunakan pendekatan Decision Tree, Artificial Neural Network, dan Support Vector Machine. Kanker payudara merupakan salah satu penyakit berbahaya yang mengancam kesehatan wanita secara serius. Data dari penelitian ini menunjukkan bahwa model klasifikasi menggunakan Artificial Neural Network dan Support Vector Machine menghasilkan akurasi yang tinggi, dengan hasil yang lebih baik setelah dilakukan hyperparameter tuning. Hasil eksperimen menunjukkan bahwa model Support Vector Machine memiliki kinerja terbaik, diikuti oleh Artificial Neural Network dan Decision Tree. Algoritma Decision Tree yang telah dituning hyperparameter menunjukkan peningkatan kinerja dibandingkan dengan algoritma Decision Tree biasa. Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan alat bantu diagnosis kanker payudara yang lebih canggih dan akurat.

Kata Kunci : Klasifikasi, Kanker Payudara, Decision Tree, ANN, SVM, Hyperparameter Tuning

PENDAHULUAN

Kanker menjadi salah satu jenis penyakit berbahaya penyebab terjadinya kematian pada manusia di seluruh dunia. [1]Kanker payudara merupakan tumor dengan insiden tertinggi yang mengancam kesehatan wanita secara serius.[2][3] Kanker payudara menempati urutan pertama terkait jumlah kanker terbanyak di Indonesia, serta menjadi salah satu penyumbang kematian pertama akibat kanker. Data Globocan tahun 2020 jumlah kasus kanker dunia mencapai total 19.292.789 kasus, untuk kasus kanker payudara berjumlah 2.261.419 (11.7%). Untuk kasus kematian yang diakibatkan oleh

penyakit kanker total 9.958.133, sedangkan kasus kematian yang diakibatkan oleh kanker payudara berjumlah 684.996 menyumbang (6.9%) dari kasus kematian yang diakibatkan oleh penyakit kanker.

Data Globocan tahun 2020, jumlah kasus baru kanker payudara mencapai 65.858 kasus (16,6%) dari total 396.914 kasus baru penyakit kanker di Indonesia. Sementara itu, untuk jumlah kematiannya mencapai lebih dari 22 ribu jiwa kasus. Kanker payudara salah satu jenis kanker yang paling umum terjadi pada perempuan di seluruh dunia.

Kanker payudara ini secara umum dibagi menjadi 2, yaitu benign atau biasa disebut jinak dan malignant atau biasa disebut juga ganas, biasanya kanker payudara jinak ditandai dengan berbentuk benjolan kecil bulat, dan lembut.[4] Kanker payudara dalam tingkat jinak biasanya akan mempunyai keadaan dan pertumbuhan yang tidak bersifat kanker. Kanker ini bisa terdeteksi tetapi tidak akan menjalar dan merusak jaringan di dekatnya.[5][6] Pada kanker payudara dalam tingkat ganas ditandai dengan bentuk yang tidak simetris, kasar, terasa nyeri, dan lainnya. Biasanya kanker payudara menjalar dan merusak jaringan dan organ lain yang ada di dekatnya.

Berdasarkan data Riskesdas, prevalensi tumor/kanker di Indonesia menunjukkan adanya peningkatan dari 1.4 per 1000 penduduk di tahun 2013 menjadi 1.79 per 1000 penduduk pada tahun 2018. Sedangkan angka kejadian untuk perempuan yang tertinggi adalah kanker payudara yaitu sebesar 42,1 per 100.000 penduduk dengan rata-rata kematian 17 per 100.000 penduduk. Kanker payudara merupakan salah satu jenis kanker yang paling sering terjadi pada wanita. Menurut data WHO, pada tahun 2020 terdapat sekitar 2,3 juta kasus baru kanker payudara di seluruh dunia. Meskipun demikian, angka kesembuhan kanker payudara juga cukup tinggi jika dideteksi pada stadium awal dan diberikan penanganan yang tepat. Oleh karena itu, deteksi dini sangat penting dalam upaya pencegahan dan pengobatan kanker payudara.

