

Teknik Klasifikasi Kanker Payudara Dengan 3 Metode; SVM, KNN, Dan Random Forest

Muhammad Vikri Fauzi, Ruth Kristiana Putri, Rizkya Putri Maulana, Adithya Kusuma Whardana S.Kom., M.Kom

Teknik Informatika, Tanri Abeng University, Jakarta, Indonesia.

Email :

Received:

Abstrak: Kanker Payudara adalah Kanker yang berasal dari penyakit non kulit yang berasal dari sel kelenjar, saluran kelenjar, dan jaringan penunjang payudara. Jurnal ini menggunakan 3 metode yaitu SVM, KNN, dan Random Forest untuk mengetahui metode mana yang lebih baik untuk digunakan sebagai pendeteksi kanker payudara. Metode SVM adalah metode yang biasa digunakan dalam klasifikasi dengan cara kerja mengklasifikasi ruang input menjadi dua kelas. Metode KNN adalah metode klasifikasi yang bekerja dengan mengambil sejumlah data terdekat sebagai acuan untuk menentukan kelas dari data baru. Metode Random Forest adalah algoritma pembelajaran *supervised* yang didalamnya ada kumpulan pohon keputusan dan dilatih dengan metode *bagging*. Dari ketiga metode tersebut hasil terbesar didapatkan dengan menggunakan metode Random Forest dengan hasil akurasi mencapai 94.7% lalu metode SVM dengan hasil akurasi 90.6% dan dengan metode KNN dengan akurasi 90%. Sumber data berasal dari Kaggle. Dengan Diagnosis “M” (malignant) dan “B” (benign). Terdiri dari 569 data dan 33 kolom.

Kata Kunci: Kanker Payudara, K-Nearest Neighbor (KNN), SVM (Support Vektor Machine), Random Forest.

I. Pendahuluan

Kanker payudara adalah kanker yang terbentuk di jaringan payudara. Kanker payudara terjadi ketika sel-sel pada jaringan yang ada di payudara tumbuh tidak terkendali dan mengambil alih jaringan payudara yang sehat dan sekitarnya. Kanker payudara bisa terbentuk di kelenjar yang menghasilkan susu (*lobulus*) atau di saluran (*duktus*) yang membawa air susu dari kelenjar ke puting payudara. Kanker juga bisa terbentuk di jaringan lemak atau jaringan ikat di dalam payudara.

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. 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. 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.

Kanker payudara termasuk kedalam kanker yang sangat berbahaya didunia, baik di negara maju atau negara berkembang. Setiap tahun 12 juta orang di seluruh dunia menderita kanker dan 7,6 juta di antaranya meninggal dunia, dari jumlah tersebut 506.000 disebabkan oleh kanker payudara ini. Hal ini menunjukkan bahwa breast cancer adalah salah satu kanker ganas di dunia. Saat ini 16 % dari semua jenis kanker pada wanita didunia adalah breast cancer. Berdasarkan data WHO 69% dari kematian breast cancer didunia terjadi di negara berkembang,

Struktur dari jurnal ini disusun sebagai sebagai Bab 1 menjelaskan tentang metode yang digunakan dan juga materi yang dipakai yang dimana metode yang digunakan ada 3 yaitu SVM (Support Vektor Machine), KNN (K-Nearest Neighbor), dan Random Forest. Materi yang digunakan pada penelitian ini merupakan dataset dari kanker payudara yang berisikan 2 label, yaitu jinak dan ganas. Kemudian di Bab 2 menjelaskan dataset yang digunakan dan juga metode yang digunakan. Terakhir Bab 3 menjelaskan hasil dan diskusi dari penelitian yang telah dilakukan.

Pada paper ini, kami melakukan perbandingan dengan menggunakan 3 metode yaitu SVM, KNN, dan Random Forest untuk klasifikasi penyakit kanker payudara.

