

**KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST***

**SKRIPSI**

Oleh:  
**RIFQI MUFIDDIN**  
**NIM. 19650004**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN METODE  
*RANDOM FOREST***

**SKRIPSI**

**Oleh:  
RIFQI MUFIDDIN  
NIM. 19650004**

**Diajukan kepada:  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana (UIN) Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**HALAMAN PERSETUJUAN**

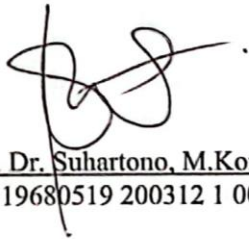
**KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN METODE  
RANDOM FOREST**

**SKRIPSI**

Oleh:  
**RIFQI MUFIDDIN**  
**NIM. 19650004**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 29 Mei 2023

Pembimbing I,



Prof. Dr. Suhartono, M.Kom  
NIP. 19680519 200312 1 001

Pembimbing II,



Agung Teguh Wibowo Almais, M.T  
NIDT. 19860301 20180201 1 235

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Firdaus Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

**HALAMAN PENGESAHAN**

**KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN METODE  
RANDOM FOREST**

**SKRIPSI**

Oleh :  
**RIFOI MUFIDDIN**  
NIM. 19650004

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal: 13 Juni 2023

**Susunan Dewan Penguji**

Ketua Penguji : Dr. Muhammad, Faisal, M.T  
NIP. 19740510 200501 1 007

Anggota Penguji I : Okta Omaruddin Aziz, M.Kom  
NIP. 19911019 201903 1 013

Anggota Penguji II : Prof. Dr. Suhartono, M.Kom  
NIP. 19680519 200312 1 001

Anggota Penguji III : Agung Teguh Wibowo Almais, M.T  
NIDT. 19860301 20180201 1 235



Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
**Dr. Fajri Kurniawan, M.MT, IPM**  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Rifqi Mufiddin

NIM : 19650004

Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Klasifikasi Kanker Payudara Menggunakan Metode Random Forest.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 13 Juni 2023

Yang membuat pernyataan,



Rifqi Mufiddin  
NIM. 19650004

## **HALAMAN MOTTO**

“Harus bisa beradaptasi dalam segala situasi”

“Menjadi berbeda dari kebanyakan orang”

“Progres diri setiap hari setidaknya 1%”

## **HALAMAN PERSEMBAHAN**

**Puji syukur kehadiran Allah Subhanahu Wa Ta'ala, shalawat dan salam bagi  
Rasul-Nya**

Skripsi ini dipersembahkan untuk kedua orang tua saya yaitu Siti Kholidah dan Ahmad Mukhlis sebagai bentuk tanggung jawab saya kepada orang tua. Perjuangan yang dilakukan oleh kedua orang tua saya semoga diberikan balasan yang setimpal oleh Allah. Terima kasih kepada seluruh orang yang terlibat dan membantu pembuatan skripsi ini secara langsung maupun tidak.

## KATA PENGANTAR

*Assalamualaikum Warahmatullahi Wabarakatuh*

Dengan menyebut nama Allah yang Maha Pengasih lagi Maha Penyayang, Segala puji dan syukur hanya bagi Allah Subhanahu Wa Ta'ala, Tuhan Semesta Alam, yang telah memberikan rahmat-Nya yang tiada terhingga serta petunjuk-Nya yang sempurna kepada hamba-Nya. Shalawat serta salam semoga tercurahkan kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam, utusan Allah yang diutus sebagai rahmat bagi seluruh alam.

Dalam kesempatan yang berbahagia ini, dengan kerendahan hati dan penuh rasa syukur, saya dengan bangga mempersembahkan kata pengantar ini sebagai bagian dari penulisan skripsi berjudul “Klasifikasi Kanker Payudara Menggunakan Metode *Random forest*”. Skripsi ini telah saya susun untuk memenuhi syarat kelulusan dan mendapatkan gelar sarjana komputer (S.Kom) di Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Dalam perjalanan penyusunan skripsi ini, saya senantiasa merasakan karunia dan bimbingan Allah Subhanahu Wa Ta'ala. Segala keberhasilan dan kemajuan yang saya raih tidak lepas dari rahmat-Nya yang melimpah serta petunjuk-Nya yang tak tergantikan. Saya merasa terhormat dan bersyukur atas kesempatan ini, yang memungkinkan saya untuk menggali lebih dalam dan memahami topik ini. Penyusunan skripsi ini juga tidak terlepas dari peran dan dukungan banyak pihak yang tak ternilai. Oleh sebab itu, penulis ingin mengucapkan terimakasih kepada:



1. Ibu, Ayah, Kakak, dan Adik tercinta yang selalu memberikan doa, dukungan, dan motivasi dalam perjalanan ini. Terima kasih atas cinta, kepercayaan, dan pengertian yang luar biasa dari keluarga saya. Semoga Allah Subhanahu Wa Ta'ala memberikan keberkahan dan kebahagiaan kepada kalian semua.
2. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi.
4. Dr. Fachrul Kurniawan, M.MT selaku Ketua Program Studi Teknik Informatika.
5. Prof. Dr. Suhartono, M.Kom dan Agung Teguh Wibowo Almais, M.T selaku dosen pembimbing yang dengan penuh kesabaran dan kebijaksanaan telah membimbing saya dalam menyelesaikan skripsi ini. Bimbingan, nasihat, dan pengarahan yang diberikan sangat berarti bagi kemajuan dan kelancaran penelitian ini.
6. Segenap civitas akademika Jurusan Teknik Informatika, Terutama seluruh dosen, Terima Kasih atas ilmu dan bimbingan yang telah diberikan selama masa perkuliahan ini.
7. Anak dari Ibu Juwita Hasni dan Bapak Taharuddin yaitu Vania Afwi yang telah berkontribusi besar dalam penulisan skripsi ini.
8. Teman-teman dekat penulis yang menjadi saksi skripsi ini dibuat.
9. Teman-teman Jurusan Teknik Informatika Angkatan 2019 "Alien" yang sama-sama berjuang mengejar gelar S.Kom.
10. Penulis sendiri yang dapat menyelesaikan skripsi dan tanggung jawab penulis.

11. Semua pihak dan tempat yang telah membantu penulis dalam menyelesaikan skripsi ini, baik secara langsung maupun tidak langsung yang tidak dapat disebutkan satu per satu.

Akhir kata, penulis menyadari bahwa penulisan pada skripsi ini masih jauh dari kata sempurna. Saya berdoa semoga skripsi ini diterima sebagai amal ibadah yang ikhlas dan bermanfaat di sisi Allah Subhanahu Wa Ta'ala. Semoga karya ini menjadi bentuk kontribusi yang tak terpisahkan dalam rangka memperkuat dan mengembangkan ilmu pengetahuan, serta menjalankan tugas sebagai hamba Allah yang bertanggung jawab.

Malang, 29 Mei 2023

Penulis

## DAFTAR ISI

<b>HALAMAN PENGAJUAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN</b> .....	<b>v</b>
<b>HALAMAN MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>xi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>ABSTRAK</b> .....	<b>xvi</b>
<b>ABSTRACT</b> .....	<b>xvii</b>
<b>خالصة</b> .....	<b>xviii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Batasan Masalah .....	5
1.5 Manfaat Penelitian .....	5
<b>BAB II STUDI PUSTAKA</b> .....	<b>6</b>
2.1 Penelitian Terkait .....	6
2.2 Kanker .....	11
2.3 Kanker Payudara .....	12
2.4 <i>Data Mining</i> .....	14
2.5 <i>Dataset</i> .....	17
2.6 <i>Machine learning</i> .....	18
2.7 <i>Random forest</i> .....	19
2.8 <i>ROC dan AUC</i> .....	22
2.9 <i>Grid Search</i> .....	23
2.10 <i>K-Fold Cross Validation</i> .....	24
2.11 <i>Confusion Matrix</i> .....	25
2.12 Bahasa Pemrograman R .....	27
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>29</b>
3.1 Tahapan Penelitian .....	29
3.2 Tahap Awal .....	30
3.2.1 Identifikasi Masalah .....	30
3.2.2 Studi Literatur .....	30
3.3 Observasi Data .....	30
3.3.1 Pengumpulan Data .....	31
3.3.2 Pemahaman Data .....	32
3.3.3 Contoh <i>Dataset</i> Penyakit Kanker Payudara .....	34
3.4 <i>Preprocessing</i> .....	35
3.4.1 <i>Data Cleaning</i> .....	35

3.4.2 Data <i>Exploration</i> .....	36
3.4.3 Split Data.....	40
3.5 Implementasi Metode <i>Random forest</i> .....	41
3.6 Evaluasi .....	43
3.7 Skenario Pengujian.....	45
3.8 Perhitungan Manual <i>Random forest</i> .....	45
<b>BAB IV UJI COBA DAN PEMBAHASAN .....</b>	<b>54</b>
4.1 Hasil Uji Coba.....	54
4.1.1 Pengujian Model A.....	54
4.1.2 Pengujian Model B.....	57
4.1.3 Pengujian Model C.....	59
4.1.4 Pengujian Model D.....	62
4.1.5 Pengujian 10-Fold Cross Validation dan ROC.....	64
4.2 Pembahasan.....	69
4.3 Integrasi Penelitian Dalam Tafsir Al-Qur'an.....	78
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>82</b>
5.1 KESIMPULAN .....	82
5.2 SARAN .....	83
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 2. 1 Data Kanker 2020 (Globocan, 2020) .....	12
Gambar 2. 2 Contoh <i>Random forest</i> .....	20
Gambar 2. 3 Contoh <i>ROC</i> dan <i>AUC</i> .....	22
Gambar 2. 4 Contoh K-Fold .....	25
Gambar 2. 5 Confusion Mmatrix .....	26
Gambar 3. 1 Tahapan Penelitian .....	29
Gambar 3. 2 <i>Citation Request</i> .....	31
Gambar 3. 3 <i>Citation Policy</i> .....	31
Gambar 3. 4 <i>Data Cleaning</i> .....	35
Gambar 3. 5 Transformasi Data .....	36
Gambar 3. 6 <i>Correlation plot</i> .....	37
Gambar 3. 7 Contoh Variabel Penting <i>Random forest</i> .....	43
Gambar 3. 8 <i>Decision tree</i> Yang Terbentuk .....	53
Gambar 4. 1 Perbandingan Diagnosis .....	70
Gambar 4. 2 Hasil Penelitian Jackins dkk.....	71
Gambar 4. 3 Hasil Confusion Matrix Peneliti Vinutha dkk.....	72
Gambar 4. 4 Hasil Presisi-Recall Peneliti Vinutha dkk .....	73
Gambar 4. 5 Hasil <i>Random forest</i> Bhardwaj dkk .....	73
Gambar 4. 6 Hasil Penelitian Bhardwaj dkk.....	74
Gambar 4. 7 Diagram Nilai Akurasi Tiap Model .....	76

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait .....	9
Tabel 3. 1 Atribut <i>Dataset</i> .....	32
Tabel 3. 2 Contoh <i>Dataset</i> .....	34
Tabel 3. 3 Contoh Data Awal.....	38
Tabel 3. 4 Contoh Hasil Normalisasi <i>Z-Score</i> .....	40
Tabel 3. 5 Data <i>Train</i> .....	45
Tabel 3. 6 Data <i>Testing</i> .....	46
Tabel 3. 7 Dataset Perhitungan Manual RF .....	46
Tabel 3. 8 <i>Boostrapped Dataset</i> Pertama.....	46
Tabel 3. 9 <i>Subset</i> 1 DT1 .....	47
Tabel 3. 10 <i>Subset</i> 2 DT1 .....	48
Tabel 3. 11 <i>Subset</i> 3 DT1 .....	48
Tabel 3. 12 <i>Subset</i> 4 DT1 .....	48
Tabel 3. 13 <i>Boostrapped Dataset</i> Kedua .....	48
Tabel 3. 14 <i>Subset</i> 1 DT2.....	49
Tabel 3. 15 <i>Subset</i> 2 DT2.....	49
Tabel 3. 16 <i>Subset</i> 3 DT2.....	49
Tabel 3. 17 <i>Subset</i> 4 DT2.....	49
Tabel 3. 18 <i>Boostrapped Dataset</i> Ketiga .....	50
Tabel 3. 19 <i>Subset</i> 1 DT3 .....	50
Tabel 3. 20 <i>Subset</i> 2 DT3 .....	50
Tabel 3. 21 <i>Subset</i> 3 DT3 .....	51
Tabel 3. 22 <i>Subset</i> 4 DT4.....	51
Tabel 3. 23 <i>Boostrapped Dataset</i> Keempat .....	51
Tabel 3. 24 <i>Subset</i> 1 DT4.....	52
Tabel 3. 25 <i>Subset</i> 2 DT4.....	52
Tabel 4. 1 Hasil Prediksi <i>Default</i> Model A.....	55
Tabel 4. 2 Confusion Matrix <i>Default</i> Model A.....	55
Tabel 4. 3 Hasil Prediksi <i>Z-Score</i> Model A.....	56
Tabel 4. 4 Confusion Matrix <i>Z-Score Tuning</i> Model A.....	56
Tabel 4. 5 Hasil Prediksi <i>Default</i> Model B .....	57
Tabel 4. 6 Confusion Matrix <i>Default</i> Model B.....	57
Tabel 4. 7 Hasil Prediksi <i>Z-Score</i> Model B .....	58
Tabel 4. 8 Confusion Matrix <i>Z-Score Tuning</i> Model B.....	59
Tabel 4. 9 Hasil Prediksi <i>Default</i> Model C.....	60
Tabel 4. 10 Confusion Matrix <i>Default</i> Model C.....	60
Tabel 4. 11 Hasil Prediksi <i>Z-Score</i> Model C .....	61
Tabel 4. 12 Confusion Matrix <i>Z-Score Tuning</i> Model C.....	61
Tabel 4. 13 Hasil Prediksi <i>Default</i> Model D.....	62
Tabel 4. 14 Confusion Matrix <i>Default</i> Model D.....	63
Tabel 4. 15 Hasil Prediksi <i>Z-Score</i> Model A.....	63
Tabel 4. 16 Confusion Matrix <i>Z-Score Tuning</i> Model D.....	64
Tabel 4. 17 Cross Validation dan ROC Model A .....	65

Tabel 4. 18 Cross Validation dan ROC Model B.....	67
Tabel 4. 19 Cross Validation dan ROC Model C.....	67
Tabel 4. 20 Cross Validation dan ROC Model D .....	68
Tabel 4. 21 Hasil Penelitian Andryan dkk .....	71
Tabel 4. 22 Hasil Penelitian Adebiyi dkk .....	72
Tabel 4. 23 Akurasi Setiap Model.....	75
Tabel 4. 24 Hasil Perbandingan Akurasi.....	76

## ABSTRAK

Mufiddin, Rifqi. 2023. **Klasifikasi Kanker Payudara Menggunakan Metode *Random forest***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Suhartono, M.Kom (II) Agung Teguh Wibowo Almais, M.T.

Kata kunci: *Klasifikasi Kanker Payudara, Random forest, Normalisasi*.

Kanker payudara adalah salah satu jenis kanker yang paling umum di kalangan wanita di seluruh dunia. Diagnosis dini dan pengobatan yang tepat sangat penting dalam meningkatkan tingkat kelangsungan hidup pasien. Banyak orang yang terlambat menyadari terkena kanker payudara dan baru menyadarinya disaat sudah tingkat stadium III dan IV. Penelitian ini memanfaatkan salah satu algoritma *supervised learning* yaitu *random forest* pada paket *caret* yang tersedia pada bahasa pemrograman R. Tujuan penelitian ini untuk mengetahui performa *random forest* dalam mengklasifikasi kanker payudara berdasarkan *Breast Cancer Wisconsin (Diagnostic) Dataset* dan menghasilkan model yang dapat mengetahui apakah seseorang mengidap kanker payudara jinak atau ganas. Data diolah menggunakan teknik *preprocessing* dengan 2 tahapan yaitu data *cleaning* dan data *exploration*, serta dilakukan pembagian data menjadi 4 bagian, yaitu model A dengan perbandingan 90% data *train* : 10% data *test*, model B dengan perbandingan 80% data *train* : 20% data *test*, model C dengan perbandingan 75% data *train* : 25% data *test*, dan model D dengan perbandingan 70% data *train* : 30% data *test*. Pada penelitian ini, terdapat pengujian dengan menggunakan *random forest* secara *default* dengan data yang tidak dinormalisasi dan *random forest* yang telah di-*tuning* memakai teknik *grid search* dengan data yang telah dinormalisasi. Didapatkan hasil terbaik pada model C yang telah dinormalisasi dan di *tuning* dengan perbandingan data train 75% dan 25 % data uji menghasilkan nilai akurasi tertinggi sebesar 98.59% yang dikategorikan sangat baik. Serta penggunaan *repeated 10-fold cross validation* dengan 3 kali pengulangan pada masing-masing model yang di-*tuning* didapatkan jumlah *mtry* terbaik yaitu 2.



## ABSTRACT

Mufiddin, Rifqi. 2023. **Classification of Breast Cancer Using *Random forest* Method.** Undergraduate Thesis. Department of Informatics Engineering Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Prof. Dr. Suhartono, M.Kom (II) Agung Teguh Wibowo Almais, M.T.

Keywords: *Classification Of Breast Cancer, Random forest, Data Scaling.*

Breast cancer is one of the most common types of cancer among women worldwide. Early diagnosis and appropriate treatment are very important in increasing the patient's survival rate. Many people are late in realizing they have breast cancer and only realize it when they are at stages III and IV. This study utilizes one of the supervised learning algorithms, namely random forest in the caret package available in the R programming language. The purpose of this study is to determine the performance of a random forest in classifying breast cancer based on the Wisconsin Breast Cancer Dataset (Diagnostic) and to produce a model that can determine whether a person has benign or malignant breast cancer. The data is processed using preprocessing techniques with 2 stages, namely data cleaning and data exploration, and the data is divided into 4 parts, namely model A with a ratio of 90% data train: 10% data test, model B with a comparison of 80% data train: 20% data test, model C with a ratio of 75% data train: 25% test data, and model D with a ratio of 70% data train: 30% test data. In this study, there is a test using a random forest by default with non-normalized data and a random forest that has been tuned using a grid search technique with normalized data. The best results were obtained in the normalized and tuned C model with a ratio of 75% train data and 25% test data resulting in the highest accuracy value of 98.59% which is categorized as very good. As well as using repeated 10-fold cross-validation with 3 repetitions for each tuned model, the best number of mtry is 2.

## خالصة

مفدين ، رقي. 2023. تصنيف سرطان الثدي بطريقة الغابة العشوائية. أطروحة جامعية. قسم هندسة المعلوماتية بكلية العلوم والتكنولوجيا مولانا مالك إبراهيم الدولة الإسلامية جامعة مالانج. المشرف: (الأول) الأستاذ الدكتور سوهارتونو كوم (الثاني) أجونج تيجوه ويوو أليس ، م

تصنيف سرطان الثدي ، غابة عشوائية ، تحجيم البيانات : الكلمات الرئيسية

سرطان الثدي أحد من أنواع السرطانات شيوعًا بين النساء في جميع العالم. التشخيص المبكر والعلاج المناسب مهمان زيادة معدل بقاء حياة المريض. كثير من الناس يتأخرون في إدراك أنهم مصابون بسرطان الثدي وأدركت عندما كانت في المرحلة الثالثة والرابعة. يستخدم هذا البحث إحدى خوارزميات التعلم الخاضعة يعني الغابة العشوائية في حزمة الإقحام المتوفرة بلغة البرمجة R. هذا البحث يهدف لتصنيف سرطان الثدي من *Breast Cancer Wisconsin (Diagnostic) Dataset*. وتحتصل النموذج لمعرفة ما إذا كان الشخص مصابًا بسرطان الثدي الحميد أو الخبيث. البيانات المعالجة باستخدام *preprocessing* بمرحلتين، هما تنظيف البيانات واستكشاف البيانات وتقسيم البيانات إلى 4 أجزاء، النموذج أ بالمقارنة ٩٠٪ مجموعة البيانات ١٠٪ بيانات الاختبار، النموذج ب بالمقارنة ٨٠٪ مجموعة البيانات ٢٠٪ بيانات الاختبار، النموذج ج بالمقارنة ٧٥٪ مجموعة البيانات ٢٥٪ بيانات الاختبار، النموذج د بالمقارنة ٧٠٪ مجموعة البيانات ٣٠٪ بيانات الاختبار. في هذا البحث، يوجد اختبار يستخدم غابة عشوائية افتراضيًا مع بيانات غير طبيعية وغابة عشوائية مضبوطة باستخدام تقنية *grid search* مع البيانات الطبيعية. نتائج هذا البحث تحصل أفضل النموذج يعني النموذج "ج" مع البيانات الطبيعية بالمقارنة ٧٥٪ مجموعة البيانات ٢٥٪ بيانات الاختبار ينتج أعلى قيمة دقة تبلغ ٩٨،٥٩٪ مع معايير جيد جدا. واستخدام *repeated 10-fold cross validation* مع ٣ تكريرات لكل نموذج مضبوط، فإن أفضل عدد من المرات هو 2.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kanker adalah suatu penyakit yang dapat menyerang pada sebagian besar organ atau jaringan tubuh (Resmiati & Arifin, 2021). Kerusakan dan pertumbuhan abnormal pada gen yang mengatur regenerasi sel dalam tubuh manusia merupakan gejala dari kondisi yang dikenal sebagai kanker (Sari, 2019). Kerusakan siklus sel dapat membuat sel membelah melebihi batas wajar yang berdampak pada rusaknya jaringan biologis disekitarnya dan menyebar ke jaringan tubuh lainnya melalui sirkulasi darah (Ma'arif & Arifin, 2017).

Kanker yang paling umum pada wanita dan penyebab utama kematian akibat kanker pada wanita di seluruh dunia adalah kanker payudara (A. Vincent, 2022). Kanker payudara yaitu penyakit yang berawal dari pembelahan sel pada jaringan payudara yang tidak terkendali dan menghasilkan benjolan atau massa (Shidqi et al., 2022). Kanker payudara juga merupakan salah satu kanker yang paling banyak terjadi di Indonesia. Baik pria maupun wanita dapat mengidap kanker payudara. Namun, wanita secara signifikan lebih mungkin mengidap penyakit ini daripada pria. Organisasi Kesehatan Dunia (WHO) memperkirakan bahwa 8-9% wanita akan terkena kanker payudara. Akibatnya, kanker payudara menjadi jenis kanker paling umum dikalangan wanita secara global (Husni et al., 2015).

Data *mining* merupakan suatu metode yang menjajikan dalam kegiatan prediksi diagnostik. Metode ini terdiri dari sejumlah alat dan teknik yang dapat

digunakan untuk mendalami kumpulan data dan membantu dalam menemukan suatu wawasan baru. (Moreira & Namen, 2018). *Data mining* adalah suatu teknik populer dalam melakukan analisis di bidang pemasaran, medis, keuangan, dan *social science*. Contohnya yaitu implementasi *machine learning* untuk mengamati perilaku tumor pada pengidap kanker payudara (Mohammed et al., 2020). *Machine learning* adalah cabang kecerdasan buatan yang berkonsentrasi pada penggunaan algoritma dan teknik tertentu untuk klasifikasi, pengenalan pola, dan prediksi (Muntiari & Hanif, 2022). *Machine learning* merupakan salah satu aspek paling penting dari kecerdasan buatan, memungkinkan terciptanya sistem komputer yang dapat belajar dari pengalaman dan riset yang telah dilakukan tanpa memerlukan pemrograman dalam setiap kasus. *Machine learning* dianggap sebagai komponen penting dari kemajuan teknologi karena mendukung otomatisasi sementara memiliki kelemahan yang minim (Aprilia et al., 2021). Teknik ini memungkinkan untuk meninjau data medis dengan cepat dan menyeluruh, juga menghindari kesalahan diagnosis yang dilakukan oleh tenaga medis (F. S. Nugraha et al., 2019).

Algoritma *Random forest* merupakan metode pengembangan dari algoritma *Decision tree*. *Random forest* terdiri dari sekumpulan *Decision tree* yang telah dilakukan *training* menggunakan sampel yang berbeda, dan setiap atribut dibagi menjadi *tree*/pohon yang dipilih dari *subset* atribut secara acak (Fauzi et al., 2020). Salah satu keunggulan *Random forest* yaitu setiap pengklasifikasi pohon seperti anggota tim dan semua anggota bekerja sama untuk mendapatkan prediksi akhir. *Random forest* cocok untuk klasifikasi biner, memiliki kemampuan untuk mengatasi kumpulan data dimana jumlah variabel melebihi jumlah pengamatan,

memiliki kemampuan untuk menangani kumpulan data dengan campuran prediktor kontinyu dan kategorikal, efisiensi tinggi dan akurasi yang sangat baik (Macaulay et al., 2021). *Random forest* merupakan salah satu algoritma *supervised learning* dalam *machine learning* (Sowah et al., 2020). *Machine learning* membutuhkan suatu algoritma agar bisa digunakan. *Supervised learning* cocok digunakan dalam mengatasi masalah klasifikasi dan regresi (A. Vincent, 2022).

Klasifikasi merupakan cara untuk mengelompokkan data dengan karakteristik yang serupa ke dalam kelas yang sudah ditentukan terlebih dahulu. Label kelas pada klasifikasi berguna untuk menamai kelompok yang memiliki pola serupa (Muntiari & Hanif, 2022). Klasifikasi pada kanker payudara digunakan untuk menentukan apakah seorang pasien mengidap kanker jinak atau ganas. Kanker jinak mengacu pada tumor yang tidak menyebar ke bagian tubuh lain dari tempat asalnya. Kanker jinak tidak menyebar ke struktur terdekat atau ke daerah lain di tubuh yang jauh. Kanker ganas memiliki pertumbuhan sel-sel di luar kendali dan menyebar secara lokal dan/atau ke tempat yang jauh (Patel, 2020). Kanker jinak memiliki bentuk bulat atau lonjong dan tepi yang halus dan tegas. Sedangkan, kanker ganas digambarkan sebagai benjolan besar, kasar, dan tidak jelas (Punitha et al., 2018).

Dalam Al Qur'an terdapat ayat yang mengajarkan tentang harus berusaha untuk meraih keberhasilan di akhirat dengan memanfaatkan ilmu yang diberikan Allah, tetapi jangan melupakan tanggung jawab masing-masing di dunia. Masing-masing orang juga harus berusaha untuk berbuat baik dan memberikan manfaat bagi

orang lain, serta tidak merusak atau menimbulkan kerugian di bumi yakni surah Al-Qashash ayat 77 yang berbunyi sebagai berikut.

وَأَتَّبِعْ فِي مَآءِثِكَ اللَّهُ الدَّارَ الْآخِرَةَ وَلَا تَنْسَ نَصِيبَكَ مِنَ الدُّنْيَا وَأَحْسِنَ كَمَا أَحْسَنَ اللَّهُ إِلَيْكَ وَلَا تَبْغِ الْفُسَادَ فِي الْأَرْضِ إِنَّ اللَّهَ لَا يُحِبُّ الْمُفْسِدِينَ

“Dan carilah dengan apa yang telah diberikan Allah kepadamu negeri akhirat, dan janganlah kamu lupakan bahagianmu dari dunia dan berbuat baiklah (kepada orang lain) sebagaimana Allah telah berbuat baik kepadamu, dan janganlah kamu berbuat kerusakan di bumi. Sesungguhnya Allah tidak menyukai orang-orang yang berbuat kerusakan.” (QS. Al-Qashash: 77).

Berdasarkan pemaparan di atas, penelitian ini akan mengimplementasikan metode *Random forest* untuk mengklasifikasi kanker payudara. Untuk penggunaan data kanker payudara diperoleh dari *UCI Machine Learning Repository*. Sehingga, penelitian ini berjudul “**Klasifikasi Kanker Payudara Menggunakan Metode *Random forest***”.

## 1.2 Pernyataan Masalah

Bagaimana performa metode *Random Forest* dalam mengklasifikasi kanker payudara berdasarkan data *Breast cancer Wisconsin (Diagnostic) Dataset*?

## 1.3 Tujuan Penelitian

Untuk mengetahui performa metode *Random Forest* dalam mengklasifikasi kanker payudara berdasarkan data *Breast cancer Wisconsin (Diagnostic) Dataset*?

#### **1.4 Batasan Masalah**

1. Menggunakan algoritma *Random forest*
2. Memanfaatkan data public yaitu data *Breast cancer Wisconsin (Diagnostic)*
3. Klasifikasi yang dilakukan hanya untuk penyakit kanker payudara

#### **1.5 Manfaat Penelitian**

1. Membantu para peneliti atau pihak akademis yang ingin mengembangkan topik dari penelitian ini.
2. Memberikan wawasan kepada pembaca tentang penggunaan metode *Random forest*.
3. Memberikan wawasan kepada pembaca tentang kanker payudara

## BAB II

### STUDI PUSTAKA

#### 2.1 Penelitian Terkait

Penelitian yang dilakukan oleh Apriliah et al., mengimplementasikan metode *Random forest* dalam memprediksi diabetes dengan menggunakan *dataset* yang didapat dari *UCI Machine learning Repository* yaitu *dataset* diabetes *Hospital in Sylhet, Bangladesh*. *Dataset* ini berisikan 520 data, 17 atribut dan 1 kelas. Peneliti menggunakan *tools* WEKA. Pada penelitian ini membandingkan 3 algoritma klasifikasi yakni *Naïve Bayes*, *SVM*, dan juga *Random forest*. Peneliti menggunakan teknik *10-fold-cross-validation*. Hasil perbandingan menunjukkan bahwa metode *Random forest* merupakan metode klasifikasi terbaik yang dapat mengklasifikasi apakah pasien dinyatakan terkena diabetes atau tidak dengan nilai akurasi dan *ROC* sebesar 97.88% dan 0,998 (Apriliah et al., 2021). Adapun penelitian yang dilakukan oleh Faisal et. al., dalam mengklasifikasi pasien diabetes retinopati dengan menggunakan metode *SVM* menghasilkan *sensitivity*, *specificity*, *positive predictive value*, *negative predictive value*, dan *AUC* sebesar 96,9%, 100%, 100%, 88.19, 0.985% (Faisal et al., 2014).

Pada penelitian Andryan et al., mengimplementasikan metode *XGBoost* dan *SVM* dalam mendiagnosa kanker payudara dengan menggunakan *dataset* yang didapat dari *UCI Machine Learning dataset Wisconsin Breast Cancer Diagnostic*. *Dataset* ini berisikan 569 data, 32 atribut dan 2 kelas. Peneliti menggunakan *tools* *Python*. Pada penelitian ini membandingkan 2 metode yakni *XGBoost* dan *SVM*. Peneliti menggunakan teknik *PCA*, *tuning* dengan *GridsearchCV*, dan membagi



data sebesar 80 untuk data latih dan 20 untuk data uji. Hasil akurasi yang didapat dengan metode *SVM* sebesar 94.28% dan *XGBoost* sebesar 96.26% (Andryan et al., 2022).

Pada penelitian Jackins et al., mengimplementasikan metode *Random forest* dalam memprediksi beberapa penyakit klinis yakni diabetes, penyakit jantung koroner, dan kanker payudara. Masing-masing *dataset* yang digunakan yaitu NIDDK untuk diabetes, *Framingham heart study* untuk penyakit jantung koroner, dan *breast cancer* Wisconsin untuk kanker payudara. Peneliti menggunakan *tools Anaconda*. Hasil yang didapatkan dari penelitian menunjukkan bahwa metode *Random forest* merupakan algoritma klasifikasi yang lebih baik dibandingkan metode *Bayesian* dengan nilai akurasi sebesar 74.03% untuk diabetes, 83.85% untuk penyakit jantung koroner, dan 92.40% untuk kanker payudara (Jackins et al., 2021).

Pada penelitian Macaulay et al., mengimplementasikan metode *Random forest* dalam memprediksi resiko kanker payudara pada perempuan di Afrika. Data yang digunakan terdiri dari partisipan pasien yang menghadiri departemen bedah *Lagos State University Teaching Hospital*, Ikeja Lagos Nigeria pada tahun 2009-2019. Total partisipan sebanyak 180 orang dimana 90 orang mengidap kanker ganas dan 90 orang lainnya mengidap kanker jinak. Hasil penelitian menunjukkan bahwa metode *Random forest* merupakan metode klasifikasi yang lebih baik dibandingkan metode *SVM* dan *Gail* dengan akurasi sebesar 91.67%, *sensitivity* 87.10%, *specificity* 96.55%, dan *AUC* 92% (Macaulay et al., 2021).

Pada penelitian Li et al., mengimplementasikan metode *Random forest* dalam memprediksi hipotensi setelah induksi anestesi untuk operasi jantung. Data yang digunakan didapatkan dari catatan kesehatan Rumah Sakit *Hainan Medical University*. Partisipan yang terlibat merupakan pasien berumur lebih dari 18 tahun yang telah menjalani operasi jantung dalam jangka waktu Desember 2007 hingga Januari 2018. Total partisipan sebanyak 3030 orang dimana 1578 memiliki pengalaman hipotensi setelah induksi anestesi. Peneliti menggunakan *Python* sebagai alat untuk menganalisis data. Hasil penelitian menunjukkan bahwa metode *Random forest* bekerja secara efektif dalam mengklasifikasi hipotensi, dengan nilai *AUC* sebesar 0.843 (0.808-0.877) (Li et al., 2021).

Pada penelitian Wang et al., mengimplementasikan metode *Random forest* dalam memprediksi Makrosomia. Data yang digunakan didapatkan dari *Shandong Multi-Center Healthcare Big Data Platform*. Partisipan yang terlibat merupakan ibu hamil yang baru melahirkan di kota Jinan dari bulan Juni 2017 hingga Mei 2018. Total partisipan sebanyak 4260 bayi baru lahir dimana 405 mengidap makrosomia dan 3855 lahir dengan berat badan normal. Peneliti menggunakan bahasa pemrograman R. Hasil penelitian menunjukkan bahwa metode *Random forest* merupakan metode klasifikasi yang lebih baik dibandingkan *Logistic Regression* dengan nilai *sensitivity* 91.7%, *specificity* 91.7%, dan *AUC* sebesar 95.3% (Wang et al., 2022).

Pada penelitian Mitra & Rajendran, mengimplementasikan metode *Random forest* dalam memprediksi pasien stroke dengan *Dataset* yang didapatkan dari website *Kaggle* yaitu *Stroke Prediction Dataset*. *Dataset* ini berisikan 5110

data pasien dan memiliki 11 fitur. Peneliti menggunakan *tools JupyterLab*, bahasa pemrograman *Python*, dan *SPSS* untuk menganalisis statistik *machine learning* dalam kualitas penelitian. Hasil penelitian menunjukkan bahwa metode *Random forest* merupakan metode klasifikasi yang lebih baik dibandingkan *SVM* dengan nilai akurasi sebesar 94.61% (Mitra & Rajendran, 2022).