Dalam era modern ini, penggunaan teknologi dan kecerdasan buatan telah membuka peluang baru untuk meningkatkan keakuratan dan efisiensi deteksi kanker payudara. Model klasifikasi seperti Decision Tree, Artificial Neural Network (ANN), dan Support Vector Machine (SVM) menjadi solusi yang menjanjikan untuk menganalisis dan mengidentifikasi kanker payudara berdasarkan data klinis dan citra medis.[7][8][9]

Decision Tree dapat memberikan interpretasi yang mudah dimengerti tentang faktor-faktor yang mempengaruhi keputusan klasifikasi, sementara Artificial Neural Network mampu menangkap pola kompleks dalam data.[10][11] Di sisi lain, Support Vector Machine dikenal dengan keunggulannya dalam menangani pemisahan kelas yang kompleks.

Meskipun ketiga pendekatan ini menawarkan potensi yang besar, namun untuk mewujudkan potensi tersebut, perlu dilakukan optimasi pada parameter-parameter kunci, serta pemilihan fitur dan pra-proses data yang efektif.[12] Pengoptimalan ini bertujuan untuk meningkatkan akurasi, presisi, recall, dan nilai F1 Score dari model klasifikasi tersebut.

Penelitian ini akan fokus pada pengembangan dan optimasi model klasifikasi kanker payudara menggunakan Decision Tree, Artificial Neural Network, dan Support Vector Machine. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan alat bantu diagnosis kanker payudara yang lebih canggih dan akurat.

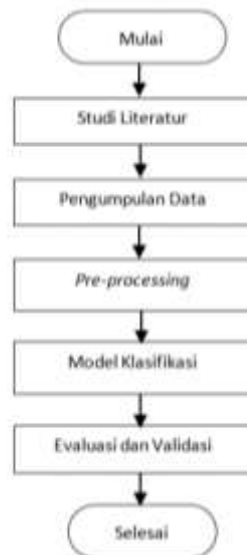
METODE

Metode penelitian yang paling umum digunakan ada empat yaitu penelitian tindakan, eksperimen, studi kasus, dan survei.[13] Pada penelitian ini dilakukan penelitian eksperimen dengan

pendekatan kuantitatif. Eksperimen yaitu jenis penelitian yang melakukan penyelidikan hubungan ketergantungan dengan menggunakan tes dimana yang mengendalikan adalah peneliti itu sendiri.

1. Tahapan Penelitian

Penelitian ini menggunakan pendekatan eksperimental dengan fokus pada pengembangan dan optimasi model klasifikasi kanker payudara. Metode penelitian terbagi menjadi beberapa tahapan, yaitu :



Gambar 1. Alur Penelitian

2. Studi Literatur

Pada tahap ini peneliti melakukan review jurnal yang berkaitan dengan judul penelitian yang akan diteliti oleh peneliti. Jurnal-jurnal yang menjadi rujukan tentu memiliki ruang lingkup dan metode yang sama dilakukan oleh peneliti.