II. Materi dan Metode

A. Material Dataset

Sumber data pada penelitian ini berasal dari UCI Machine Learning Repository. UCI Machine Learning adalah sebuah web koleksi database, domain teori, dan data generator yang digunakan untuk mencari sebuah database yang mempelajari *machine learning* untuk keperluan analisis. UCI Machine Learning digunakan oleh pendidik maupun peneliti sebagai sumber utama dari data set pada *machine learning*.

Machine Learning merupakan teknik untuk melakukan inferensi (menitikberatkan ranah hubungan variabel) terhadap data dengan pendekatan matematis. Intinya untuk membuat model (matematis) yang merefleksikan pola – pola data. Machine Learning sendiri adalah cabang ilmu *Artificial Intelligence* yang memiliki konsep dimana komputer sebagai mesin memiliki kemampuan untuk melakukan adaptasi terhadap lingkungan yang baru dan mampu mendeteksi pola dari fakta yang ada.

Data yang digunakan berjudul Breast Cancer Wisconsin (Diagnostic Data Set) yang berasal dari *image* digital Fine Needle Aspirate(FNA) dari *breast mass*. Pada data ini digambarkan karakteristik dari inti sel yang ada pada gambar. Yang dibuat oleh Dr. William H. Wolberg, W. Nick Street, dan Olvi L. Mangasarian.

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>
1

Atribut yang terdapat dalam dataset ini adalah :

1. ID number
2. Diagnosis (M = *Malignant*, B = *Benign*)

10 *real-valued features* yang ada di dataset :

- a. *Radius* (berisikan rata – rata jarak dari pusat ke point pada perimeter)
- b. *Textures* (standar deviasi nilai *gray-scale*)
- c. Perimeter
- d. Area
- e. *Smoothness* (rata-rata variasi lokal dalam panjang radius)
- f. *Compactness* (rata-rata keliling² / luas - 1,0)
- g. *Cocavity* (rata-rata keparahan bagian cekung dari kontur)
- h. *Concave Points* (rata-rata untuk jumlah bagian cekung dari kontur)
- i. *Symmetry*
- j. *Fractal Dimension*

B. Metode

1. Support Vector Machine (SVM)

SVM adalah kelas algoritma untuk pembelajaran mesin yang bisa digunakan sebagai pengenalan pola dan regresi yang berdasarkan teori dari pembelajaran statistik dan prinsip minimalisasi risiko struktural. Metode SVM ini dibuat untuk mencari *hyperplane* yang memisahkan satu set contoh positif dari satu set contoh yang negatif dengan margin maksimum. Margin ini didefinisikan oleh jarak *hyperplane* ke terdekat dari contoh positif dan negatif.

Metode SVM ini juga banyak digunakan sebagai metode klasifikasi dengan akurasi yang tinggi. Lalu untuk kelebihan menggunakan SVM ini adalah metode ini mampu menghasilkan model klasifikasi yang baik meski dengan data yang tidak terlalu banyak, akan tetapi metode ini juga memiliki kekurangan, yaitu metode ini sulit diaplikasikan kedalam sistem dengan jumlah data yang sangat banyak.

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$$

Dimana $\mathbf{x} = (x_1, x_2, \dots, x_D)^T$ adalah variabel input, dan $\mathbf{w} = (w_0, w_1, \dots, w_D)^T$ adalah parameter, $\boldsymbol{\phi}(\mathbf{x})$ adalah fungsi basis, M adalah jumlah total parameter dari model. Biasanya, $\phi_0(\mathbf{x}) = 1$, sehingga w_0 berfungsi sebagai bias.

Support Vector Machine (SVM) menggunakan model linear sebagai decision boundary dengan bentuk umum sbb:

$$y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b$$

Dimana \mathbf{x} adalah vektor input, \mathbf{w} adalah parameter bobot, $\boldsymbol{\phi}(\mathbf{x})$ adalah fungsi basis, dan b adalah suatu bias.