Tabel 2. 1 Penelitian Terkait

No	Referensi	Metode Penelitian	Hasil Penelitian	Perbedaan
1	Aprilia et al., (2021)	<i>Support Vector Machine, Naïve Bayes, dan Random forest</i>	Hasil perbandingan menunjukkan bahwa metode <i>Random forest</i> merupakan metode klasifikasi terbaik yang dapat mengklasifikasi apakah pasien dinyatakan terkena diabetes atau tidak dengan nilai akurasi dan <i>ROC</i> sebesar 97.88% dan 0.998	Data yang digunakan berbeda yaitu <i>dataset diabetes Hospital in Sylhet, Bangladesh</i> akan tetapi salah satu metode yang digunakan sama yaitu <i>Random forest</i> . Peneliti menggunakan <i>Breast Cancer Wisconsin (Diagnostic) Dataset</i>
2	Andryan et al., (2022)	<i>XGBoost dan Support Vector Machine</i>	Hasil akurasi yang didapat dengan metode <i>SVM</i> sebesar 94.28% dan <i>XGBoost</i> sebesar 96.26%.	Dataset yang digunakan sama, akan tetapi metode dan teknik pengolahan data yang digunakan berbeda. Pada jurnal menggunakan <i>PCA</i> dan evaluasi menggunakan <i>ROC</i> . Sedangkan, Peneliti tidak menggunakan <i>PCA</i> , menggunakan metode <i>Random forest</i> dan evaluasi performa model menggunakan <i>ROC, K-Fold, dan Confusion Matrix</i> .
3	Jackins et al., (2021)	<i>Random forest dan Naïve Bayes</i>	Hasil yang didapatkan dari penelitian menunjukkan bahwa metode <i>Random forest</i> merupakan algoritma klasifikasi yang lebih baik dibandingkan metode <i>Bayesian</i> dengan nilai	Salah satu dataset dan metode yang digunakan sama. Perbedaan terletak pada pengolahan dataset. Pada Jurnal menggunakan teknik <i>DBSCAN</i> . Sedangkan, peneliti tidak menggunakan teknik

No	Referensi	Metode Penelitian	Hasil Penelitian	Perbedaan
			akurasi sebesar 74.03% untuk diabetes, 83.85% untuk penyakit jantung koroner, dan 92.40% untuk kanker payudara.	<i>DBSCAN</i> , hanya menggunakan <i>tuning hyperparameter</i> dan membagi dataset menjadi 4 bagian.
4	Macaulay et al., (2021)	<i>Random forest, SVM, dan Gail Model</i>	Hasil penelitian menunjukkan bahwa metode <i>Random forest</i> merupakan metode klasifikasi yang lebih baik dibandingkan metode <i>SVM</i> dan <i>Gail</i> dengan akurasi sebesar 91.67%, <i>sensitivity</i> 87.10%, <i>specificity</i> 96.55%, dan <i>AUC</i> 92%.	Data kanker payudara yang digunakan berbeda. Pada jurnal menggunakan data partisipan pasien yang menghadiri departemen bedah <i>Lagos State University Teaching Hospital</i> . Sedangkan peneliti menggunakan dataset <i>Breast Cancer Wisconsin (Diagnostic) Dataset</i> .
5	Li et al., (2021)	<i>Random forest</i>	Hasil penelitian menunjukkan bahwa metode <i>Random forest</i> bekerja secara efektif dalam mengklasifikasi hipotensi, dengan nilai <i>AUC</i> sebesar 0.843 (0.808-0.877).	Data yang digunakan berbeda yaitu data dari rumah sakit Hainan dan penyakit yang diteliti yaitu hipotensi. Peneliti menggunakan <i>Breast Cancer Wisconsin (Diagnostic) Dataset</i>
6	Wang et al., (2022)	<i>Random forest dan Logistic Regression</i>	Hasil penelitian menunjukkan bahwa metode <i>Random forest</i> merupakan metode klasifikasi yang lebih baik dibandingkan <i>Logistic Regression</i> dengan nilai <i>sensitivity</i> 91.7%, <i>specificity</i> 91.7%, dan <i>AUC</i> sebesar 95.3%.	Data yang digunakan berbeda yaitu data dari <i>Shandong Multi-Center Healthcare Big Data Platform</i> dan penyakit yang diteliti yaitu Makrosomia. Peneliti menggunakan <i>Breast Cancer Wisconsin (Diagnostic) Dataset</i>
7	Mitra & Rajendran, (2022)	<i>Random forest dan Support Vector Machine</i>	Hasil penelitian menunjukkan bahwa metode <i>Random forest</i> merupakan metode klasifikasi yang lebih baik dibandingkan <i>SVM</i> dengan nilai akurasi sebesar 94.61%.	Data yang digunakan berbeda yaitu <i>Stroke prediction dataset</i> dan penyakit yang diteliti yaitu <i>Stroke</i> . Peneliti menggunakan <i>Breast Cancer Wisconsin (Diagnostic) Dataset</i>

Dari beberapa penelitian yang telah dipaparkan, penggunaan metode *Random forest* dalam mengklasifikasi penyakit menghasilkan akurasi yang baik. Oleh sebab itu, penelitian ini ditulis untuk mengetahui performa *random forest* dalam mengklasifikasi penyakit kanker payudara, bahasa pemrograman R dan juga *dataset* yang didapat dari *UCI Machine Learning Repository* <http://archive.ics.uci.edu/ml> yaitu *Breast cancer Wisconsin (Diagnostic)*.

## 2.2 Kanker

Kanker merupakan suatu penyakit yang diakibatkan oleh pertumbuhan sel abnormal yang tumbuh dan berkembang pada tubuh dapat mengakibatkan kematian (Rahayuwati et al., 2020). Kanker merupakan penyakit tidak menular yang mempengaruhi kesehatan masyarakat. Dibandingkan pria, wanita memiliki peluang 100 kali lebih besar terkena kanker payudara (Azmi et al., 2020). Faktor genetik, faktor karsinogenik (seperti bahan kimia, radiasi, virus, hormon, dan iritasi yang menetap), dan variabel perilaku atau gaya hidup adalah penyebab sebenarnya dari kanker (merokok, pola makan yang tidak sehat, alkohol, dan kurangnya aktivitas fisik). Namun, perilaku dan makanan memiliki peluang lebih besar dari 30% menyebabkan kematian akibat kanker. Indeks massa tubuh yang tinggi, konsumsi buah dan sayuran yang tidak mencukupi, tidak aktif, merokok, dan penggunaan alkohol adalah beberapa di antaranya. Terutama unsur merokok, yang menyumbang hampir 70% dari semua penyebab kematian yang dapat dicegah di seluruh dunia (Rahayuwati et al., 2020). Berikut merupakan data kanker di Indonesia pada tahun 2020 (Globocan, 2020).

Indonesia Source: Globocan										
Incidence, Mortality and Prevalence by cancer site										
Cancer	New cases				Deaths				5-year prevalence (all ages)	
	Number	Rank	(%)	Cum.risk	Number	Rank	(%)	Cum.risk	Number	Prop. (per 100 000)
Breast	65 858	1	16.6	4.83	22 430	2	9.6	1.78	201 143	148.11
Cervix uteri	36 633	2	9.2	2.69	21 003	3	9.0	1.73	92 930	68.43
Lung	34 783	3	8.8	1.54	30 843	1	13.2	1.39	37 663	13.77
Liver	21 392	4	5.4	0.92	20 920	4	8.9	0.91	22 530	8.24
Nasopharynx	19 943	5	5.0	0.75	13 399	5	5.7	0.56	54 670	19.99
Colon	17 368	6	4.4	0.74	9 444	8	4.0	0.38	39 217	14.34
Non-Hodgkin lymphoma	16 125	7	4.1	0.66	9 024	9	3.8	0.38	43 175	15.78
Rectum	16 059	8	4.0	0.68	8 342	10	3.6	0.35	39 268	14.36
Leukaemia	14 979	9	3.8	0.52	11 530	6	4.9	0.42	41 701	15.25
Ovary	14 896	10	3.8	1.09	9 581	7	4.1	0.77	37 533	27.64
Prostate	13 563	11	3.4	1.47	4 863	13	2.1	0.38	42 772	31.06
Thyroid	13 114	12	3.3	0.49	2 224	19	0.95	0.07	38 650	14.13
Bladder	7 828	13	2.0	0.37	3 885	14	1.7	0.18	20 053	7.33
Corpus uteri	7 773	14	2.0	0.60	2 626	18	1.1	0.20	22 087	16.26
Brain, central nervous system	5 964	15	1.5	0.22	5 298	12	2.3	0.20	15 310	5.60
Pancreas	5 781	16	1.5	0.24	5 690	11	2.4	0.24	4 913	1.80
Lip, oral cavity	5 780	17	1.5	0.24	3 087	15	1.3	0.13	14 197	5.19
Larynx	3 663	18	0.92	0.16	2 146	20	0.92	0.09	9 346	3.42
Stomach	3 484	19	0.88	0.15	2 946	16	1.3	0.12	4 661	1.70
Multiple myeloma	3 151	20	0.79	0.15	2 743	17	1.2	0.13	7 319	2.68
Salivary glands	2 666	21	0.67	0.12	1 229	23	0.52	0.06	7 517	2.75
Kidney	2 394	22	0.60	0.10	1 358	21	0.58	0.06	6 144	2.25
Melanoma of skin	1 609	23	0.41	0.07	699	25	0.30	0.03	4 331	1.58
Testis	1 497	24	0.38	0.09	430	28	0.18	0.03	5 214	3.79
Oropharynx	1 482	25	0.37	0.06	848	24	0.36	0.04	3 490	1.28
Vulva	1 328	26	0.33	0.11	442	27	0.19	0.04	3 962	2.92
Oesophagus	1 327	27	0.33	0.06	1 283	22	0.55	0.06	1 422	0.52
Hodgkin lymphoma	1 188	28	0.30	0.05	363	30	0.15	0.02	3 659	1.34
Penis	1 017	29	0.26	0.09	347	31	0.15	0.03	2 954	2.14
Anus	762	30	0.19	0.03	366	29	0.16	0.02	1 868	0.68
Gallbladder	657	31	0.17	0.03	461	26	0.20	0.02	828	0.30
Vagina	473	32	0.12	0.04	218	32	0.09	0.02	1 196	0.88
Hypopharynx	264	33	0.07	0.02	152	34	0.06	0.01	408	0.15
Mesothelioma	231	34	0.06	0.01	208	33	0.09	0.01	274	0.10
Kaposi sarcoma	108	35	0.03	0.01	60	35	0.03	0.00	284	0.10
All cancer sites	396 914	-	-	14.87	234 511	-	-	9.38	946 088	345.9

Gambar 2. 1 Data Kanker 2020 (Globocan, 2020)

### 2.3 Kanker Payudara

Kanker payudara atau *Ca Mammar* adalah suatu kondisi di mana sel-sel kehilangan kendali atas mekanisme normalnya, yang mengakibatkan sel-sel tumbuh secara tidak normal dengan cepat dan tidak terkendali (Nurhayati et al., 2019). Terjadi ketika sel-sel berkembang secara tidak normal di lobulus, yaitu jaringan kelenjar yang menghasilkan susu, atau di saluran, yang mengalirkan susu dari kelenjar ke puting susu. Selain itu, jaringan ikat atau lemak di payudara dapat menjadi tempat bagi kanker untuk berkembang (Edriss et al., 2016). Kanker payudara umumnya menyerang wanita. Akan tetapi, tidak menutup kemungkinan

dapat menyerang pria dengan perbandingan 1 dari 1000 (Saputra Liambo et al., 2022).

Meski tumor tidak selalu kanker, masih banyak orang yang percaya bahwa kanker dan tumor adalah hal yang sama. Untuk menentukan jenis kanker payudara dapat digunakan biopsi dan mamografi untuk pemeriksaan. Apakah suatu sel kanker payudara bersifat ganas atau jinak dapat diketahui dari hasil pemeriksaan biopsi dengan menggunakan *Fine Needle Aspiration (FNA)* (Chazar & Erawan, 2020). *FNA* merupakan proses pengambilan sel atau cairan dengan menggunakan jarum ke dalam massa dan diaspirasi, dan massa yang didapatkan dimanfaatkan untuk didiagnosis secara sitologi berdasarkan temuan sitomorfologi (Kamelia & Agus, 2021). Kanker payudara yang berkembang dari tumor dibagi menjadi beberapa fase, mulai dari stadium 0 hingga stadium IV. Banyak pasien kanker payudara yang baru mengetahui kondisinya setelah stadium lanjut ke tingkat yang tinggi akibat keterlambatan dalam mengidentifikasi gejala (rata-rata stadium III dan IV). Peluang kematian sangat besar dalam kondisi seperti ini (Chazar & Erawan, 2020).

Faktor risiko kanker payudara dibagi menjadi faktor risiko yang *valid* dan faktor potensial. Geografi, usia, riwayat keluarga, riwayat menstruasi, kehamilan, dan penyakit payudara jinak adalah beberapa faktor yang valid. Estrogen eksogen, obesitas, kontrasepsi oral, konsumsi alkohol, diet tinggi lemak, dan merokok merupakan faktor potensial (Hero, 2020). Berdasarkan gambar 2.1 penyakit kanker payudara merupakan penyakit kanker dengan jumlah kasus terbanyak di Indonesia

dan menduduki peringkat ke 2 dalam kasus kematian yang disebabkan oleh kanker (Globocan, 2020).

#### **2.4 Data Mining**

Data *Mining* adalah proses untuk mengidentifikasi dan mengekstraksi informasi yang berguna dan pengetahuan terkait bersumber dari basis data besar menggunakan pendekatan matematika, statistik, kecerdasan buatan, dan pembelajaran mesin. Data *mining* merupakan sekumpulan prosedur yang digunakan untuk menemukan nilai tambah dari sumber data berupa pengetahuan yang sebelumnya tidak diketahui (Yuli Mardi, 2019).

Teknik dasar yang digunakan dalam data *mining* memungkinkan seseorang untuk mengekstraksi pengetahuan dan wawasan penting dari sejumlah besar data. Ini adalah bidang interdisipliner yang menggabungkan ide-ide dari disiplin terkait seperti pengenalan pola, statistik, sistem basis data, dan *machine learning*. Data *mining* sebenarnya adalah langkah dalam proses penemuan pengetahuan yang lebih besar yang juga mencakup aktivitas pra-pemrosesan seperti ekstraksi data, pembersihan, fusi, dan konstruksi fitur, serta aktivitas pasca-pemrosesan seperti interpretasi pola dan model, pembuatan hipotesis, dan aktivitas lainnya. Proses penemuan pengetahuan dan penambangan data seringkali cukup interaktif dan berulang (Zaki & Meira, 2014).

Tugas yang dapat diselesaikan oleh data *mining* terbagi menjadi enam bagian (Larose, 2005), yaitu:



a) *Description*

Deskripsi bertujuan untuk mencari cara dalam mengidentifikasi pola (*pattern*) dan tren yang terdapat dalam suatu data. Hasil dari identifikasi model data *mining* harus menggambarkan pola (*pattern*) yang jelas dan dapat menerima interpretasi serta penjelasan intuitif. Deskripsi berkualitas tinggi biasanya dapat dicapai dengan menganalisis data eksplorasi, metode grafis digunakan untuk mengeksplorasi data dalam mencari pola (*pattern*) dan tren.

b) *Estimation*

Dengan pengecualian variabel target bersifat numerik dan bukan kategorikal maka estimasi dan klasifikasi mirip. Dengan *record* "lengkap", yang mencakup nilai variabel target dan prediktor, model dapat dibangun. Nilai variabel target kemudian diperkirakan untuk pengamatan baru berdasarkan nilai prediktor. Misalnya, berdasarkan usia pasien, jenis kelamin, indeks massa tubuh, dan kadar natrium darah, dapat menjadi bahan untuk menentukan pengukuran tekanan darah sistolik pasien rumah sakit. Model estimasi didapatkan berkat hubungan antara tekanan darah sistolik dan variabel prediktor dalam *training* set. Kemudian, model dapat diterapkan dalam kasus baru.

c) *Prediction*

Prediksi identik dengan klasifikasi dan estimasi, yang membedakan yaitu hasilnya digunakan untuk memprediksi masa depan. Memprediksi kemungkinan bahwa suatu molekul atau senyawa tertentu dapat membantu dalam pembuatan obat baru yang menguntungkan bagi perusahaan farmasi merupakan contoh dari tugas prediksi.

d) *Classification*

Dalam klasifikasi, terdapat variabel kategori target, contohnya kelompok pendapatan, yang dapat dibagi menjadi tiga kategori atau kelas yaitu berpenghasilan tinggi, berpenghasilan menengah, dan berpenghasilan rendah. Terdapat sebuah permasalahan yaitu peneliti ingin mengklasifikasikan kelompok pendapatan berdasarkan karakteristik lain seperti pekerjaan, jenis kelamin, dan usia maka klasifikasi cocok untuk mengklasifikasikan data yang belum ada di dalam *database*. Contoh, seorang profesor wanita berusia 63 tahun dapat diklasifikasikan dalam kelompok berpenghasilan tinggi. Menemukan model atau fungsi yang mendeskripsikan dan memisahkan data ke dalam kelas-kelas adalah proses klasifikasi. Memeriksa kualitas objek dapat membantu dalam mengklasifikasikannya ke dalam salah satu kategori yang telah ditentukan sebelumnya.

e) *Clustering*

*Clustering* adalah proses pengelompokan data ke dalam kelas objek yang sama tanpa menggunakan kelas data tertentu sebagai basis. *Cluster* adalah sekelompok *record* yang berbeda dari *record* di *cluster* lain namun dapat dibandingkan dengan *record* lain. Memproduksi pengelompokan objek yang sebanding satu sama lain adalah tujuannya. Semakin banyak objek yang ada di setiap *cluster* dan semakin berbeda setiap *cluster*, semakin tinggi kualitas analisis *cluster*.

f) *Association*

Menemukan atribut yang muncul pada periode tertentu merupakan tugas asosiasi dalam data *mining*. Hal ini lebih sering disebut sebagai analisis keranjang belanja (*market basket analysis*). Menemukan pedoman untuk menilai hubungan antara dua atau lebih atribut adalah tujuan dari *task* asosiasi.

## 2.5 *Dataset*

*Dataset* dalam pembelajaran mesin sering digunakan untuk membentuk suatu model yang dapat memprediksi atau mengklasifikasikan data baru. Dalam buku "*Pattern Recognition and Machine learning*", *dataset* terdiri dari beberapa objek yang mewakili suatu observasi atau item yang ingin dianalisis. Setiap objek dalam *dataset* dicirikan oleh beberapa atribut atau fitur yang mewakili karakteristik dari objek tersebut (Christopher M. Bishop, 2006).

*Dataset* dapat diperoleh dari berbagai sumber, seperti *database*, *file*, atau hasil pengamatan. *Dataset* juga dapat berasal dari suatu proses pengumpulan data seperti survei atau eksperimen. Dalam pembelajaran mesin, *dataset* biasanya dibagi menjadi dua bagian: data latih (*training data*) dan data uji (*test data*). Data latih digunakan untuk membentuk model pembelajaran mesin, sedangkan data uji digunakan untuk menguji performa model yang telah dibentuk (Christopher M. Bishop, 2006).

Kualitas *dataset* sangat mempengaruhi hasil akhir dari model pembelajaran mesin. Hal ini termasuk memastikan keakuratan, konsistensi, dan relevansi dari data. Jika *dataset* tidak memenuhi syarat, maka model pembelajaran mesin yang dibentuk tidak akan memiliki performa yang baik. Oleh karena itu,

pemilihan dan pengolahan *dataset* yang tepat sangat penting dalam proses pembelajaran mesin (Christopher M. Bishop, 2006).

## **2.6 Machine learning**

Artificial Intelligence adalah teknik yang dimanfaatkan guna meniru kecerdasan makhluk hidup maupun makhluk tak hidup dalam memecahkan masalah. Ada tiga metode yang dapat digunakan untuk melakukan tugas ini: *Fuzzy Logic (FL)*, *Evolutionary Computing (EC)*, dan *Machine learning (ML)*. *Machine learning* atau yang biasa dikenal dengan pembelajaran mesin adalah teknik yang paling banyak digunakan karena sering digunakan untuk menirukan atau menggantikan penalaran manusia untuk memecahkan masalah. Sesuai dengan namanya, pembelajaran mesin bertujuan untuk mengidentifikasi bagaimana proses pembelajaran dan generalisasi manusia bekerja. *Training* merupakan ciri khas dari pembelajaran mesin. Pembelajaran mesin memerlukan data guna dipelajari sebagai data *training*. Metode dalam pembelajaran mesin yang digunakan untuk memilah objek berdasarkan ciri tertentu dalam membedakan benda satu dengan yang lain merupakan metode Klasifikasi (Ahmad Hania, 2017). Pembelajaran mesin dapat digunakan dalam bidang pertanian dan kehutanan seperti memodelkan pertumbuhan tanaman dan membangun model tanaman virtual yang dipengaruhi oleh lingkungan (Suhartono et al., 2012)(Suhartono et al., 2018). Pembelajaran mesin sangat berpengaruh dalam dunia medis contohnya yaitu pengambilan keputusan pada suatu penyakit dan penemuan obat (Lestandy, 2022).

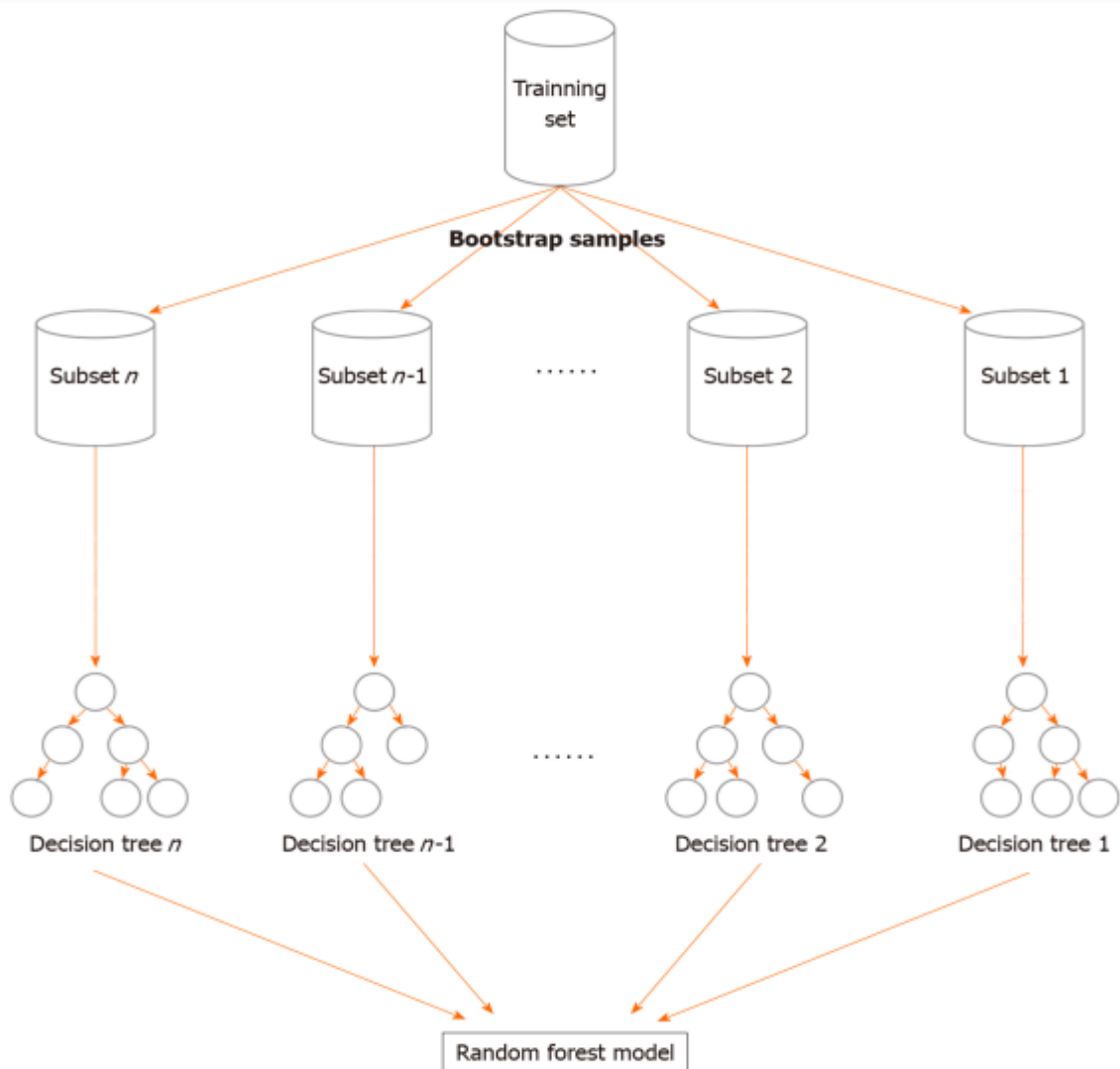
Pembelajaran mesin terbagi menjadi 3 kategori yaitu *Supervised learning*, *Unsupervised Learning*, dan *Reinforcement Learning* (Jovel & Greiner, 2021).

*Supervised learning* cocok untuk mengatasi masalah klasifikasi dan regresi. Ketika variabel *output* berbentuk kategori, contohnya penyakit dan tidak ada penyakit termasuk dalam masalah klasifikasi. Ketika variabel *output* bernilai riil, contohnya berat atau dollar termasuk dalam masalah regresi. *Supervised learning* menggunakan metode klasifikasi guna memberikan label terhadap kumpulan data dalam mengklasifikasikan kelas yang tidak dikenal. Terdapat beberapa algoritma populer dalam *supervised learning* meliputi *Random forest*, *Decision tree*, *SVM*, *Naïve Bayesian*, *Linear Regression*, dan *Neural Network* (Roihan et al., 2020).

## **2.7 Random forest**

Breiman awalnya memperkenalkan *Random forest* pada tahun 2001. Dalam penelitiannya, penulis mendemonstrasikan manfaat *Random forest* antara lain kemampuan untuk membuat kesalahan yang lebih rendah, memberikan hasil klasifikasi yang bagus, dapat mengelola data *training* yang sangat besar secara efisien, dan digunakan untuk mengestimasi *missing data* (Breiman, 2014).

*Random forest* terdiri dari sekumpulan *Decision tree* yang telah dilakukan *training* menggunakan sampel yang berbeda, dan setiap atribut dibagi menjadi *tree/pohon* yang dipilih dari *subset* atribut secara acak (Fauzi et al., 2020). Semakin banyak *tree* yang digunakan, akan meningkatkan keakuratan hasil. *Random forest* digunakan untuk mengklasifikasikan data berdasarkan hasil dari pemilihan *tree* yang dibentuk. *Tree* yang dipilih merupakan *tree* yang terbaik. Pembuatan *tree* pada *Random forest* dibuat hingga mencapai ukuran maksimum dari *tree* data (R. H. Nugraha et al., 2022).



Gambar 2. 2 Contoh *Random forest*

*Pseudocode* metode *Random forest* (Jackins et al., 2021):

1. Pilih fitur “n” secara acak dari total fitur “k”. Dimana  $n \ll k$
2. Di antara fitur “n”, hitung *node* “n” menggunakan titik pisah terbaik.
3. Mengkategorikan *node* menjadi *node* anak menggunakan pemisahan terbaik.
4. Ulangi langkah 1 sampai 3 hingga jumlah *node* “l” tercapai.
5. Bangun *random forest* dengan mengulangi langkah 1 sampai 4 sebanyak “n” kali untuk membuat “n” jumlah pohon.

Berikut adalah langkah-langkah pembentukan *random forest* (Breiman, 2014):

1. Ambil dataset acak sebanyak  $n$  (jumlah total data) dari dataset latih dengan penggantian, sehingga dataset yang diambil bisa memiliki data yang sama.
2. Dari dataset acak tersebut, buatlah sebuah pohon keputusan dengan menggunakan algoritma CART (Classification and Regression Trees) dan aturan splitting Gini Index atau Entropy. Pohon keputusan ini akan menjadi sebuah decision tree.

$$Gini(A) = 1 - \sum_{i=1}^n P_i^2 \quad (2.1)$$

$$Entropy(A) = - \sum_{i=1}^n P_i \log_2(P_i) \quad (2.2)$$

Dimana:

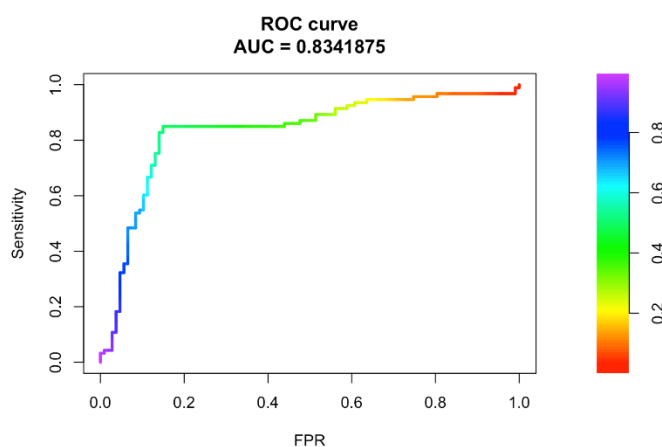
$n$  : jumlah kelas target

$P_i$  : proporsi jumlah sampel kelas  $i$  terhadap jumlah total sampel

3. Ulangi langkah 1 dan 2 sebanyak  $M$  kali untuk membuat  $M$  decision tree. Setiap decision tree dibuat dengan dataset acak yang berbeda-beda.
4. Klasifikasikan setiap observasi dengan menggunakan setiap decision tree yang telah dibuat. Klasifikasi ini menghasilkan  $M$  prediksi untuk setiap observasi.
5. Gabungkan hasil prediksi dari  $M$  decision tree untuk menghasilkan satu prediksi akhir. Jika menggunakan regresi, dapat diambil rata-rata prediksi dari  $M$  decision tree, sedangkan jika menggunakan klasifikasi, dapat digunakan metode voting atau weighted voting untuk menentukan hasil akhir.
6. Evaluasi akurasi model pada dataset uji

## 2.8 ROC dan AUC

*Receiver operating characteristic curves* merupakan alat visual yang berfungsi untuk membedakan dua model klasifikasi. Teori deteksi sinyal, yang dibuat selama Perang Dunia II untuk interpretasi gambar radar, adalah asal mula kurva *ROC* (Junge & Dettori, 2018). *Trade-off* antara *true positive rate* (TPR) dan *false positive rate* ditampilkan pada kurva *ROC* untuk model tertentu (FPR). *ROC* adalah kurva yang digunakan untuk mengevaluasi kinerja model klasifikasi pada berbagai *threshold*. *ROC* menggambarkan *sensitivity* (*True Positive Rate*) pada sumbu Y dan *specificity* (*False Positive Rate*) pada sumbu X. Semakin dekat kurva *ROC* ke sudut kiri atas, semakin baik performa model klasifikasi. *AUC* adalah metrik evaluasi kinerja model klasifikasi yang dihitung berdasarkan luas area di bawah kurva *ROC*. *AUC* menggambarkan seberapa baik model dapat membedakan antara kelas positif dan negatif. Nilai *AUC* berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan model yang sempurna dalam membedakan antara kelas positif dan negatif, dan nilai 0,5 menunjukkan model yang sama baiknya dengan tebakan acak (Fawcett, 2006).



Gambar 2. 3 Contoh *ROC* dan *AUC*



Berikut merupakan rumus *ROC & AUC*.

$$TPR = \frac{TP}{TP + FN} \quad (2.3)$$

$$FPR = \frac{FP}{FP + FN} \quad (2.4)$$

$$AUC = \frac{1}{2} \sum_{i=1}^n (x_{i+1} - x_i)(y_{i+1} - y_i) \quad (2.5)$$

## 2.9 Grid Search

Pada klasifikasi terdapat optimasi yang dapat dilakukan untuk meningkatkan hasil prediksi dari sebuah model yaitu dengan teknik *Grid Search* (W. Nugraha & Sasongko, 2022). *Grid Search* merupakan algoritma pengoptimalan yang mencari semua kemungkinan kombinasi dari parameter model dalam *machine learning*. Optimasi parameter atau *tuning hyperparameter* dengan *Grid Search* pada *random forest* dapat dilakukan dengan mengontrol struktur dari setiap pohon, jumlah fitur dari setiap pohon, ukuran, dan keacakan hutan yang dapat meningkatkan akurasi prediksi dari model (Togatorop et al., 2022). Berikut contoh langkah-langkah *grid search* pada *random forest* untuk mencari *mtry* terbaik (Siji George & Sumathi, 2020).

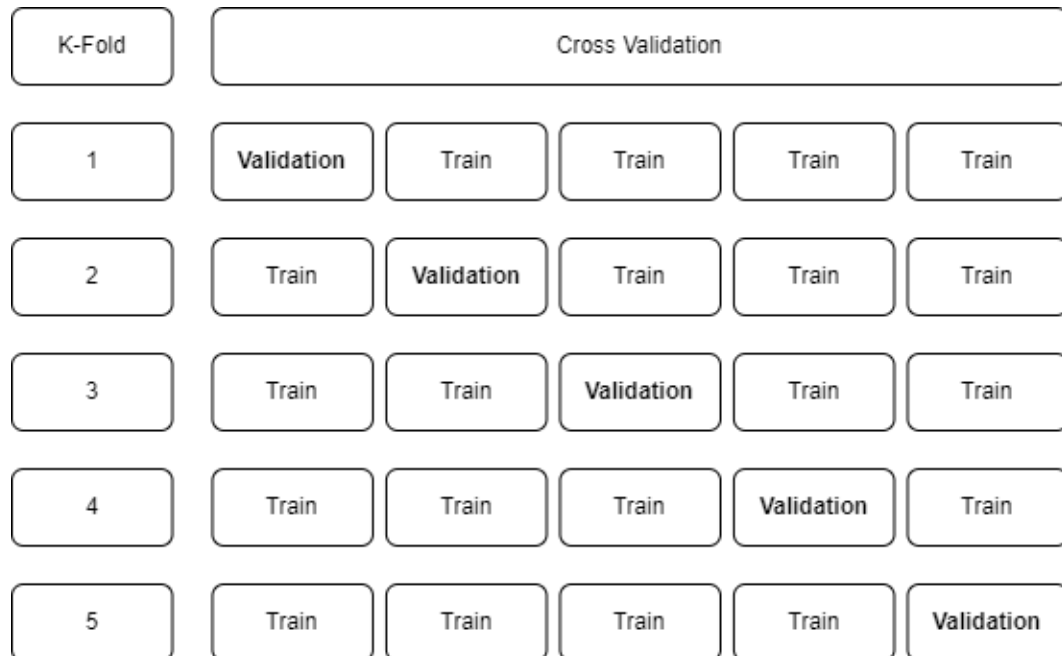
1. Langkah pertama adalah menentukan rentang nilai yang akan dieksplorasi untuk parameter "*mtry*". Misalnya jika ingin mencari nilai terbaik antara 2 hingga 10.
2. Membuat sebuah *grid* yang berisi kombinasi nilai-nilai yang mungkin untuk "*mtry*". Dapat membuat *grid* dengan nilai-nilai [2, 3, 4, 5, 6, 7, 8, 9, 10].

3. Membagi data menjadi *k-fold* untuk validasi silang (*cross-validation*) dengan menggunakan *10-fold cross-validation*.
4. Melakukan iterasi melalui setiap kombinasi nilai "*mtry*" dalam *grid* yang telah dibuat.
5. Pada setiap iterasi, melatih model random forest menggunakan nilai "*mtry*" dari *grid* pada *training set*.
6. Setelah model dilatih, lakukan evaluasi menggunakan metrik evaluasi ROC.
7. Langkah ini diulang untuk setiap kombinasi "*mtry*" dalam *grid*.
8. Memilih kombinasi "*mtry*" yang memberikan hasil terbaik berdasarkan nilai ROC tertinggi.
9. Setelah kombinasi "*mtry*" terbaik dipilih, dapat melatih ulang model random forest dengan menggunakan nilai "*mtry*" tersebut pada seluruh data training.
10. Model yang dilatih dengan parameter "*mtry*" terbaik ini dapat digunakan untuk melakukan prediksi pada data baru.

### **2.10 K-Fold Cross Validation**

Cross validation merupakan suatu teknik validasi model yang digunakan untuk menilai bagaimana hasil statistik analisis menyamaratakan data independen. Cross validation digunakan untuk mengevaluasi performa dari sebuah model untuk melakukan prediksi. K-Fold merupakan salah satu teknik cross validation yang memecah data menjadi K bagian set dengan ukuran data yang sama. Dalam setiap iterasi, salah satu fold digunakan sebagai data validasi (*validation set*), sementara yang lainnya digunakan sebagai data pelatihan (*training set*). Proses ini diulang

sebanyak K kali, dengan setiap fold digunakan sebagai data validasi secara bergantian (Azis et al., 2020). Berikut contoh dari *K-Fold cross validation*.



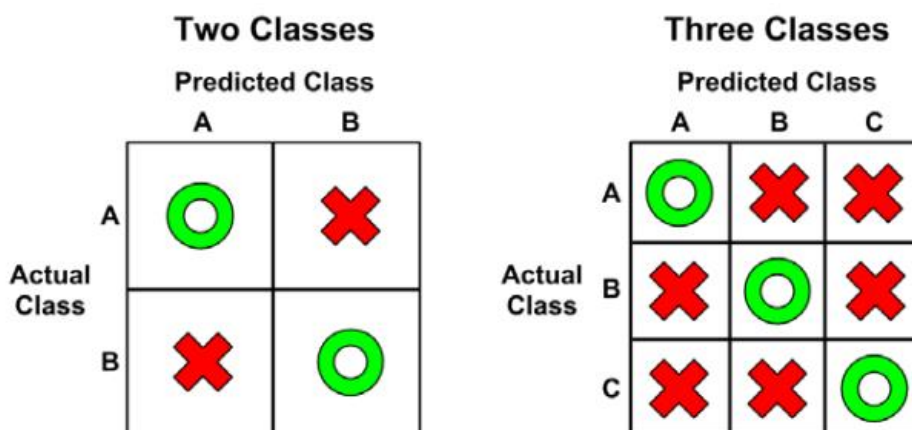
Gambar 2. 4 Contoh K-Fold

### 2.11 Confusion Matrix

*Confusion matrix* yaitu tabel yang mengkategorikan prediksi berdasarkan seberapa dekat kesesuaiannya dengan nilai aktual data. Salah satu dimensi tabel menunjukkan kemungkinan kategori nilai prediksi sementara dimensi lainnya menunjukkan hal yang sama untuk nilai sebenarnya. Meskipun sampai saat ini matrix yang sering digunakan yaitu *confusion matrix* berukuran 2 x 2, sejumlah kelas dapat diprediksi oleh model menggunakan matriks. Matriks konfusi 3 x 3 model tiga kelas ditunjukkan pada gambar terlampir bersama dengan matriks konfusi yang terkenal untuk model biner dua kelas.

Klasifikasi yang tepat yaitu ketika nilai prediksi sama dengan nilai sebenarnya. *Confusion matrix* menunjukkan bahwa prediksi yang benar terletak

pada diagonal (dilambangkan dengan O). Prediksi yang salah yaitu ketika nilai prediksi dan nilai sebenarnya berbeda (dilambangkan dengan X). Metrik kinerja model klasifikasi dihitung berdasarkan jumlah prediksi dalam gambar berikut yang berada di atas dan di luar diagonal (Lantz, 2013)



Gambar 2. 5 Confusion Mmatrix

*Confusion matrix* mengandung informasi tentang *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN). TP adalah jumlah kasus yang diklasifikasikan sebagai positif dan juga benar-benar positif. FP adalah jumlah kasus yang diklasifikasikan sebagai positif tetapi sebenarnya negatif. FN adalah jumlah kasus yang diklasifikasikan sebagai negatif tetapi sebenarnya positif. TN adalah jumlah kasus yang diklasifikasikan sebagai negatif dan juga benar-benar negatif (Zeng, 2020).