3. Pengumpulan Data

Data yang digunakan merupakan dataset publik berupa data yang diperoleh dari Kaggle dengan link <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>. yang terdiri dari informasi demografis, kebiasaan, dan catatan medis historis. Beberapa pasien memutuskan untuk tidak menjawab beberapa pertanyaan karena masalah privasi (missing value).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                    569 non-null    int64
1   diagnosis                            569 non-null    object
2   radius_mean                          569 non-null    float64
3   texture_mean                         569 non-null    float64
4   perimeter_mean                       569 non-null    float64
5   area_mean                           569 non-null    float64
6   smoothness_mean                      569 non-null    float64
7   compactness_mean                     569 non-null    float64
8   concavity_mean                       569 non-null    float64
9   concave points_mean                  569 non-null    float64
10  symmetry_mean                        569 non-null    float64
11  fractal_dimension_mean               569 non-null    float64
12  radius_se                            569 non-null    float64
13  texture_se                           569 non-null    float64
14  perimeter_se                         569 non-null    float64
15  area_se                              569 non-null    float64
16  smoothness_se                        569 non-null    float64
17  compactness_se                       569 non-null    float64
18  concavity_se                         569 non-null    float64
19  concave points_se                    569 non-null    float64
20  symmetry_se                          569 non-null    float64
21  fractal_dimension_se                 569 non-null    float64
22  radius_worst                         569 non-null    float64
23  texture_worst                        569 non-null    float64
24  perimeter_worst                      569 non-null    float64
25  area_worst                           569 non-null    float64
26  smoothness_worst                     569 non-null    float64