Hasil yang didapatkan dengan menggunakan metode SVM untuk data *train* mendapatkan 87.44% dan untuk testing data mendapatkan 90.06% dengan jumlah *Benign* dari data *train* yaitu 249 dan *Malignant* 149, lalu untuk data test mendapatkan data *Benign* berjumlah 108 dan *Malignant* 63.

2. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan metode algoritma untuk melakukan klasifikasi dengan jumlah tetangga paling dekat dengan objek. Data ini diproyeksikan ke kelompok yang banyak, yang dimana dalam masing – masing kelompok mempresentasikan fitur dari data yang diambil.

Nilai k pada algoritma KNN mendefinisikan berapa banyak tetangga yang akan diperiksa untuk menentukan klasifikasi titik kueri tertentu. Misalnya, jika k=1 maka akan ditugaskan ke kelas yang sama dengan tetangga terdekatnya. Pengujian dengan nilai k=1 mendapatkan tingkat akurasi lebih tinggi, perbedaan ukuran piksel data citra, penambahan noise, dan perbedaan resolusi kamera mempengaruhi tingkat akurasi yang di peroleh.

KNN sendiri merupakan algoritma machine learning yang bersifat non-parametric. Metode yang bersifat non-parametric memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, terlepas data tersebut berukuran kecil ataupun besar.

Algoritma non-parametric seperti KNN menggunakan sejumlah parameter yang fleksibel, dan jumlah parameter seringkali bertambah seiring data yang semakin banyak. Algoritma non-parametric secara komputasi lebih lambat, tetapi membuat lebih sedikit asumsi tentang data.

Untuk menentukan titik data mana yang paling dekat dengan titik kueri tertentu, jarak antara titik kueri dan titik data lainnya perlu dihitung. Metrik jarak ini membantu membentuk batasan keputusan, yang mengarahkan kueri partisi ke kelas yang berbeda.

Perhitungan jarak ketetanggaan menggunakan algoritma euclidien dirumuskan:

$$euc = \sqrt{((a_1 - b_1)^2 + \dots + (a_n - b_n)^2)}$$

Dimana a = a₁, a₂, .. , a_n dan b = b₁, b₂, .. , b_n mewakili n nilai atribut dari dua record.

Untuk menghitung kemiripan kasus, digunakan rumus (kusrini, 2009) :

$$\begin{aligned} & Similarity(p, q) \\ &= \frac{n \sum_{i=1}^n f(p_i, q_i) \times w_i}{w_i} \end{aligned}$$

Keterangan :

p = Kasus baru

q = Kasus yang ada dalam penyimpanan (data training)

N = jumlah atribut dalam setiap kasus

i = Atribut individu antara 1 sampai dengan n
 f = Fungsi similarity atribut i antara kasus p dan kasus q
 w = bobot yang diberikan pada atribut ke- i

Adapun langkah – langkah klasifikasi KNN sebagai berikut :

1. Tentukan parameter nilai k = banyaknya jumlah tetangga terdekat
2. Hitung jarak antara data baru dengan semua data training
3. Urutkan jarak dan tetapkan tetangga terdekat berdasarkan jarak minimum ke- k
4. Periksa kelas dari tetangga terdekat
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru

3. Random Forest

Random Forest adalah algoritma pembelajaran mesin (*machine learning*) yang menggabungkan hasil dari beberapa pohon keputusan untuk sampai pada satu hasil. Seperti namanya, hutan terdiri dari banyak pohon (*trees*) yang diperoleh dengan proses agregasi *bagging* atau *bootstrap*. Setiap pohon pada Random Forest memancarkan prediksi kelas. Kelas prediksi dengan suara terbanyak menjadi kandidat prediksi model.

Random Forest adalah algoritma pembelajaran mesin yang digunakan untuk mengklasifikasikan set data besar. Karena fungsinya bisa digunakan untuk banyak ukuran dengan skala berbeda dan performa tinggi. Klasifikasi ini dilakukan dengan menggabungkan pohon-pohon menjadi pohon keputusan menggunakan dataset pelatihan yang Anda miliki. Algoritma Random Forest, yang memiliki kelebihan sebagai versatile, algoritma ini biasanya menghasilkan akurasi yang bagus tanpa di-tune. Semakin tingginya jumlah pohon, semakin tinggi akurasinya dan semakin terhindar dari masalah overfitting dan bisa diatasi.