Berikut merupakan rumus untuk menghitung *accuracy* dari *confusion matrix* (Normawati & Prayogi, 2021).

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.6)$$

## 2.12 Bahasa Pemrograman R

Menurut buku "*An Introduction to R*", bahasa R adalah bahasa pemrograman yang fleksibel, gratis, dan sumber terbuka yang dirancang khusus untuk analisis data dan visualisasi. Bahasa R memiliki sintaks yang mudah dipahami dan memiliki berbagai fungsi *built-in* (bawaan) untuk melakukan tugas-tugas statistik seperti: menghitung statistik deskriptif, uji hipotesis, analisis regresi, dan lain-lain. Selain fungsi *built-in*, bahasa R juga memiliki berbagai paket tambahan (*CRAN packages*) yang dapat ditemukan dan digunakan oleh pengguna. Paket-paket ini memungkinkan pengguna untuk melakukan tugas-tugas spesifik seperti: *machine learning* dan pemrosesan data terdistribusi (Shah, 2013).

Bahasa R juga memiliki kemampuan untuk memvisualisasikan data secara interaktif dan menghasilkan visualisasi yang berkualitas dan menarik. Bahasa R dapat membuat berbagai jenis visualisasi data seperti diagram batang, diagram pie, dan garis. Secara keseluruhan, bahasa R adalah bahasa pemrograman yang sangat berguna dan banyak digunakan oleh para analis data dan ilmuwan. Bahasa R memiliki berbagai fungsi dan paket tambahan yang membuatnya sangat fleksibel dan bisa digunakan untuk berbagai aplikasi analisis data dan visualisasi (Shah, 2013).

Menurut buku "*The Art of R Programming*", bahasa R memiliki komunitas yang sangat aktif dan menyediakan bantuan dan sumber daya yang berlimpah bagi para penggunanya. Banyak organisasi dan individu menggunakan bahasa R untuk memecahkan masalah-masalah bisnis, ilmiah, dan akademis yang kompleks.

Bahasa R memiliki fitur-fitur unik yang membuatnya sangat cocok untuk analisis data dan pemodelan statistik. Beberapa fitur tersebut termasuk (Matloff, 2011):

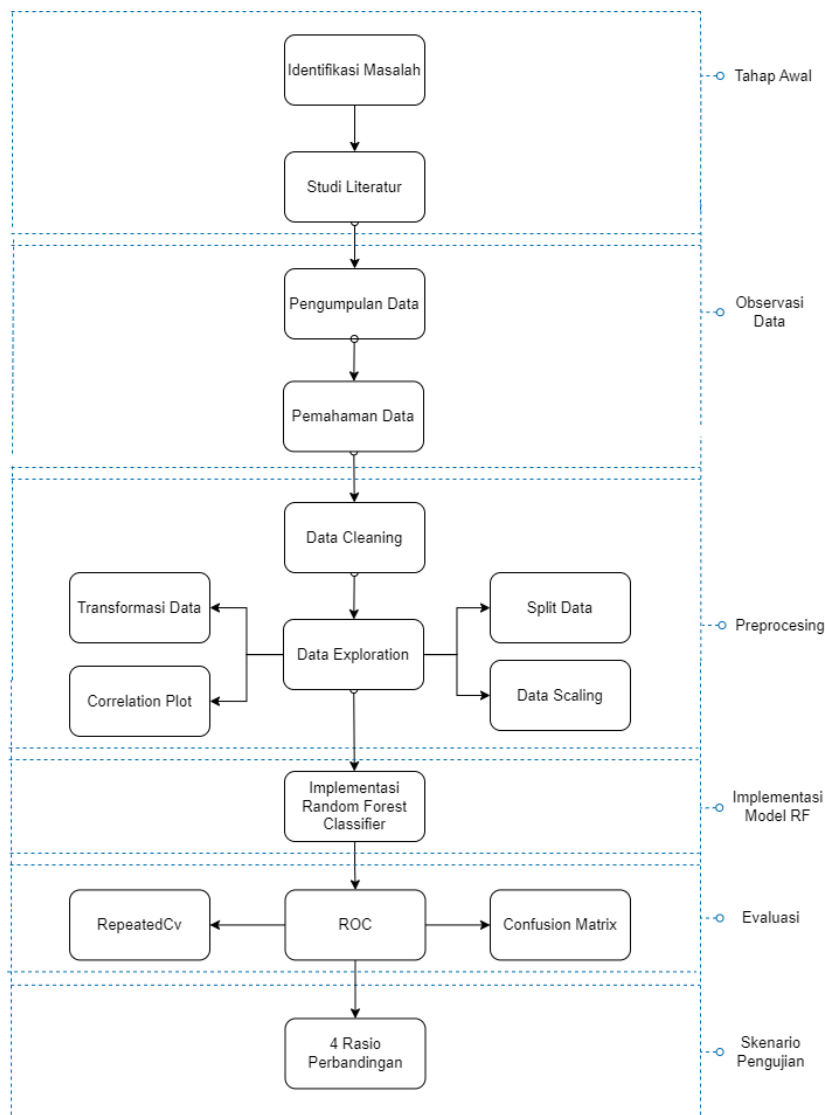
1. Fungsi-fungsi *built-in* yang kuat untuk melakukan manipulasi data, analisis statistik, dan visualisasi.
2. Kemampuan untuk menambahkan fungsi-fungsi baru melalui paket tambahan (*packages*) yang dapat diunduh dari repositori publik.
3. Kemampuan untuk bekerja dengan data yang besar dan kompleks dengan memanfaatkan memori eksternal.
4. Kemampuan untuk mengelola data dengan mudah menggunakan struktur data yang berbeda, seperti vektor, matriks, dan data *frame*.
5. Kemampuan untuk melakukan komunikasi dengan perangkat lunak lain melalui antarmuka *API* atau menggunakan paket tambahan.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Tahapan Penelitian

Untuk menyusun tugas akhir, diperlukan beberapa data dan informasi untuk mendukung validitas dalam materi dan pembahasan. Oleh karena itu, dalam bab ini peneliti akan menjelaskan metode dan langkah-langkah yang digunakan dalam penyusunan tugas akhir agar penulisan terstruktur dengan baik.



Gambar 3. 1 Tahapan Penelitian

## **3.2 Tahap Awal**

Pada tahapan ini terdapat 2 tahapan yang dilakukan yaitu identifikasi masalah dan studi literatur. Identifikasi masalah dilakukan guna memfokuskan tujuan dan cakupan penelitian. Sedangkan, studi literatur dilakukan guna mengumpulkan dan mengevaluasi informasi yang sesuai dengan topik penelitian serta memperdalam pemahaman tentang topik yang diteliti.

### **3.2.1 Identifikasi Masalah**

Pada tahapan ini peneliti mengidentifikasi masalah yang akan diteliti dan dibahas pada tugas akhir ini. Analisis dan evaluasi terhadap data dan informasi dilakukan guna peneliti dapat menemukan masalah-masalah terkait kanker payudara. Identifikasi masalah dapat membantu dalam membuat perencanaan dan penelitian yang lebih efektif serta membantu dalam menentukan arah dan tujuan dari penelitian yang akan dilakukan.

### **3.2.2 Studi Literatur**

Pada tahap ini dilakukan pengkajian terhadap berbagai literatur melalui buku, jurnal, maupun website yang mendukung topik penelitian, seperti *Breast cancer*, *Random forest*, *Data Mining*, *Confusion matrix*, dan *Dataset*.

## **3.3 Observasi Data**

Pada tahapan ini terdapat 2 tahapan yaitu pengumpulan data dan pemahaman data. Pada tahapan pengumpulan data, data dikumpulkan dari berbagai sumber dan format, seperti survei, sumber data publik, atau pengambilan data dari sistem informasi yang ada. Sedangkan, tahapan pemahaman data dilakukan guna



mengekplorasi data dan karakteristik data, variabel yang ada, dan hubungan antara variabel.

### 3.3.1 Pengumpulan Data

Pada tahap ini pengumpulan data dan informasi menggunakan data sekunder penderita kanker payudara yang didapat dari *UCI Machine Learning Repository* dengan nama dataset *Breast Cancer Wisconsin (Diagnostic)* dengan alamat ([https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(diagnostic\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic))).

Total data yang didapat berjumlah 569 dan memiliki 32 atribut. Dataset tersebut dapat digunakan dengan mengikuti *citation request* dari dataset tersebut dan merujuk kepada *citation policy* dari *UCI Machine Learning Repository* yang dapat dilihat pada alamat berikut ([https://archive.ics.uci.edu/ml/citation\\_policy.html](https://archive.ics.uci.edu/ml/citation_policy.html)).

#### Citation Request:

Please refer to the Machine Learning Repository's [citation policy](#).

Gambar 3. 2 *Citation Request*

#### Citation Policy:

If you publish material based on databases obtained from this repository, then, in your acknowledgements, please note the assistance you received by using this repository. This will help others to obtain the same data sets and replicate your experiments. We suggest the following pseudo-APA reference format for referring to this repository:

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Here is a BibTeX citation as well:

```
@misc{Dua:2019 ,
author = "Dua, Dheeru and Graff, Casey",
year = "2017",
title = "{UCI} Machine Learning Repository",
uri = "http://archive.ics.uci.edu/ml",
institution = "University of California, Irvine, School of Information and Computer Sciences" }
```

A few data sets have additional citation requests. These requests can be found on the bottom of each data set's web page.

Gambar 3. 3 *Citation Policy*

### 3.3.2 Pemahaman Data

Data yang didapat memiliki 32 atribut yang didapat berdasarkan operasi biopsi atau *fine needle aspirate (FNA)* menghasilkan 10 atribut utama yaitu *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, *fractal dimension*. Masing-masing 10 atribut utama tadi memiliki 3 indikator yaitu *mean*, *standard error/se*, dan *worst*. Kemudian untuk 2 atribut lainnya yaitu *ID number* dan juga *Diagnosis*. Untuk lebih detailnya dapat dilihat pada tabel berikut.

Tabel 3. 1 Atribut *Dataset*

No	Atribut	Keterangan
1	<i>ID number</i>	ID unik untuk setiap pasien
2	<i>Diagnosis</i>	Variabel target dalam data set ini, dengan dua kelas: "B" untuk kanker payudara <i>benign</i> (tidak ganas) dan "M" untuk kanker payudara <i>malignan</i> (ganas).
3	<i>Radius</i>	Rata-rata jarak dari titik pusat tumor ke titik-titik pada perimeter.
4	<i>Texture</i>	Standar deviasi dari nilai skala abu-abu. Ukuran karakteristik tekstur pada tumor dan bisa menunjukkan bagaimana jaringan tumor terdistribusi dalam bentuk dan konsistensi.
5	<i>Perimeter</i>	Ukuran panjang total dari garis yang mengelilingi tumor dan menunjukkan batas antara tumor dan jaringan sekitarnya.
6	<i>Area</i>	Luas dari tumor dan bisa menunjukkan ukuran tumor. Tumor dengan area yang lebih besar cenderung lebih besar dalam ukuran dan lebih tidak teratur dalam bentuk, dan bisa menjadi tanda bahwa tumor adalah kanker payudara <i>malignan</i> (ganas).
7	<i>Smoothness</i>	<i>Smoothness</i> dihitung dengan membandingkan luas tumor dengan panjang rata-rata garis yang mengelilingi tumor. Semakin besar nilai <i>smoothness</i> , berarti permukaan tumor lebih halus dan lebih tidak beraturan, dan sebaliknya.
8	<i>Compactness</i>	<i>Compactness</i> adalah suatu ukuran yang digunakan untuk mengukur bagaimana tumor memiliki konsentrasi massa. <i>Compactness</i> dihitung dengan membagi luas tumor dengan (perimeter kuadrat dikurangi dengan satu).
9	<i>Concavity</i>	Tingkat kemiringan dari kontur tumor dan bisa menunjukkan bagaimana jaringan tumor terdistribusi dalam bentuk. Kontur adalah garis yang mengelilingi tumor dan menunjukkan batas antara tumor dan jaringan sekitarnya.
10	<i>Concave Points</i>	Jumlah bagian <i>concave</i> dari kontur.
11	<i>Symmetry</i>	Simetri dari tumor.
12	<i>Fractal Dimension</i>	Dimensi fraktal dari tumor.

No	Atribut	Keterangan
13	<i>Mean</i>	Ukuran yang mengukur nilai rata-rata dari suatu fitur seperti ukuran, kontur, atau simetri tumor.
14	<i>Standard Error/SE</i>	Ukuran yang mengukur tingkat ketidakpastian dari nilai rata-rata fitur tumor.
15	<i>Worst</i>	Ukuran yang mengukur nilai maksimal dari suatu fitur seperti ukuran, kontur, atau simetri tumor.

### 3.3.3 Contoh *Dataset* Penyakit Kanker Payudara

Tabel 3. 2 Contoh *Dataset*

<i>Diagnosis</i>	<i>Radius_mean</i>	<i>Texture_mean</i>	<i>Perimeter_mean</i>	<i>Area_mean</i>	<i>Smoothness_mean</i>	<i>Compactness_mean</i>	<i>Concavity_mean</i>	<i>Concave points_mean</i>	<i>Symmetry_mean</i>
M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419
M	20.57	17.77	132.9	1326	0.0874	0.07864	0.08869	0.07017	0.1812
M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069
M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597
M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809
M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087
M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794
M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196
M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235

### 3.4 Preprocessing

*Preprocessing* adalah tahapan yang dilakukan untuk memeriksa data yang akan digunakan dengan tujuan untuk memaksimalkan hasil. Peneliti menggunakan 2 tahapan yakni data *cleaning* dan data *exploration*.

#### 3.4.1 Data Cleaning

Data *cleaning* dilakukan untuk memperbaiki dan memodifikasi data untuk membuatnya siap digunakan dalam analisis, seperti menghapus duplikat, menghapus data tidak relevan, dan mengisi data yang hilang. Pada *dataset* ini terdapat kolom X yang berisi nilai *NA (Not Available)* dan kolom tersebut perlu untuk dihapus. Kolom *ID* juga perlu untuk dihapus dikarenakan tidak digunakan untuk penelitian kali ini.

#### Sebelum

```
'data.frame': 569 obs. of 33 variables:
 $ id          : int  842302 842517 84300903 84348301
844981 84501001 ...
 $ diagnosis   : chr  "M" "M" "M" "M" ...
 $ radius_mean : num  18 20.6 19.7 11.4 20.3 ...
 $ texture_mean : num  10.4 17.8 21.2 20.4 14.3 ...
 $ perimeter_mean : num  122.8 132.9 130 77.6 135.1 ...
 $ symmetry_worst : num  0.46 0.275 0.361 0.66
 $ fractal_dimension_worst: num  0.1189 0.089 0.0876 0
 $ X           : logi  NA NA NA NA NA NA ..
```

#### Setelah

```
'data.frame': 569 obs. of 31 variables:
 $ diagnosis   : chr  "M" "M" "M" "M"
 $ radius_mean : num  18 20.6 19.7 11
 $ texture_mean : num  10.4 17.8 21.2
 $ perimeter_mean : num  122.8 132.9 130
 $ symmetry_worst : num  0.46 0.275 0.361 0.664 0.236 ...
 $ fractal_dimension_worst: num  0.1189 0.089 0.0876 0.173 0.0761
```

Gambar 3. 4 Data *Cleaning*

### 3.4.2 Data Exploration

Memahami dan menjelajahi data melalui visualisasi dan analisis statistik, seperti memahami jenis data, variabel, dan sumber data, memeriksa data untuk menentukan apakah ada data yang hilang, duplikat, atau tidak sesuai, memvisualisasikan data, menghitung statistik seperti rata-rata dan standar deviasi, dan melakukan analisis univariat, bivariat, atau multivariat.

#### 3.4.2.1 Transformasi Data

Mengubah tipe data *diagnosis* dari *character/string* menjadi *factor/kategori*. Konversi dari tipe data "*chr*" (*character*) ke tipe data "*factor*" membuat variabel target "*diagnosis*" menjadi variabel biner, yaitu memiliki dua *level*. *Level-level* tersebut adalah "M" dan "B". Tipe data "*factor*" digunakan untuk menangani variabel yang memiliki beberapa *level* dan memiliki makna yang berbeda. Tipe data *chr* harus diubah menjadi tipe data *factor* agar bisa digunakan sebagai target dalam analisis data.

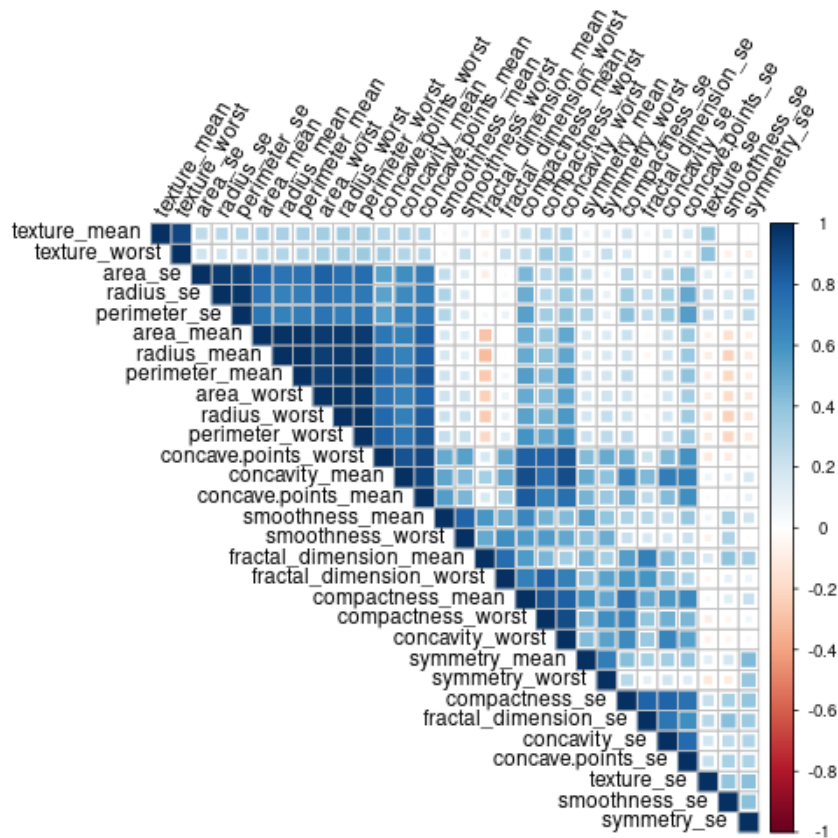
```
$ diagnosis          : chr  "M" "M" "M" "M" ..
Menjadi
$ diagnosis          : Factor w/ 2 levels "B","M": 2 2 2 2 2 2 2 2 2 2 ..
```

Gambar 3. 5 Transformasi Data

#### 3.4.2.2 Correlation plot

*Correlation plot* digunakan untuk mengeksplorasi hubungan antara variabel-variabel dalam suatu data. Plot ini berguna untuk mengetahui apakah ada hubungan linear antara variabel-variabel dan bagaimana tingkat kuatnya hubungan

tersebut. Dapat dilihat pada gambar berikut, variabel '*mean*' dan *worst* sangat tinggi korelasinya.



Gambar 3. 6 *Correlation plot*

### 3.4.2.3 Data Scaling

Data *scaling* merupakan proses transformasi data numerik menjadi rentang nilai tertentu atau skala tertentu agar lebih mudah untuk diolah atau diproses. Data *scaling* diperlukan karena perbedaan skala dan *range* data numerik dapat mempengaruhi hasil analisis dan interpretasi data. Contohnya, data yang memiliki range atau skala yang besar dapat memiliki pengaruh lebih besar pada analisis dibandingkan dengan data yang memiliki skala yang lebih kecil. Oleh karena itu, data *scaling* digunakan untuk mengubah skala data sehingga dapat dibandingkan

dengan lebih mudah dan akurat. Implementasi yang digunakan pada sistem menggunakan normalisasi *z-score*. Teknik *z-score* mengubah nilai variabel menjadi memiliki *mean* 0 dan standar deviasi 1. Rumus normalisasi *z-score* sebagai berikut.

$$x' = \frac{x_i - \text{mean}(x)}{\text{std}(x)} \quad (3.1)$$

Dimana:

$x'$  : nilai dari data yang akan dinormalisasi

$\text{mean}(x)$  : nilai rata-rata dari semua data yang akan dinormalisasi

$\text{std}(x)$  : nilai standar deviasi dari semua data yang akan dinormalisasi

Peneliti memberikan contoh dari pengimplementasian data *scaling* menggunakan normalisasi *z-score* pada 5 fitur tidak termasuk diagnosis dan 20 data.

Tabel 3. 3 Contoh Data Awal

<i>diagnosis</i> <i>s</i>	<i>Radius_mea</i> <i>n</i>	<i>Texture_mea</i> <i>n</i>	<i>Perimeter_mea</i> <i>n</i>	<i>Area_mea</i> <i>n</i>	<i>Smoothness_mea</i> <i>n</i>
M	19.810	22.15	130.00	1260.0	0.09831
B	13.540	14.36	87.46	566.3	0.09779
B	13.080	15.71	85.63	520.0	0.10750
B	9.504	12.44	60.34	273.9	0.10240
M	15.340	14.26	102.50	704.4	0.10730
M	21.160	23.04	137.20	1404.0	0.09428
M	16.650	21.38	110.00	904.6	0.11210
B	13.030	18.42	82.61	523.8	0.08983
B	8.196	16.84	51.71	201.9	0.08600
B	12.050	14.63	78.04	449.3	0.10310
B	11.760	21.60	74.72	427.9	0.08637
M	18.220	18.70	120.30	1033.0	0.11480
M	15.100	22.02	97.26	712.8	0.09056
M	14.710	21.59	95.55	656.9	0.11370
B	8.618	11.79	54.34	224.5	0.09752
B	10.170	14.88	64.55	311.9	0.11340



<i>diagnosi</i> <i>s</i>	<i>Radius_mea</i> <i>n</i>	<i>Texture_mea</i> <i>n</i>	<i>Perimeter_mea</i> <i>n</i>	<i>Area_mea</i> <i>n</i>	<i>Smoothness_mea</i> <i>n</i>
B	8.598	20.98	54.66	221.8	0.12430
M	12.680	23.84	82.69	499.0	0.11220
M	14.780	23.94	97.40	668.3	0.11720
M	18.940	21.31	123.60	1130.0	0.09009

Jika ingin mengimplementasikan persamaan (3.1) pada Tabel 3.3. Langkah pertama yaitu menghitung masing-masing *mean* dan *standard deviasi* dari tiap atribut. Berikut contoh pengubahan nilai pada atribut *radius\_mean*. Langkah pertama yaitu mencari *mean* dari atribut *radius\_mean* didapatkan hasil sebesar  $275.936/20 = 13.7968$ . Setelah itu mencari standar deviasinya. Berikut rumus standar deviasi.

$$std(x) = \sqrt{\frac{(\sum(Xi - \bar{X})^2)}{(n - 1)}} \quad (3.2)$$

Dimana:

$Xi$  : nilai dari setiap data

$\bar{X}$  : nilai rata-rata dari semua data

$n$  : jumlah data

$std(x)$  yang didapatkan yaitu  $= \sqrt{((19.810-13.7968)^2 + (13.540-13.7968)^2 + (13.080-13.7968)^2 + (9.504-13.7968)^2 + (15.340 -13.7968)^2 + (21.160-13.7968)^2 + (16.650-13.7968)^2 + (13.030-13.7968)^2 + (8.196-13.7968)^2 + (12.050-13.7968)^2 + (11.760-13.7968)^2 + (18.220-13.7968)^2 + (15.100 -13.7968)^2 + (14.710-13.7968)^2 / (20-1) = 3.720394839$ . Setelah itu substitusikan nilai *mean* dan standar deviasi pada masing persamaan (3.1) dan didapatkan  $x' = (19.810 - 13.7968) /$

$3.720 = 1.5754$ . Lakukan langkah yang sama sebanyak jumlah data pada masing-masing atribut dan hasil yang didapatkan sebagai berikut.

Tabel 3. 4 Contoh Hasil Normalisasi Z-Score

<i>diagnosis</i>	<i>Radius_mean</i>	<i>Texture_mean</i>	<i>Perimeter_mean</i>	<i>Area_mean</i>	<i>Smoothness_mean</i>
M	1.5754	0.8687	1.5723	1.7748	-0.4084
B	-0.0673	-1.0894	-0.0803	-0.1942	-0.4543
B	-0.1878	-0.7501	-0.1514	-0.3256	0.4027
B	-1.1246	-1.5720	-1.1339	-1.0241	-0.0474
M	0.4043	-1.1145	0.5039	0.1978	0.3850
M	1.9290	1.0924	1.8520	2.1835	-0.7641
M	0.7475	0.6752	0.7953	0.7660	0.8086
B	-0.2009	-0.0689	-0.2688	-0.3148	-1.1568
B	-1.4673	-0.4660	-1.4692	-1.2285	-1.4948
B	-0.4576	-1.0215	-0.4463	-0.5263	0.0143
B	-0.5336	0.7305	-0.5753	-0.5870	-1.4622
M	1.1588	0.0015	1.1955	1.1305	1.0469
M	0.3414	0.8360	0.3004	0.2216	-1.0924
M	0.2392	0.7279	0.2339	0.0630	0.9498
B	-1.3568	-1.7354	-1.3670	-1.1644	-0.4781
B	-0.9502	-0.9587	-0.9704	-0.9163	0.9234
B	-1.3620	0.5746	-1.3546	-1.1720	1.8854
M	-0.2926	1.2935	-0.2656	-0.3852	0.8175
M	0.2576	1.3186	0.3058	0.0953	1.2587
M	1.3474	0.6576	1.3237	1.4058	-1.1339

### 3.4.3 Split Data

*Split* data berguna untuk membagi data menjadi dua bagian atau lebih yaitu data latih (*training data*) dan data uji (*testing data*). Hal ini sangat penting dilakukan karena dalam pemodelan data *machine learning*, model harus dibangun dan diuji dengan data yang berbeda. Data latih digunakan untuk membangun model, sementara data uji digunakan untuk mengevaluasi kinerja model. *Split* data yang tepat.

### 3.4.3.1 Training dan Testing

Peneliti membagi data menjadi 4 bagian, yaitu model A dengan perbandingan 90% data *train* : 10% data *test*, model B dengan perbandingan 80% data *train* : 20% data *test*, model C dengan perbandingan 75% data *train* : 25% data *test*, dan model D dengan perbandingan 70% data *train* : 30% data *test*. Dengan pembagian beberapa dataset yang berbeda akan didapatkan nilai ketepatan prediksi terbaik (Almais et al., 2022). Peneliti menggunakan teknik *split* data *Cross-Validation* atau biasa dikenal dengan teknik *k-fold cross validation*. Pembagian model A menghasilkan 513 data *train* dan 56 data *test*. Model B menghasilkan 456 data *train* dan 113 data *test*. Model C menghasilkan 427 data *train* dan 142 data *test*. Model D menghasilkan 399 data *test* dan 170 data *uji*.

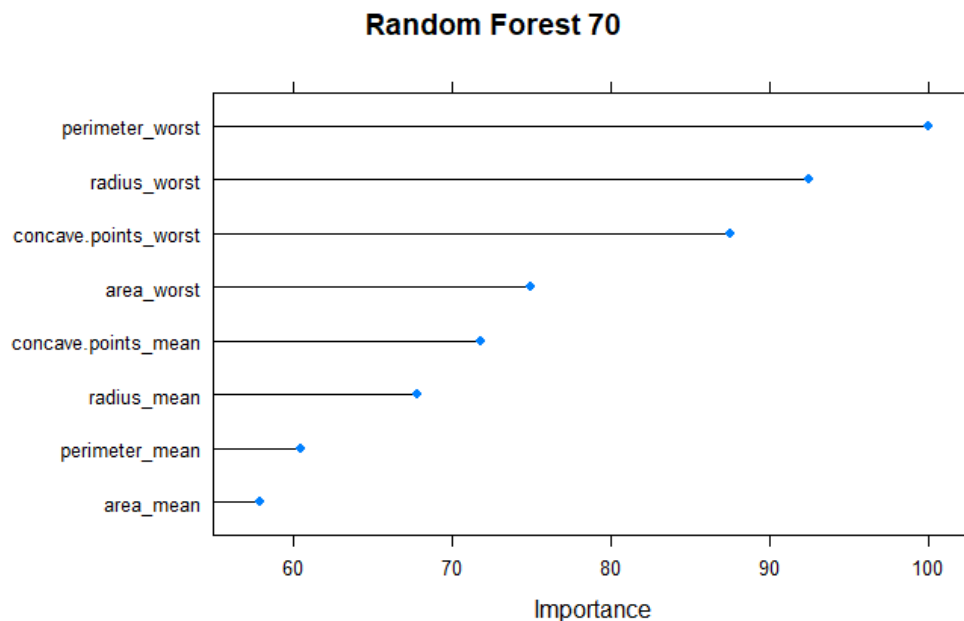
## 3.5 Implementasi Metode *Random forest*

Pada masing-masing model *tuning* proses pelatihan metode *Random forest* dibentuk menggunakan fungsi *train* dengan beberapa parameter sebagai berikut.

1. Mengatur *trainControl* digunakan untuk mengatur beberapa parameter kontrol pada proses pelatihan model, termasuk metode validasi silang, jumlah iterasi, dan jenis tampilan ringkasan. Parameter-parameter yang digunakan dalam *trainControl* yaitu:
  - a) Memilih metode: metode yang digunakan yaitu “*repeatedcv*” atau validasi silang dengan *K-fold cross-validation* dengan jumlah lipatan (*fold*) sebanyak 10 dengan 3 kali perulangan dan menggunakan fungsi *grid* pada argumen “*search*” untuk menemukan parameter terbaik. Menentukan Probabilitas kelas akan dihitung atau tidak, peneliti memilih untuk

dihitung yang nantinya akan digunakan pada prediksi data uji dan juga dapat mengambil keputusan atau membuat kurva ROC dan menghitung AUC. Kemudian, menggunakan fungsi *summaryFunction* yang berguna untuk menghitung ringkasan statistik pada model seperti akurasi, *sensitivity*, *specificity*, dan area di bawah kurva ROC.

2. Mengatur *train* digunakan untuk melatih model pada data yang telah diproses sebelumnya. Parameter-parameter yang digunakan dalam *train*:
  - a) Formula **diagnosis~.** = mengindikasikan bahwa kolom diagnosis adalah variabel target dan semua kolom lainnya adalah variabel fitur (input).
  - b) Model akan dilatih menggunakan beberapa model yang sudah dibagi.
  - c) Parameter *method* = "**rf**" merupakan fungsi untuk menggunakan model *Random forest* dari paket *caret*.
  - d) Parameter *metric* = "**ROC**" menunjukkan bahwa evaluasi model akan dilakukan dengan menghitung ROC AUC.
  - e) Parameter *tuneLength* = **20** menentukan panjang *tuning grid*, yaitu jumlah nilai yang akan diuji untuk tiap parameter.
  - f) Parameter *ntree* = **500** menentukan jumlah pohon dalam model *Random forest*
  - g) Memanggil fungsi *trainControl*
3. Melihat variabel yang dianggap penting oleh *Random forest* dengan fungsi *varImp*. Contoh yang digunakan pada model D *Tuning*.



Gambar 3. 7 Contoh Variabel Penting *Random forest*

### 3.6 Evaluasi

Terdapat beberapa evaluasi dalam penelitian ini, yaitu dengan memanfaatkan kombinasi *K-fold Cross Validation*, *metric ROC*, dan *confusion matrix*. *Repeated K-fold Cross Validation* yaitu metode yang membagi dataset menjadi beberapa bagian, lalu menggunakan beberapa bagian untuk training dan beberapa bagian untuk testing. Dalam kasus ini, peneliti menggunakan *Repeated 10-fold cross-validation*, artinya dataset dibagi menjadi 10 bagian dengan 3 kali pengulangan. *Metric ROC* merupakan metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi. *ROC (Receiver Operating Characteristic) curve* adalah plot yang menunjukkan performa model klasifikasi di berbagai tingkat *cutoff threshold*. *ROC curve* mengukur seberapa baik model membedakan antara kelas positif dan negatif. Area di bawah kurva *ROC (AUC-ROC)* juga sering digunakan sebagai ukuran kinerja model. *Confusion Matrix* merupakan tabel yang

digunakan untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* menunjukkan jumlah *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) yang dihasilkan oleh model.

Dalam penelitian ini, ketiga kombinasi tersebut digunakan bersama-sama untuk mengevaluasi performa model *random forest* dengan mencari jumlah *mtry* terbaik dengan teknik *Grid Search*. *K-fold Cross Validation* digunakan untuk membagi dataset menjadi 10 *fold* dengan 3 kali pengulangan. *Metric ROC* digunakan untuk mengukur performa model *random forest* dalam membedakan antara kelas positif dan negatif (kanker jinak atau kanker ganas) dan akan menghasilkan nilai ROC untuk setiap *fold*. Kemudian, nilai ROC dari setiap *fold* akan diambil rata-ratanya untuk mendapatkan nilai ROC akhir untuk penentuan *mtry* model *random forest*. *Confusion matrix* digunakan untuk mengevaluasi kinerja model *random forest* secara detail dengan melihat TP, TN, FP, dan FN yang dihasilkan oleh model *random forest*. Pertama-tama data dibagi menjadi beberapa *fold*, pada setiap *fold* model *random forest* di *training* dengan mencari kombinasi *mtry* terbaik kemudian diuji pada *fold* tersebut dan dievaluasi dengan *metric ROC*. Setelah model *random forest* dilatih, *confusion matrix* digunakan untuk menunjukkan jumlah prediksi yang benar dan salah dari model *random forest* pada setiap kelas, dibandingkan dengan label kelas aktual dari data uji. Berikut rumus *accuracy* menggunakan *Confusion Matrix* sesuai dengan persamaan pada sub bab 2.10 persamaan (2.6).

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.3)$$

### 3.7 Skenario Pengujian

Peneliti membagi data menjadi 4 bagian, yaitu model A dengan perbandingan 90% data *train* : 10% data *test*, model B dengan perbandingan 80% data *train* : 20% data *test*, model C dengan perbandingan 75% data *train* : 25% data *test*, dan model D dengan perbandingan 70% data *train* : 30% data *test*. Dengan pembagian beberapa dataset yang berbeda akan didapatkan nilai ketepatan prediksi terbaik (Almais et al., 2022). Skenario pengujian pada penelitian ini, menggunakan 31 variabel atau seluruh fitur pada *dataset*. Pada masing-masing pembagian 4 model tadi akan dilakukan perbandingan antara model *random forest* menggunakan data yang tidak dinormalisasi atau secara *default* dengan model *random forest* menggunakan data yang dinormalisasi menggunakan *z-score* dan di *tuning* menggunakan teknik *Grid Search* untuk mencari jumlah *mtry* terbaik.

### 3.8 Perhitungan Manual *Random forest*

Peneliti akan menggunakan data contoh normalisasi *z-score* pada Tabel 3.4 dan menggunakan model D. Membagi dataset menjadi data *train* dan *test* dengan perbandingan 70:30.

Tabel 3. 5 Data *Train*

<i>diagnosi</i> <i>s</i>	<i>Radius_mea</i> <i>n</i>	<i>Texture_mea</i> <i>n</i>	<i>Perimeter_mea</i> <i>n</i>	<i>Area_mea</i> <i>n</i>	<i>Smoothness_mea</i> <i>n</i>
M	1.5754	0.8687	1.5723	1.7748	-0.4084
B	-0.0673	-1.0894	-0.0803	-0.1942	-0.4543
B	-0.1878	-0.7501	-0.1514	-0.3256	0.4027
B	-1.1246	-1.5720	-1.1339	-1.0241	-0.0474
M	0.4043	-1.1145	0.5039	0.1978	0.3850
M	1.9290	1.0924	1.8520	2.1835	-0.7641
M	0.7475	0.6752	0.7953	0.7660	0.8086
B	-0.2009	-0.0689	-0.2688	-0.3148	-1.1568
B	-1.4673	-0.4660	-1.4692	-1.2285	-1.4948

<i>diagnosis</i>	<i>Radius_mean</i>	<i>Texture_mean</i>	<i>Perimeter_mean</i>	<i>Area_mean</i>	<i>Smoothness_mean</i>
B	-0.4576	-1.0215	-0.4463	-0.5263	0.0143
B	-0.5336	0.7305	-0.5753	-0.5870	-1.4622
M	1.1588	0.0015	1.1955	1.1305	1.0469
M	0.3414	0.8360	0.3004	0.2216	-1.0924
M	0.2392	0.7279	0.2339	0.0630	0.9498

Tabel 3. 6 Data *Testing*

<i>diagnosis</i>	<i>Radius_mean</i>	<i>Texture_mean</i>	<i>Perimeter_mean</i>	<i>Area_mean</i>	<i>Smoothness_mean</i>
B	-1.3568	-1.7354	-1.3670	-1.1644	-0.4781
B	-0.9502	-0.9587	-0.9704	-0.9163	0.9234
B	-1.3620	0.5746	-1.3546	-1.1720	1.8854
M	-0.2926	1.2935	-0.2656	-0.3852	0.8175
M	0.2576	1.3186	0.3058	0.0953	1.2587
M	1.3474	0.6576	1.3237	1.4058	-1.1339

Membuat 5 *bootstrapped dataset*, peneliti membagi lagi agar lebih mudah dalam perhitungan dengan menggunakan 5 data train dan 5 fitur untuk contoh perhitungan.