27  compactness_worst                    569 non-null    float64
28  concavity_worst                      569 non-null    float64
29  concave points_worst                 569 non-null    float64
30  symmetry_worst                       569 non-null    float64
31  fractal_dimension_worst              569 non-null    float64
dtypes: float64(30), int64(1), object(1)
memory usage: 142.4+ KB
```

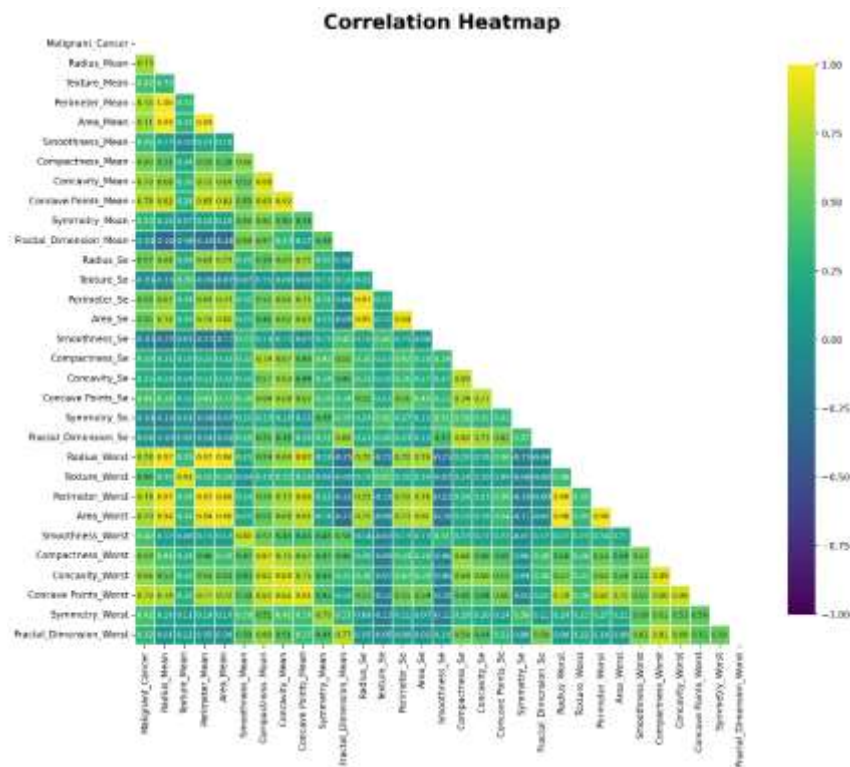
Gambar 2. Data Frame

Pada gambar 2 dapat dilihat bahwa dataset memiliki 569 data dan 32 atribut yang berupa kolom.

4. Pre-processing dan Pengolahan Data (Cleaning Data).

Preprocessing dan pengolahan data dilakukan pada Google Colab dengan melakukan normalisasi, penghapusan fitur yang tidak berguna, dan treatment data yang hilang. Data berjumlah 569 record data dengan 32 atribut. Pada tahapan selanjutnya peneliti melakukan perubahan nilai dalam kolom "diagnosis" dengan menggunakan fungsi map untuk mengganti menjadi "M" dengan 1 dan "B" dengan 0. Setelah itu peneliti mengubah nama kolom "diagnosis" menjadi "malignant_cancer". Hal tersebut dilakukan untuk memperjelas keberadaan kanker ganas.

Selanjutnya peneliti berfokus pada perhitungan korelasi antar kolom dalam DataFrame yang berisi kolom "malignant_cancer" Bersama dengan nilai korelasinya terhadap kolom lain. Kemudian data tersebut diurutkan berdasarkan nilai korelasi terhadap kolom "malignant_cancer" secara menurun. Seperti dapat dilihat pada gambar 3 dibawah ini.



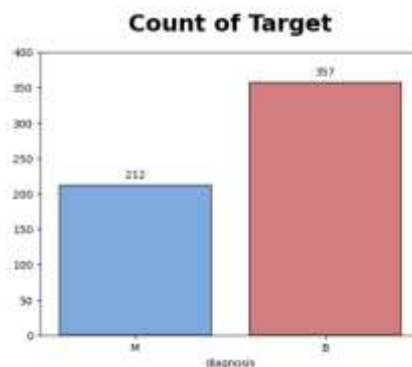
Gambar 3. Corelation Heatmap



Gambar 4. Visualisasi Corelation Heatmap

Pada Gambar 4 menunjukan heatmap korelasi yang memvisualisasikan seberapa erat hubungan antara fitur-fitur dalam DataFrame data_cancer dengan status kanker ganas. Titik fokusnya adalah warna pada heatmap, di mana nilai korelasi tinggi ditunjukkan oleh warna yang lebih terang,

sementara nilai korelasi rendah ditunjukkan oleh warna yang lebih gelap. Juga, nilai korelasi spesifik ditampilkan di setiap sel heatmap untuk memberikan informasi lebih lanjut.

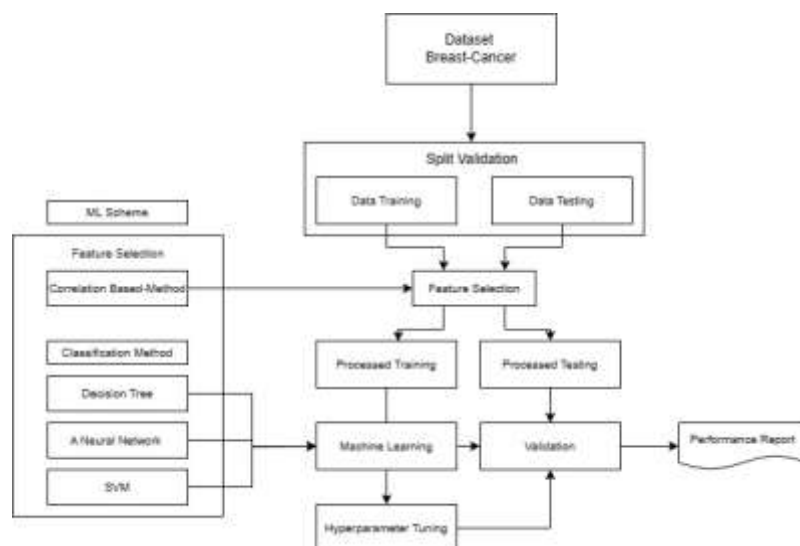


Gambar 5. Count Of Target