Algoritma Random Forest meningkatkan keacakan model saat pohon tumbuh. Alih-alih menemukan fitur yang paling penting saat memisahkan sebuah node, hutan acak mencari fitur terbaik di antara sub-fitur acak. Akibatnya, ia menawarkan banyak variasi dan seringkali menghasilkan model yang lebih baik.

Algoritma Random Forest diperkenalkan oleh Leo Breiman dan Adele Cutler. Algoritma ini didasarkan pada konsep pembelajaran sinkron, yang melibatkan penggabungan beberapa pengklasifikasi untuk memecahkan masalah yang kompleks dan meningkatkan kinerja model. Random Forest beroperasi dalam dua fase. Tahap pertama terdiri dari menggabungkan sejumlah N pohon keputusan (*decision tree*) untuk membuat Random Forest. Kemudian tahap kedua adalah membuat prediksi untuk setiap pohon yang dibuat pada tahap pertama.

Cara kerja Algoritma Random Forest dapat dijabarkan dalam langkah – langkah berikut :

1. Membangun *decision tree* untuk setiap sampel yang dipilih. Kemudian akan diperoleh hasil prediksi dari setiap pohon keputusan (*decision tree*) yang telah dibuat.

2. Proses voting dilakukan untuk setiap hasil prediksi. Untuk masalah klasifikasi, gunakan modus (nilai yang paling sering muncul), sedangkan untuk masalah regresi gunakan rata-rata (mean).
3. Algoritma memilih sampel acak dari kumpulan data yang disediakan
4. Algoritma akan memilih prediksi dengan suara terbanyak (*most votes*) sebagai prediksi akhir.

Kelebihan Random Forest :

1. Bekerja dengan baik menggunakan data non-linear.
2. Resiko overfitting sangat rendah.
3. Akurasinya lebih tinggi dari algoritma klasifikasi lainnya.

Kekurangan Random Forest :

1. Hutan acak (Random Forest) cenderung bias ketika berhadapan dengan variabel kategori.
2. Waktu komputasi pada kumpulan data berskala besar relatif lambat.
3. Tidak cocok untuk metode linier dengan banyak fitur jarang (sparse).

$$P(c|I, \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T P_t(c|I, \mathbf{x}) .$$

Setiap pohon dilatih pada satu set gambar yang disintesis secara acak. Subkumpulan data pencitraan acak untuk memastikan distribusi seragam di seluruh payudara dengan kanker. Setiap pohon dilatih menggunakan algoritma berikut :

1. Usulkan secara acak satu set kandidat pemecah $\phi = (\theta, \tau)$ (parameter fungsi θ dan ambang batas τ).
2. Partisi himpunan sampel $Q = \{(I, \mathbf{x})\}$ menjadi himpunan bagian kiri dan kanan dengan ϕ berturut-turut

$$\begin{aligned} Q_l(\phi) &= \{ (I, \mathbf{x}) \mid f_{\theta}(I, \mathbf{x}) < \tau \} \\ Q_r(\phi) &= Q \setminus Q_l(\phi) \end{aligned}$$

3. Menghitung ϕ memberikan perolehan informasi terbesar di mana Shannon entropi $H(Q)$ dihitung pada grafik normalisasi fraksi label $I(\mathbf{x})$ untuk semua $(I, \mathbf{x}) \in Q$.