Tabel 3. 7 Dataset Perhitungan Manual RF

<b>I D</b>	<i>Diagnosis</i>	<i>Radius_mean</i>	<i>Texture_mean</i>	<i>Perimeter_mean</i>	<i>Area_mean</i>	<i>Smoothness_mean</i>
1	M	0.4043	-1.1145	0.5039	0.1978	0.385
2	M	1.929	1.0924	1.852	2.1835	-0.7641
3	M	0.7475	0.6752	0.7953	0.766	0.8086
4	B	-0.2009	-0.0689	-0.2688	-0.3148	-1.1568
5	B	-1.4673	-0.466	-1.4692	-1.2285	-1.4948

Tabel 3. 8 *Boostrapped Dataset* Pertama

<i>Radius_mean</i>	<i>Texture_mean</i>	<b>Diagnosis</b>
1.9290	1.0924	M
0.4043	-1.1145	M
0.7475	0.6752	B
-0.2009	-0.0689	B
-0.2009	-0.0689	B



ID yang digunakan pada *bootstrapped dataset* pertama yaitu (2 1 3 4 4) ID yang tidak digunakan yaitu 5. Fitur yang digunakan adalah *Radius\_mean* dan *Texture\_mean*. Setelah itu menghitung Gini Impurity (GI) dengan persamaan sebagai berikut.

$$GI = 1 - (p1^2 + p2^2) \quad (3.4)$$

Dimana:

P : proporsi jumlah observasi

GI(*Radius\_mean*) yang didapat yaitu  $GI(\text{Radius\_mean}) = 1 - ((2/5)^2 + (3/5)^2) = 0.48$ . Begitupun dengan GI(*Texture\_mean*). Karena GI(X1) dan GI(X2) memiliki nilai yang sama, maka dapat memilih salah satu variabel sebagai *root node*, misalnya *Radius\_mean*. Selanjutnya, perlu untuk membuat split pada *Radius\_mean* dengan *threshold* tertentu. Nilai *threshold* dapat ditentukan dengan nilai *mean* atau *median*, *Information Gain*, dan *Grid Search*. Hal terpenting untuk menentukan nilai *threshold* dapat menghasilkan *decision tree* yang optimal, pada contoh kali ini peneliti menggunakan nilai *mean* sebagai *threshold*. *Mean Radius\_mean* yang didapat 0.5358.

Kemudian, data akan terbagi menjadi 2 *subset*:

*Subset 1* ( $\text{Radius\_mean} \leq 0.5358$ ):

Tabel 3. 9 *Subset 1* DT1

<i>Radius_mean</i>	<i>Texture_mean</i>	Diagnosis
0.4043	-1.1145	M
-0.2009	-0.0689	B
-0.2009	-0.0689	B

*Subset 2* ( $Radius\_mean \geq 0.5358$ )

Tabel 3. 10 *Subset 2* DT1

<i>Radius_mean</i>	<i>Texture_mean</i>	<b>Diagnosis</b>
0.7475	0.6752	B
1.9290	1.0924	M

Hitung *Gini Impurity*-nya pada masing-masing *subset*,  $GI(1) = 1 - ((1/3)^2 + (2/3)^2) = 0.44444444$ . Sedangkan,  $GI(2) = 1 - ((1/2)^2 + (1/2)^2) = 0.5$ . Jadi  $GI(1)$  yang digunakan sebagai *internal node*. Kemudian, menghitung *mean Texture\_mean* didapatkan hasil  $-0.417433$ . Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 3* ( $Texture\_mean \leq -0.417433$ ):

Tabel 3. 11 *Subset 3* DT1

<i>Radius_mean</i>	<i>Texture_mean</i>	<b>Diagnosis</b>
-0.2009	-0.0689	B
-0.2009	-0.0689	B

*Subset 4* ( $Texture\_mean \geq -0.417433$ )

Tabel 3. 12 *Subset 4* DT1

<i>Radius_mean</i>	<i>Texture_mean</i>	<b>Diagnosis</b>
0.4043	-1.1145	M

Jadi jika nilai  $Texture\_mean \leq -0.417433$ , maka hasil *leaf node* nya B dan jika jika nilai  $Texture\_mean \geq 0.417433$ , maka hasil *leaf nodenya* M.

Tabel 3. 13 *Boostrapped Dataset* Kedua

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
-0.4660	-1.4692	B
-0.4660	-1.4692	B
-0.0689	-0.2688	B
0.6752	0.7953	M
-1.1145	0.5039	M

ID yang digunakan pada *bootstrapped dataset* kedua yaitu (5 5 4 3 1), ID yang tidak digunakan yaitu 2. Fitur yang digunakan adalah *Texture\_mean* dan *Perimeter\_mean*. Lakukan langkah yang sama seperti tabel 3.9. Dikarenakan nilai GI sama, peneliti akan menggunakan *Perimeter\_mean* sebagai *root node*. *Mean Perimeter\_mean* sebesar -0.3816. Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 1* ( $Perimeter\_mean \leq -0.3816$ ):

Tabel 3. 14 *Subset 1 DT2*

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
-0.4660	-1.4692	B
-0.4660	-1.4692	B

*Subset 2* ( $Perimeter\_mean \geq -0.3816$ ):

Tabel 3. 15 *Subset 2 DT2*

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
-0.0689	-0.2688	B
0.6752	0.7953	M
-1.1145	0.5039	M

*Leaf node* pada subset 1 sudah terbentuk yaitu B dikarenakan homogen. Selanjutnya yaitu mencari *mean* dari *Texture\_mean* dari subset 2, didapatkan *mean* sebesar -0.1694. Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 3* ( $Texture\_mean \leq -0.1694$ ):

Tabel 3. 16 *Subset 3 DT2*

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
-1.1145	0.5039	M

*Subset 4* ( $Texture\_mean \geq -0.1694$ )

Tabel 3. 17 *Subset 4 DT2*

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
-0.0689	-0.2688	B

<i>Texture_mean</i>	<i>Perimeter_mean</i>	<b>Diagnosis</b>
0.6752	0.7953	M

Sehingga *Leaf node* pada subset 3 menjadi jika  $Texture\_mean \leq -0.1694$  maka *leaf node* nya M dan jika  $Texture\_mean \geq -0.1694$  maka *leaf node* antara B dan M.

Tabel 3. 18 *Boostrapped Dataset* Ketiga

<i>Perimeter_mean</i>	<i>Area_mean</i>	<b>Diagnosis</b>
1.852	2.1835	M
1.852	2.1835	M
0.7953	0.766	M
-0.2688	-0.3148	B
-0.2688	-0.3148	B

ID yang digunakan pada *boostrapped dataset* ketiga yaitu (2 2 3 4 4), ID yang tidak digunakan yaitu 1 dan 5. Fitur yang digunakan adalah *Perimeter\_mean* dan *Area\_mean*. *Area\_mean* sebagai *root node*. Kemudian, mencari *Mean* dari *Area\_mean* yang didapatkan yaitu 0.90068. Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 1* ( $Area\_mean \leq 0.90068$ ):

Tabel 3. 19 *Subset 1 DT3*

<i>Perimeter_mean</i>	<i>Area_mean</i>	<b>Diagnosis</b>
0.7953	0.766	M
-0.2688	-0.3148	B
-0.2688	-0.3148	B

*Subset 2* ( $Area\_mean \geq 0.90068$ ):

Tabel 3. 20 *Subset 2 DT3*

<i>Perimeter_mean</i>	<i>Area_mean</i>	<b>Diagnosis</b>
1.852	2.1835	M
1.852	2.1835	M

*Leaf node* pada subset 2 sudah terbentuk yaitu M dikarenakan homogen. Selanjutnya yaitu mencari *mean* dari *Perimeter\_mean* dari subset 1, didapatkan *mean* sebesar 0.0859. Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 3* ( $Perimeter\_mean \leq 0.0859$ ):

Tabel 3. 21 *Subset 3* DT3

<i>Perimeter_mean</i>	<i>Area_mean</i>	<b>Diagnosis</b>
-0.2688	-0.3148	B
-0.2688	-0.3148	B

*Subset 4* ( $Perimeter\_mean \geq 0.0859$ ):

Tabel 3. 22 *Subset 4* DT4

<i>Perimeter_mean</i>	<i>Area_mean</i>	<b>Diagnosis</b>
0.7953	0.766	M

Jadi jika nilai  $Perimeter\_mean \leq -0.4962$ , maka hasil *leaf node* nya B dan jika nilai  $Perimeter\_mean \geq -0.4962$ , maka hasil *leaf node*nya M.

Tabel 3. 23 *Boostrapped Dataset* Keempat

<i>Area_mean</i>	<i>Smoothness_mean</i>	<b>Diagnosis</b>
0.1978	0.385	M
0.766	0.8086	M
-0.3148	-1.1568	B
-0.3148	-1.1568	B
-1.2285	-1.4948	B

ID yang digunakan pada *boostrapped dataset* kedua yaitu (1 3 4 4 5), ID yang tidak digunakan yaitu 2. Fitur yang digunakan adalah *Area\_mean* dan *Smoothness\_mean*. *Mean Smoothness\_mean* sebesar -0.52296. Sehingga *subset* yang dihasilkan sebagai berikut.

*Subset 1* ( $Smoothness\_mean \leq -0.52296$ ):

Tabel 3. 24 *Subset 1* DT4

<i>Area_mean</i>	<i>Smoothness_mean</i>	<b>Diagnosis</b>
-0.3148	-1.1568	B
-0.3148	-1.1568	B
-1.2285	-1.4948	B

*Subset 2* ( $Smoothness\_mean \geq -0,52296$ ):

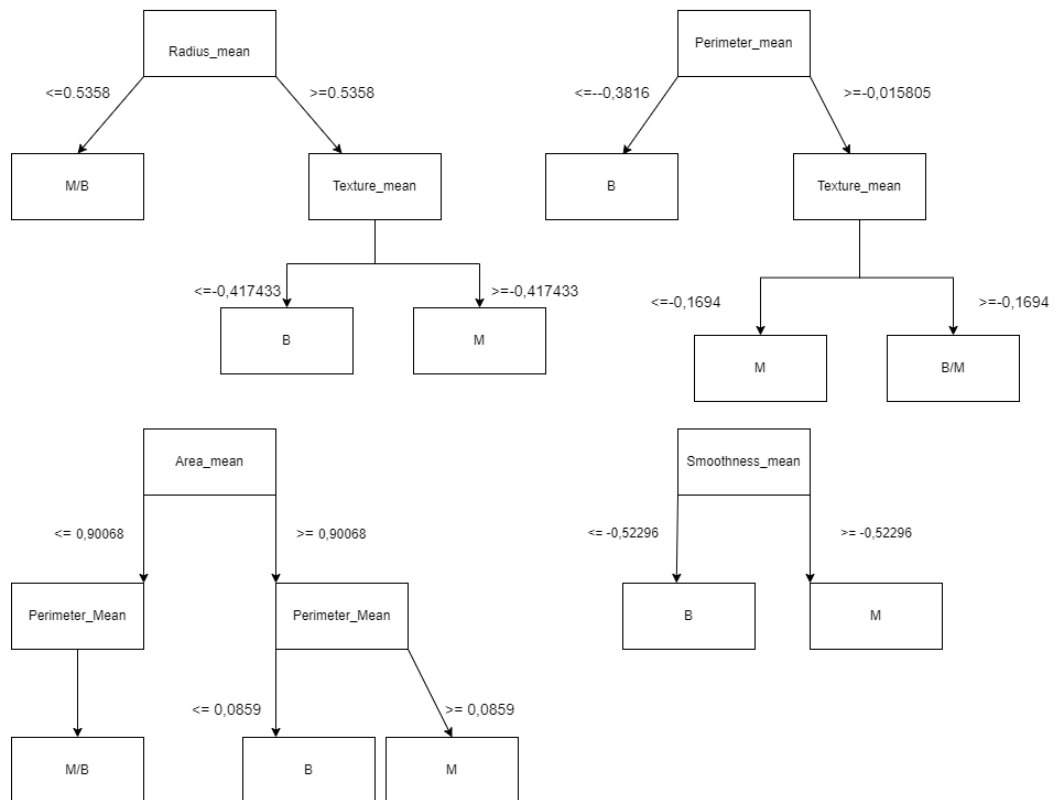
Tabel 3. 25 *Subset 2* DT4

<i>Area_mean</i>	<i>Smoothness_mean</i>	<b>Diagnosis</b>
0.1978	0.385	M
0.766	0.8086	M

Oleh karena kedua *subset* hanya memiliki 1 kelas, tidak perlu dilakukan split lebih lanjut dan dapat mengambil keputusan bahwa *subset 1* ( $Smoothness\_mean \leq -0.52296$ ) memiliki diagnosis B, sedangkan *subset 2* ( $Smoothness\_mean \geq -0.52296$ ) memiliki diagnosis M.

Dari *Decision tree* yang telah dibentuk dapat dimasukkan data testing yang telah dibagi pada Tabel 3.6, kemudian hasil akhir data testing berdasarkan voting dari masing-masing *decision tree* yang telah dibentuk. Akan tetapi, pada penggunaannya jumlah *decision tree* dalam sebuah *random forest* tergantung pada kompleksitas masalah dan ukuran data.

Untuk lebih jelasnya hasil *decision tree* yang dibentuk sebagai berikut.



Gambar 3. 8 *Decision tree* Yang Terbentuk

## BAB IV

### UJI COBA DAN PEMBAHASAN

#### 4.1 Hasil Uji Coba

Pada Sub Bab ini berisi penjabaran dan penjelasan tentang analisis hasil pengujian sistem berdasarkan skenario pengujian yang berguna untuk mengetahui performa metode *Random forest* pada sistem dalam mengklasifikasi kanker payudara dengan menggunakan 569 dataset yang telah di *split* menjadi 4 bagian. Setelah skenario pengujian dilakukan didapatkan beberapa nilai akurasi dalam menggunakan metode *Random forest*. Setelah itu, sistem yang telah dibangun perlu untuk dievaluasi guna mengetahui kinerja metode *Random forest* dalam mengklasifikasi kanker payudara menggunakan kombinasi 3 teknik evaluasi yang telah dipaparkan pada sub bab 3.6.

##### 4.1.1 Pengujian Model A

Pada tahap ini menggunakan data *train* sebanyak 513 data dan data *test* sebanyak 56 data atau rasio perbandingan sebesar 90:10. Pada pengujian ini terdapat beberapa pembagian skenario pengujian yakni *default* atau tanpa dinormalisasi dan normalisasi *z-score* ditambah dengan *tuning hyperparameter* menggunakan *Grid Search*.

##### 4.1.1.1 Default

Pada tahapan uji coba *default* atau tanpa normalisasi dengan 31 variabel didapatkan hasil sebagai berikut.



Tabel 4. 1 Hasil Prediksi *Default* Model A

<i>radius_me an</i>	<i>texture_me an</i>	<i>perimeter_me an</i>	<i>area_me an</i>	<i>smoothness_me an</i>	<i>Actu al</i>	<i>Predicte d</i>
13.44	21.58	86.18	563	0.08162	M	B
10.95	21.35	71.9	371.1	0.1227	M	B
13.17	21.81	85.42	531.5	0.09714	M	B
12.77	22.47	81.72	506.3	0.09055	M	B
14.26	19.65	97.83	629.9	0.07837	B	M
...	...	...	...	...	...	...
11.9	14.65	78.11	432.8	0.1152	B	B
8.219	20.7	53.27	203.9	0.09405	B	B
12.34	22.22	79.85	464.5	0.1012	B	B
23.27	22.04	152.1	1686	0.08439	M	M
17.47	24.68	116.1	984.6	0.1049	M	M

Tabel 4. 2 Confusion Matrix *Default* Model A

<b>Prediksi</b>	<b>Data Aktual</b>	
	<b>B</b>	<b>M</b>
<b>B</b>	34	4
<b>M</b>	1	17

Dari Tabel 4.1 menunjukkan bahwa model *random forest* memprediksi terdapat 34 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 17 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 4 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 1 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.1 sebagai berikut.

$$Akurasi = \frac{34 + 17}{34 + 1 + 17 + 4} = 91.07\%$$

#### 4.1.1.2 Z-Score Tuning

Pada tahapan uji coba normalisasi menggunakan z-score dikombinasikan dengan *tuning* dan menggunakan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 3 Hasil Prediksi Z-Score Model A

<i>radius_melan</i>	<i>texture_melan</i>	<i>perimeter_melan</i>	<i>area_melan</i>	<i>smoothness_melan</i>	<i>Actual</i>	<i>Predicted</i>
-0.2234	-0.7974	-0.2255	-0.3833	0.8134	B	M
1.4309	1.2812	1.6639	1.3302	0.0739	M	M
0.4888	1.0835	0.4828	0.3632	-0.8781	M	M
-0.2716	0.5860	-0.2695	-0.3506	0.0554	M	M
-1.1229	-1.0253	-1.1284	-0.9746	1.2116	B	B
...	...	...	...	...	...	...
0.1852	1.0812	0.2235	0.0381	1.4818	M	M
-0.5157	-0.6439	-0.5259	-0.5228	-0.3299	B	B
-0.1695	-1.9413	-0.1670	-0.2719	2.3279	B	B
1.3884	1.2323	1.2359	1.1966	-0.4316	M	M
1.0847	0.1675	0.9149	0.9295	-0.8774	M	M

Tabel 4. 4 Confusion Matrix Z-Score Tuning Model A

Prediksi	Data Aktual	
	B	M
B	34	0
M	1	21

Dari Tabel 4.2 menunjukkan bahwa model *random forest* memprediksi terdapat 34 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 20 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 0 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 1 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.2 sebagai berikut.

$$Akurasi = \frac{34 + 21}{34 + 1 + 21 + 0} = 98.21\%$$

#### 4.1.2 Pengujian Model B

Pada tahap ini menggunakan data *train* sebanyak 456 data dan data *test* sebanyak 113 data atau rasio perbandingan sebesar 80:20. Pada pengujian ini terdapat beberapa pembagian skenario pengujian yakni *default* atau tanpa dinormalisasi dan normalisasi z-score ditambah dengan *tuning hyperparameter* menggunakan *Grid Search*.

##### 4.1.2.1 Default

Pada tahapan uji coba *default* atau tanpa normalisasi dengan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 5 Hasil Prediksi *Default* Model B

<i>radius_mean</i>	<i>texture_mean</i>	<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	<i>Actual</i>	<i>Predicted</i>
14.99	25.2	95.54	698.8	0.09387	M	B
13.34	15.86	86.49	520	0.1078	B	M
13.96	17.05	91.43	602.4	0.1096	M	B
15.73	11.28	102.8	747.2	0.1043	B	M
19.69	21.25	130	1203	0.1096	M	M
...	...	...	...	...	...	...
11.42	20.38	77.58	386.1	0.1425	M	M
13	21.82	87.5	519.8	0.1273	M	M
16.02	23.24	102.7	797.8	0.08206	M	M
13.73	22.61	93.6	578.3	0.1131	M	M
14.54	27.54	96.73	658.8	0.1139	M	M

Tabel 4. 6 Confusion Matrix *Default* Model B

Prediksi	Data Aktual	
	B	M
B	69	2

Prediksi	Data Aktual	
	B	M
M	2	40

Dari Tabel 4.3 menunjukkan bahwa model *random forest* memprediksi terdapat 69 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 40 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 2 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 2 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.3 sebagai berikut.

$$Akurasi = \frac{69 + 40}{69 + 2 + 40 + 2} = 96.46\%$$

#### 4.1.2.2 Z-Score Tuning

Pada tahapan uji coba normalisasi menggunakan z-score dikombinasikan dengan *tuning* dan menggunakan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 7 Hasil Prediksi Z-Score Model B

<i>radius_melan</i>	<i>texture_melan</i>	<i>perimeter_melan</i>	<i>area_melan</i>	<i>smoothness_melan</i>	Actual	Predicted
0.7698	0.0396	0.6762	0.6402	-1.5579	B	M
-0.0475	-0.5207	-0.0222	-0.1492	0.9414	M	B
0.0405	0.7580	0.0741	-0.0713	0.5290	M	B
0.8208	1.0905	0.8573	0.6945	1.6595	M	M
0.7414	0.5348	0.7462	0.6098	-0.0185	M	M
...	...	...	...	...	...	...
-0.3114	-0.2022	-0.3852	-0.3725	-0.4643	B	B
-0.2716	-0.1464	-0.2465	-0.3418	1.3822	M	M
-0.5895	-1.0834	-0.5732	-0.5842	0.4792	B	B
1.1614	-0.1371	1.1659	1.0744	1.3111	M	M
-0.7399	-0.1255	-0.7667	-0.6987	-0.0797	B	B

Tabel 4. 8 Confusion Matrix Z-Score *Tuning* Model B

Prediksi	Data Aktual	
	B	M
B	70	2
M	1	40

Dari Tabel 4.4 menunjukkan bahwa model *random forest* memprediksi terdapat 70 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 40 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 2 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 1 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.4 sebagai berikut.

$$Akurasi = \frac{70 + 40}{70 + 1 + 40 + 2} = 97.35\%$$

#### 4.1.3 Pengujian Model C

Pada tahap ini menggunakan data *train* sebanyak 427 data dan data *test* sebanyak 142 data atau rasio perbandingan sebesar 75:25. Pada pengujian ini terdapat beberapa pembagian skenario pengujian yakni *default* atau tanpa dinormalisasi dan normalisasi z-score ditambah dengan *tuning hyperparameter* menggunakan *Grid Search*.

#### 4.1.3.1 Default

Pada tahapan uji coba *default* atau tanpa normalisasi dengan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 9 Hasil Prediksi *Default* Model C

<i>radius_melan</i>	<i>texture_melan</i>	<i>perimeter_melan</i>	<i>area_melan</i>	<i>smoothness_melan</i>	<i>Actual</i>	<i>Predicted</i>
13.44	21.58	86.18	563	0.08162	M	B
10.95	21.35	71.9	371.1	0.1227	M	B
13.17	21.81	85.42	531.5	0.09714	M	B
13.34	15.86	86.49	520	0.1078	B	M
16.84	19.46	108.4	880.2	0.07445	B	M
...	...	...	...	...	...	...
13.96	17.05	91.43	602.4	0.1096	M	B
17.85	13.23	114.6	992.1	0.07838	B	M
17.99	10.38	122.8	1001	0.1184	M	M
11.42	20.38	77.58	386.1	0.1425	M	M
18.25	19.98	119.6	1040	0.09463	M	M

Tabel 4. 10 Confusion Matrix *Default* Model C

<b>Prediksi</b>	<b>Data Aktual</b>	
	<b>B</b>	<b>M</b>
<b>B</b>	86	4
<b>M</b>	3	49

Dari Tabel 4.5 menunjukkan bahwa model *random forest* memprediksi terdapat 86 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 49 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 4 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 3 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.5 sebagai berikut.

$$Akurasi = \frac{86 + 49}{86 + 3 + 49 + 4} = 95.07\%$$

#### 4.1.3.2 Z-Score Tuning

Pada tahapan uji coba normalisasi menggunakan z-score dikombinasikan dengan *tuning* dan menggunakan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 11 Hasil Prediksi Z-Score Model C

<i>radius_mean</i>	<i>texture_mean</i>	<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	<i>Actual</i>	<i>Predicted</i>
0.3271	0.7255	0.2861	0.1583	-0.4117	M	B
0.4207	0.021	0.3305	0.2947	-1.277	M	B
0.4888	1.0835	0.4828	0.3632	-0.8781	M	M
1.6126	0.665	1.5651	1.7195	0.1386	M	M
-0.1667	-1.1462	-0.1856	-0.2517	0.1017	B	B
...	...	...	...	...	...	...
-1.3119	-1.5926	-1.3017	-1.0826	0.4294	B	B
0.3441	-1.1694	0.4334	0.1407	0.7778	M	M
1.2777	1.3532	1.3511	1.2307	0.7139	M	M
0.7414	0.5348	0.7462	0.6098	-0.0185	M	M
-0.2404	0.2303	-0.1913	-0.3117	0.5503	M	M

Tabel 4. 12 Confusion Matrix Z-Score Tuning Model C

Prediksi	Data Aktual	
	B	M
B	89	2
M	0	51

Dari Tabel 4.6 menunjukkan bahwa model *random forest* memprediksi terdapat 89 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 51 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 2 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest*

memprediksi terdapat 0 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.6 sebagai berikut.

$$Akurasi = \frac{89 + 51}{89 + 0 + 51 + 2} = 98.59\%$$

#### 4.1.4 Pengujian Model D

Pada tahap ini menggunakan data *train* sebanyak 399 data dan data *test* sebanyak 170 data atau rasio perbandingan sebesar 70:30. Pada pengujian ini terdapat beberapa pembagian skenario pengujian yakni *default* atau tanpa dinormalisasi dan normalisasi z-score ditambah dengan *tuning hyperparameter* menggunakan *Grid Search*.

##### 4.1.4.1 Default

Pada tahapan uji coba *default* atau tanpa normalisasi dengan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 13 Hasil Prediksi *Default* Model D

<i>radius_mean</i>	<i>texture_mean</i>	<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	<i>Actual</i>	<i>Predicted</i>
14.64	15.24	95.77	651.9	0.1132	B	M
15.1	16.39	99.58	674.5	0.115	B	M
11.76	18.14	75.0	431.1	0.09968	M	B
16.5	18.29	106.6	838.1	0.09686	B	M
14.6	23.29	93.97	664.7	0.08682	M	B
...	...	...	...	...	...	...
15.13	29.81	96.71	719.5	0.0832	M	B
14.27	22.55	93.77	629.8	0.1038	M	B
13.8	15.79	90.43	584.1	0.1007	M	B
15.85	23.95	103.7	782.7	0.08401	M	M
13.54	14.36	87.46	566.3	0.09779	B	B



Tabel 4. 14 Confusion Matrix *Default* Model D

Prediksi	Data Aktual	
	B	M
B	104	5
M	3	58

Dari Tabel 4.7 menunjukkan bahwa model *random forest* memprediksi terdapat 104 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 58 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 5 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 3 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.7 sebagai berikut.

$$Akurasi = \frac{104 + 58}{104 + 3 + 58 + 5} = 95.29\%$$

#### 4.1.4.2 Z-Score Tuning

Pada tahapan uji coba normalisasi menggunakan z-score dikombinasikan dengan *tuning* dan menggunakan 31 variabel didapatkan hasil sebagai berikut.

Tabel 4. 15 Hasil Prediksi Z-Score Model A

<i>radius_mean</i>	<i>texture_mean</i>	<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	Actual	Predicted
-0.9016	0.479	-0.8259	-0.8064	1.8728	M	B
-0.1468	1.2556	-0.1732	-0.2338	-0.2688	M	B
0.7698	0.0396	0.6762	0.6402	-1.5579	B	M
-0.0475	-0.5207	-0.0222	-0.1492	0.9414	M	B
-0.6718	-0.2673	-0.6983	-0.6359	0.236	M	B
...	...	...	...	...	...	...
0.0405	0.758	0.0741	-0.0713	0.529	M	B
-0.4731	1.1045	-0.3292	-0.5086	1.5813	M	M
0.5371	0.9185	0.4416	0.4061	-1.0168	M	M

<i>radius_mean</i>	<i>texture_mean</i>	<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	<i>Actual</i>	<i>Predicted</i>
0.1171	1.9182	0.1959	0.0111	1.2471	M	M
-1.3119	-1.5926	-1.3017	-1.0826	0.4294	B	B

Tabel 4. 16 Confusion Matrix Z-Score *Tuning* Model D

Prediksi	Data Aktual	
	B	M
B	106	5
M	1	58

Dari Tabel 4.8 menunjukkan bahwa model *random forest* memprediksi terdapat 106 data kanker payudara jinak dan hasil sebenarnya terdeteksi kanker payudara jinak, model *random forest* memprediksi terdapat 58 data kanker payudara ganas dan hasil sebenarnya terdeteksi kanker payudara ganas, model *random forest* memprediksi terdapat 5 data kanker payudara jinak dan hasil sebenarnya ternyata terdeteksi kanker payudara ganas, dan model *random forest* memprediksi terdapat 1 data kanker payudara ganas dan hasil sebenarnya ternyata kanker payudara jinak.

Perhitungan akurasi dari tabel 4.8 sebagai berikut.

$$Akurasi = \frac{106 + 58}{106 + 1 + 58 + 5} = 96.47\%$$

#### 4.1.5 Pengujian 10-Fold Cross Validation dan ROC

Pada penelitian ini *K-fold Cross Validation* digunakan untuk membagi dataset menjadi 10 *fold* dengan 3 kali pengulangan. *Metric ROC* digunakan untuk mengukur performa model *random forest* dalam membedakan antara kelas positif dan negatif (kanker jinak atau kanker ganas) dan akan menghasilkan nilai ROC untuk setiap *fold*. Kemudian, nilai ROC dari setiap *fold* akan diambil rata-ratanya

untuk mendapatkan nilai ROC akhir untuk penentuan berapa banyak jumlah *tuning mtry* terbaik untuk model *random forest*.

#### 4.1.5.1 Model A

Didapatkan jumlah *mtry* terbaik untuk model *random forest* model A yaitu 2 dengan ROC sebesar 99.11%, *Sensitivity* sebesar 97.51%, dan *Specificity* sebesar 93.34%. Untuk detailnya dari hasil *cross validation dan roc* sebagai berikut.

Tabel 4. 17 Cross Validation dan ROC Model A

<i>mtry</i>	ROC(SD)	Sens(SD)	Spec(SD)
2	0.9911 (0.0103)	0.9751 (0.0298)	0.9334 (0.0648)
3	0.9905 (0.0109)	0.9752 (0.0309)	0.9369 (0.0654)
4	0.9911 (0.0102)	0.9721 (0.0308)	0.9317 (0.0636)
6	0.9905 (0.0109)	0.9710 (0.0304)	0.9404 (0.0566)
7	0.9882 (0.0158)	0.9721 (0.0307)	0.9404 (0.0582)
9	0.9892 (0.0135)	0.9710 (0.0292)	0.9404 (0.0598)
10	0.9888 (0.0140)	0.9700 (0.0308)	0.9404 (0.0566)
12	0.9891 (0.0135)	0.9751 (0.0264)	0.9404 (0.0566)
13	0.9881 (0.0148)	0.9720 (0.0320)	0.9404 (0.0582)
15	0.9888 (0.0127)	0.9710 (0.0281)	0.9404 (0.0566)
16	0.9885 (0.0139)	0.9680 (0.0357)	0.9404 (0.0566)
18	0.9876 (0.0159)	0.9689 (0.0316)	0.9404 (0.0566)
19	0.9874 (0.0151)	0.9648 (0.0342)	0.9404 (0.0614)
21	0.9878 (0.0142)	0.9679 (0.0331)	0.9404 (0.0566)
22	0.9877 (0.0144)	0.9658 (0.0367)	0.9369 (0.0609)
24	0.9890 (0.0118)	0.9669 (0.0381)	0.9439 (0.0497)
25	0.9876 (0.0140)	0.9648 (0.0371)	0.9369 (0.0609)
27	0.9881 (0.0127)	0.9648 (0.0334)	0.9404 (0.0566)
28	0.9875 (0.0131)	0.9648 (0.0361)	0.9422 (0.0524)
30	0.9866 (0.0155)	0.9648 (0.0352)	0.9335 (0.0600)

ROC mengukur seberapa baik model membedakan antara kanker ganas dan kanker jinak. Semakin tinggi nilai ROC, semakin baik model dalam memprediksi kanker ganas dan kanker jinak. *Sensitivity* menunjukkan sejauh mana

model mampu mengidentifikasi kanker ganas yang sebenarnya. *Sensitivity* dinyatakan sebagai rasio antara jumlah kanker ganas yang benar diklasifikasikan dengan jumlah keseluruhan kanker ganas yang ada dalam data aktual. Semakin tinggi *Sensitivity*, semakin baik model dalam mengenali kanker ganas. *Specificity* menunjukkan sejauh mana model mampu mengidentifikasi kanker jinak yang sebenarnya. *Specificity* dinyatakan sebagai rasio antara jumlah kanker jinak yang benar diklasifikasikan dengan jumlah keseluruhan kanker jinak yang ada dalam data aktual. Semakin tinggi nilai *Specificity*, semakin baik model dalam mengenali kanker jinak. Tanda kurung pada *ROC*, *Sens*, dan *Spec* merupakan hasil standar deviasi. *ROC*, *Sensitivity*, dan *Specificity* merupakan metrik evaluasi klasifikasi yang digunakan untuk mengukur kinerja model. Sedangkan nilai standar deviasi pada *ROC*, *Sens*, dan *Spec* memberikan informasi tentang variasi nilai evaluasi dalam setiap kombinasi parameter. Jika nilai standar deviasi lebih kecil dari rata-rata pengujian dan semakin kecil nilai standar deviasinya maka semakin konsisten atau stabil performa model dalam membedakan antara kanker ganas dan kanker jinak pada setiap iterasi *repeated cross-validation*. Dengan melihat standar deviasi, dapat memperoleh pengetahuan tentang variasi performa model *random forest* berdasarkan jumlah parameter *mtry* yang diuji. Hal ini dapat membantu dalam pemilihan jumlah parameter *mtry* terbaik yang dapat menghasilkan performa yang lebih konsisten atau optimal untuk model *random forest* yang dibangun. Akan tetapi, pada penelitian ini acuan utama yang digunakan untuk menentukan jumlah *mtry* terbaik yaitu dengan melihat nilai ROC tertinggi.

#### 4.1.5.2 Model B

Didapatkan jumlah *mtry* terbaik untuk model *random forest* model B yaitu 2 dengan ROC sebesar 99.04%, *Sensitivity* sebesar 97.90%, dan *Specificity* sebesar 93.73%. Untuk detailnya dari hasil *cross validation dan roc* sebagai berikut.

Tabel 4. 18 Cross Validation dan ROC Model B

<i>mtry</i>	ROC(SD)	Sens(SD)	Spec(SD)
2	0.9904(0.0136)	0.9790(0.0270)	0.9373(0.0636)
3	0.9891(0.0136)	0.9766(0.0311)	0.9392(0.0588)
4	0.9886(0.0148)	0.9755(0.0349)	0.9392(0.0588)
6	0.9875(0.0165)	0.9778(0.0341)	0.9412(0.0618)
7	0.9862(0.0179)	0.9767(0.0350)	0.9353(0.0605)
9	0.9853(0.0194)	0.9779(0.0325)	0.9373(0.0617)
10	0.9853(0.0197)	0.9791(0.0313)	0.9412(0.0598)
12	0.9843(0.0208)	0.9755(0.0348)	0.9353(0.0643)
13	0.9844(0.0199)	0.9744(0.0329)	0.9353(0.0643)
15	0.9848(0.0195)	0.9755(0.0333)	0.9353(0.0643)
16	0.9839(0.0204)	0.9733(0.0353)	0.9412(0.0637)
18	0.9848(0.0192)	0.9779(0.0324)	0.9333(0.0613)
19	0.9844(0.0201)	0.9755(0.0359)	0.9353(0.0605)
21	0.9844(0.0198)	0.9756(0.0332)	0.9353(0.0605)
22	0.9830(0.0220)	0.9743(0.0368)	0.9333(0.0593)
24	0.9822(0.0222)	0.9732(0.0353)	0.9314(0.0657)
25	0.9840(0.0214)	0.9733(0.0326)	0.9314(0.0638)
27	0.9829(0.0217)	0.9756(0.0332)	0.9314(0.0638)
28	0.9841(0.0197)	0.9721(0.0361)	0.9294(0.0626)
30	0.9827(0.0214)	0.9721(0.0322)	0.9294(0.0698)

#### 4.1.5.3 Model C

Didapatkan jumlah *mtry* terbaik untuk model *random forest* model C yaitu 2 dengan ROC sebesar 99.04%, *Sensitivity* sebesar 97.90%, dan *Specificity* sebesar 93.73%. Untuk detailnya dari hasil *cross validation dan roc* sebagai berikut.

Tabel 4. 19 Cross Validation dan ROC Model C

<i>mtry</i>	ROC(SD)	Sens(SD)	Spec(SD)
2	0.9904(0.0136)	0.9790(0.0270)	0.9373(0.0636)

<i>mtry</i>	ROC(SD)	Sens(SD)	Spec(SD)
3	0.9891(0.0136)	0.9766(0.0311)	0.9392(0.0588)
4	0.9886(0.0148)	0.9755(0.0349)	0.9392(0.0588)
6	0.9875(0.0165)	0.9778(0.0341)	0.9412(0.0618)
7	0.9862(0.0179)	0.9767(0.0350)	0.9353(0.0605)
9	0.9853(0.0194)	0.9779(0.0325)	0.9373(0.0617)
10	0.9853(0.0197)	0.9791(0.0313)	0.9412(0.0598)
12	0.9843(0.0208)	0.9755(0.0348)	0.9353(0.0643)
13	0.9844(0.0199)	0.9744(0.0329)	0.9353(0.0643)
15	0.9848(0.0195)	0.9755(0.0333)	0.9353(0.0643)
16	0.9839(0.0204)	0.9733(0.0353)	0.9412(0.0637)
18	0.9848(0.0192)	0.9779(0.0324)	0.9333(0.0613)
19	0.9844(0.0201)	0.9755(0.0359)	0.9353(0.0605)
21	0.9844(0.0198)	0.9756(0.0332)	0.9353(0.0605)
22	0.9830(0.0220)	0.9743(0.0368)	0.9333(0.0593)
24	0.9822(0.0222)	0.9732(0.0353)	0.9314(0.0657)
25	0.9840(0.0214)	0.9733(0.0326)	0.9314(0.0638)
27	0.9829(0.0217)	0.9756(0.0332)	0.9314(0.0638)
28	0.9841(0.0197)	0.9721(0.0361)	0.9294(0.0626)
30	0.9827(0.0214)	0.9721(0.0322)	0.9294(0.0698)

#### 4.1.5.4 Model D

Didapatkan jumlah *mtry* terbaik untuk model *random forest* model D yaitu 2 dengan ROC sebesar 99.16%, *Sensitivity* sebesar 97.33%, dan *Specificity* sebesar 95.32%. Untuk detailnya dari hasil *cross validation dan roc* sebagai berikut.