Pada Gambar 5 menunjukkan Jumlah sampel untuk setiap kategori dapat dilihat secara visual, dan perbandingannya terlihat jelas. target diagnosis antara kedua kategori yaitu kategori M dengan jumlah 212 data dan kategori B 357 data.

5. Model Klasifikasi

Model yang akan digunakan pada penelitian ini adalah metode klasifikasi Decision Tree, Artificial Neural Network dan Support Vector Machine dengan menggunakan Correlation Based-Method sebagai seleksi fitur. Model klasifikasi yang diusulkan dapat dilihat pada Gambar 6.



Gambar 6. Model Klasifikasi

Pada Gambar 6 menunjukkan skema machine learning yang akan di terapkan yaitu dataset Breast-Cancer dibagi menjadi dua yaitu data training dan data testing dengan menggunakan seleksi fitur Correlation Based-Method, selanjutnya melakukan pengujian model kinerja ketiga Algoritma tersebut mencari nilai Akurasi, Presisi, Recall dan F1 Score. Setelah mendapatkan hasil dari kinerja ketiga algoritma tersebut peneliti melakukan hyperparamter tuning yang bertujuan untuk meningkatkan permorma model dari kegita algoritma tersebut.

6. Evaluasi dan Validasi

Langkah terakhir adalah mengevaluasi dan memvalidasi hasil dari pelatihan. Data yang diperoleh dari proses ini akan menunjukkan nilai akurasi dan kehilangan, yang dapat digunakan untuk menentukan apakah teknik yang kami gunakan untuk penelitian ini berhasil atau tidak.

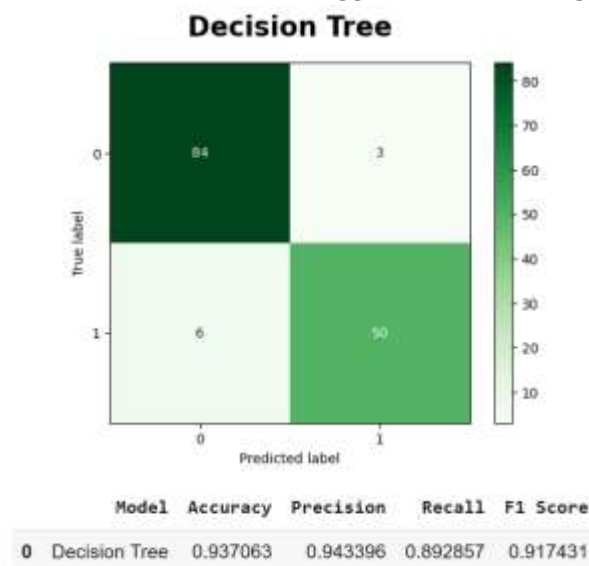
HASIL

Penerapan metode Decision Tree, Artificial Neural Network dan Support Vector Machine dengan menggunakan seleksi fitur Correlation Based-Method pada dataset Breast Cancer bertujuan untuk mengetahui dan meningkatkan nilai akurasi dari penelitian-penelitian sebelumnya, sehingga akan dihasilkan set atribut terbaik untuk klasifikasi pasien kanker payudara. Meningkat atau tidaknya nilai akurasi dapat dilihat pada hasil eksperimen. Eksperimen ini dilakukan dengan dua tahap, pertama adalah eksperimen menggunakan metode Decision Tree, Artificial Neural Network dan Support Vector Machine dan kedua adalah eksperimen menggunakan metode Decision Tree, Artificial Neural Network dan Support Vector Machine dengan menggunakan hypeparameter tuning.

Hasil Eksperimen

Metode Decision Tree

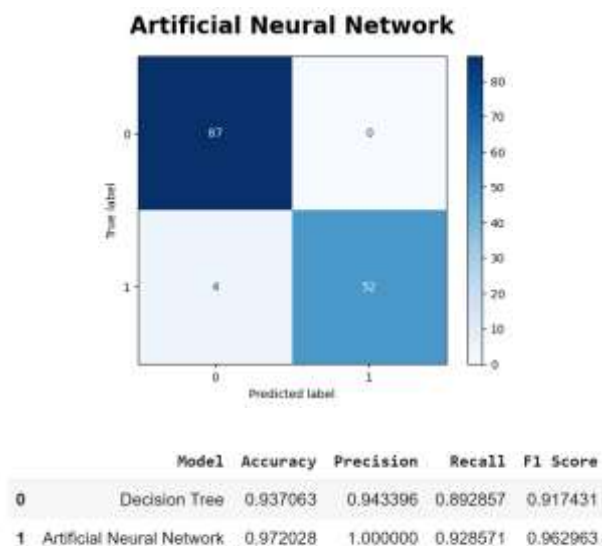