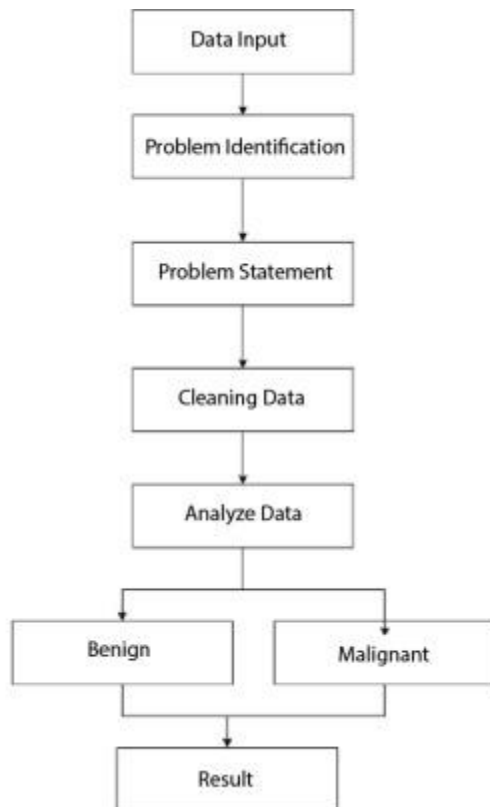
$$\phi^* = \underset{\phi}{\operatorname{argmax}} G(\phi)$$

$$G(\phi) = H(Q) - \sum_{s \in \{l, r\}} \frac{|Q_s(\phi)|}{|Q|} H(Q_s(\phi))$$

4. Jika gain maksimum $G(\phi^*)$ cukup dan kedalaman pohon kurang dari maksimum, maka iterasi untuk himpunan bagian kiri dan kanan $Q_l(\phi^*)$ dan $Q_r(\phi^*)$.

C. Langkah – Langkah Penelitian

1. Flowchart



Keterangan :

Data Input : langkah pertama yang dilakukan adalah memasukkan data mentah yang akan di proses

Problem Identification : identifikasi masalah dalam hal ini adalah menentukan antara *Benign* dan *Malignant*

Problem Statement : dalam hal ini pernyataan masalah di jurnal ini adalah untuk menentukan mana yang lebih tinggi presentasinya dengan menggunakan 3 metode

Cleaning Data : dalam tahap ini data dibersihkan dalam artian tidak semua data yang ada akan terpakai.

Benign dan *Malignant* : 2 hal ini adalah tujuan dari hasil data yang dipakai dalam jurnal ini.

Result : merupakan sekumpulan hasil dari proses pada tahap sebelumnya.

Tahapan yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Menggunakan dataset yang berjudul Breast Cancer Wisconsin yang diambil dari UCI Machine Learning.
2. Kemudian menentukan variabel X dan Y nya dimana X sebagai Rasio dan Y sebagai Nominal (Label), dimana label yang digunakan di data ada 2 yaitu Malignant(M) dan Benign(B).
3. Membagi data menjadi data training dan data testing.
4. Mengklasifikasikan dataset menggunakan algoritma SVM.
 - a. Mengimpor package SVM
 - b. Membuat grafik dari target Malignant (M) dan Benign(B)
 - c. Fungsi klasifikasi report dari X train dan X test

- d. Melihat hasil akurasi klasifikasi dari SVM
5. Mengklasifikasikan dataset menggunakan algoritma KNN.
 - a. Mengimpor package KNN
 - b. Mengaktifkan fungsi klasifikasi KNN dimana $n=4$
 - c. Memasukkan data training pada fungsi klasifikasi untuk KNN
 - d. Matrix hasil prediksi
 - e. Melihat hasil prediksi akurasi menggunakan KNN
6. Mengklasifikasikan dataset menggunakan algoritma Random Forest.
 - a. Target variabel, dimana yang digunakan itu adalah label Malignant(M) dan Benign(B)
 - b. Melakukan drop ke beberapa variabel yang tidak dipakai
 - c. Binarisasi variabel target
 - d. Mengimpor package Random Forest
 - e. Melakukan hasil akurasi menggunakan Random Forest
7. Melakukan perbandingan pemilihan algoritma mana yang paling terbaik berdasarkan performa klasifikasi.