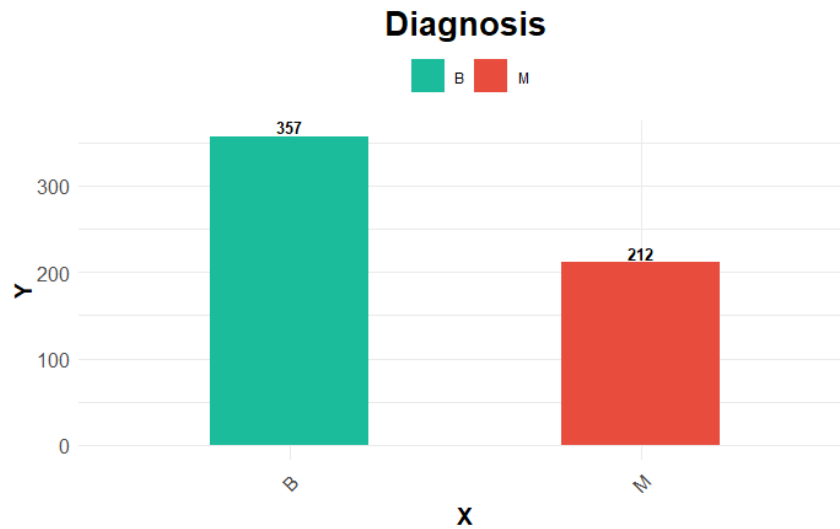
Tabel 4. 20 Cross Validation dan ROC Model D

<i>mtry</i>	ROC(SD)	Sens(SD)	Spec(SD)
2	0.9916(0.0123)	0.9733(0.0338)	0.9532(0.0558)
3	0.9910(0.0131)	0.9733(0.0338)	0.9486(0.0603)
4	0.9902(0.0151)	0.9760(0.0325)	0.9441(0.0684)
6	0.9892(0.0164)	0.9747(0.0323)	0.9397(0.0690)
7	0.9895(0.0163)	0.9760(0.0325)	0.9463(0.0645)
9	0.9887(0.0171)	0.9760(0.0325)	0.9419(0.0698)
10	0.9880(0.0181)	0.9760(0.0325)	0.9419(0.0698)
12	0.9886(0.0171)	0.9747(0.0323)	0.9419(0.0676)
13	0.9881(0.0176)	0.9747(0.0323)	0.9397(0.0690)
15	0.9881(0.0176)	0.9733(0.0321)	0.9397(0.0667)

<i>mtry</i>	<b>ROC(SD)</b>	<b>Sens(SD)</b>	<b>Spec(SD)</b>
16	0.9884(0.0168)	0.9747(0.0306)	0.9397(0.0667)
18	0.9875(0.0177)	0.9747(0.0306)	0.9463(0.0691)
19	0.9878(0.0169)	0.9760(0.0270)	0.9419(0.0698)
21	0.9870(0.0178)	0.9747(0.0340)	0.9419(0.0653)
22	0.9873(0.0170)	0.9760(0.0290)	0.9419(0.0653)
24	0.9866(0.0180)	0.9787(0.0273)	0.9463(0.0713)
25	0.9867(0.0181)	0.9760(0.0308)	0.9419(0.0720)
27	0.9863(0.0175)	0.9800(0.0229)	0.9463(0.0645)
28	0.9864(0.0182)	0.9773(0.0272)	0.9441(0.0684)
30	0.9865(0.0171)	0.9760(0.0290)	0.9463(0.0669)

## 4.2 Pembahasan

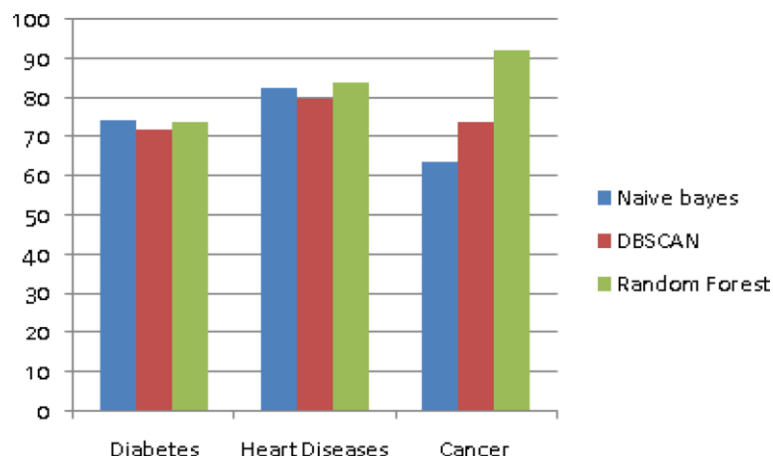
Pada tahapan ini didapatkan dari hasil uji coba pada sistem. Dataset diperoleh dari *UCI Machine Learning Repository* dengan nama dataset *Breast Cancer Wisconsin (Diagnostic)*. Total data yang didapat berjumlah 569 dan memiliki 32 atribut. 10 atribut utama yaitu *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, *fractal dimension*. Masing-masing 10 atribut utama tadi memiliki 3 indikator yaitu *mean*, *standard error/se*, dan *worst*. Kemudian untuk 2 atribut lainnya yaitu *ID number* dan juga *Diagnosis*. *Diagnosis* memiliki 2 kelas dengan nilai M sebagai kanker ganas dan B sebagai kanker jinak. Pada gambar berikut terlihat adanya ketidakseimbangan antar jumlah kelas M dan B, sebanyak 212 pasien mengidap kanker ganas dan 357 pasien mengidap kanker jinak.



Gambar 4. 1 Perbandingan Diagnosis

Penelitian yang dilakukan oleh Jackins et al., dalam memprediksi beberapa penyakit klinis yang salah satunya menggunakan dataset yang sama yaitu *Breast Cancer Wisconsin (Diagnostic)* dengan menggunakan metode *Naïve Bayes* dan *Random forest*. Untuk metode *random forest* menghasilkan akurasi 92.40% dan *Naïve Bayes* menghasilkan akurasi 63.74%. Pada jurnal peneliti menggunakan *K-means clustering* dan *DBSCAN (density-based spatial clustering of applications with noise)* untuk menemukan keefektifan algoritma yang diusulkan pada masing-masing dataset (Jackins et al., 2021). Untuk detailnya dapat dilihat pada gambar berikut.





Gambar 4. 2 Hasil Penelitian Jackins dkk

Penelitian yang dilakukan oleh Andryan et al., pada dataset yang sama yaitu *Breast Cancer Wisconsin (Diagnostic)* dengan menggunakan metode *XGBoost* dan *SVM*. Pada jurnal peneliti teknik *KDD (Knowledge Discovery in Databases)* yang merupakan metode saintifik pada data mining. Peneliti juga menggunakan teknik reduksi dimensi yaitu *PCA*. Untuk metode *XGBoost* menghasilkan akurasi 96.26% dan *SVM* menghasilkan akurasi 94.28% (Andryan et al., 2022). Untuk detailnya dapat dilihat pada tabel berikut.

Tabel 4. 21 Hasil Penelitian Andryan dkk

	<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>roc_auc</i>
<i>SVM</i>	0.942857143	0.936738737	0.913445378	0.975800215
<i>XGB</i>	0.962637	0.969548	0.935798	0.986499

Penelitian yang dilakukan oleh Adebityi et al., pada dataset yang sama yaitu *Breast Cancer Wisconsin (Diagnostic)* dengan menggunakan metode *Random forest* dan *SVM*. Peneliti menggunakan *tools python*. Kemudian usulan yang dilakukan peneliti yaitu dengan membandingkan 2 metode tadi dengan teknik fitur ekstraksi *LDA (Linear Discriminant Analysis)*. Untuk metode *Random forest default* yaitu 94.7% dan *Random forest* dengan *LDA* sebesar 95.6%. Kemudian metode

SVM default 92.1% dan SVM dengan LDA sebesar 96.4% (Adebiyi et al., 2022).

Untuk detailnya dapat dilihat pada tabel berikut.

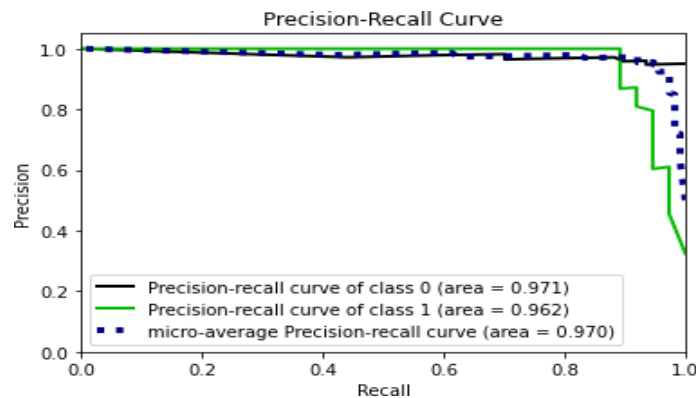
Tabel 4. 22 Hasil Penelitian Adebiyi dkk

Performance Metrics (%)	Random forest	SVM	LDA + Random forest	LDA + SVM	Formula
Accuracy	94.7	92.1	95.6	96.4	$(TP + TN)/(P + N)$
Sensitivity	95.5	93.9	95.6	95.7	$TP/(TP + FN)$
Specificity	93.6	89.5	95.7	97.8	$TN/(FP + TN)$
Precision	97.0	92.5	97.0	96.4	$TP/(TP + FP)$
F1 score	95.5	92.2	96.3	97.8	$2 TP/(2 TP + FP + FN)$

Penelitian yang dilakukan oleh Vinutha et al., pada dataset yang sama yaitu *Breast Cancer Wisconsin (Diagnostic)* dengan menggunakan metode *Random forest*. Kemudian optimasi yang dilakukan peneliti yaitu dengan *grid search* untuk mencari *n\_estimator* atau jumlah pohon terbaik pada model *random forest* menggunakan *python*. Hasil yang didapat dari metode *Random forest default* yaitu akurasi sebesar 95.61%, *presisi-recallnya* sebesar 96.2% untuk memprediksi kanker payudara ganas dan sebesar 97.1% untuk memprediksi kanker payudara jinak (Vinutha et al., 2022). Untuk detailnya dapat dilihat pada gambar berikut.

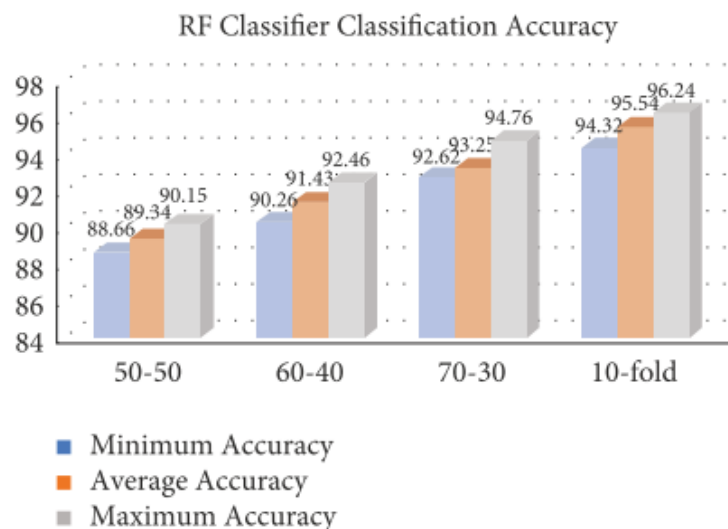


Gambar 4. 3 Hasil Confusion Matrix Peneliti Vinutha dkk



Gambar 4. 4 Hasil Presisi-Recall Peneliti Vinutha dkk

Penelitian yang dilakukan oleh Bhardwaj et al., pada dataset yang sama yaitu *Breast Cancer Wisconsin (Diagnostic)* dengan menggunakan metode *MLP*, *KNN*, *GP*, dan *Random forest*. Peneliti menggunakan skenario pengujian dengan cara membagi data train dan test menjadi beberapa bagian yaitu 50:50, 60:40, 70:30, dan *10-fold*. Kemudian metode yang ditentukan dicoba dalam setiap skenario pengujian yang peneliti inginkan. Untuk detailnya dapat dilihat pada gambar berikut (Bhardwaj et al., 2022).



Gambar 4. 5 Hasil *Random forest* Bhardwaj dkk

Classifier	Partition scheme	Sensitivity (%)	Precision (%)	Specificity (%)	F1 score (%)
MLP	50-50	66.73	65.42	64.58	65.36
	60-40	68.51	67.35	66.34	67.48
	70-30	70.36	69.63	68.21	69.52
	10-fold	72.46	71.38	70.62	71.58
KNN	50-50	74.63	73.26	72.37	73.34
	60-40	76.39	75.57	74.18	75.68
	70-30	78.82	77.26	76.43	77.34
	10-fold	78.62	79.34	78.62	79.28
GP	50-50	84.73	83.62	82.15	83.46
	60-40	86.93	85.76	84.83	85.62
	70-30	88.46	87.25	86.53	87.44
	10-fold	90.22	89.62	88.72	89.48
RF	50-50	90.86	89.49	88.51	89.36
	60-40	92.63	91.38	90.68	91.52
	70-30	94.27	93.67	92.64	93.74
	10-fold	96.29	95.45	94.48	95.56

Gambar 4. 6 Hasil Penelitian Bhardwaj dkk

Pada penelitian ini menggunakan metode *random forest* pada dataset yang sama. Dilakukan *preprocessing* dengan beberapa tahapan dalam proses pengolahan data yaitu dengan menggunakan data cleaning dan data exploration. Pertama, dengan menghapus kolom ID dan kolom X yang berisi nilai *NA (Not Available)*. Kedua, dengan mengubah tipe data *diagnosis* dari *character/string* menjadi *factor/kategori*. Terakhir dengan *Data Scaling* menggunakan teknik normalisasi *z-score* atau *zero-mean* yang berguna untuk meningkatkan kecepatan dan akurasi dari sistem (Ambarwari et al., 2020).

Kemudian dilakukan pembagian data menjadi 4 bagian, yaitu model A dengan perbandingan 90% data *train* : 10% data *test*, model B dengan perbandingan 80% data *train* : 20% data *test*, model C dengan perbandingan 75% data *train* : 25% data *test*, dan model D dengan perbandingan 70% data *train* : 30% data *test*. Dengan pembagian beberapa dataset yang berbeda akan didapatkan nilai ketepatan prediksi terbaik. (Almais et al., 2022). Langkah berikutnya membandingkan metode *random forest* pada data latih tanpa dinormalisasi dengan metode *random forest* pada data latih yang telah di normalisasi dan juga dilakukan *tuning*. Untuk model yang

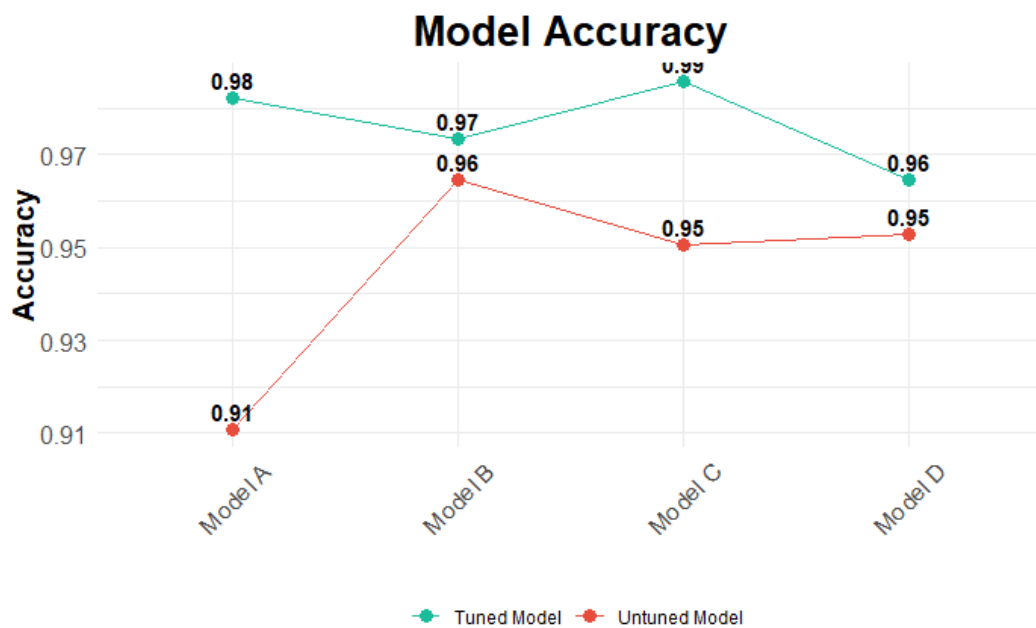
dituning data dibagi menjadi beberapa fold, pada setiap *fold* model *random forest* di *training* dengan mencari kombinasi *mtry* terbaik kemudian diuji pada *fold* tersebut dan dievaluasi dengan *metric ROC*. Setelah model *random forest* dilatih, *confusion matrix* digunakan untuk menunjukkan jumlah prediksi yang benar dan salah dari model *random forest* pada setiap kelas, dibandingkan dengan label kelas aktual dari data uji. Berikut tabel hasil akurasi dari masing-masing model.

Tabel 4. 23 Akurasi Setiap Model

Model	Banyaknya Data = 569				Seed	Akurasi	
	Train		Test			Default	Scaling + Tuning
	Jumlah	Persentase	Jumlah	Persentase			
A	513	90%	56	10%	1234	91.07%	98.21%
B	456	80%	113	20%	1234	96.46%	97.35%
C	427	75%	142	25%	1234	95.07%	98.59%
D	399	70%	170	30%	1234	95.29%	96.47%

Pada Tabel 4.23 model *Random forest* memberikan performa yang baik pada dataset yang digunakan, hal ini menunjukkan kemampuan model untuk mengklasifikasikan data dengan baik. Hasil akurasi terbaik diperoleh dari model C yang telah dinormalisasi dan di *tuning*, dengan menggunakan rasio perbandingan 75% data *train* (427 data *train*) dan data *test* 25% (142 data *test*) diperoleh hasil akurasi sebesar 98.59%. Dapat dilihat pada Tabel 4.23 data yang telah dinormalisasi mengalami peningkatan akurasi hal ini dikarenakan data akan memiliki skala yang seragam dengan rata-rata 0 dan standar deviasi 1, pada dataset asli memiliki skala yang beragam. Dalam *Random forest*, penggunaan skala yang seragam dapat menghindari beberapa fitur dengan skala besar yang akan mempengaruhi dalam pembentukan pohon keputusan. Untuk hasil akurasi dari *random forest* secara *default* mengalami kenaikan akan tetapi pada model A terjadi penurunan, hal ini

dapat terjadi akibat *overfitting* dimana pembentukan pohon pada penelitian ini yaitu sebesar 500 dan data uji yang digunakan hanya sebanyak 56. Untuk hasil akurasi dari *random forest* dengan menggunakan normalisasi dan *tuning* lebih stabil dikarenakan memanfaatkan penggunaan *repeated k-fold cross validation* dan *ROC*. Berikut visualisasi diagram garis dari masing-masing model.



Gambar 4. 7 Diagram Nilai Akurasi Tiap Model

Berikut merupakan hasil perbandingan akurasi yang didapatkan oleh peneliti dengan peneliti lainnya.

Tabel 4. 24 Hasil Perbandingan Akurasi

Referensi	Metode	Pengolahan Data	Hasil Akurasi Terbaik
Jackins et al., (2021)	<i>Naïve Bayes &amp; Random forest</i>	<i>K-Means</i> dan <i>DBSCAN</i>	92.40% <i>Random forest</i>
Andryan et al., (2022)	<i>SVM &amp; XGBoost</i>	<i>PCA</i>	96.28% <i>XGBoost</i>
Adebiyi et al., (2022)	<i>Random forest &amp; SVM</i>	<i>LDA</i>	95.60% <i>Random forest</i>
Vinutha et al., (2022)	<i>Random forest &amp; SVM</i>	<i>Grid Search</i> <i>n_estimator</i>	95.61% <i>Random forest</i>
Bhardwaj et al., (2022)	<i>MLP, KNN, GP, dan Random forest</i>	<i>10 Fold</i>	96.24% <i>Random forest</i>

Referensi	Metode	Pengolahan Data	Hasil Akurasi Terbaik
Peneliti	<i>Random forest</i>	<i>Repeated K-Fold Cross Validation, ROC, Normalisasi Z-Score, Grid Search mtry</i>	<b>98.59%</b> <i>Random forest</i>

Pada Tabel 4.24 dapat dilihat bahwa terdapat beberapa teknik usulan yang dilakukan oleh para peneliti. Mulai dari teknik *clustering* yaitu *k-means* dan *dbscan*, teknik reduksi dimensi yaitu *PCA* dan *LDA*, *Grid Search* untuk mencari jumlah pohon yang terbaik, *Cross Validation 10 Fold*. *Repeated k-fold cross validation* digunakan untuk membagi data menjadi beberapa fold, pada setiap *fold* model *random forest* di *training* dengan mencari kombinasi *mtry* terbaik kemudian diuji pada *fold* tersebut dan dievaluasi dengan *metric ROC*, dengan menggunakan data validasi yang berbeda pada setiap iterasi, *k-fold cross-validation* memberikan perkiraan yang lebih stabil dan lebih objektif tentang kinerja model, serta mengatasi bias dan *overfitting* untuk model *random forest* (Mahesh et al., 2022). Normalisasi *z-score* digunakan untuk skala yang seragam pada dataset dengan rata-rata 0 dan standar deviasi 1 (Ambarwari et al., 2020). *Grid search* digunakan untuk mencari *mtry* terbaik atau berapa banyak jumlah fitur yang digunakan pada masing-masing *split node* saat pembentukan pohon *random forest* dengan menggunakan kombinasi *repeated k-fold* dan *ROC*. Dapat dikatakan bahwa kinerja *Random forest* dengan menggunakan seluruh fitur pada dataset Breast Cancer Wisconsin (Diagnostic) bekerja dengan baik tanpa memerlukan teknik *clustering* maupun teknik reduksi dimensi dengan menghasilkan akurasi sebesar 98.59%.

Nilai klasifikasi dapat dikelompokkan menjadi beberapa kelompok berdasarkan rentang nilainya. Jika rentang nilai adalah 90% - 100%, maka klasifikasi dikatakan sangat baik. Rentang nilai 80% - 90% diklasifikasikan sebagai baik, rentang nilai 70% - 80% dikategorikan sebagai cukup baik, rentang nilai 60% - 70% dikelompokkan dalam klasifikasi kurang baik, sedangkan rentang nilai 50% - 60% dikategorikan sebagai gagal (Gorunescu, 2011). Berdasarkan rentang nilai klasifikasi tersebut, maka keseluruhan model termasuk dalam kelompok klasifikasi sangat baik.

Untuk hasil pengujian *repeated 10-fold cross validation* dengan 3 kali pengulangan pada masing-masing model didapatkan jumlah *mtry* terbaik yaitu 2. Model A dengan ROC sebesar 99.11%, *Sensitivity* sebesar 97.51%, dan *Specificity* sebesar 93.34%. Model B dengan ROC sebesar 99,04%, *Sensitivity* sebesar 97.90%, dan *Specificity* sebesar 93.73%. Model C dengan ROC sebesar 99.04%, *Sensitivity* sebesar 97.90%, dan *Specificity* sebesar 93.73%. Model D dengan ROC sebesar 99.16%, *Sensitivity* sebesar 97.33%, dan *Specificity* sebesar 95.32%. Terdapat perbedaan pada akurasi yang didapatkan dikarenakan perbedaan pada jumlah observasi data dan perbedaan model dalam memilih data latih.

### **4.3 Integrasi Penelitian Dalam Tafsir Al-Qur'an**

Dalam Al-Qur'an terdapat ayat yang mengajarkan tentang pentingnya saling membantu dan berbuat kebajikan untuk kepentingan bersama, serta menunjukkan bahwa segala bentuk bantuan dan manfaat yang diberikan haruslah didasarkan pada kebaikan dan ketakwaan kepada Allah yaitu QS. Al-Ma'idah ayat 2:.



وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ ۗ وَاتَّقُوا اللَّهَ ۖ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

“Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan pelanggaran. Dan bertakwalah kamu kepada Allah, sesungguhnya Allah amat berat siksa-Nya.” (QS. Al-Maidah: 2)

Penelitian ini dibangun sebagai salah satu cara dalam mengimplementasikan ayat diatas, dengan cara memberikan manfaat tentang bagaimana mengklasifikasi kanker payudara menggunakan metode *Random forest* yang peneliti harapkan guna mengetahui seseorang mengidap kanker payudara jinak atau kanker payudara ganas, juga mengetahui kinerja dari metode tersebut dan penelitian yang telah dilakukan dapat memberikan informasi terkait hal-hal yang dapat menyebabkan seseorang terkena kanker payudara. Sistem yang telah dibangun dapat membantu mengantisipasi kanker payudara sejak dini dan dapat dilakukan penanganan yang sesuai. Terdapat hadits yang berkaitan dengan penelitian ini yaitu hadits riwayat Muslim, Dari Abu Hurairah radhiyallahu ‘anhu, Rasulullah shallallahu ‘alaihi wa sallam bersabda:

مَنْ نَفَسَ عَنْ مُؤْمِنٍ كَرْبَةً مِنْ كَرْبِ الدُّنْيَا نَفَسَ اللَّهُ عَنْهُ كَرْبَةً مِنْ كَرْبِ يَوْمِ الْقِيَامَةِ وَمَنْ يَسَّرَ عَلَىٰ مُعْسِرٍ يَسَّرَ اللَّهُ عَلَيْهِ فِي الدُّنْيَا وَالْآخِرَةِ وَمَنْ سَتَرَ مُسْلِمًا سَتَرَهُ اللَّهُ فِي الدُّنْيَا وَالْآخِرَةِ وَاللَّهُ فِي عَوْنِ الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ أَخِيهِ وَمَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ وَمَا اجْتَمَعَ قَوْمٌ فِي بَيْتٍ مِنْ بُيُوتِ اللَّهِ يَتْلُونَ كِتَابَ اللَّهِ وَيَتَدَارَسُونَهُ بَيْنَهُمْ إِلَّا نَزَلَتْ عَلَيْهِمُ السَّكِينَةُ وَعَشِيَتْهُمُ الرَّحْمَةُ وَحَفَّتْهُمُ الْمَلَائِكَةُ وَذَكَرَهُمُ اللَّهُ فِيمَنْ عِنْدَهُ وَمَنْ بَطَأَ بِهِ عَمَلُهُ لَمْ يُسْرِعْ بِهِ نَسَبُهُ.

“Barangsiapa membebaskan seorang mukmin dari suatu kesulitan dunia, maka Allah akan membebaskannya dari suatu kesulitan pada hari kiamat. Barangsiapa memberi kemudahan kepada orang yang berada dalam kesulitan, maka Allah akan memberikan kemudahan di dunia dan akhirat. Barangsiapa menutupi aib seorang muslim, maka Allah akan menutup aibnya di dunia dan akhirat. Allah akan selalu menolong hamba-Nya selama hamba tersebut menolong saudaranya sesama

*muslim. Barangsiapa menempuh jalan untuk mencari ilmu, maka Allah akan memudahkan jalan ke surga baginya. Tidaklah sekelompok orang berkumpul di suatu masjid (rumah Allah) untuk membaca al-Qur'an, melainkan mereka akan diliputi ketenangan, rahmat, dan dikelilingi para malaikat, serta Allah akan menyebut-nyebut mereka pada malaikat-malaikat yang berada di sisi-Nya. Barangsiapa yang ketinggalan amalnya, maka nasibnya tidak juga meninggikannya.”* (Hadits Riwayat Muslim, Shahih Muslim, juz VIII, hal. 71, hadits no. 7028, dari Abu Hurairah r.a.)

Hadits ini menekankan pentingnya memberikan manfaat bagi sesama dan membantu meringankan beban mereka. Sebagai umat Muslim harus berusaha untuk selalu berbuat kebaikan dan memberikan manfaat bagi orang lain, termasuk dalam hal pencegahan dan pengobatan penyakit seperti kanker payudara. Dengan demikian, akan mendapatkan keberkahan dan bantuan dari Allah dalam menjalani kehidupan di dunia maupun di akhirat.

Dalam Al-Qur'an juga diajarkan untuk menjaga kesehatan diri terdapat pada surah Al-Baqarah ayat 195:

وَأَنْفِقُوا فِي سَبِيلِ اللَّهِ وَلَا تُلْقُوا بِأَيْدِيكُمْ إِلَى التَّهْلُكَةِ وَأَحْسِنُوا إِنَّ اللَّهَ يُحِبُّ الْمُحْسِنِينَ

*“Dan belanjakanlah (harta bendamu) di jalan Allah, dan janganlah kamu menjatuhkan dirimu sendiri ke dalam kebinasaan, dan berbuat baiklah, karena sesungguhnya Allah menyukai orang-orang yang berbuat baik.”* (QS. Al-Baqarah: 195)

Mereka yang tidak memperhatikan kesehatannya termasuk dalam kelompok yang merugikan diri sendiri dan mengalami kehancuran. Hal ini karena mereka tidak memelihara nikmat yang telah diberikan oleh Allah kepada mereka. Sebab, tidak merawat apa yang telah diberikan oleh Allah. Karena tubuh yang sehat adalah modal untuk melakukan ibadah dan amal yang baik. Oleh karena itu, harus

memelihara dan merawat tubuh dengan baik, sehingga dapat terhindar dari berbagai penyakit dan masalah kesehatan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 KESIMPULAN

Berdasarkan pengujian yang telah dilakukan pada bab sebelumnya, untuk mengetahui performa metode *Random forest* pada sistem dalam mengklasifikasi kanker payudara. Setelah dilakukan pembagian data menjadi 4 bagian dengan perbandingan 90% data *train* : 10% data *test*, 80% data *train* : 20% data *test*, 75% data *train* : 25% data *test*, 70% data *train* : 30% data *test* dan dilakukan perbandingan antara model *random forest* menggunakan data yang tidak dinormalisasi atau secara *default* dengan model *random forest* menggunakan data yang dinormalisasi menggunakan *z-score* dan di *tuning* menggunakan teknik *Grid Search*. Model C yang telah dinormalisasi dan di *tuning* dengan perbandingan data *train* 75% dan 25 % data uji menghasilkan nilai akurasi tertinggi sebesar 98.59%. Model C berhasil mengklasifikasi 140 data uji dari 142 data uji dengan benar, Model C salah mengklasifikasi 2 data kanker payudara yang aslinya ganas sebagai kanker payudara jinak. Berdasarkan hasil akurasi dari Model C menunjukkan bahwa model dapat mengklasifikasi data target dengan data prediksi dengan kategori sangat baik. Untuk hasil pengujian *repeated 10-fold cross validation* dengan 3 kali pengulangan pada masing-masing model didapatkan jumlah *mtry* terbaik yaitu 2. Untuk performa *Random forest* dengan menggunakan seluruh fitur pada dataset *Breast Cancer Wisconsin (Diagnostic)* bekerja dengan baik tanpa memerlukan teknik clustering maupun teknik reduksi dimensi.

## 5.2 SARAN

Penulis menyadari bahwa penelitian yang dilakukan tentu belum sempurna. Berikut adalah beberapa saran yang dapat penulis berikan untuk penelitian selanjutnya:

- a. Dapat mencoba menggunakan paket *Random forest* yang lainnya dari bahasa pemrograman R maupun bahasa pemrograman lainnya, dikarenakan masing-masing paket *Random forest* memiliki *default* parameter yang berbeda.
- b. Dapat mencoba teknik *tuning hyperparameter* lainnya untuk *Random forest*
- c. Untuk metode *Random forest* dapat mencoba teknik normalisasi selain z-score, teknik clustering selain K-Means dan DBSCAN, teknik reduksi selain LDA, teknik class imbalance dikarenakan jumlah kanker ganas dan kanker jinak pada dataset tidak seimbang.
- d. Dapat mencoba dan menjelajahi paket *caret* untuk *method* pada fungsi *trainControl* selain *method repeatedcv* dan *metric* pada fungsi *train* selain *ROC*
- e. Dapat mencoba metode Machine Learning lain untuk klasifikasi kanker payudara pada dataset *Breast Cancer Wisconsin (Diagnostic)* atau data kanker payudara yang berbeda.

## DAFTAR PUSTAKA

- A. Vincent, J. P. J. (2022). Komparasi Tingkat Akurasi Random Forest Dan Knn Untuk Mendiagnosis Penyakit Kanker Payudara. *Universitas Pelita Harapan PSDKU Medan Jurusan Sistem Informasi*, 7(1), 49–61.
- Adebiyi, M. O., Arowolo, M. O., Mshelia, M. D., & Olugbara, O. O. (2022). A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. *Applied Sciences (Switzerland)*, 12(22). <https://doi.org/10.3390/app122211455>
- Ahmad Hania, A. (2017). Mengenal Artificial Intelligence, Machine Learning, & Deep Learning. *Jurnal Teknologi Indonesia*, 1(June), 1–6. <https://amt-it.com/mengenal-perbedaan-artificial-intelligence-machine-learning-deep-learning/>
- Almais, A. T. W., Cahyo Crysdiyan, Holle, K. F. H., & Akbar Roihan. (2022). Smart Assessment Menggunakan Backpropagation Neural Network Smart Assessment using Backpropagation Neural Network. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(3), 525–538. <https://doi.org/DOI: 10.30812/matrik.v21i3.1382>
- Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 117–122.
- Andryan, M. R., Fajri, M., & Sulistyowati, N. (2022). Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (Svm) Untuk Diagnosis Penyakit Kanker Payudara. *JIKO (Jurnal Informatika Dan Komputer)*, 6(1), 1. <https://doi.org/10.26798/jiko.v6i1.500>
- Aprilia, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi*, 10(1), 163. <https://doi.org/10.32520/stmsi.v10i1.1129>
- Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81–86. <https://doi.org/10.33096/ilkom.v12i2.507.81-86>
- Azmi, A. N., Kurniawan, B., Siswandi, A., & Detty, A. U. (2020). Hubungan Faktor Keturunan Dengan Kanker Payudara DI RSUD Abdoel Moeloek. *Jurnal Ilmiah Kesehatan Sandi Husada*, 12(2), 702–707.

- <https://doi.org/10.35816/jiskh.v12i2.373>
- Bhardwaj, A., Bhardwaj, H., Sakalle, A., Uddin, Z., Sakalle, M., & Ibrahim, W. (2022). Tree-Based and Machine Learning Algorithm Analysis for Breast Cancer Classification. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/6715406>
- Breiman, L. (2014). Random Forests. In *Machine Learning* (Vol. 45, Issue 1). Cambridge University Press; 1st edition. <https://doi.org/doi.org/10.1023/A:1010933404324>
- Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67–80. <https://doi.org/10.37424/informasi.v12i1.48>
- Christopher M. Bishop. (2006). Pattern Recognition and Machine Learning. In *EAI/Springer Innovations in Communication and Computing*. [https://doi.org/10.1007/978-3-030-57077-4\\_11](https://doi.org/10.1007/978-3-030-57077-4_11)
- Edriss, E., Ali, E., & Feng, W. Z. (2016). Breast Cancer Classification using Support Vector Machine and Neural Network. *International Journal of Science and Research (IJSR)*, 5(3), 1–6. <https://doi.org/10.21275/v5i3.nov161719>
- Faisal, M., Wahono, D., Purnama, I. K. E., Hariadi, M., & Purnomo, M. H. (2014). Classification of diabetic retinopathy patients using support vector machines (SVM) based on digital retinal image. *Journal of Theoretical and Applied Information Technology*, 59(1), 197–204.
- Fauzi, A., Supriyadi, R., & Maulidah, N. (2020). Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest. *Jurnal Infortech*, 2(1), 96–101. <https://doi.org/10.31294/infortech.v2i1.8079>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Globocan. (2020). *Cancer in Indonesia*. Global Cancer Observatory. <https://doi.org/10.1001/jama.247.22.3087>
- Gorunescu, F. (2011). *Data Mining (Concepts, Models and Techniques)*. Springer Berlin, Heidelberg. <https://doi.org/https://doi.org/10.1007/978-3-642-19721-5>
- Hero, S. K. (2020). Faktor Resiko Kanker Payudara. *Jurnal Bagus*, 02(01), 402–406.