Eksperimen menggunakan metode Decision Tree menggunakan tools Google Colab.



Gambar 7. Hasil Eksperimen Decision Tree

Pada Gambar 7 Menunjukan Hasil eksperimen pada Algoritma Decision Tree menghasilkan hasil akurasi 93%, Presisi 94% , Recall 89% dan F1 Score 91%.

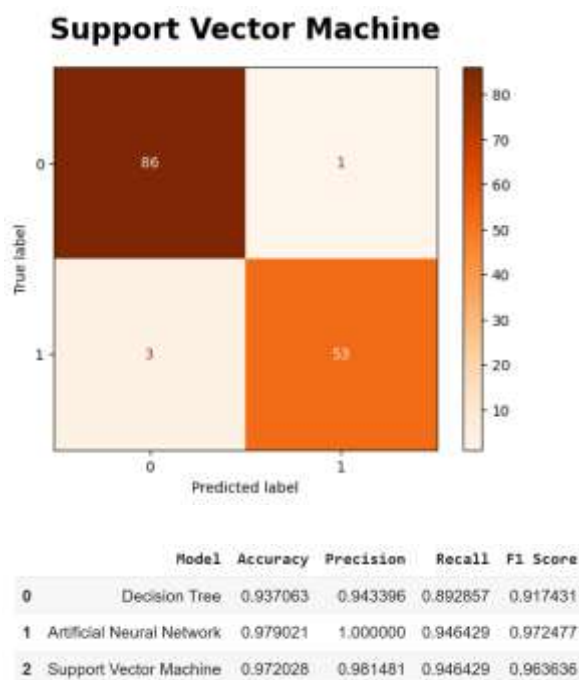
Metode Artificial Neural Network



Gambar 8. Hasil Eksperimen Artificial Neural Network

Pada Gambar 8 menunjukan hasil dari eksperimen pada Algoritma Artificial Neural Network menghasilkan hasil akurasi 97%, Presisi 100% , Recall 92% dan F1 Score 96%.

Metode Support Vector Machine



Gambar 9. Hasil Eksperimen Support Vector Machine

Pada Gambar 9 menunjukan hasil dari eksperimen pada Algoritma Support Vector Machine menghasilkan hasil akurasi 97%, Presisi 98% , Recall 94% dan F1 Score 96%.

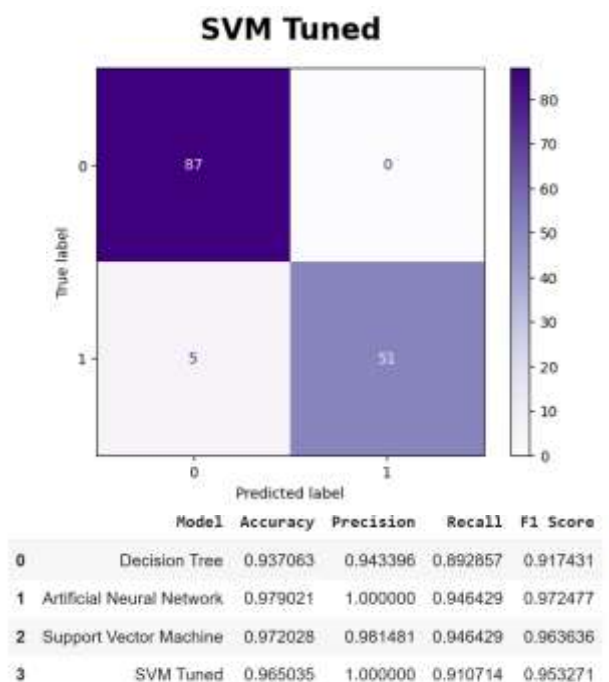
Setelah melakukan eksperimen pada ketiga model algoritma tersebut selanjutnya melakukan pengujian eksperimen tiga model algoritma dengan menggunakan hyperparameter tuning. Dengan melakukan hyperparameter tuning, kita berusaha untuk membuat model yang lebih baik sesuai dengan karakteristik dan kebutuhan khusus dari tugas atau data Breast-Cancer. Proses ini melibatkan eksplorasi berbagai kombinasi hyperparameter, pengujian model, dan pemilihan parameter terbaik yang menghasilkan model yang optimal

PEMBAHASAN

Hasil Ekperimen Menggunakan Hyperparameter Tuning

Support Vector Machine Hyperparameter Tuning

Dengan menggunakan parameter grid untuk SVM, kita dapat secara sistematis mengeksplorasi berbagai kombinasi nilai parameter untuk mencari yang paling optimal dalam hal performa model SVM.

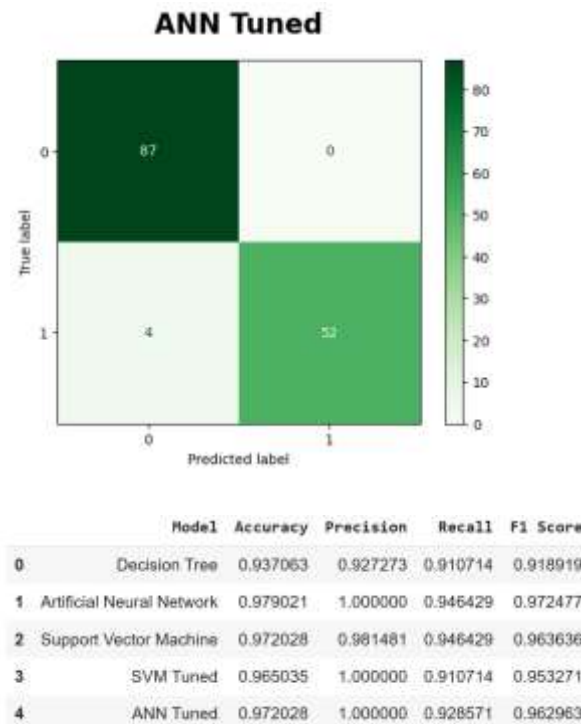


Gambar 10. Hasil Eksperimen SVM Tuned

Pada Gambar 10 menunjukan hasil dari eksperimen pada Algoritma Support Vector Machine yang menggunakan hyperparameter tuning grid menghasilkan hasil akurasi 96%, Presisi 100% , Recall 91% dan F1 Score 95%.

Artificial Neural Netrwok Hyperparameter Tuning

Dengan menggunakan parameter grid untuk ANN ini, kita dapat melakukan eksplorasi sistematis terhadap berbagai kombinasi nilai parameter untuk mencari yang paling optimal dalam hal performa model Artificial Neural Network.