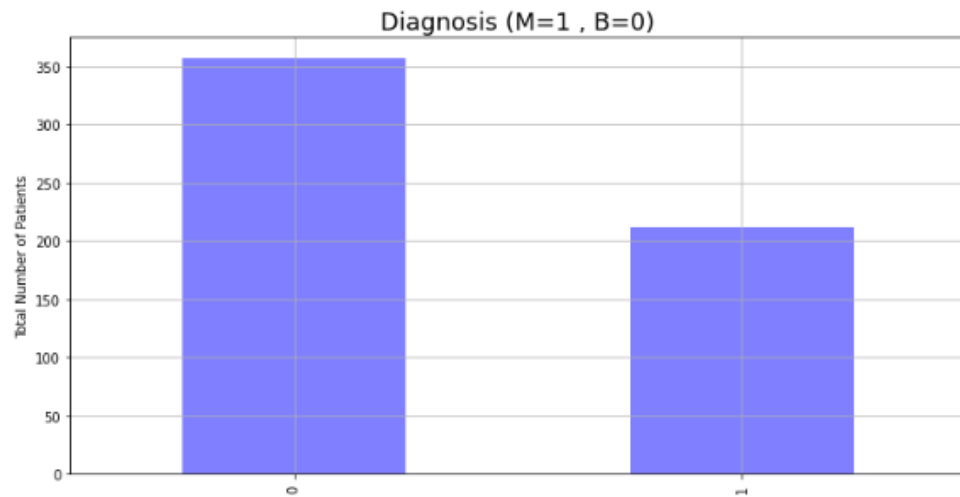
III. Kesimpulan

Berdasarkan hasil klasifikasi kanker payudara dari perbandingan 3 metode yang digunakan (SVM, KNN dan Random Forest) menghasilkan kesimpulan dimana terlihat dari tabel 1.0. dimana tabel tersebut berisikan dari data Benign(Jinak), Malignant(Ganas) dan akurasi dari setiap metode yang digunakan.

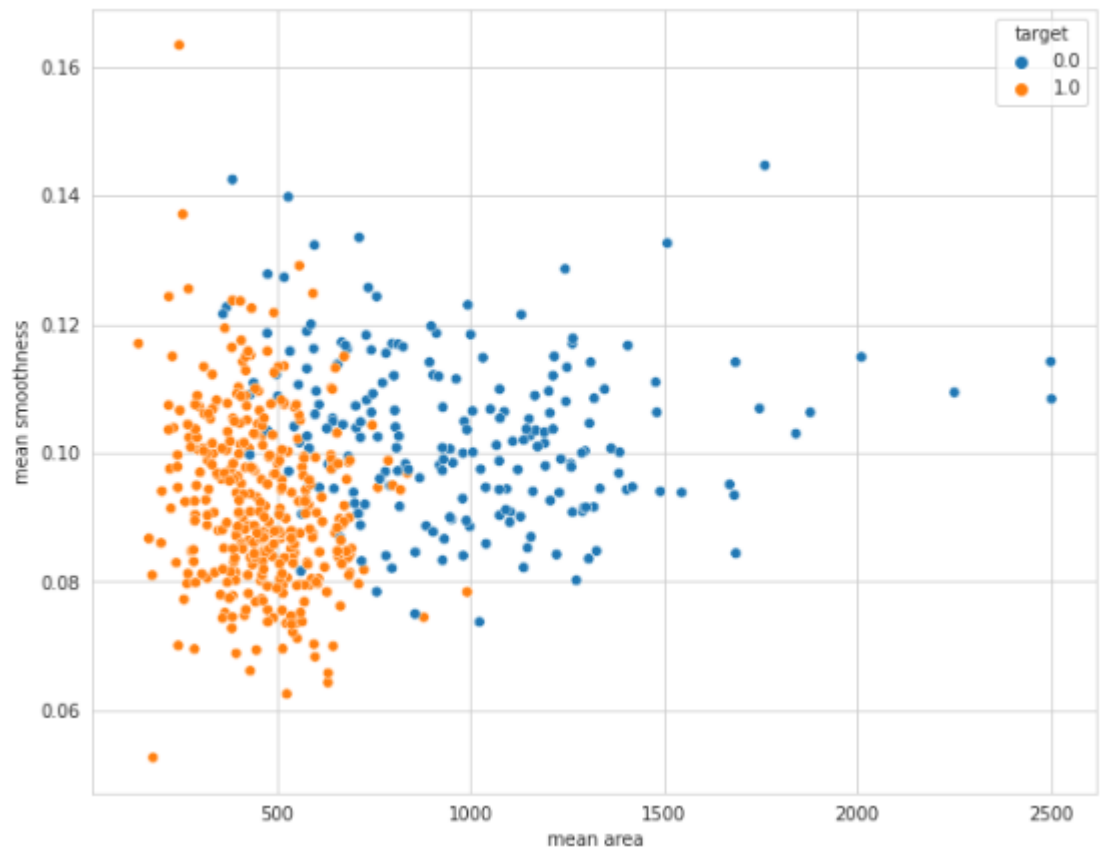
Metode	Benign	Malignant	Accuracy Score
SVM	108	63	90.06%
KNN	75	39	90.00%
Random Forest	67	47	94.07%