- Husni, M., Romadoni, S., & Rukiyati, D. (2015). Hubungan Dukungan Keluarga dengan Kualitas Hidup Pasien Kanker Payudara di Instalasi Rawat Inap Bedah RSUP Dr. Mohammad Hoesin Palembang Tahun 2012. *Jurnal Keperawatan Sriwijaya*, 2(2), 77–83.
- Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *Journal of Supercomputing*, 77(5), 5198–5219. <https://doi.org/10.1007/s11227-020-03481-x>
- Jovel, J., & Greiner, R. (2021). An Introduction to Machine Learning Approaches for Biomedical Research. *Frontiers in Medicine*, 8(December), 1–15. <https://doi.org/10.3389/fmed.2021.771607>
- Junge, M. R. J., & Dettori, J. R. (2018). ROC Solid: Receiver Operator Characteristic (ROC) Curves as a Foundation for Better Diagnostic Tests. *Global Spine Journal*, 8(4), 424–429. <https://doi.org/10.1177/2192568218778294>
- Kamelia, M., & Agus, S. (2021). Fine Needle Aspiration Biopsy (FNAB) Massa Intraabdomen dipandu Ultrasonografi. *Health and Medical Journal*, 4(1), 55–61. <https://doi.org/10.33854/heme.v4i1.819>
- Lantz, B. (2013). *Machine Learning With R*. <https://doi.org/10.4018/978-1-7998-2718-4.ch015>
- Larose, D. T. (2005). Discovering knowledge in data: an introduction to data mining. In *Wiley-Interscience*. <https://doi.org/10.1002/9781118874059>
- Lestandy, M. (2022). Deteksi Dini Kanker Payudara Menggunakan Metode Convolution Neural Network (CNN). *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 12(1), 65. <https://doi.org/10.35585/inspir.v12i1.2667>
- Li, X. F., Huang, Y. Z., Tang, J. Y., Li, R. C., & Wang, X. Q. (2021). Development of a random forest model for hypotension prediction after anesthesia induction for cardiac surgery. *World Journal of Clinical Cases*, 9(29), 8729–8739. <https://doi.org/10.12998/wjcc.v9.i29.8729>
- Ma'arif, F., & Arifin, T. (2017). Optimasi Fitur Menggunakan Backward Elimination dan Algoritma SVM untuk Klasifikasi Kanker Payudara. *Jurnal Informatika*, 4(1), 46–53.
- Macaulay, B. O., Aribisala, B. S., Akande, S. A., Akinnuwesi, B. A., & Olabanjo, O. A. (2021). Breast cancer risk prediction in African women using Random Forest Classifier. *Cancer Treatment and Research Communications*, 28, 7. <https://doi.org/10.1016/j.ctarc.2021.100396>



- Mahesh, T. R., Kaladevi, A. C., Balajee, J. M., Vivek, V., Prabu, M., & Muthukumar, V. (2022). An Efficient Ensemble Method Using K-Fold Cross Validation for the Early Detection of Benign and Malignant Breast Cancer. *International Journal of Integrated Engineering*, 14(7), 204–216. <https://doi.org/10.30880/ijie.2022.14.07.015>
- Matloff, N. (2011). *The Art of R Programming: A Tour of Statistical Software Design*.
- Mitra, R., & Rajendran, T. (2022). Efficient Prediction of Stroke Patients Using Random Forest Algorithm in Comparison to Support Vector Machine. In *Advances in Parallel Computing Algorithms, Tools and Paradigms* (pp. 530–536). <https://doi.org/10.3233/apc220075>
- Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In *Communications in Computer and Information Science: Vol. 1234 CCIS*. Springer Singapore. [https://doi.org/10.1007/978-981-15-7205-0\\_10](https://doi.org/10.1007/978-981-15-7205-0_10)
- Moreira, L. B., & Namen, A. A. (2018). A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Computer Methods and Programs in Biomedicine*, 165, 139–149. <https://doi.org/10.1016/j.cmpb.2018.08.016>
- Muntiari, N. R., & Hanif, K. H. (2022). Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning. *Jurnal Ilmu Komputer Dan Teknologi*, 3(1), 1–6. <https://doi.org/10.35960/ikomti.v3i1.766>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711. <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- Nugraha, F. S., Shidiq, M. J., & Rahayu, S. (2019). Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara. *Jurnal Pilar Nusa Mandiri*, 15(2), 149–156. <https://doi.org/10.33480/pilar.v15i2.601>
- Nugraha, R. H., Yuwono, E., Prasetyohadi, L., Arief, Y. B., & Patria, H. (2022). Analisis Konsumsi Energi Listrik Pelanggan Dan Biaya Pokok Produksi Penyediaan Energi Listrik dengan Machine Learning. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), 47–56. <http://dx.doi.org/10.30645/jsakti.v6i1.424>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search. *SISTEMASI : Jurnal Sistem Informasi*, 11(2),

391–401.

- Nurhayati, N., Arifin, Z., & Hardono, H. (2019). Kejadian Kanker Payudara (Studi Retrospektif) Di Lampung, Indonesia. *Holistik Jurnal Kesehatan*, 13(2), 172–183. <https://doi.org/10.33024/hjk.v13i2.1052>
- Patel, A. (2020). Benign vs Malignant Tumors. *JAMA Oncology*, 6(9), 1488. <https://doi.org/10.1001/jamaoncol.2020.2592>
- Punitha, S., Amuthan, A., & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal*, 3(2), 348–358. <https://doi.org/10.1016/j.fcij.2018.10.005>
- Rahayuwati, L., Rizal, Iqbal, A., Lukman, M., & Juniarti, N. (2020). Pendidikan Kesehatan tentang Pencegahan Penyakit Kanker dan Menjaga Kualitas Kesehatan. *Media Karya Kesehatan*, 3(1), 59–69.
- Resmiati, R., & Arifin, T. (2021). Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination. *Sistemasi*, 10(2), 381. <https://doi.org/10.32520/stmsi.v10i2.1238>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Saputra Liambo, I., Frisithody, A., & Hajrul Malaka, M. (2022). Review: Patofisiologi, Epidemiologi, dan Lini Sel Kanker Payudara Review: Pathophysiology, Epidemiology, and Cell Line of Breast Cancer. *Pharmauho: Jurnal Farmasi*, 8(1), 17–22. <https://doi.org/10.33772/pharmauho.v8i1>
- Sari, N. P. W. P. (2019). Women Living With Breast and Cervical Cancer in the Community: The Face of Surabaya Nowadays. *Indonesian Journal of Cancer*, 12(4), 116. <https://doi.org/10.33371/ijoc.v12i4.605>
- Shah, N. (2013). An Introduction to R. *Practical Graph Mining with R*, 2, 27–52. <https://doi.org/10.1201/b15352-7>
- Shidqi, Z. N., Saraswati, L. D., Kusariana, N., Sutiningsih, D., & Udiyono, A. (2022). Faktor-Faktor Keterlambatan Diagnosis Kanker Pada Pasien Kanker Payudara : Systematic Review. *Jurnal Epidemiologi Kesehatan Komunitas*, 7(2), 471–481. <https://doi.org/10.14710/jekk.v7i2.14911>
- Siji George, C. G., & Sumathi, B. (2020). Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction. *International Journal of Advanced Computer Science and Applications*, 11(9),

- 173–178. <https://doi.org/10.14569/IJACSA.2020.0110920>
- Sowah, R. A., Bampoe-Addo, A. A., Armoo, S. K., Saalia, F. K., Gatsi, F., & Sarkodie-Mensah, B. (2020). Design and Development of Diabetes Management System Using Machine Learning. *International Journal of Telemedicine and Applications*, 2020. <https://doi.org/10.1155/2020/8870141>
- Suhartono, Hariadi, M., & Purnomo, M. H. (2012). Integration of artificial neural networks into an lindenmayer system based plant modeling environment with mathematics. *Studia Universitatis Vasile Goldis, Seria Stiintele Vietii*, 22(3), 411–418.
- Suhartono, Kurniawan, F., & Imran, B. (2018). Identification of virtual plants using bayesian networks based on parametric L-system. *International Journal of Advances in Intelligent Informatics*, 4(1), 40–52. <https://doi.org/10.26555/ijain.v4i1.157>
- Togatorop, P. R., Sianturi, M., Simamora, D., & Silaen, D. (2022). Optimizing Random Forest using Genetic Algorithm for Heart Disease Classification. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 13(1), 60. <https://doi.org/10.24843/lkjiti.2022.v13.i01.p06>
- Vinutha, D. C., Sushma, S. J., Assegie, T. A., Paramesha, K., & Geethashree, A. (2022). Optimization of RFM for automated breast cancer detection. *International Journal of Health Sciences*, 6(March), 5973–5982. <https://doi.org/10.53730/ijhs.v6ns1.6218>
- Wang, F., Wang, Y., Ji, X., & Wang, Z. (2022). Effective Macrosomia Prediction Using Random Forest Algorithm. *International Journal of Environmental Research and Public Health*, 19(6). <https://doi.org/10.3390/ijerph19063245>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika. *Jurnal Edik Informatika*, 2(2), 213–219.
- Zaki, M. J., & Meira, M. J. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press. <https://books.google.com.tr/books?id=Gh9GAwAAQBAJ&lpg=PR9&dq=Data Mining and Analysis: Foundations and Algorithms&hl=tr&pg=PR9#v=onepage&q=Data Mining and Analysis: Foundations and Algorithms&f=false>
- Zeng, G. (2020). On the confusion matrix in credit scoring and its analytical properties. *Communications in Statistics - Theory and Methods*, 49(9), 2080–2093. <https://doi.org/10.1080/03610926.2019.1568485>

## **LAMPIRAN**

## LAMPIRAN

### Lampiran I Dataset Kanker Payudara

Catatan: Dikarenakan fitur yang terdapat pada dataset terlalu banyak, penulis hanya akan menampilkan fitur mean saja.

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	17,99	10,38	122,8	1001	0,1184	0,2776	0,3001	0,1471	0,2419	0,07871
M	20,57	17,77	132,9	1326	0,08474	0,07864	0,0869	0,07017	0,1812	0,05667
M	19,69	21,25	130	1203	0,1096	0,1599	0,1974	0,1279	0,2069	0,05999
M	11,42	20,38	77,58	386,1	0,1425	0,2839	0,2414	0,1052	0,2597	0,09744
M	20,29	14,34	135,1	1297	0,1003	0,1328	0,198	0,1043	0,1809	0,05883
M	12,45	15,7	82,57	477,1	0,1278	0,17	0,1578	0,08089	0,2087	0,07613
M	18,25	19,98	119,6	1040	0,09463	0,109	0,1127	0,074	0,1794	0,05742
M	13,71	20,83	90,2	577,9	0,1189	0,1645	0,09366	0,05985	0,2196	0,07451
M	13	21,82	87,5	519,8	0,1273	0,1932	0,1859	0,09353	0,235	0,07389
M	12,46	24,04	83,97	475,9	0,1186	0,2396	0,2273	0,08543	0,203	0,08243
M	16,02	23,24	102,7	797,8	0,08206	0,06669	0,03299	0,03323	0,1528	0,05697
M	15,78	17,89	103,6	781	0,0971	0,1292	0,09954	0,06606	0,1842	0,06082
M	19,17	24,8	132,4	1123	0,0974	0,2458	0,2065	0,1118	0,2397	0,078
M	15,85	23,95	103,7	782,7	0,08401	0,1002	0,09938	0,05364	0,1847	0,05338
M	13,73	22,61	93,6	578,3	0,1131	0,2293	0,2128	0,08025	0,2069	0,07682
M	14,54	27,54	96,73	658,8	0,1139	0,1595	0,1639	0,07364	0,2303	0,07077
M	14,68	20,13	94,74	684,5	0,09867	0,072	0,07395	0,05259	0,1586	0,05922

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	16,13	20,68	108,1	798,8	0,117	0,2022	0,1722	0,1028	0,2164	0,07356
M	19,81	22,15	130	1260	0,09831	0,1027	0,1479	0,09498	0,1582	0,05395
B	13,54	14,36	87,46	566,3	0,09779	0,08129	0,06664	0,04781	0,1885	0,05766
B	13,08	15,71	85,63	520	0,1075	0,127	0,04568	0,0311	0,1967	0,06811
B	9,504	12,44	60,34	273,9	0,1024	0,06492	0,02956	0,02076	0,1815	0,06905
M	15,34	14,26	102,5	704,4	0,1073	0,2135	0,2077	0,09756	0,2521	0,07032
M	21,16	23,04	137,2	1404	0,09428	0,1022	0,1097	0,08632	0,1769	0,05278
M	16,65	21,38	110	904,6	0,1121	0,1457	0,1525	0,0917	0,1995	0,0633
M	17,14	16,4	116	912,7	0,1186	0,2276	0,2229	0,1401	0,304	0,07413
M	14,58	21,53	97,41	644,8	0,1054	0,1868	0,1425	0,08783	0,2252	0,06924
M	18,61	20,25	122,1	1094	0,0944	0,1066	0,149	0,07731	0,1697	0,05699
M	15,3	25,27	102,4	732,4	0,1082	0,1697	0,1683	0,08751	0,1926	0,0654
M	17,57	15,05	115	955,1	0,09847	0,1157	0,09875	0,07953	0,1739	0,06149
M	18,63	25,11	124,8	1088	0,1064	0,1887	0,2319	0,1244	0,2183	0,06197
M	11,84	18,7	77,93	440,6	0,1109	0,1516	0,1218	0,05182	0,2301	0,07799
M	17,02	23,98	112,8	899,3	0,1197	0,1496	0,2417	0,1203	0,2248	0,06382
M	19,27	26,47	127,9	1162	0,09401	0,1719	0,1657	0,07593	0,1853	0,06261
M	16,13	17,88	107	807,2	0,104	0,1559	0,1354	0,07752	0,1998	0,06515
M	16,74	21,59	110,1	869,5	0,0961	0,1336	0,1348	0,06018	0,1896	0,05656
M	14,25	21,72	93,63	633	0,09823	0,1098	0,1319	0,05598	0,1885	0,06125
B	13,03	18,42	82,61	523,8	0,08983	0,03766	0,02562	0,02923	0,1467	0,05863
M	14,99	25,2	95,54	698,8	0,09387	0,05131	0,02398	0,02899	0,1565	0,05504
M	13,48	20,82	88,4	559,2	0,1016	0,1255	0,1063	0,05439	0,172	0,06419
M	13,44	21,58	86,18	563	0,08162	0,06031	0,0311	0,02031	0,1784	0,05587

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	10,95	21,35	71,9	371,1	0,1227	0,1218	0,1044	0,05669	0,1895	0,0687
M	19,07	24,81	128,3	1104	0,09081	0,219	0,2107	0,09961	0,231	0,06343
M	13,28	20,28	87,32	545,2	0,1041	0,1436	0,09847	0,06158	0,1974	0,06782
M	13,17	21,81	85,42	531,5	0,09714	0,1047	0,08259	0,05252	0,1746	0,06177
M	18,65	17,6	123,7	1076	0,1099	0,1686	0,1974	0,1009	0,1907	0,06049
B	8,196	16,84	51,71	201,9	0,086	0,05943	0,01588	0,005917	0,1769	0,06503
M	13,17	18,66	85,98	534,6	0,1158	0,1231	0,1226	0,0734	0,2128	0,06777
B	12,05	14,63	78,04	449,3	0,1031	0,09092	0,06592	0,02749	0,1675	0,06043
B	13,49	22,3	86,91	561	0,08752	0,07698	0,04751	0,03384	0,1809	0,05718
B	11,76	21,6	74,72	427,9	0,08637	0,04966	0,01657	0,01115	0,1495	0,05888
B	13,64	16,34	87,21	571,8	0,07685	0,06059	0,01857	0,01723	0,1353	0,05953
B	11,94	18,24	75,71	437,6	0,08261	0,04751	0,01972	0,01349	0,1868	0,0611
M	18,22	18,7	120,3	1033	0,1148	0,1485	0,1772	0,106	0,2092	0,0631
M	15,1	22,02	97,26	712,8	0,09056	0,07081	0,05253	0,03334	0,1616	0,05684
B	11,52	18,75	73,34	409	0,09524	0,05473	0,03036	0,02278	0,192	0,05907
M	19,21	18,57	125,5	1152	0,1053	0,1267	0,1323	0,08994	0,1917	0,05961
M	14,71	21,59	95,55	656,9	0,1137	0,1365	0,1293	0,08123	0,2027	0,06758
B	13,05	19,31	82,61	527,2	0,0806	0,03789	0,000692	0,004167	0,1819	0,05501
B	8,618	11,79	54,34	224,5	0,09752	0,05272	0,02061	0,007799	0,1683	0,07187
B	10,17	14,88	64,55	311,9	0,1134	0,08061	0,01084	0,0129	0,2743	0,0696
B	8,598	20,98	54,66	221,8	0,1243	0,08963	0,03	0,009259	0,1828	0,06757
M	14,25	22,15	96,42	645,7	0,1049	0,2008	0,2135	0,08653	0,1949	0,07292
B	9,173	13,86	59,2	260,9	0,07721	0,08751	0,05988	0,0218	0,2341	0,06963
M	12,68	23,84	82,69	499	0,1122	0,1262	0,1128	0,06873	0,1905	0,0659

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	14,78	23,94	97,4	668,3	0,1172	0,1479	0,1267	0,09029	0,1953	0,06654
B	9,465	21,01	60,11	269,4	0,1044	0,07773	0,02172	0,01504	0,1717	0,06899
B	11,31	19,04	71,8	394,1	0,08139	0,04701	0,03709	0,0223	0,1516	0,05667
B	9,029	17,33	58,79	250,5	0,1066	0,1413	0,313	0,04375	0,2111	0,08046
B	12,78	16,49	81,37	502,5	0,09831	0,05234	0,03653	0,02864	0,159	0,05653
M	18,94	21,31	123,6	1130	0,09009	0,1029	0,108	0,07951	0,1582	0,05461
B	8,888	14,64	58,79	244	0,09783	0,1531	0,08606	0,02872	0,1902	0,0898
M	17,2	24,52	114,2	929,4	0,1071	0,183	0,1692	0,07944	0,1927	0,06487
M	13,8	15,79	90,43	584,1	0,1007	0,128	0,07789	0,05069	0,1662	0,06566
B	12,31	16,52	79,19	470,9	0,09172	0,06829	0,03372	0,02272	0,172	0,05914
M	16,07	19,65	104,1	817,7	0,09168	0,08424	0,09769	0,06638	0,1798	0,05391
B	13,53	10,94	87,91	559,2	0,1291	0,1047	0,06877	0,06556	0,2403	0,06641
M	18,05	16,15	120,2	1006	0,1065	0,2146	0,1684	0,108	0,2152	0,06673
M	20,18	23,97	143,7	1245	0,1286	0,3454	0,3754	0,1604	0,2906	0,08142
B	12,86	18	83,19	506,3	0,09934	0,09546	0,03889	0,02315	0,1718	0,05997
B	11,45	20,97	73,81	401,5	0,1102	0,09362	0,04591	0,02233	0,1842	0,07005
B	13,34	15,86	86,49	520	0,1078	0,1535	0,1169	0,06987	0,1942	0,06902
M	25,22	24,91	171,5	1878	0,1063	0,2665	0,3339	0,1845	0,1829	0,06782
M	19,1	26,29	129,1	1132	0,1215	0,1791	0,1937	0,1469	0,1634	0,07224
B	12	15,65	76,95	443,3	0,09723	0,07165	0,04151	0,01863	0,2079	0,05968
M	18,46	18,52	121,1	1075	0,09874	0,1053	0,1335	0,08795	0,2132	0,06022
M	14,48	21,46	94,25	648,2	0,09444	0,09947	0,1204	0,04938	0,2075	0,05636
M	19,02	24,59	122	1076	0,09029	0,1206	0,1468	0,08271	0,1953	0,05629
B	12,36	21,8	79,78	466,1	0,08772	0,09445	0,06015	0,03745	0,193	0,06404



diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	14,64	15,24	95,77	651,9	0,1132	0,1339	0,09966	0,07064	0,2116	0,06346
B	14,62	24,02	94,57	662,7	0,08974	0,08606	0,03102	0,02957	0,1685	0,05866
M	15,37	22,76	100,2	728,2	0,092	0,1036	0,1122	0,07483	0,1717	0,06097
B	13,27	14,76	84,74	551,7	0,07355	0,05055	0,03261	0,02648	0,1386	0,05318
B	13,45	18,3	86,6	555,1	0,1022	0,08165	0,03974	0,0278	0,1638	0,0571
M	15,06	19,83	100,3	705,6	0,1039	0,1553	0,17	0,08815	0,1855	0,06284
M	20,26	23,03	132,4	1264	0,09078	0,1313	0,1465	0,08683	0,2095	0,05649
B	12,18	17,84	77,79	451,1	0,1045	0,07057	0,0249	0,02941	0,19	0,06635
B	9,787	19,94	62,11	294,5	0,1024	0,05301	0,006829	0,007937	0,135	0,0689
B	11,6	12,84	74,34	412,6	0,08983	0,07525	0,04196	0,0335	0,162	0,06582
M	14,42	19,77	94,48	642,5	0,09752	0,1141	0,09388	0,05839	0,1879	0,0639
M	13,61	24,98	88,05	582,7	0,09488	0,08511	0,08625	0,04489	0,1609	0,05871
B	6,981	13,43	43,79	143,5	0,117	0,07568	0	0	0,193	0,07818
B	12,18	20,52	77,22	458,7	0,08013	0,04038	0,02383	0,0177	0,1739	0,05677
B	9,876	19,4	63,95	298,3	0,1005	0,09697	0,06154	0,03029	0,1945	0,06322
B	10,49	19,29	67,41	336,1	0,09989	0,08578	0,02995	0,01201	0,2217	0,06481
M	13,11	15,56	87,21	530,2	0,1398	0,1765	0,2071	0,09601	0,1925	0,07692
B	11,64	18,33	75,17	412,5	0,1142	0,1017	0,0707	0,03485	0,1801	0,0652
B	12,36	18,54	79,01	466,7	0,08477	0,06815	0,02643	0,01921	0,1602	0,06066
M	22,27	19,67	152,8	1509	0,1326	0,2768	0,4264	0,1823	0,2556	0,07039
B	11,34	21,26	72,48	396,5	0,08759	0,06575	0,05133	0,01899	0,1487	0,06529
B	9,777	16,99	62,5	290,2	0,1037	0,08404	0,04334	0,01778	0,1584	0,07065
B	12,63	20,76	82,15	480,4	0,09933	0,1209	0,1065	0,06021	0,1735	0,0707
B	14,26	19,65	97,83	629,9	0,07837	0,2233	0,3003	0,07798	0,1704	0,07769

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	10,51	20,19	68,64	334,2	0,1122	0,1303	0,06476	0,03068	0,1922	0,07782
B	8,726	15,83	55,84	230,9	0,115	0,08201	0,04132	0,01924	0,1649	0,07633
B	11,93	21,53	76,53	438,6	0,09768	0,07849	0,03328	0,02008	0,1688	0,06194
B	8,95	15,76	58,74	245,2	0,09462	0,1243	0,09263	0,02308	0,1305	0,07163
M	14,87	16,67	98,64	682,5	0,1162	0,1649	0,169	0,08923	0,2157	0,06768
M	15,78	22,91	105,7	782,6	0,1155	0,1752	0,2133	0,09479	0,2096	0,07331
M	17,95	20,01	114,2	982	0,08402	0,06722	0,07293	0,05596	0,2129	0,05025
B	11,41	10,82	73,34	403,3	0,09373	0,06685	0,03512	0,02623	0,1667	0,06113
M	18,66	17,12	121,4	1077	0,1054	0,11	0,1457	0,08665	0,1966	0,06213
M	24,25	20,2	166,2	1761	0,1447	0,2867	0,4268	0,2012	0,2655	0,06877
B	14,5	10,89	94,28	640,7	0,1101	0,1099	0,08842	0,05778	0,1856	0,06402
B	13,37	16,39	86,1	553,5	0,07115	0,07325	0,08092	0,028	0,1422	0,05823
B	13,85	17,21	88,44	588,7	0,08785	0,06136	0,0142	0,01141	0,1614	0,0589
M	13,61	24,69	87,76	572,6	0,09258	0,07862	0,05285	0,03085	0,1761	0,0613
M	19	18,91	123,4	1138	0,08217	0,08028	0,09271	0,05627	0,1946	0,05044
B	15,1	16,39	99,58	674,5	0,115	0,1807	0,1138	0,08534	0,2001	0,06467
M	19,79	25,12	130,4	1192	0,1015	0,1589	0,2545	0,1149	0,2202	0,06113
B	12,19	13,29	79,08	455,8	0,1066	0,09509	0,02855	0,02882	0,188	0,06471
M	15,46	19,48	101,7	748,9	0,1092	0,1223	0,1466	0,08087	0,1931	0,05796
M	16,16	21,54	106,2	809,8	0,1008	0,1284	0,1043	0,05613	0,216	0,05891
B	15,71	13,93	102	761,7	0,09462	0,09462	0,07135	0,05933	0,1816	0,05723
M	18,45	21,91	120,2	1075	0,0943	0,09709	0,1153	0,06847	0,1692	0,05727
M	12,77	22,47	81,72	506,3	0,09055	0,05761	0,04711	0,02704	0,1585	0,06065
B	11,71	16,67	74,72	423,6	0,1051	0,06095	0,03592	0,026	0,1339	0,05945

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	11,43	15,39	73,06	399,8	0,09639	0,06889	0,03503	0,02875	0,1734	0,05865
M	14,95	17,57	96,85	678,1	0,1167	0,1305	0,1539	0,08624	0,1957	0,06216
B	11,28	13,39	73	384,8	0,1164	0,1136	0,04635	0,04796	0,1771	0,06072
B	9,738	11,97	61,24	288,5	0,0925	0,04102	0	0	0,1903	0,06422
M	16,11	18,05	105,1	813	0,09721	0,1137	0,09447	0,05943	0,1861	0,06248
B	11,43	17,31	73,66	398	0,1092	0,09486	0,02031	0,01861	0,1645	0,06562
B	12,9	15,92	83,74	512,2	0,08677	0,09509	0,04894	0,03088	0,1778	0,06235
B	10,75	14,97	68,26	355,3	0,07793	0,05139	0,02251	0,007875	0,1399	0,05688
B	11,9	14,65	78,11	432,8	0,1152	0,1296	0,0371	0,03003	0,1995	0,07839
M	11,8	16,58	78,99	432	0,1091	0,17	0,1659	0,07415	0,2678	0,07371
B	14,95	18,77	97,84	689,5	0,08138	0,1167	0,0905	0,03562	0,1744	0,06493
B	14,44	15,18	93,97	640,1	0,0997	0,1021	0,08487	0,05532	0,1724	0,06081
B	13,74	17,91	88,12	585	0,07944	0,06376	0,02881	0,01329	0,1473	0,0558
B	13	20,78	83,51	519,4	0,1135	0,07589	0,03136	0,02645	0,254	0,06087
B	8,219	20,7	53,27	203,9	0,09405	0,1305	0,1321	0,02168	0,2222	0,08261
B	9,731	15,34	63,78	300,2	0,1072	0,1599	0,4108	0,07857	0,2548	0,09296
B	11,15	13,08	70,87	381,9	0,09754	0,05113	0,01982	0,01786	0,183	0,06105
B	13,15	15,34	85,31	538,9	0,09384	0,08498	0,09293	0,03483	0,1822	0,06207
B	12,25	17,94	78,27	460,3	0,08654	0,06679	0,03885	0,02331	0,197	0,06228
M	17,68	20,74	117,4	963,7	0,1115	0,1665	0,1855	0,1054	0,1971	0,06166
B	16,84	19,46	108,4	880,2	0,07445	0,07223	0,0515	0,02771	0,1844	0,05268
B	12,06	12,74	76,84	448,6	0,09311	0,05241	0,01972	0,01963	0,159	0,05907
B	10,9	12,96	68,69	366,8	0,07515	0,03718	0,00309	0,006588	0,1442	0,05743
B	11,75	20,18	76,1	419,8	0,1089	0,1141	0,06843	0,03738	0,1993	0,06453

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	19,19	15,94	126,3	1157	0,08694	0,1185	0,1193	0,09667	0,1741	0,05176
M	19,59	18,15	130,7	1214	0,112	0,1666	0,2508	0,1286	0,2027	0,06082
B	12,34	22,22	79,85	464,5	0,1012	0,1015	0,0537	0,02822	0,1551	0,06761
M	23,27	22,04	152,1	1686	0,08439	0,1145	0,1324	0,09702	0,1801	0,05553
B	14,97	19,76	95,5	690,2	0,08421	0,05352	0,01947	0,01939	0,1515	0,05266
B	10,8	9,71	68,77	357,6	0,09594	0,05736	0,02531	0,01698	0,1381	0,064
M	16,78	18,8	109,3	886,3	0,08865	0,09182	0,08422	0,06576	0,1893	0,05534
M	17,47	24,68	116,1	984,6	0,1049	0,1603	0,2159	0,1043	0,1538	0,06365
B	14,97	16,95	96,22	685,9	0,09855	0,07885	0,02602	0,03781	0,178	0,0565
B	12,32	12,39	78,85	464,1	0,1028	0,06981	0,03987	0,037	0,1959	0,05955
M	13,43	19,63	85,84	565,4	0,09048	0,06288	0,05858	0,03438	0,1598	0,05671
M	15,46	11,89	102,5	736,9	0,1257	0,1555	0,2032	0,1097	0,1966	0,07069
B	11,08	14,71	70,21	372,7	0,1006	0,05743	0,02363	0,02583	0,1566	0,06669
B	10,66	15,15	67,49	349,6	0,08792	0,04302	0	0	0,1928	0,05975
B	8,671	14,45	54,42	227,2	0,09138	0,04276	0	0	0,1722	0,06724
B	9,904	18,06	64,6	302,4	0,09699	0,1294	0,1307	0,03716	0,1669	0,08116
M	16,46	20,11	109,3	832,9	0,09831	0,1556	0,1793	0,08866	0,1794	0,06323
B	13,01	22,22	82,01	526,4	0,06251	0,01938	0,001595	0,001852	0,1395	0,05234
B	12,81	13,06	81,29	508,8	0,08739	0,03774	0,009193	0,0133	0,1466	0,06133
M	27,22	21,87	182,1	2250	0,1094	0,1914	0,2871	0,1878	0,18	0,0577
M	21,09	26,57	142,7	1311	0,1141	0,2832	0,2487	0,1496	0,2395	0,07398
M	15,7	20,31	101,2	766,6	0,09597	0,08799	0,06593	0,05189	0,1618	0,05549
B	11,41	14,92	73,53	402	0,09059	0,08155	0,06181	0,02361	0,1167	0,06217
M	15,28	22,41	98,92	710,6	0,09057	0,1052	0,05375	0,03263	0,1727	0,06317

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	10,08	15,11	63,76	317,5	0,09267	0,04695	0,001597	0,002404	0,1703	0,06048
M	18,31	18,58	118,6	1041	0,08588	0,08468	0,08169	0,05814	0,1621	0,05425
B	11,71	17,19	74,68	420,3	0,09774	0,06141	0,03809	0,03239	0,1516	0,06095
B	11,81	17,39	75,27	428,9	0,1007	0,05562	0,02353	0,01553	0,1718	0,0578
B	12,3	15,9	78,83	463,7	0,0808	0,07253	0,03844	0,01654	0,1667	0,05474
M	14,22	23,12	94,37	609,9	0,1075	0,2413	0,1981	0,06618	0,2384	0,07542
B	12,77	21,41	82,02	507,4	0,08749	0,06601	0,03112	0,02864	0,1694	0,06287
B	9,72	18,22	60,73	288,1	0,0695	0,02344	0	0	0,1653	0,06447
M	12,34	26,86	81,15	477,4	0,1034	0,1353	0,1085	0,04562	0,1943	0,06937
M	14,86	23,21	100,4	671,4	0,1044	0,198	0,1697	0,08878	0,1737	0,06672
B	12,91	16,33	82,53	516,4	0,07941	0,05366	0,03873	0,02377	0,1829	0,05667
M	13,77	22,29	90,63	588,9	0,12	0,1267	0,1385	0,06526	0,1834	0,06877
M	18,08	21,84	117,4	1024	0,07371	0,08642	0,1103	0,05778	0,177	0,0534
M	19,18	22,49	127,5	1148	0,08523	0,1428	0,1114	0,06772	0,1767	0,05529
M	14,45	20,22	94,49	642,7	0,09872	0,1206	0,118	0,0598	0,195	0,06466
B	12,23	19,56	78,54	461	0,09586	0,08087	0,04187	0,04107	0,1979	0,06013
M	17,54	19,32	115,1	951,6	0,08968	0,1198	0,1036	0,07488	0,1506	0,05491
M	23,29	26,67	158,9	1685	0,1141	0,2084	0,3523	0,162	0,22	0,06229
M	13,81	23,75	91,56	597,8	0,1323	0,1768	0,1558	0,09176	0,2251	0,07421
B	12,47	18,6	81,09	481,9	0,09965	0,1058	0,08005	0,03821	0,1925	0,06373
M	15,12	16,68	98,78	716,6	0,08876	0,09588	0,0755	0,04079	0,1594	0,05986
B	9,876	17,27	62,92	295,4	0,1089	0,07232	0,01756	0,01952	0,1934	0,06285
M	17,01	20,26	109,7	904,3	0,08772	0,07304	0,0695	0,0539	0,2026	0,05223
B	13,11	22,54	87,02	529,4	0,1002	0,1483	0,08705	0,05102	0,185	0,0731

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	15,27	12,91	98,17	725,5	0,08182	0,0623	0,05892	0,03157	0,1359	0,05526
M	20,58	22,14	134,7	1290	0,0909	0,1348	0,164	0,09561	0,1765	0,05024
B	11,84	18,94	75,51	428	0,08871	0,069	0,02669	0,01393	0,1533	0,06057
M	28,11	18,47	188,5	2499	0,1142	0,1516	0,3201	0,1595	0,1648	0,05525
M	17,42	25,56	114,5	948	0,1006	0,1146	0,1682	0,06597	0,1308	0,05866
M	14,19	23,81	92,87	610,7	0,09463	0,1306	0,1115	0,06462	0,2235	0,06433
M	13,86	16,93	90,96	578,9	0,1026	0,1517	0,09901	0,05602	0,2106	0,06916
B	11,89	18,35	77,32	432,2	0,09363	0,1154	0,06636	0,03142	0,1967	0,06314
B	10,2	17,48	65,05	321,2	0,08054	0,05907	0,05774	0,01071	0,1964	0,06315
M	19,8	21,56	129,7	1230	0,09383	0,1306	0,1272	0,08691	0,2094	0,05581
M	19,53	32,47	128	1223	0,0842	0,113	0,1145	0,06637	0,1428	0,05313
B	13,65	13,16	87,88	568,9	0,09646	0,08711	0,03888	0,02563	0,136	0,06344
B	13,56	13,9	88,59	561,3	0,1051	0,1192	0,0786	0,04451	0,1962	0,06303
B	10,18	17,53	65,12	313,1	0,1061	0,08502	0,01768	0,01915	0,191	0,06908
M	15,75	20,25	102,6	761,3	0,1025	0,1204	0,1147	0,06462	0,1935	0,06303
B	13,27	17,02	84,55	546,4	0,08445	0,04994	0,03554	0,02456	0,1496	0,05674
B	14,34	13,47	92,51	641,2	0,09906	0,07624	0,05724	0,04603	0,2075	0,05448
B	10,44	15,46	66,62	329,6	0,1053	0,07722	0,006643	0,01216	0,1788	0,0645
B	15	15,51	97,45	684,5	0,08371	0,1096	0,06505	0,0378	0,1881	0,05907
B	12,62	23,97	81,35	496,4	0,07903	0,07529	0,05438	0,02036	0,1514	0,06019
M	12,83	22,33	85,26	503,2	0,1088	0,1799	0,1695	0,06861	0,2123	0,07254
M	17,05	19,08	113,4	895	0,1141	0,1572	0,191	0,109	0,2131	0,06325
B	11,32	27,08	71,76	395,7	0,06883	0,03813	0,01633	0,003125	0,1869	0,05628
B	11,22	33,81	70,79	386,8	0,0778	0,03574	0,004967	0,006434	0,1845	0,05828