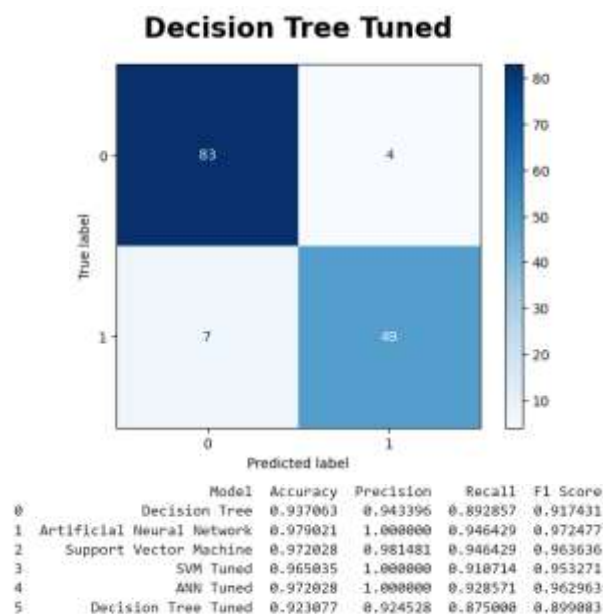


Gambar 11. Hasil Eksperimen ANN Tuned

Pada Gambar 11 menunjukkan hasil dari eksperimen pada Algoritma Artificial Neural Network yang menggunakan hyperparameter tuning grid menghasilkan hasil akurasi 97%, Presisi 100% , Recall 92% dan F1 Score 96%.

Decision Tree Hyperparameter Tuning

Dengan menggunakan parameter grid untuk Decision Tree ini, kita dapat melakukan eksplorasi sistematis terhadap berbagai kombinasi nilai parameter untuk mencari yang paling optimal dalam hal performa model Decision Tree.



Gambar 12. Hasil Eksperimen Decision Tree Tuned

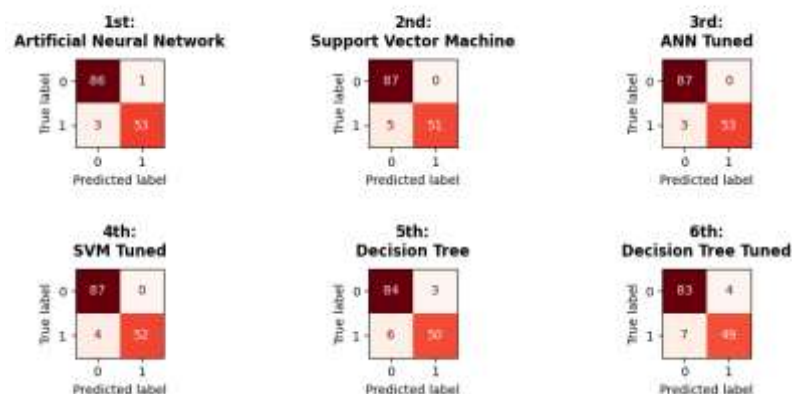
Pada Gambar 12 menunjukkan hasil dari eksperimen pada Algoritma Decision Tree yang menggunakan hyperparameter tuning grid menghasilkan hasil akurasi 92%, Presisi 92% , Recall 87% dan F1 Score 89%.



Gambar 13. Model Comparison

Pada Gambar 13 menunjukkan model komparasi dari kinerja dari setiap algoritma namun hasil menunjukkan Support Vector Machine dengan Akurasi Tertinggi dengan 97% Presisi 98% Recall 95% F1 Score 96% , Artificial Neural Network Akurasi 97%, Presisi 100%, Recall 93%, F1 Score 96%, ANN Tuned Akurasi 97%, Presisi 100%, Recall 93%, F1 Score 96%, SVM Tuned Akurasi 97%, Presisi 100%, Recall 91%, F1 Score 95%, Decision Tree 94%, Presisi 94%, Recall 89%, F1 Score 92%, Decision Tree Tuned Akurasi 92%, Presisi 92%, Recall 88%, F1 Score 90%.

Machine Learning Model Ranking



Gambar 14. Machine Learning Model Ranking

Pada Gambar 14 Menunjukkan peringkat Ranking dari Model yang digunakan yaitu Algoritma Support Vector Machine dengan peringkat pertama, kedua Artificial Neural Network, ketiga Algoritma ANN dengan Hyperparameter Tuning, Keempat Algoritma SVM dengan Hyperparameter Tuning, Kelima Algoritma Decision Tree dan yang terakhir keenam Algoritma DT dengan Hyperparameter tuning.

KESIMPULAN

Penelitian ini menunjukkan bahwa model klasifikasi menggunakan Artificial Neural Network dan Support Vector Machine menghasilkan akurasi yang tinggi dalam klasifikasi kanker payudara. Hasil eksperimen menunjukkan bahwa model Support Vector Machine memiliki kinerja terbaik, diikuti oleh Artificial Neural