Terlihat dari hasil dari tabel 1.0, bahwa nilai akurasi algoritma Random Forest lebih tinggi, ini membuktikan bahwa algoritma Random Forest memiliki metode terbaik di antara dua metode yang tersisa, termasuk algoritma KNN dan SVM, dalam dataset yang sama digunakan.

Pada gambar 1.0 berisikan jumlah dataset Malignant yang diberi label 1 dan Benign yang diberi label 0. Terlihat bahwa lebih banyak data malignant.



Kemudian pada gambar 1.2 berisikan dari pesebaran Malignant dan Benign di dalam database Breast Cancer Wisconsin.



Hasil yang diperoleh dengan penggunaan kombinasi ketiga metode mendiganosis dataset yang dihasilkan lebih akurat. Pemilihan algoritma yang tepat juga dapat berkontribusi pada keakuratan hasil prediksi. Pada penelitian selanjutnya dapat dikembangkan dengan menggunakan metode lain dan kumpulan dataset lainnya sebagai bahan uji untuk menentukan model terbaik untuk melakukan penelitian. Metode yang digunakan ternyata lebih unggul daripada algoritma yang diterapkan di Database Breast Cancer Wisconsin dan menunjukkan bahwa ini bisa menjadi alternatif yang menarik.

Sebelumnya penelitian ini sangat berterima kasih kepada pembuat database yang mana digunakan sebagai bahan dalam penelitian ini sehingga penelitian ini bisa terlaksanakan.

Penelitian lanjutan dapat dianjurkan untuk mengembangkan algoritma yang ada, maupun memakai metode yang lainnya untuk mendapatkan hasil yang lebih bagus.

Referensi

- [1] Rafika Harrabi and Ezzedine Ben Braiek. Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images. EURASIP Journal on Image and Video Processing, 2012:1-11.
- [2] Ahmad Fauzi, Riki Supriyadi, Nurlaelatul Maulidah. Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest. Jurnal Infortech, 2020;2:1-6.
- [3] Farizi Rachman, Santi Wulan Purnami. Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM). JURNAL SAINS DAN SENI ITS, 2012;1:1-6.
- [4] Cuong Nguyen, Yong Wang, Ha Nam Nguyen. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. Journal of Biomedical Science and Engineering, 2013;6:1-10.
- [5] Shakil Mahmud Bobby and Shaela Sharmin. Medical Image Denoising Techniques against Hazardous Noises: An IQA Metrics Based Comparative Analysis. I.J. Image, Graphics and Signal Processing. 2021;2:25-43.
- [6] Vincent Angkasa and Jefri Junifer Pangaribuan. KOMPARASI TINGKAT AKURASI RANDOM FOREST DAN KNN UNTUK MENDIAGNOSIS PENYAKIT KANKER PAYUDARA. Universitas Pelita Harapan PSDKU Medan Jurusan Sistem Informasi. 2022;7:49-61.
- [7] Runi Hari Bagus Saputra, Roy Mubarak. Implementasi Algoritma Random Forest Untuk Mendiagnosis Kejadian Berulang (Kekambuhan) Pada Kanker Payudara Berbasis Web. OKTAL : Jurnal Ilmu Komputer dan Sains. 2022;1(06):564-572.
- [8] Widhi Ramdhani , David Bona, Rafi Bagus Musyaffa, Chaerur Rozikin. Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor. Jurnal Ilmiah Wahana Pendidikan. 2022;8(12):445-452.