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	20,51	27,81	134,4	1319	0,09159	0,1074	0,1554	0,0834	0,1448	0,05592
B	9,567	15,91	60,21	279,6	0,08464	0,04087	0,01652	0,01667	0,1551	0,06403
B	14,03	21,25	89,79	603,4	0,0907	0,06945	0,01462	0,01896	0,1517	0,05835
M	23,21	26,97	153,5	1670	0,09509	0,1682	0,195	0,1237	0,1909	0,06309
M	20,48	21,46	132,5	1306	0,08355	0,08348	0,09042	0,06022	0,1467	0,05177
B	14,22	27,85	92,55	623,9	0,08223	0,1039	0,1103	0,04408	0,1342	0,06129
M	17,46	39,28	113,4	920,6	0,09812	0,1298	0,1417	0,08811	0,1809	0,05966
B	13,64	15,6	87,38	575,3	0,09423	0,0663	0,04705	0,03731	0,1717	0,0566
B	12,42	15,04	78,61	476,5	0,07926	0,03393	0,01053	0,01108	0,1546	0,05754
B	11,3	18,19	73,93	389,4	0,09592	0,1325	0,1548	0,02854	0,2054	0,07669
B	13,75	23,77	88,54	590	0,08043	0,06807	0,04697	0,02344	0,1773	0,05429
M	19,4	23,5	129,1	1155	0,1027	0,1558	0,2049	0,08886	0,1978	0,06
B	10,48	19,86	66,72	337,7	0,107	0,05971	0,04831	0,0307	0,1737	0,0644
B	13,2	17,43	84,13	541,6	0,07215	0,04524	0,04336	0,01105	0,1487	0,05635
B	12,89	14,11	84,95	512,2	0,0876	0,1346	0,1374	0,0398	0,1596	0,06409
B	10,65	25,22	68,01	347	0,09657	0,07234	0,02379	0,01615	0,1897	0,06329
B	11,52	14,93	73,87	406,3	0,1013	0,07808	0,04328	0,02929	0,1883	0,06168
M	20,94	23,56	138,9	1364	0,1007	0,1606	0,2712	0,131	0,2205	0,05898
B	11,5	18,45	73,28	407,4	0,09345	0,05991	0,02638	0,02069	0,1834	0,05934
M	19,73	19,82	130,7	1206	0,1062	0,1849	0,2417	0,0974	0,1733	0,06697
M	17,3	17,08	113	928,2	0,1008	0,1041	0,1266	0,08353	0,1813	0,05613
M	19,45	19,33	126,5	1169	0,1035	0,1188	0,1379	0,08591	0,1776	0,05647
M	13,96	17,05	91,43	602,4	0,1096	0,1279	0,09789	0,05246	0,1908	0,0613
M	19,55	28,77	133,6	1207	0,0926	0,2063	0,1784	0,1144	0,1893	0,06232

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	15,32	17,27	103,2	713,3	0,1335	0,2284	0,2448	0,1242	0,2398	0,07596
M	15,66	23,2	110,2	773,5	0,1109	0,3114	0,3176	0,1377	0,2495	0,08104
M	15,53	33,56	103,7	744,9	0,1063	0,1639	0,1751	0,08399	0,2091	0,0665
M	20,31	27,06	132,9	1288	0,1	0,1088	0,1519	0,09333	0,1814	0,05572
M	17,35	23,06	111	933,1	0,08662	0,0629	0,02891	0,02837	0,1564	0,05307
M	17,29	22,13	114,4	947,8	0,08999	0,1273	0,09697	0,07507	0,2108	0,05464
M	15,61	19,38	100	758,6	0,0784	0,05616	0,04209	0,02847	0,1547	0,05443
M	17,19	22,07	111,6	928,3	0,09726	0,08995	0,09061	0,06527	0,1867	0,0558
M	20,73	31,12	135,7	1419	0,09469	0,1143	0,1367	0,08646	0,1769	0,05674
B	10,6	18,95	69,28	346,4	0,09688	0,1147	0,06387	0,02642	0,1922	0,06491
B	13,59	21,84	87,16	561	0,07956	0,08259	0,04072	0,02142	0,1635	0,05859
B	12,87	16,21	82,38	512,2	0,09425	0,06219	0,039	0,01615	0,201	0,05769
B	10,71	20,39	69,5	344,9	0,1082	0,1289	0,08448	0,02867	0,1668	0,06862
B	14,29	16,82	90,3	632,6	0,06429	0,02675	0,00725	0,00625	0,1508	0,05376
B	11,29	13,04	72,23	388	0,09834	0,07608	0,03265	0,02755	0,1769	0,0627
M	21,75	20,99	147,3	1491	0,09401	0,1961	0,2195	0,1088	0,1721	0,06194
B	9,742	15,67	61,5	289,9	0,09037	0,04689	0,01103	0,01407	0,2081	0,06312
M	17,93	24,48	115,2	998,9	0,08855	0,07027	0,05699	0,04744	0,1538	0,0551
B	11,89	17,36	76,2	435,6	0,1225	0,0721	0,05929	0,07404	0,2015	0,05875
B	11,33	14,16	71,79	396,6	0,09379	0,03872	0,001487	0,003333	0,1954	0,05821
M	18,81	19,98	120,9	1102	0,08923	0,05884	0,0802	0,05843	0,155	0,04996
B	13,59	17,84	86,24	572,3	0,07948	0,04052	0,01997	0,01238	0,1573	0,0552
B	13,85	15,18	88,99	587,4	0,09516	0,07688	0,04479	0,03711	0,211	0,05853
M	19,16	26,6	126,2	1138	0,102	0,1453	0,1921	0,09664	0,1902	0,0622



diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	11,74	14,02	74,24	427,3	0,07813	0,0434	0,02245	0,02763	0,2101	0,06113
M	19,4	18,18	127,2	1145	0,1037	0,1442	0,1626	0,09464	0,1893	0,05892
M	16,24	18,77	108,8	805,1	0,1066	0,1802	0,1948	0,09052	0,1876	0,06684
B	12,89	15,7	84,08	516,6	0,07818	0,0958	0,1115	0,0339	0,1432	0,05935
B	12,58	18,4	79,83	489	0,08393	0,04216	0,00186	0,002924	0,1697	0,05855
B	11,94	20,76	77,87	441	0,08605	0,1011	0,06574	0,03791	0,1588	0,06766
B	12,89	13,12	81,89	515,9	0,06955	0,03729	0,0226	0,01171	0,1337	0,05581
B	11,26	19,96	73,72	394,1	0,0802	0,1181	0,09274	0,05588	0,2595	0,06233
B	11,37	18,89	72,17	396	0,08713	0,05008	0,02399	0,02173	0,2013	0,05955
B	14,41	19,73	96,03	651	0,08757	0,1676	0,1362	0,06602	0,1714	0,07192
B	14,96	19,1	97,03	687,3	0,08992	0,09823	0,0594	0,04819	0,1879	0,05852
B	12,95	16,02	83,14	513,7	0,1005	0,07943	0,06155	0,0337	0,173	0,0647
B	11,85	17,46	75,54	432,7	0,08372	0,05642	0,02688	0,0228	0,1875	0,05715
B	12,72	13,78	81,78	492,1	0,09667	0,08393	0,01288	0,01924	0,1638	0,061
B	13,77	13,27	88,06	582,7	0,09198	0,06221	0,01063	0,01917	0,1592	0,05912
B	10,91	12,35	69,14	363,7	0,08518	0,04721	0,01236	0,01369	0,1449	0,06031
M	11,76	18,14	75	431,1	0,09968	0,05914	0,02685	0,03515	0,1619	0,06287
B	14,26	18,17	91,22	633,1	0,06576	0,0522	0,02475	0,01374	0,1635	0,05586
B	10,51	23,09	66,85	334,2	0,1015	0,06797	0,02495	0,01875	0,1695	0,06556
M	19,53	18,9	129,5	1217	0,115	0,1642	0,2197	0,1062	0,1792	0,06552
B	12,46	19,89	80,43	471,3	0,08451	0,1014	0,0683	0,03099	0,1781	0,06249
M	20,09	23,86	134,7	1247	0,108	0,1838	0,2283	0,128	0,2249	0,07469
B	10,49	18,61	66,86	334,3	0,1068	0,06678	0,02297	0,0178	0,1482	0,066
B	11,46	18,16	73,59	403,1	0,08853	0,07694	0,03344	0,01502	0,1411	0,06243

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	11,6	24,49	74,23	417,2	0,07474	0,05688	0,01974	0,01313	0,1935	0,05878
B	13,2	15,82	84,07	537,3	0,08511	0,05251	0,001461	0,003261	0,1632	0,05894
B	9	14,4	56,36	246,3	0,07005	0,03116	0,003681	0,003472	0,1788	0,06833
B	13,5	12,71	85,69	566,2	0,07376	0,03614	0,002758	0,004419	0,1365	0,05335
B	13,05	13,84	82,71	530,6	0,08352	0,03735	0,004559	0,008829	0,1453	0,05518
B	11,7	19,11	74,33	418,7	0,08814	0,05253	0,01583	0,01148	0,1936	0,06128
B	14,61	15,69	92,68	664,9	0,07618	0,03515	0,01447	0,01877	0,1632	0,05255
B	12,76	13,37	82,29	504,1	0,08794	0,07948	0,04052	0,02548	0,1601	0,0614
B	11,54	10,72	73,73	409,1	0,08597	0,05969	0,01367	0,008907	0,1833	0,061
B	8,597	18,6	54,09	221,2	0,1074	0,05847	0	0	0,2163	0,07359
B	12,49	16,85	79,19	481,6	0,08511	0,03834	0,004473	0,006423	0,1215	0,05673
B	12,18	14,08	77,25	461,4	0,07734	0,03212	0,01123	0,005051	0,1673	0,05649
M	18,22	18,87	118,7	1027	0,09746	0,1117	0,113	0,0795	0,1807	0,05664
B	9,042	18,9	60,07	244,5	0,09968	0,1972	0,1975	0,04908	0,233	0,08743
B	12,43	17	78,6	477,3	0,07557	0,03454	0,01342	0,01699	0,1472	0,05561
B	10,25	16,18	66,52	324,2	0,1061	0,1111	0,06726	0,03965	0,1743	0,07279
M	20,16	19,66	131,1	1274	0,0802	0,08564	0,1155	0,07726	0,1928	0,05096
B	12,86	13,32	82,82	504,8	0,1134	0,08834	0,038	0,034	0,1543	0,06476
M	20,34	21,51	135,9	1264	0,117	0,1875	0,2565	0,1504	0,2569	0,0667
B	12,2	15,21	78,01	457,9	0,08673	0,06545	0,01994	0,01692	0,1638	0,06129
B	12,67	17,3	81,25	489,9	0,1028	0,07664	0,03193	0,02107	0,1707	0,05984
B	14,11	12,88	90,03	616,5	0,09309	0,05306	0,01765	0,02733	0,1373	0,057
B	12,03	17,93	76,09	446	0,07683	0,03892	0,001546	0,005592	0,1382	0,0607
M	16,27	20,71	106,9	813,7	0,1169	0,1319	0,1478	0,08488	0,1948	0,06277

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	16,26	21,88	107,5	826,8	0,1165	0,1283	0,1799	0,07981	0,1869	0,06532
M	16,03	15,51	105,8	793,2	0,09491	0,1371	0,1204	0,07041	0,1782	0,05976
B	12,98	19,35	84,52	514	0,09579	0,1125	0,07107	0,0295	0,1761	0,0654
B	11,22	19,86	71,94	387,3	0,1054	0,06779	0,005006	0,007583	0,194	0,06028
B	11,25	14,78	71,38	390	0,08306	0,04458	0,0009737	0,002941	0,1773	0,06081
B	12,3	19,02	77,88	464,4	0,08313	0,04202	0,007756	0,008535	0,1539	0,05945
M	17,06	21	111,8	918,6	0,1119	0,1056	0,1508	0,09934	0,1727	0,06071
B	12,99	14,23	84,08	514,3	0,09462	0,09965	0,03738	0,02098	0,1652	0,07238
M	18,77	21,43	122,9	1092	0,09116	0,1402	0,106	0,0609	0,1953	0,06083
B	10,05	17,53	64,41	310,8	0,1007	0,07326	0,02511	0,01775	0,189	0,06331
M	23,51	24,27	155,1	1747	0,1069	0,1283	0,2308	0,141	0,1797	0,05506
B	14,42	16,54	94,15	641,2	0,09751	0,1139	0,08007	0,04223	0,1912	0,06412
B	9,606	16,84	61,64	280,5	0,08481	0,09228	0,08422	0,02292	0,2036	0,07125
B	11,06	14,96	71,49	373,9	0,1033	0,09097	0,05397	0,03341	0,1776	0,06907
M	19,68	21,68	129,9	1194	0,09797	0,1339	0,1863	0,1103	0,2082	0,05715
B	11,71	15,45	75,03	420,3	0,115	0,07281	0,04006	0,0325	0,2009	0,06506
B	10,26	14,71	66,2	321,6	0,09882	0,09159	0,03581	0,02037	0,1633	0,07005
B	12,06	18,9	76,66	445,3	0,08386	0,05794	0,00751	0,008488	0,1555	0,06048
B	14,76	14,74	94,87	668,7	0,08875	0,0778	0,04608	0,03528	0,1521	0,05912
B	11,47	16,03	73,02	402,7	0,09076	0,05886	0,02587	0,02322	0,1634	0,06372
B	11,95	14,96	77,23	426,7	0,1158	0,1206	0,01171	0,01787	0,2459	0,06581
B	11,66	17,07	73,7	421	0,07561	0,0363	0,008306	0,01162	0,1671	0,05731
M	15,75	19,22	107,1	758,6	0,1243	0,2364	0,2914	0,1242	0,2375	0,07603
M	25,73	17,46	174,2	2010	0,1149	0,2363	0,3368	0,1913	0,1956	0,06121

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	15,08	25,74	98	716,6	0,1024	0,09769	0,1235	0,06553	0,1647	0,06464
B	11,14	14,07	71,24	384,6	0,07274	0,06064	0,04505	0,01471	0,169	0,06083
B	12,56	19,07	81,92	485,8	0,0876	0,1038	0,103	0,04391	0,1533	0,06184
B	13,05	18,59	85,09	512	0,1082	0,1304	0,09603	0,05603	0,2035	0,06501
B	13,87	16,21	88,52	593,7	0,08743	0,05492	0,01502	0,02088	0,1424	0,05883
B	8,878	15,49	56,74	241	0,08293	0,07698	0,04721	0,02381	0,193	0,06621
B	9,436	18,32	59,82	278,6	0,1009	0,05956	0,0271	0,01406	0,1506	0,06959
B	12,54	18,07	79,42	491,9	0,07436	0,0265	0,001194	0,005449	0,1528	0,05185
B	13,3	21,57	85,24	546,1	0,08582	0,06373	0,03344	0,02424	0,1815	0,05696
B	12,76	18,84	81,87	496,6	0,09676	0,07952	0,02688	0,01781	0,1759	0,06183
B	16,5	18,29	106,6	838,1	0,09686	0,08468	0,05862	0,04835	0,1495	0,05593
B	13,4	16,95	85,48	552,4	0,07937	0,05696	0,02181	0,01473	0,165	0,05701
M	20,44	21,78	133,8	1293	0,0915	0,1131	0,09799	0,07785	0,1618	0,05557
M	20,2	26,83	133,7	1234	0,09905	0,1669	0,1641	0,1265	0,1875	0,0602
B	12,21	18,02	78,31	458,4	0,09231	0,07175	0,04392	0,02027	0,1695	0,05916
M	21,71	17,25	140,9	1546	0,09384	0,08562	0,1168	0,08465	0,1717	0,05054
M	22,01	21,9	147,2	1482	0,1063	0,1954	0,2448	0,1501	0,1824	0,0614
M	16,35	23,29	109	840,4	0,09742	0,1497	0,1811	0,08773	0,2175	0,06218
B	15,19	13,21	97,65	711,8	0,07963	0,06934	0,03393	0,02657	0,1721	0,05544
M	21,37	15,1	141,3	1386	0,1001	0,1515	0,1932	0,1255	0,1973	0,06183
M	20,64	17,35	134,8	1335	0,09446	0,1076	0,1527	0,08941	0,1571	0,05478
B	13,69	16,07	87,84	579,1	0,08302	0,06374	0,02556	0,02031	0,1872	0,05669
B	16,17	16,07	106,3	788,5	0,0988	0,1438	0,06651	0,05397	0,199	0,06572
B	10,57	20,22	70,15	338,3	0,09073	0,166	0,228	0,05941	0,2188	0,0845

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	13,46	28,21	85,89	562,1	0,07517	0,04726	0,01271	0,01117	0,1421	0,05763
B	13,66	15,15	88,27	580,6	0,08268	0,07548	0,04249	0,02471	0,1792	0,05897
M	11,08	18,83	73,3	361,6	0,1216	0,2154	0,1689	0,06367	0,2196	0,0795
B	11,27	12,96	73,16	386,3	0,1237	0,1111	0,079	0,0555	0,2018	0,06914
B	11,04	14,93	70,67	372,7	0,07987	0,07079	0,03546	0,02074	0,2003	0,06246
B	12,05	22,72	78,75	447,8	0,06935	0,1073	0,07943	0,02978	0,1203	0,06659
B	12,39	17,48	80,64	462,9	0,1042	0,1297	0,05892	0,0288	0,1779	0,06588
B	13,28	13,72	85,79	541,8	0,08363	0,08575	0,05077	0,02864	0,1617	0,05594
M	14,6	23,29	93,97	664,7	0,08682	0,06636	0,0839	0,05271	0,1627	0,05416
B	12,21	14,09	78,78	462	0,08108	0,07823	0,06839	0,02534	0,1646	0,06154
B	13,88	16,16	88,37	596,6	0,07026	0,04831	0,02045	0,008507	0,1607	0,05474
B	11,27	15,5	73,38	392	0,08365	0,1114	0,1007	0,02757	0,181	0,07252
M	19,55	23,21	128,9	1174	0,101	0,1318	0,1856	0,1021	0,1989	0,05884
B	10,26	12,22	65,75	321,6	0,09996	0,07542	0,01923	0,01968	0,18	0,06569
B	8,734	16,84	55,27	234,3	0,1039	0,07428	0	0	0,1985	0,07098
M	15,49	19,97	102,4	744,7	0,116	0,1562	0,1891	0,09113	0,1929	0,06744
M	21,61	22,28	144,4	1407	0,1167	0,2087	0,281	0,1562	0,2162	0,06606
B	12,1	17,72	78,07	446,2	0,1029	0,09758	0,04783	0,03326	0,1937	0,06161
B	14,06	17,18	89,75	609,1	0,08045	0,05361	0,02681	0,03251	0,1641	0,05764
B	13,51	18,89	88,1	558,1	0,1059	0,1147	0,0858	0,05381	0,1806	0,06079
B	12,8	17,46	83,05	508,3	0,08044	0,08895	0,0739	0,04083	0,1574	0,0575
B	11,06	14,83	70,31	378,2	0,07741	0,04768	0,02712	0,007246	0,1535	0,06214
B	11,8	17,26	75,26	431,9	0,09087	0,06232	0,02853	0,01638	0,1847	0,06019
M	17,91	21,02	124,4	994	0,123	0,2576	0,3189	0,1198	0,2113	0,07115

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	11,93	10,91	76,14	442,7	0,08872	0,05242	0,02606	0,01796	0,1601	0,05541
B	12,96	18,29	84,18	525,2	0,07351	0,07899	0,04057	0,01883	0,1874	0,05899
B	12,94	16,17	83,18	507,6	0,09879	0,08836	0,03296	0,0239	0,1735	0,062
B	12,34	14,95	78,29	469,1	0,08682	0,04571	0,02109	0,02054	0,1571	0,05708
B	10,94	18,59	70,39	370	0,1004	0,0746	0,04944	0,02932	0,1486	0,06615
B	16,14	14,86	104,3	800	0,09495	0,08501	0,055	0,04528	0,1735	0,05875
B	12,85	21,37	82,63	514,5	0,07551	0,08316	0,06126	0,01867	0,158	0,06114
M	17,99	20,66	117,8	991,7	0,1036	0,1304	0,1201	0,08824	0,1992	0,06069
B	12,27	17,92	78,41	466,1	0,08685	0,06526	0,03211	0,02653	0,1966	0,05597
B	11,36	17,57	72,49	399,8	0,08858	0,05313	0,02783	0,021	0,1601	0,05913
B	11,04	16,83	70,92	373,2	0,1077	0,07804	0,03046	0,0248	0,1714	0,0634
B	9,397	21,68	59,75	268,8	0,07969	0,06053	0,03735	0,005128	0,1274	0,06724
B	14,99	22,11	97,53	693,7	0,08515	0,1025	0,06859	0,03876	0,1944	0,05913
M	15,13	29,81	96,71	719,5	0,0832	0,04605	0,04686	0,02739	0,1852	0,05294
B	11,89	21,17	76,39	433,8	0,09773	0,0812	0,02555	0,02179	0,2019	0,0629
B	9,405	21,7	59,6	271,2	0,1044	0,06159	0,02047	0,01257	0,2025	0,06601
M	15,5	21,08	102,9	803,1	0,112	0,1571	0,1522	0,08481	0,2085	0,06864
B	12,7	12,17	80,88	495	0,08785	0,05794	0,0236	0,02402	0,1583	0,06275
B	11,16	21,41	70,95	380,3	0,1018	0,05978	0,008955	0,01076	0,1615	0,06144
B	11,57	19,04	74,2	409,7	0,08546	0,07722	0,05485	0,01428	0,2031	0,06267
B	14,69	13,98	98,22	656,1	0,1031	0,1836	0,145	0,063	0,2086	0,07406
B	11,61	16,02	75,46	408,2	0,1088	0,1168	0,07097	0,04497	0,1886	0,0632
B	13,66	19,13	89,46	575,3	0,09057	0,1147	0,09657	0,04812	0,1848	0,06181
B	9,742	19,12	61,93	289,7	0,1075	0,08333	0,008934	0,01967	0,2538	0,07029

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	10,03	21,28	63,19	307,3	0,08117	0,03912	0,00247	0,005159	0,163	0,06439
B	10,48	14,98	67,49	333,6	0,09816	0,1013	0,06335	0,02218	0,1925	0,06915
B	10,8	21,98	68,79	359,9	0,08801	0,05743	0,03614	0,01404	0,2016	0,05977
B	11,13	16,62	70,47	381,1	0,08151	0,03834	0,01369	0,0137	0,1511	0,06148
B	12,72	17,67	80,98	501,3	0,07896	0,04522	0,01402	0,01835	0,1459	0,05544
M	14,9	22,53	102,1	685	0,09947	0,2225	0,2733	0,09711	0,2041	0,06898
B	12,4	17,68	81,47	467,8	0,1054	0,1316	0,07741	0,02799	0,1811	0,07102
M	20,18	19,54	133,8	1250	0,1133	0,1489	0,2133	0,1259	0,1724	0,06053
M	18,82	21,97	123,7	1110	0,1018	0,1389	0,1594	0,08744	0,1943	0,06132
B	14,86	16,94	94,89	673,7	0,08924	0,07074	0,03346	0,02877	0,1573	0,05703
M	13,98	19,62	91,12	599,5	0,106	0,1133	0,1126	0,06463	0,1669	0,06544
B	12,87	19,54	82,67	509,2	0,09136	0,07883	0,01797	0,0209	0,1861	0,06347
B	14,04	15,98	89,78	611,2	0,08458	0,05895	0,03534	0,02944	0,1714	0,05898
B	13,85	19,6	88,68	592,6	0,08684	0,0633	0,01342	0,02293	0,1555	0,05673
B	14,02	15,66	89,59	606,5	0,07966	0,05581	0,02087	0,02652	0,1589	0,05586
B	10,97	17,2	71,73	371,5	0,08915	0,1113	0,09457	0,03613	0,1489	0,0664
M	17,27	25,42	112,4	928,8	0,08331	0,1109	0,1204	0,05736	0,1467	0,05407
B	13,78	15,79	88,37	585,9	0,08817	0,06718	0,01055	0,009937	0,1405	0,05848
B	10,57	18,32	66,82	340,9	0,08142	0,04462	0,01993	0,01111	0,2372	0,05768
M	18,03	16,85	117,5	990	0,08947	0,1232	0,109	0,06254	0,172	0,0578
B	11,99	24,89	77,61	441,3	0,103	0,09218	0,05441	0,04274	0,182	0,0685
M	17,75	28,03	117,3	981,6	0,09997	0,1314	0,1698	0,08293	0,1713	0,05916
B	14,8	17,66	95,88	674,8	0,09179	0,0889	0,04069	0,0226	0,1893	0,05886
B	14,53	19,34	94,25	659,7	0,08388	0,078	0,08817	0,02925	0,1473	0,05746

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	21,1	20,52	138,1	1384	0,09684	0,1175	0,1572	0,1155	0,1554	0,05661
B	11,87	21,54	76,83	432	0,06613	0,1064	0,08777	0,02386	0,1349	0,06612
M	19,59	25	127,7	1191	0,1032	0,09871	0,1655	0,09063	0,1663	0,05391
B	12	28,23	76,77	442,5	0,08437	0,0645	0,04055	0,01945	0,1615	0,06104
B	14,53	13,98	93,86	644,2	0,1099	0,09242	0,06895	0,06495	0,165	0,06121
B	12,62	17,15	80,62	492,9	0,08583	0,0543	0,02966	0,02272	0,1799	0,05826
B	13,38	30,72	86,34	557,2	0,09245	0,07426	0,02819	0,03264	0,1375	0,06016
B	11,63	29,29	74,87	415,1	0,09357	0,08574	0,0716	0,02017	0,1799	0,06166
B	13,21	25,25	84,1	537,9	0,08791	0,05205	0,02772	0,02068	0,1619	0,05584
B	13	25,13	82,61	520,2	0,08369	0,05073	0,01206	0,01762	0,1667	0,05449
B	9,755	28,2	61,68	290,9	0,07984	0,04626	0,01541	0,01043	0,1621	0,05952
M	17,08	27,15	111,2	930,9	0,09898	0,111	0,1007	0,06431	0,1793	0,06281
M	27,42	26,27	186,9	2501	0,1084	0,1988	0,3635	0,1689	0,2061	0,05623
B	14,4	26,99	92,25	646,1	0,06995	0,05223	0,03476	0,01737	0,1707	0,05433
B	11,6	18,36	73,88	412,7	0,08508	0,05855	0,03367	0,01777	0,1516	0,05859
B	13,17	18,22	84,28	537,3	0,07466	0,05994	0,04859	0,0287	0,1454	0,05549
B	13,24	20,13	86,87	542,9	0,08284	0,1223	0,101	0,02833	0,1601	0,06432
B	13,14	20,74	85,98	536,9	0,08675	0,1089	0,1085	0,0351	0,1562	0,0602
B	9,668	18,1	61,06	286,3	0,08311	0,05428	0,01479	0,005769	0,168	0,06412
M	17,6	23,33	119	980,5	0,09289	0,2004	0,2136	0,1002	0,1696	0,07369
B	11,62	18,18	76,38	408,8	0,1175	0,1483	0,102	0,05564	0,1957	0,07255
B	9,667	18,49	61,49	289,1	0,08946	0,06258	0,02948	0,01514	0,2238	0,06413
B	12,04	28,14	76,85	449,9	0,08752	0,06	0,02367	0,02377	0,1854	0,05698
B	14,92	14,93	96,45	686,9	0,08098	0,08549	0,05539	0,03221	0,1687	0,05669



diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	12,27	29,97	77,42	465,4	0,07699	0,03398	0	0	0,1701	0,0596
B	10,88	15,62	70,41	358,9	0,1007	0,1069	0,05115	0,01571	0,1861	0,06837
B	12,83	15,73	82,89	506,9	0,0904	0,08269	0,05835	0,03078	0,1705	0,05913
B	14,2	20,53	92,41	618,4	0,08931	0,1108	0,05063	0,03058	0,1506	0,06009
B	13,9	16,62	88,97	599,4	0,06828	0,05319	0,02224	0,01339	0,1813	0,05536
B	11,49	14,59	73,99	404,9	0,1046	0,08228	0,05308	0,01969	0,1779	0,06574
M	16,25	19,51	109,8	815,8	0,1026	0,1893	0,2236	0,09194	0,2151	0,06578
B	12,16	18,03	78,29	455,3	0,09087	0,07838	0,02916	0,01527	0,1464	0,06284
B	13,9	19,24	88,73	602,9	0,07991	0,05326	0,02995	0,0207	0,1579	0,05594
B	13,47	14,06	87,32	546,3	0,1071	0,1155	0,05786	0,05266	0,1779	0,06639
B	13,7	17,64	87,76	571,1	0,0995	0,07957	0,04548	0,0316	0,1732	0,06088
B	15,73	11,28	102,8	747,2	0,1043	0,1299	0,1191	0,06211	0,1784	0,06259
B	12,45	16,41	82,85	476,7	0,09514	0,1511	0,1544	0,04846	0,2082	0,07325
B	14,64	16,85	94,21	666	0,08641	0,06698	0,05192	0,02791	0,1409	0,05355
M	19,44	18,82	128,1	1167	0,1089	0,1448	0,2256	0,1194	0,1823	0,06115
B	11,68	16,17	75,49	420,5	0,1128	0,09263	0,04279	0,03132	0,1853	0,06401
M	16,69	20,2	107,1	857,6	0,07497	0,07112	0,03649	0,02307	0,1846	0,05325
B	12,25	22,44	78,18	466,5	0,08192	0,052	0,01714	0,01261	0,1544	0,05976
B	17,85	13,23	114,6	992,1	0,07838	0,06217	0,04445	0,04178	0,122	0,05243
M	18,01	20,56	118,4	1007	0,1001	0,1289	0,117	0,07762	0,2116	0,06077
B	12,46	12,83	78,83	477,3	0,07372	0,04043	0,007173	0,01149	0,1613	0,06013
B	13,16	20,54	84,06	538,7	0,07335	0,05275	0,018	0,01256	0,1713	0,05888
B	14,87	20,21	96,12	680,9	0,09587	0,08345	0,06824	0,04951	0,1487	0,05748
B	12,65	18,17	82,69	485,6	0,1076	0,1334	0,08017	0,05074	0,1641	0,06854

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	12,47	17,31	80,45	480,1	0,08928	0,0763	0,03609	0,02369	0,1526	0,06046
M	18,49	17,52	121,3	1068	0,1012	0,1317	0,1491	0,09183	0,1832	0,06697
M	20,59	21,24	137,8	1320	0,1085	0,1644	0,2188	0,1121	0,1848	0,06222
B	15,04	16,74	98,73	689,4	0,09883	0,1364	0,07721	0,06142	0,1668	0,06869
M	13,82	24,49	92,33	595,9	0,1162	0,1681	0,1357	0,06759	0,2275	0,07237
B	12,54	16,32	81,25	476,3	0,1158	0,1085	0,05928	0,03279	0,1943	0,06612
M	23,09	19,83	152,1	1682	0,09342	0,1275	0,1676	0,1003	0,1505	0,05484
B	9,268	12,87	61,49	248,7	0,1634	0,2239	0,0973	0,05252	0,2378	0,09502
B	9,676	13,14	64,12	272,5	0,1255	0,2204	0,1188	0,07038	0,2057	0,09575
B	12,22	20,04	79,47	453,1	0,1096	0,1152	0,08175	0,02166	0,2124	0,06894
B	11,06	17,12	71,25	366,5	0,1194	0,1071	0,04063	0,04268	0,1954	0,07976
B	16,3	15,7	104,7	819,8	0,09427	0,06712	0,05526	0,04563	0,1711	0,05657
M	15,46	23,95	103,8	731,3	0,1183	0,187	0,203	0,0852	0,1807	0,07083
B	11,74	14,69	76,31	426	0,08099	0,09661	0,06726	0,02639	0,1499	0,06758
B	14,81	14,7	94,66	680,7	0,08472	0,05016	0,03416	0,02541	0,1659	0,05348
M	13,4	20,52	88,64	556,7	0,1106	0,1469	0,1445	0,08172	0,2116	0,07325
B	14,58	13,66	94,29	658,8	0,09832	0,08918	0,08222	0,04349	0,1739	0,0564
M	15,05	19,07	97,26	701,9	0,09215	0,08597	0,07486	0,04335	0,1561	0,05915
B	11,34	18,61	72,76	391,2	0,1049	0,08499	0,04302	0,02594	0,1927	0,06211
M	18,31	20,58	120,8	1052	0,1068	0,1248	0,1569	0,09451	0,186	0,05941
M	19,89	20,26	130,5	1214	0,1037	0,131	0,1411	0,09431	0,1802	0,06188
B	12,88	18,22	84,45	493,1	0,1218	0,1661	0,04825	0,05303	0,1709	0,07253
B	12,75	16,7	82,51	493,8	0,1125	0,1117	0,0388	0,02995	0,212	0,06623
B	9,295	13,9	59,96	257,8	0,1371	0,1225	0,03332	0,02421	0,2197	0,07696

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
M	24,63	21,6	165,5	1841	0,103	0,2106	0,231	0,1471	0,1991	0,06739
B	11,26	19,83	71,3	388,1	0,08511	0,04413	0,005067	0,005664	0,1637	0,06343
B	13,71	18,68	88,73	571	0,09916	0,107	0,05385	0,03783	0,1714	0,06843
B	9,847	15,68	63	293,2	0,09492	0,08419	0,0233	0,02416	0,1387	0,06891
B	8,571	13,1	54,53	221,3	0,1036	0,07632	0,02565	0,0151	0,1678	0,07126
B	13,46	18,75	87,44	551,1	0,1075	0,1138	0,04201	0,03152	0,1723	0,06317
B	12,34	12,27	78,94	468,5	0,09003	0,06307	0,02958	0,02647	0,1689	0,05808
B	13,94	13,17	90,31	594,2	0,1248	0,09755	0,101	0,06615	0,1976	0,06457
B	12,07	13,44	77,83	445,2	0,11	0,09009	0,03781	0,02798	0,1657	0,06608
B	11,75	17,56	75,89	422,9	0,1073	0,09713	0,05282	0,0444	0,1598	0,06677
B	11,67	20,02	75,21	416,2	0,1016	0,09453	0,042	0,02157	0,1859	0,06461
B	13,68	16,33	87,76	575,5	0,09277	0,07255	0,01752	0,0188	0,1631	0,06155
M	20,47	20,67	134,7	1299	0,09156	0,1313	0,1523	0,1015	0,2166	0,05419
B	10,96	17,62	70,79	365,6	0,09687	0,09752	0,05263	0,02788	0,1619	0,06408
M	20,55	20,86	137,8	1308	0,1046	0,1739	0,2085	0,1322	0,2127	0,06251
M	14,27	22,55	93,77	629,8	0,1038	0,1154	0,1463	0,06139	0,1926	0,05982
B	11,69	24,44	76,37	406,4	0,1236	0,1552	0,04515	0,04531	0,2131	0,07405
B	7,729	25,49	47,98	178,8	0,08098	0,04878	0	0	0,187	0,07285
B	7,691	25,44	48,34	170,4	0,08668	0,1199	0,09252	0,01364	0,2037	0,07751
B	11,54	14,44	74,65	402,9	0,09984	0,112	0,06737	0,02594	0,1818	0,06782
B	14,47	24,99	95,81	656,4	0,08837	0,123	0,1009	0,0389	0,1872	0,06341
B	14,74	25,42	94,7	668,6	0,08275	0,07214	0,04105	0,03027	0,184	0,0568
B	13,21	28,06	84,88	538,4	0,08671	0,06877	0,02987	0,03275	0,1628	0,05781
B	13,87	20,7	89,77	584,8	0,09578	0,1018	0,03688	0,02369	0,162	0,06688

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
B	13,62	23,23	87,19	573,2	0,09246	0,06747	0,02974	0,02443	0,1664	0,05801
B	10,32	16,35	65,31	324,9	0,09434	0,04994	0,01012	0,005495	0,1885	0,06201
B	10,26	16,58	65,85	320,8	0,08877	0,08066	0,04358	0,02438	0,1669	0,06714
B	9,683	19,34	61,05	285,7	0,08491	0,0503	0,02337	0,009615	0,158	0,06235
B	10,82	24,21	68,89	361,6	0,08192	0,06602	0,01548	0,00816	0,1976	0,06328
B	10,86	21,48	68,51	360,5	0,07431	0,04227	0	0	0,1661	0,05948
B	11,13	22,44	71,49	378,4	0,09566	0,08194	0,04824	0,02257	0,203	0,06552
B	12,77	29,43	81,35	507,9	0,08276	0,04234	0,01997	0,01499	0,1539	0,05637
B	9,333	21,94	59,01	264	0,0924	0,05605	0,03996	0,01282	0,1692	0,06576
B	12,88	28,92	82,5	514,3	0,08123	0,05824	0,06195	0,02343	0,1566	0,05708
B	10,29	27,61	65,67	321,4	0,0903	0,07658	0,05999	0,02738	0,1593	0,06127
B	10,16	19,59	64,73	311,7	0,1003	0,07504	0,005025	0,01116	0,1791	0,06331
B	9,423	27,88	59,26	271,3	0,08123	0,04971	0	0	0,1742	0,06059
B	14,59	22,68	96,39	657,1	0,08473	0,133	0,1029	0,03736	0,1454	0,06147
B	11,51	23,93	74,52	403,5	0,09261	0,1021	0,1112	0,04105	0,1388	0,0657
B	14,05	27,15	91,38	600,4	0,09929	0,1126	0,04462	0,04304	0,1537	0,06171
B	11,2	29,37	70,67	386	0,07449	0,03558	0	0	0,106	0,05502
M	15,22	30,62	103,4	716,9	0,1048	0,2087	0,255	0,09429	0,2128	0,07152
M	20,92	25,09	143	1347	0,1099	0,2236	0,3174	0,1474	0,2149	0,06879
M	21,56	22,39	142	1479	0,111	0,1159	0,2439	0,1389	0,1726	0,05623
M	20,13	28,25	131,2	1261	0,0978	0,1034	0,144	0,09791	0,1752	0,05533
M	16,6	28,08	108,3	858,1	0,08455	0,1023	0,09251	0,05302	0,159	0,05648
M	20,6	29,33	140,1	1265	0,1178	0,277	0,3514	0,152	0,2397	0,07016
B	7,76	24,54	47,92	181	0,05263	0,04362	0	0	0,1587	0,05884