DAFTAR PUSTAKA

- R. Andini and Y. P. Astuti, "Math Unesa," *J. Ilm. Mat.*, vol. 9, no. 2, 2021
- E. Marfianti, "Peningkatan Pengetahuan Kanker Payudara dan Ketrampilan Periksa Payudara Sendiri (SADARI) untuk Deteksi Dini Kanker Payudara di Semutan Jatimulyo Dlingo," *J. Abdimas Madani dan Lestari*, vol. 3, no. 1, 2021, doi: 10.20885/jamali.vol3.iss1.art4.
- N. Sari, "Karakteristik Penyebab Kanker Payudara," *J. Ilm. PANNMED (Pharmacist, Anal. Nurse, Nutr. Midwifery, Environ. Dent.*, vol. 16, no. 1, 2021, doi: 10.36911/pannmed.v16i1.1002.
- A. Mardiana and L. Kurniasari, "Hubungan Pengetahuan Pemeriksaan Payudara Sendiri (SADARI) dengan Kejadian Kanker Payudara di Kalimantan Timur," *Borneo Student Res.*, vol. 2, no. 2, 2021.
- I. Fajri, D. Nurhamsyah, S. Aisyah, K. A. Mudrikah, and A. R. Azjurnia, "Terapi Non-Farmakologi dalam Mengurangi Tingkat Nyeri pada Pasien Kanker Payudara Stadium 2-4: Literature Review," *J. Ilm. Keperawatan Indones.*, vol. 5, no. 2, 2022, doi: 10.31000/jiki.v5i2.6139.
- R. Rukinah and S. Luba, "Pengetahuan Wanita Usia Subur Tentang Pencegahan Kanker Payudara," *J. Ilm. Kesehat. Sandi Husada*, vol. 10, no. 1, 2021, doi: 10.35816/jiskh.v10i1.597.
- P. Purwono, P. Dewi, S. K. Wibisono, and B. P. Dewa, "Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network," *Insect (Informatics Secur. J. Tek. Inform.*, vol. 7, no. 2, 2022, doi: 10.33506/insect.v7i2.1828.
- R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, 2023, doi: 10.34128/jsi.v9i1.622.
- M. A. Jabbar, E. Hasmin, C. Susanto, W. Musu, and I. Artikel, "Komparasi Algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara," *CSRID J.*, vol. 14, no. 3, 2022.
- L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1851.
- I. M. Karo Karo and H. Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score," *J. Teknol. Terpadu*, vol. 8, no. 2, 2022, doi: 10.54914/jtt.v8i2.564.
- W. Lestari and S. Sumarlinda, "Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma Machine Learning," *SATIN - Sains dan Teknol. Inf.*, vol. 9, no. 1, 2023, doi: 10.33372/stn.v9i1.918.
- A. Muhson, "Teknik Analisis Kuantitatif," *Makal. Tek. Anal. II*, 2006.