- [9] H. Harafani, H. Aji Al-Kautsar. MENINGKATKAN KINERJA K-NN UNTUK KLASIFIKASI KANKER PAYUDARA DENGAN SELEKSI FITUR. Jurnal Pendidikan Teknologi dan Kejuruan. 2021;18(1):99-110.
- [10] Fourina Ayu Novianti dan Santi Wulan Purnami . Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. JURNAL SAINS DAN SENI ITS. 2012;1(1):147-152.
- [11] Helmi Imaduddin, Brian Aditya Hermansyah and Frischa Aura Salsabilla B. ARISON OF SUPPORT VECTOR MACHINE AND DECISION TREE METHODS IN THE CLASSIFICATION OF BREAST CANCER. Cyberspace: Jurnal Pendidikan Teknologi Informasi. 2021; 5(1):22-30.
- [12] Chalifa Chazar, Bagus Erawan Widhiaputra. Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. INFORMASI (Jurnal Informatika dan Sistem Informasi). 2020;12(1):67-80.2
- [13] Permana Putra, Akim M H Pardede , Siswan Syahputra. ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA. Jurnal Teknik Informatika Kaputama (JTIK), 2022;6:297-305.
- [14] Harun Al Azies, Gangga Anuraga. Klasifikasi Daerah Tertinggal di Indonesia Menggunakan Algoritma SVM dan K-NN. Jurnal ILMU DASAR, 2021;22:31-38.
- [15] Wiyli Yustanti. Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. Jurnal Matematika, Statistika & Komputasi (JMSK), 2012;9:57-68.
- [16] Retno Paras Rasmi. PENINGKATAN HASIL DIAGNOSIS KANKER PAYUDARA DARI HASIL CITRA MAMMOGRAM MENGGUNAKAN METODE EKSTRASI CIRI DAN KLASIFIKASI. Jurnal Teknik Elektro FTI Universitas Islam Indonesia Yogyakarta, 2020;1:5-13.
- [17] Emi Susilowati, Amelia Tri Hapsari, Muhammad Efendi, Priadhana Edi Kresnha. DIAGNOSA PENYAKIT KANKER PAYUDARA MENGGUNAKAN METODE K-MEANS CLUSTERING. Jurnal Sistem Informasi, Teknologi Informatika dan Komputer, 2020;10:27-32.
- [18] Sri Rezeki Candra Nursari, Nanda Mahya Barokatun Nisa. SERVICES CANCER DETECTION SYSTEM USING K-NEAREST NEIGHBOURS(K-NN) METHOD AND NAÏVE BAYES CLASSIFIER. International Journal Information System and Computer Science (IJISCS). 2020;4(1):40-43.

- [19] Dewi Cahyantia, Alifah Rahmayania, Syafira Ainy Husniara. Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. Indonesian Journal of Data and Science. 2020;1(2):39-43.
- [20] Rina Resmiati, Toni Arifin. Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination. SISTEMASI: Jurnal Sistem Informasi.2021;10(2):381-393.
- [21] Ahmad Roihan, Po Abas Sunarya, Ageng Setiani Rafika. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. IJCIT (Indonesian Journal on Computer and Information Technology), 2020;5(1):75-82.