## Lampiran II Hasil Pengujian Model C

radius_ mean	texture_ mean	perimeter_ mean	area_m ean	smoothness _mean	compactness _mean	concavity_ mean	concave.point s_mean	symmetry_ mean	fractal_dimensio n_mean	Act ual	Predic ted
17,99	10,38	122,8	1001	0,1184	0,2776	0,3001	0,1471	0,2419	0,07871	M	M
11,42	20,38	77,58	386,1	0,1425	0,2839	0,2414	0,1052	0,2597	0,09744	M	M
18,25	19,98	119,6	1040	0,09463	0,109	0,1127	0,074	0,1794	0,05742	M	M
13,71	20,83	90,2	577,9	0,1189	0,1645	0,09366	0,05985	0,2196	0,07451	M	M
12,46	24,04	83,97	475,9	0,1186	0,2396	0,2273	0,08543	0,203	0,08243	M	M
15,78	17,89	103,6	781	0,0971	0,1292	0,09954	0,06606	0,1842	0,06082	M	M
15,85	23,95	103,7	782,7	0,08401	0,1002	0,09938	0,05364	0,1847	0,05338	M	M
16,13	20,68	108,1	798,8	0,117	0,2022	0,1722	0,1028	0,2164	0,07356	M	M
13,54	14,36	87,46	566,3	0,09779	0,08129	0,06664	0,04781	0,1885	0,05766	B	B
13,08	15,71	85,63	520	0,1075	0,127	0,04568	0,0311	0,1967	0,06811	B	B
15,34	14,26	102,5	704,4	0,1073	0,2135	0,2077	0,09756	0,2521	0,07032	M	M
18,61	20,25	122,1	1094	0,0944	0,1066	0,149	0,07731	0,1697	0,05699	M	M
17,02	23,98	112,8	899,3	0,1197	0,1496	0,2417	0,1203	0,2248	0,06382	M	M
13,44	21,58	86,18	563	0,08162	0,06031	0,0311	0,02031	0,1784	0,05587	M	B
10,95	21,35	71,9	371,1	0,1227	0,1218	0,1044	0,05669	0,1895	0,0687	M	B
13,28	20,28	87,32	545,2	0,1041	0,1436	0,09847	0,06158	0,1974	0,06782	M	M
13,17	21,81	85,42	531,5	0,09714	0,1047	0,08259	0,05252	0,1746	0,06177	M	B
13,17	18,66	85,98	534,6	0,1158	0,1231	0,1226	0,0734	0,2128	0,06777	M	M
18,22	18,7	120,3	1033	0,1148	0,1485	0,1772	0,106	0,2092	0,0631	M	M
13,05	19,31	82,61	527,2	0,0806	0,03789	0,000692	0,004167	0,1819	0,05501	B	B
8,618	11,79	54,34	224,5	0,09752	0,05272	0,02061	0,007799	0,1683	0,07187	B	B
10,17	14,88	64,55	311,9	0,1134	0,08061	0,01084	0,0129	0,2743	0,0696	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
8,598	20,98	54,66	221,8	0,1243	0,08963	0,03	0,009259	0,1828	0,06757	B	B
9,465	21,01	60,11	269,4	0,1044	0,07773	0,02172	0,01504	0,1717	0,06899	B	B
13,53	10,94	87,91	559,2	0,1291	0,1047	0,06877	0,06556	0,2403	0,06641	B	B
13,34	15,86	86,49	520	0,1078	0,1535	0,1169	0,06987	0,1942	0,06902	B	M
13,27	14,76	84,74	551,7	0,07355	0,05055	0,03261	0,02648	0,1386	0,05318	B	B
20,26	23,03	132,4	1264	0,09078	0,1313	0,1465	0,08683	0,2095	0,05649	M	M
11,6	12,84	74,34	412,6	0,08983	0,07525	0,04196	0,0335	0,162	0,06582	B	B
6,981	13,43	43,79	143,5	0,117	0,07568	0	0	0,193	0,07818	B	B
13,11	15,56	87,21	530,2	0,1398	0,1765	0,2071	0,09601	0,1925	0,07692	M	M
11,34	21,26	72,48	396,5	0,08759	0,06575	0,05133	0,01899	0,1487	0,06529	B	B
12,63	20,76	82,15	480,4	0,09933	0,1209	0,1065	0,06021	0,1735	0,0707	B	B
10,51	20,19	68,64	334,2	0,1122	0,1303	0,06476	0,03068	0,1922	0,07782	B	B
8,726	15,83	55,84	230,9	0,115	0,08201	0,04132	0,01924	0,1649	0,07633	B	B
14,87	16,67	98,64	682,5	0,1162	0,1649	0,169	0,08923	0,2157	0,06768	M	M
11,41	10,82	73,34	403,3	0,09373	0,06685	0,03512	0,02623	0,1667	0,06113	B	B
13,37	16,39	86,1	553,5	0,07115	0,07325	0,08092	0,028	0,1422	0,05823	B	B
13,85	17,21	88,44	588,7	0,08785	0,06136	0,0142	0,01141	0,1614	0,0589	B	B
15,46	19,48	101,7	748,9	0,1092	0,1223	0,1466	0,08087	0,1931	0,05796	M	M
11,43	17,31	73,66	398	0,1092	0,09486	0,02031	0,01861	0,1645	0,06562	B	B
10,75	14,97	68,26	355,3	0,07793	0,05139	0,02251	0,007875	0,1399	0,05688	B	B
9,731	15,34	63,78	300,2	0,1072	0,1599	0,4108	0,07857	0,2548	0,09296	B	B
11,15	13,08	70,87	381,9	0,09754	0,05113	0,01982	0,01786	0,183	0,06105	B	B
16,84	19,46	108,4	880,2	0,07445	0,07223	0,0515	0,02771	0,1844	0,05268	B	M
14,97	19,76	95,5	690,2	0,08421	0,05352	0,01947	0,01939	0,1515	0,05266	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
17,47	24,68	116,1	984,6	0,1049	0,1603	0,2159	0,1043	0,1538	0,06365	M	M
8,671	14,45	54,42	227,2	0,09138	0,04276	0	0	0,1722	0,06724	B	B
27,22	21,87	182,1	2250	0,1094	0,1914	0,2871	0,1878	0,18	0,0577	M	M
11,41	14,92	73,53	402	0,09059	0,08155	0,06181	0,02361	0,1167	0,06217	B	B
12,3	15,9	78,83	463,7	0,0808	0,07253	0,03844	0,01654	0,1667	0,05474	B	B
14,22	23,12	94,37	609,9	0,1075	0,2413	0,1981	0,06618	0,2384	0,07542	M	M
19,18	22,49	127,5	1148	0,08523	0,1428	0,1114	0,06772	0,1767	0,05529	M	M
12,23	19,56	78,54	461	0,09586	0,08087	0,04187	0,04107	0,1979	0,06013	B	B
17,01	20,26	109,7	904,3	0,08772	0,07304	0,0695	0,0539	0,2026	0,05223	M	M
15,27	12,91	98,17	725,5	0,08182	0,0623	0,05892	0,03157	0,1359	0,05526	B	B
11,84	18,94	75,51	428	0,08871	0,069	0,02669	0,01393	0,1533	0,06057	B	B
28,11	18,47	188,5	2499	0,1142	0,1516	0,3201	0,1595	0,1648	0,05525	M	M
11,89	18,35	77,32	432,2	0,09363	0,1154	0,06636	0,03142	0,1967	0,06314	B	B
19,8	21,56	129,7	1230	0,09383	0,1306	0,1272	0,08691	0,2094	0,05581	M	M
19,53	32,47	128	1223	0,0842	0,113	0,1145	0,06637	0,1428	0,05313	M	M
13,27	17,02	84,55	546,4	0,08445	0,04994	0,03554	0,02456	0,1496	0,05674	B	B
14,34	13,47	92,51	641,2	0,09906	0,07624	0,05724	0,04603	0,2075	0,05448	B	B
20,51	27,81	134,4	1319	0,09159	0,1074	0,1554	0,0834	0,1448	0,05592	M	M
20,48	21,46	132,5	1306	0,08355	0,08348	0,09042	0,06022	0,1467	0,05177	M	M
13,75	23,77	88,54	590	0,08043	0,06807	0,04697	0,02344	0,1773	0,05429	B	B
12,89	14,11	84,95	512,2	0,0876	0,1346	0,1374	0,0398	0,1596	0,06409	B	B
20,94	23,56	138,9	1364	0,1007	0,1606	0,2712	0,131	0,2205	0,05898	M	M
17,3	17,08	113	928,2	0,1008	0,1041	0,1266	0,08353	0,1813	0,05613	M	M
13,96	17,05	91,43	602,4	0,1096	0,1279	0,09789	0,05246	0,1908	0,0613	M	B



radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
19,55	28,77	133,6	1207	0,0926	0,2063	0,1784	0,1144	0,1893	0,06232	M	M
20,31	27,06	132,9	1288	0,1	0,1088	0,1519	0,09333	0,1814	0,05572	M	M
12,87	16,21	82,38	512,2	0,09425	0,06219	0,039	0,01615	0,201	0,05769	B	B
9,742	15,67	61,5	289,9	0,09037	0,04689	0,01103	0,01407	0,2081	0,06312	B	B
17,93	24,48	115,2	998,9	0,08855	0,07027	0,05699	0,04744	0,1538	0,0551	M	M
11,74	14,02	74,24	427,3	0,07813	0,0434	0,02245	0,02763	0,2101	0,06113	B	B
19,4	18,18	127,2	1145	0,1037	0,1442	0,1626	0,09464	0,1893	0,05892	M	M
12,89	15,7	84,08	516,6	0,07818	0,0958	0,1115	0,0339	0,1432	0,05935	B	B
11,94	20,76	77,87	441	0,08605	0,1011	0,06574	0,03791	0,1588	0,06766	B	B
14,26	18,17	91,22	633,1	0,06576	0,0522	0,02475	0,01374	0,1635	0,05586	B	B
19,53	18,9	129,5	1217	0,115	0,1642	0,2197	0,1062	0,1792	0,06552	M	M
10,49	18,61	66,86	334,3	0,1068	0,06678	0,02297	0,0178	0,1482	0,066	B	B
11,6	24,49	74,23	417,2	0,07474	0,05688	0,01974	0,01313	0,1935	0,05878	B	B
9	14,4	56,36	246,3	0,07005	0,03116	0,003681	0,003472	0,1788	0,06833	B	B
12,18	14,08	77,25	461,4	0,07734	0,03212	0,01123	0,005051	0,1673	0,05649	B	B
14,11	12,88	90,03	616,5	0,09309	0,05306	0,01765	0,02733	0,1373	0,057	B	B
16,27	20,71	106,9	813,7	0,1169	0,1319	0,1478	0,08488	0,1948	0,06277	M	M
16,26	21,88	107,5	826,8	0,1165	0,1283	0,1799	0,07981	0,1869	0,06532	M	M
16,03	15,51	105,8	793,2	0,09491	0,1371	0,1204	0,07041	0,1782	0,05976	M	M
11,22	19,86	71,94	387,3	0,1054	0,06779	0,005006	0,007583	0,194	0,06028	B	B
12,3	19,02	77,88	464,4	0,08313	0,04202	0,007756	0,008535	0,1539	0,05945	B	B
11,06	14,96	71,49	373,9	0,1033	0,09097	0,05397	0,03341	0,1776	0,06907	B	B
11,71	15,45	75,03	420,3	0,115	0,07281	0,04006	0,0325	0,2009	0,06506	B	B
15,75	19,22	107,1	758,6	0,1243	0,2364	0,2914	0,1242	0,2375	0,07603	M	M

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
22,01	21,9	147,2	1482	0,1063	0,1954	0,2448	0,1501	0,1824	0,0614	M	M
10,57	20,22	70,15	338,3	0,09073	0,166	0,228	0,05941	0,2188	0,0845	B	B
13,46	28,21	85,89	562,1	0,07517	0,04726	0,01271	0,01117	0,1421	0,05763	B	B
11,27	12,96	73,16	386,3	0,1237	0,1111	0,079	0,0555	0,2018	0,06914	B	B
12,39	17,48	80,64	462,9	0,1042	0,1297	0,05892	0,0288	0,1779	0,06588	B	B
10,26	12,22	65,75	321,6	0,09996	0,07542	0,01923	0,01968	0,18	0,06569	B	B
8,734	16,84	55,27	234,3	0,1039	0,07428	0	0	0,1985	0,07098	B	B
12,1	17,72	78,07	446,2	0,1029	0,09758	0,04783	0,03326	0,1937	0,06161	B	B
11,06	14,83	70,31	378,2	0,07741	0,04768	0,02712	0,007246	0,1535	0,06214	B	B
11,93	10,91	76,14	442,7	0,08872	0,05242	0,02606	0,01796	0,1601	0,05541	B	B
12,94	16,17	83,18	507,6	0,09879	0,08836	0,03296	0,0239	0,1735	0,062	B	B
10,94	18,59	70,39	370	0,1004	0,0746	0,04944	0,02932	0,1486	0,06615	B	B
11,04	16,83	70,92	373,2	0,1077	0,07804	0,03046	0,0248	0,1714	0,0634	B	B
10,03	21,28	63,19	307,3	0,08117	0,03912	0,00247	0,005159	0,163	0,06439	B	B
10,48	14,98	67,49	333,6	0,09816	0,1013	0,06335	0,02218	0,1925	0,06915	B	B
20,18	19,54	133,8	1250	0,1133	0,1489	0,2133	0,1259	0,1724	0,06053	M	M
10,97	17,2	71,73	371,5	0,08915	0,1113	0,09457	0,03613	0,1489	0,0664	B	B
13,78	15,79	88,37	585,9	0,08817	0,06718	0,01055	0,009937	0,1405	0,05848	B	B
17,75	28,03	117,3	981,6	0,09997	0,1314	0,1698	0,08293	0,1713	0,05916	M	M
12,62	17,15	80,62	492,9	0,08583	0,0543	0,02966	0,02272	0,1799	0,05826	B	B
13	25,13	82,61	520,2	0,08369	0,05073	0,01206	0,01762	0,1667	0,05449	B	B
9,755	28,2	61,68	290,9	0,07984	0,04626	0,01541	0,01043	0,1621	0,05952	B	B
11,6	18,36	73,88	412,7	0,08508	0,05855	0,03367	0,01777	0,1516	0,05859	B	B
12,83	15,73	82,89	506,9	0,0904	0,08269	0,05835	0,03078	0,1705	0,05913	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
14,2	20,53	92,41	618,4	0,08931	0,1108	0,05063	0,03058	0,1506	0,06009	B	B
13,9	16,62	88,97	599,4	0,06828	0,05319	0,02224	0,01339	0,1813	0,05536	B	B
13,7	17,64	87,76	571,1	0,0995	0,07957	0,04548	0,0316	0,1732	0,06088	B	B
19,44	18,82	128,1	1167	0,1089	0,1448	0,2256	0,1194	0,1823	0,06115	M	M
16,69	20,2	107,1	857,6	0,07497	0,07112	0,03649	0,02307	0,1846	0,05325	M	M
17,85	13,23	114,6	992,1	0,07838	0,06217	0,04445	0,04178	0,122	0,05243	B	M
18,01	20,56	118,4	1007	0,1001	0,1289	0,117	0,07762	0,2116	0,06077	M	M
12,54	16,32	81,25	476,3	0,1158	0,1085	0,05928	0,03279	0,1943	0,06612	B	B
11,06	17,12	71,25	366,5	0,1194	0,1071	0,04063	0,04268	0,1954	0,07976	B	B
13,4	20,52	88,64	556,7	0,1106	0,1469	0,1445	0,08172	0,2116	0,07325	M	M
19,89	20,26	130,5	1214	0,1037	0,131	0,1411	0,09431	0,1802	0,06188	M	M
12,88	18,22	84,45	493,1	0,1218	0,1661	0,04825	0,05303	0,1709	0,07253	B	B
24,63	21,6	165,5	1841	0,103	0,2106	0,231	0,1471	0,1991	0,06739	M	M
13,71	18,68	88,73	571	0,09916	0,107	0,05385	0,03783	0,1714	0,06843	B	B
13,94	13,17	90,31	594,2	0,1248	0,09755	0,101	0,06615	0,1976	0,06457	B	B
12,07	13,44	77,83	445,2	0,11	0,09009	0,03781	0,02798	0,1657	0,06608	B	B
20,55	20,86	137,8	1308	0,1046	0,1739	0,2085	0,1322	0,2127	0,06251	M	M
7,729	25,49	47,98	178,8	0,08098	0,04878	0	0	0,187	0,07285	B	B
10,32	16,35	65,31	324,9	0,09434	0,04994	0,01012	0,005495	0,1885	0,06201	B	B
9,683	19,34	61,05	285,7	0,08491	0,0503	0,02337	0,009615	0,158	0,06235	B	B
9,333	21,94	59,01	264	0,0924	0,05605	0,03996	0,01282	0,1692	0,06576	B	B
9,423	27,88	59,26	271,3	0,08123	0,04971	0	0	0,1742	0,06059	B	B
14,05	27,15	91,38	600,4	0,09929	0,1126	0,04462	0,04304	0,1537	0,06171	B	B
7,76	24,54	47,92	181	0,05263	0,04362	0	0	0,1587	0,05884	B	B



### Lampiran III Hasil Pengujian Model C Tuning + Normalisasi

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-0,4760	-0,8346	-0,3868	-0,5052	2,2355	1,2432	0,8655	0,8239	1,0045	1,8883	M	M
0,4888	1,0835	0,4828	0,3632	-0,8781	-0,0784	0,1327	0,1217	0,1291	-1,3339	M	M
1,6126	0,6650	1,5651	1,7195	0,1386	-0,0311	0,7414	1,1870	-0,8376	-1,2531	M	M
-0,1667	-1,1462	-0,1856	0,2517	0,1017	-0,4365	-0,2780	-0,0286	0,2677	-0,7277	B	B
-1,3119	-1,5926	-1,3017	1,0826	0,4294	-0,7464	-0,7431	-0,7257	0,0123	0,8856	B	B
0,3441	-1,1694	0,4334	0,1407	0,7778	2,0669	1,4915	1,2535	2,5876	1,0654	M	M
1,2777	1,3532	1,3511	1,2307	0,7139	1,5973	1,7950	1,9452	1,3547	-0,1172	M	M
0,7414	0,5348	0,7462	0,6098	-0,0185	0,5540	0,5770	0,2902	0,3078	-0,8835	M	M
-0,2404	0,2303	-0,1913	0,3117	0,5503	0,7434	0,1213	0,3263	0,5923	0,7114	M	M
-0,2716	0,5860	-0,2695	0,3506	0,0554	0,0068	-0,0779	0,0928	-0,2394	-0,1455	M	M
-0,5895	-1,0834	-0,5732	0,5842	0,4792	-0,2541	-0,2870	-0,5523	-0,4983	-0,3353	B	B
-0,1808	0,6999	-0,2082	0,2668	-0,6286	-0,5181	-0,5179	-0,3886	-0,0096	-0,7957	B	B
1,1614	-0,1371	1,1659	1,0744	1,3111	0,8361	1,1089	1,4710	1,0228	0,0428	M	M
-0,3057	0,0047	-0,3852	0,3628	-1,1206	-1,2582	-1,1052	-1,1533	0,0269	-1,1030	B	B
-1,5633	-1,7437	-1,5486	1,2230	0,0825	-0,9774	-0,8554	-1,0597	-0,4692	1,2850	B	B
-0,4107	1,0580	-0,3819	0,4430	1,1262	0,4139	0,3011	0,5106	0,3406	0,4394	M	M
-1,3230	0,4000	-1,3111	1,0954	0,5716	-0,5039	-0,8414	-0,8731	-0,3451	0,8771	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
0,5513	0,0838	0,4992	0,4626	-0,3328	-0,3806	0,1115	0,4500	-0,0497	-1,2588	M	M
-0,1695	-1,9413	-0,1670	-0,2719	2,3279	0,0068	-0,2512	0,4289	2,1572	0,5116	B	B
1,1131	-0,7300	1,1618	0,9977	0,7210	2,0877	0,9985	1,5226	1,2416	0,5570	M	M
1,3884	1,2323	1,2359	1,1966	-0,4316	0,3079	0,7276	0,8708	0,5157	-0,9217	M	M
0,1455	-0,9416	0,1564	0,0085	1,1974	0,5597	0,1362	0,5598	1,1103	0,0938	B	B
-0,1922	-0,2301	-0,2210	0,2836	0,4152	-0,4296	-0,6154	-0,5443	-0,6333	-0,8070	B	B
-2,0279	-1,3624	-1,9828	1,4532	1,4675	-0,5427	-1,1139	-1,2607	0,4318	2,1787	B	B
-1,0321	0,0001	-1,0107	0,9059	0,2510	-0,3514	-0,7382	-0,9512	1,4787	0,2850	B	B
-0,7058	-0,2231	-0,6913	0,6888	1,2685	-0,0500	-0,2270	-0,3626	-0,0387	0,3403	B	B
-1,5327	-0,8044	-1,4869	1,2048	1,3253	-0,4228	-0,5956	-0,7649	-0,5932	1,9167	B	B
-1,4691	-0,8207	-1,3675	1,1642	-0,1237	0,3779	0,0481	-0,6659	-1,8480	1,2510	B	B
1,0847	0,1675	0,9149	0,9295	-0,8774	-0,7029	-0,1991	0,1815	1,1577	-1,7772	M	M
1,6069	1,3556	1,5816	1,5263	0,3654	1,0331	2,0785	1,7004	1,4240	-0,2362	M	M
0,3782	0,0443	0,4005	0,2671	0,9129	0,3401	0,7250	0,8234	0,4355	-0,6852	M	M
-1,2455	-1,7018	-1,2646	1,0411	-0,2745	-1,1990	-1,1139	-1,2607	0,3333	0,2015	B	B
0,5626	-0,2882	0,5404	0,4493	0,0604	0,1772	0,0711	0,2709	0,1801	-0,0450	M	M
0,0887	-0,9555	0,0823	0,0420	0,2375	-0,0424	-0,0493	0,1650	-0,3196	-0,2815	B	B
-1,6766	0,3279	-1,5926	1,2815	-0,1643	0,4953	0,5432	-0,7020	1,4970	2,8061	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-1,2475	-0,9183	-1,1601	-1,0079	0,7707	1,0520	4,0392	0,7641	2,6861	4,2721	B	B
-0,9442	-2,2273	-0,9547	-0,8448	-0,0299	-0,8896	-0,7964	-0,8231	-1,5708	0,1703	B	B
0,9485	1,2533	0,9931	0,9369	0,6072	1,0596	1,5943	1,4272	-0,9981	0,1207	M	M
-0,5128	-1,6042	-0,5399	-0,5421	0,4579	-0,6538	-0,6138	-0,3072	0,5376	-0,4600	B	B
-0,1979	0,0791	-0,2522	-0,2543	-0,4181	-0,7851	-0,3791	-0,3747	-0,7792	-0,8622	M	M
-0,8647	-1,0648	-0,8955	-0,8019	0,3015	-0,8883	-0,8175	-0,5950	-0,8960	0,5513	B	B
-1,5483	-1,1252	-1,5453	-1,2153	-0,3541	-1,1660	-1,1139	-1,2607	-0,3269	0,6292	B	B
-0,3170	0,6813	-0,4099	-0,3651	-2,4069	-1,6087	-1,0939	-1,2130	-1,5197	-1,4812	B	B
0,3271	0,7255	0,2861	0,1583	-0,4117	0,0163	-0,4397	-0,4198	-0,3087	0,0527	M	B
0,2079	0,9115	0,3470	0,0469	0,5716	1,7734	1,0148	1,0273	-0,2722	0,5556	M	M
1,1216	0,5930	1,0466	1,0489	-1,6105	-0,3393	0,2697	0,2284	-0,1518	-1,3310	M	M
0,0916	0,2163	0,1037	0,0346	0,1678	0,3079	0,3663	0,2804	0,5048	0,2638	M	M
-0,0900	1,0370	-0,0168	-0,1622	2,5554	1,3720	0,8405	1,1041	1,6027	1,6164	M	M
0,8180	0,2256	0,7297	0,7087	-0,6143	-0,5927	-0,2421	0,1284	0,7820	-1,4968	M	M
-0,2887	0,7557	-0,2037	-0,3566	0,2730	0,8324	-0,0219	0,0541	0,1400	1,4592	B	B
0,3243	-1,4833	0,2552	0,2006	-1,0339	-0,7960	-0,3748	-0,4471	-1,6510	-1,0676	B	B
1,8310	0,6627	1,7585	1,8047	-0,3882	0,5767	0,9433	1,2033	-0,1701	-1,7786	M	M
-0,6491	-0,0813	-0,6774	-0,6447	-0,5440	-0,6692	-0,7791	-0,9017	-1,0163	-0,3155	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
0,0178	1,0510	0,0371	-0,1256	-0,1230	0,4972	0,2848	0,4046	1,5444	0,2170	M	M
-1,1144	-0,4207	-1,1078	-0,9482	-1,1249	-0,8572	-0,3896	-0,9847	0,5558	0,0499	B	B
0,4605	0,2233	0,4375	0,3024	0,4366	0,3041	0,3249	0,4046	0,4501	0,0329	M	M
-0,2433	-0,5277	-0,3053	-0,3083	-0,8469	-1,0301	-0,6681	-0,6278	-1,1513	-0,8580	B	B
-0,3681	0,7069	-0,2761	-0,4310	0,8845	1,4307	1,0123	0,5075	1,1358	1,3799	M	M
0,8294	-0,0487	0,8820	0,6823	1,2613	1,0009	1,2820	1,5484	1,1650	0,0641	M	M
-1,2940	-0,7858	-1,3070	-1,0664	-0,8333	-1,2018	-0,9067	-0,8311	-0,9507	0,1746	B	B
0,9457	4,6478	0,8820	0,7550	0,1251	0,4821	0,6636	1,0100	-0,0096	-0,4444	M	M
-0,9867	1,3788	-0,9860	-0,8749	0,0149	-0,6059	-0,8155	-0,8445	0,3114	0,0697	B	B
1,9332	0,9929	1,9314	2,0150	0,3086	1,0653	2,2880	2,1153	1,4350	-0,5407	M	M
1,5388	2,2042	1,7133	1,5689	-0,2674	1,9306	1,1239	1,6875	0,2969	-0,0676	M	M
0,4207	0,0210	0,3305	0,2947	-1,2770	-0,9123	-0,5859	-0,5270	-0,9653	-1,1852	M	B
-0,9697	0,2558	-0,9247	-0,8809	0,8418	0,4650	-0,0542	-0,5218	-0,5239	0,8247	B	B
1,0791	1,2068	0,9560	0,9775	-0,5553	-0,6451	-0,3990	-0,0381	-0,9981	-1,0903	M	M
-0,7938	-1,1927	-0,8304	-0,7340	-0,1828	-1,2425	-1,0952	-1,1748	0,5194	-0,6498	B	B
-0,6774	-1,2252	-0,7296	-0,6467	-1,2962	-1,1539	-0,8323	-0,5486	1,0556	-0,2362	B	B
-0,4391	-0,2068	-0,4996	-0,4714	-0,8838	-1,1774	-1,0906	-1,1854	-0,4181	-0,6016	B	B
-0,6207	0,3419	-0,5802	-0,6078	-0,7331	-0,0614	-0,2893	-0,2837	-0,8157	0,6887	B	B
0,2363	-0,0441	0,2083	0,0921	-0,4579	-0,1157	-0,3688	-0,0188	0,2458	-0,6059	B	B



radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-0,3341	-0,7602	-0,3633	-0,4012	0,2943	-0,4717	-0,3418	-0,3922	-0,2977	0,2694	B	B
1,5331	-0,0906	1,5445	1,5973	1,3253	1,1334	1,6420	1,4762	-0,0716	0,3856	M	M
-1,0321	-0,1580	-1,0333	-0,9110	0,7423	-0,7112	-0,8258	-0,8020	-1,2024	0,4536	B	B
-0,3880	-1,3763	-0,3983	-0,4285	-0,5987	-0,4707	-0,6056	-0,6041	-0,7683	-0,1980	B	B
-0,5526	-1,2113	-0,6057	-0,5498	-1,3524	-1,3675	-0,9730	-1,1305	-0,5056	-0,8934	B	B
-0,5469	-0,9485	-0,5745	-0,5598	-0,6847	-0,7364	-0,8638	-0,8247	-0,6333	-0,2135	B	B
-0,4135	-0,4626	-0,4411	-0,4688	0,4579	-0,5245	-0,7134	-0,7177	-0,3816	-0,4189	B	B
-0,0049	-1,4903	-0,0798	-0,1091	-0,2325	-0,9710	-0,8925	-0,5564	-1,6000	-0,8211	B	B
-0,5951	-0,3161	-0,6535	-0,5936	-1,3887	-1,2387	-1,0945	-1,1166	-1,5671	-0,2971	B	B
-1,1570	-0,4091	-1,1342	-0,9778	0,3086	-0,5885	-0,7989	-0,8033	0,2859	0,0726	B	B
-0,8704	-1,0067	-0,8428	-0,7985	0,4934	-0,2532	-0,4369	-0,3997	-0,1299	0,8884	B	B
-1,4896	-0,8834	-1,4498	-1,1761	-0,9549	-0,5181	-0,5217	-0,6471	0,4318	0,4833	B	B
-1,3312	-0,2254	-1,3231	-1,0693	0,3228	-0,8479	-0,7740	-0,8984	-1,1148	0,9620	B	B
-0,3880	-0,1045	-0,4156	-0,4498	0,0284	-0,4700	-0,7767	-0,8017	-0,1919	-0,1370	B	B
1,7913	0,5790	1,7215	1,8133	-0,3456	0,1659	0,1153	0,7456	-0,7063	-1,0237	M	M
2,2368	0,6069	2,2730	2,3503	0,7067	1,7242	1,9569	2,6076	0,0452	-0,1980	M	M
0,5796	-0,7486	0,5898	0,3797	0,1735	0,7471	-0,2796	0,1302	0,6507	0,4139	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-0,8761	-1,0136	-0,8765	-0,8019	-1,1725	-0,6353	-0,6691	-0,7262	0,6981	-0,0478	B	B
-0,0702	-0,7276	-0,1481	-0,1656	-1,8558	-1,0609	-0,8574	-1,0415	-0,7464	-1,1412	B	B
1,5388	0,9115	1,5199	1,4751	0,3299	0,5199	1,2143	1,3705	0,6470	-0,5605	M	M
-1,0974	-1,6437	-1,0790	-0,9471	0,2560	-0,5476	-0,8727	-0,7535	-0,0424	0,4097	B	B
0,3867	0,1582	0,4293	0,2552	1,3964	0,9819	1,2582	1,0878	0,4282	0,6575	M	M
-0,5753	-0,3649	-0,5720	-0,5930	0,4650	-0,1280	-0,5139	-0,4036	0,4574	-0,1682	B	B
-0,0191	-0,4905	-0,0913	-0,1301	-1,1313	-0,9606	-0,7776	-0,4229	-0,6224	-0,7305	B	B
-0,1752	-0,0929	-0,1592	-0,2750	0,6783	0,1961	-0,0376	0,1260	-0,0205	-0,2843	B	B
-0,6604	-0,4719	-0,6876	-0,6336	-0,3904	-0,7957	-0,7560	-0,8386	0,1291	-0,3693	B	B
-0,6235	-1,9483	-0,6514	-0,6030	-0,5432	-0,9831	-0,7870	-0,7979	-0,7683	-1,0463	B	B
-0,3369	-0,7253	-0,3617	-0,4185	0,1728	-0,3026	-0,7004	-0,6448	-0,2795	-0,1130	B	B
0,3895	0,4163	0,4499	0,4212	1,1120	0,9990	0,7953	0,9250	0,9972	0,8275	M	M
-0,1326	-0,0371	-0,1033	-0,2262	-0,4117	0,1961	0,0975	-0,0206	0,1327	-0,1399	B	B
-1,2444	-0,0394	-1,2362	-1,0377	0,7921	-0,3978	-1,0018	-0,7538	2,6496	1,0612	B	B
-0,8505	-0,6207	-0,8848	-0,7780	-1,0559	-1,2497	-0,9422	-0,9076	-1,0966	-0,1866	B	B
-0,3993	-0,3766	-0,4522	-0,4364	-1,2372	-1,1194	-0,9380	-0,7878	-1,2863	-1,0421	B	B
1,7175	0,0582	1,7215	1,6911	1,2045	0,8437	1,5617	1,9839	-0,3196	-0,3212	M	M
1,3316	0,6232	1,3059	1,2932	0,3868	0,6544	0,8856	0,9927	0,4792	-0,2093	M	M

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-0,3568	0,0582	-0,3827	-0,4140	-0,3555	-0,4830	-0,8885	-0,7221	0,1801	0,0952	B	B
1,9786	0,2861	1,8985	2,0718	0,0341	0,2492	0,8580	1,7159	-0,9397	-0,8764	M	M
-0,6405	0,5232	-0,6230	-0,6334	-2,1495	0,0390	-0,0129	-0,6458	-1,6875	0,4706	B	B
-0,4277	-0,4975	-0,4671	-0,4603	-0,7487	-0,9475	-0,7418	-0,6752	-0,0460	-0,6427	B	B
-0,2603	1,3858	-0,3238	-0,3324	-0,6008	-0,9901	-0,7662	-0,7278	-0,7026	-0,9854	B	B
-0,7115	-0,2580	-0,6416	-0,6993	1,5031	0,8324	0,1656	0,1732	0,5303	1,3813	B	B
-1,2657	-0,1859	-1,2543	-1,0394	-0,4906	-0,7907	-0,7441	-0,8705	1,5553	0,1887	B	B
-0,5270	2,4832	-0,5988	-0,5385	-1,3773	-1,3323	-1,1139	-1,2607	-0,4035	-0,4529	B	B
-0,0645	-0,6207	-0,1234	-0,1577	-1,9966	-0,9685	-0,8349	-0,9156	0,0050	-1,0534	B	B
0,6023	0,0512	0,7338	0,4572	0,4437	1,6087	1,6909	1,1087	1,2380	0,4224	M	M
-0,5582	-0,2929	-0,5629	-0,5672	-0,3904	-0,4916	-0,7481	-0,8672	-1,2680	0,0060	B	B
-0,1213	-0,3835	-0,1732	-0,2381	0,2232	-0,4690	-0,5434	-0,4463	-0,2904	-0,2716	B	B
1,5076	-0,1092	1,4869	1,4552	0,8916	0,7661	1,7160	1,8164	0,0415	-0,2334	M	M
0,2108	0,2140	0,1708	0,0739	-0,0349	-0,3956	-0,2579	0,0152	-1,1841	-0,7532	B	B
-0,4192	-0,2603	-0,3819	-0,4811	0,7992	0,5502	-0,1082	0,0469	-0,6224	0,8133	B	B
-0,0872	1,2091	0,0149	-0,1676	1,4107	1,2073	0,5883	0,4812	1,6903	1,3558	M	M
2,5433	0,1256	2,4746	2,9186	-0,2091	0,4385	0,9885	1,3242	-1,1185	-1,1271	M	M
-0,6774	-1,0694	-0,6444	-0,6504	-1,0929	-0,1464	-0,2702	-0,5806	-1,1403	0,6774	B	B

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean	Actual	Predicted
-0,2064	0,2861	-0,1370	-0,2790	1,0125	0,8058	0,6987	0,8453	1,1103	1,4804	M	M
0,1285	-1,3089	0,0955	0,0111	0,1393	-0,2871	-0,0825	-0,1399	-0,2649	-0,9061	B	B
0,2618	-0,0511	0,2177	0,1336	-0,2994	-0,3479	-0,1749	-0,1435	-0,9142	-0,5166	M	M
-0,8136	0,1256	-0,8506	-0,7581	-0,7999	-1,1401	-1,0503	-1,1147	-0,6370	0,0896	B	B
-0,1894	-0,1255	-0,1864	-0,2949	0,7921	0,1791	-0,5869	-0,4484	-0,3233	0,0527	B	B
-0,5072	-1,6321	-0,5362	-0,5296	-0,4501	-0,7815	-0,7428	-0,5785	-0,4473	-0,6682	B	B
-0,8988	-0,3882	-0,8716	-0,8220	0,0362	-0,1292	-0,4537	-0,5422	-0,7026	0,1816	B	B
1,8225	0,3651	1,8861	1,8559	0,5859	1,3171	1,5015	2,1463	1,1504	-0,0407	M	M
-1,8156	1,4416	-1,8103	-1,3529	-1,0936	-1,0520	-1,1139	-1,2607	0,2130	1,4238	B	B
-0,2603	2,0391	-0,2917	-0,3310	-0,6862	-0,6735	-0,7392	-0,4167	-0,6698	-0,7064	B	B
-0,0730	0,3279	-0,0905	-0,1992	-0,0413	-0,0481	-0,6513	-0,6502	-0,6990	0,5782	B	B
-0,9271	0,5093	-0,9654	-0,8365	-1,5678	-1,1753	-1,1139	-1,2607	-0,5494	-0,4699	B	B
-0,8505	0,7325	-0,8428	-0,7857	-0,0498	-0,4242	-0,5088	-0,6791	0,7966	0,3856	B	B
-1,3349	1,9973	-1,3461	-1,0900	-1,0758	-1,0344	-1,1139	-1,2607	-0,2540	-0,3127	B	B
-0,7427	1,0789	-0,7181	-0,7143	-0,2667	-0,0424	0,2810	-0,2028	-1,5452	0,4111	B	B
2,1091	0,7208	2,0590	2,3418	1,0409	0,2189	1,9456	2,3189	-0,3123	-0,9302	M	M
1,7034	2,0833	1,6145	1,7223	0,1024	-0,0178	0,6924	1,2626	-0,2175	-1,0577	M	M