KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST* (RF)

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

Disusun Oleh
SITI TARWIYATUL FIFI ISTAMMAROH
H72216045

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2020

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama	:	SITI TARWIYATUL FIFI ISTAMMAROH
NIM	:	H72216045
Program Studi	:	Matematika
Angkatan	:	2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul " KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST* (RF) ". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 09 Maret 2020

Yang menyatakan,

METERAI TEMPEL

450C0AHF557382253

6000

ENAM RIBU RUPIAH

SITI TARWIYATUL FIFI ISTAMMAROH

NIM. H72216045

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama	:	SITI TARWIYATUL FIFI ISTAMMAROH
NIM	:	H72216045
Judul Skripsi	:	KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST* (RF)

telah diperiksa dan disetujui untuk diujikan.

Surabaya, 09 Maret 2020

Pembimbing

Wika Dianita Utami, M.Sc

NIP. 199206102018012003

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : SITI TARWIYATUL FIFI ISTAMMAROH
NIM : H72216045
Judul Skripsi : KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST* (RF)

Telah dipertahankan di depan Tim Penguji
pada tanggal 20 April 2020

Mengesahkan,
Tim Penguji

Penguji I

Wika Dianita Utami, M.Sc
NIP. 199206102018012003

Penguji II

Dr. Moh. Hafiyusholeh, M.Pmat
NIP. 198002042014031001

Penguji III

Nurissaidah Ulinnuha, M.Kom
NIP. 199011022014032004

Penguji IV

Putroue Keumala Intan, M.Si
NIP. 198805282018012001

Mengetahui,
Plt. Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. Evi Fatimatur Rusydiyah, M.Ag
NIP. 197312272005012003

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : SITI TARWIYATUL FIFI ISTAMMAROH
NIM : H72216045
Fakultas/Jurusan : SAINTEK/MATEMATIKA
E-mail address : fifiistammaroh@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
■ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (……)
yang berjudul :
KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE RANDOM FOREST (RF)

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 28 Juli 2020

Penulis

(SITI TARWIYATUL FIFI ISTAMMAROH)

ABSTRAK

KLASIFIKASI REKURENSI PASIEN KANKER PAYUDARA MENGGUNAKAN METODE *RANDOM FOREST* (RF)

Kanker payudara adalah tumor ganas yang berasal dari sel-sel yang terdapat pada payudara. Pengobatan kanker payudara tergantung pada besarnya ukuran dan stadium kanker. Kasus rekurensi sering terjadi walaupun pengobatan awal telah dilakukan. Penelitian ini bertujuan untuk mengetahui bagaimana klasifikasi pasien kanker payudara di Rumah Sakit F di Surabaya yang rekuren dan tidak rekuren berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker payudara dan mengetahui faktor-faktor yang berpengaruh terhadap rekurensi kanker payudara. Rekurensi yang dimaksud dalam penelitian ini yaitu kembalinya pasien kanker payudara di Rumah Sakit F di Surabaya karena penyakit yang sama. Adapun variabel yang digunakan yaitu variabel responnya adalah rekuren dan tidak rekuren, sedangkan variabel penjelasnya adalah usia, jenis pengobatan yang dijalani, status pernikahan, riwayat kanker dalam keluarga, ukuran tumor, dan stadium klinik. Metode *random forest* merupakan suatu algoritma yang digunakan pada klasifikasi data. Data yang digunakan yaitu data sekunder, yang diperoleh dari rekam medis pasien kanker payudara di Rumah Sakit F di Surabaya pada tahun 2017 sampai Desember 2018 dengan jumlah dan sebanyak 350 pasien. Hasil pengklasifikasian pada metode *random forest* adalah variabel yang paling berpengaruh terhadap rekuren kanker payudara adalah variabel usia, variabel stadium klinik, ukuran tumor dan jenis pengobatan yang dijalani pasien. Akurasi yang diperoleh untuk data prediksi sebesar 51,76%.

Kata kunci: Klasifikasi, *Random Forest*, Rekurensi, Kanker Payudara

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

RECURRENCE RECURRENCE OF BREAST CANCER PATIENTS USING THE RANDOM FOREST METHOD

Breast cancer is a malignant tumor originating from cells found in the breast. Treatment of breast cancer depends on the size and stage of the cancer. Cases of recurrence often occur despite initial treatment. The recurrence referred to in this study is the return of breast cancer patients at F Hospital in Surabaya due to the same disease. This study aims to determine the factors that influence the recurrence of breast cancer. The variables used are the response variables are recurrences and not recurrences, while the explanatory variables are age, type of treatment undertaken, marital status, family history of cancer, tumor size, and clinical stage. The random forest method is an algorithm used in data classification. The data used are secondary data, obtained from the medical records of breast cancer patients at F Hospital in Surabaya in 2017 to December 2018 with a number and as many as 350 patients. The results of the classification in the random forest method are the variables that most influence the recurrence of breast cancer are the age variable, clinical stage variable, tumor size and the type of treatment the patient is undergoing. The accuracy obtained for predictive data is 51.76%.

Keywords: Classification, Random Forest, Recurrence, Breast Cancer

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

HALAMAN JUDUL . . . . . **i**
HALAMAN PERNYATAAN . . . . . **ii**
LEMBAR PERSETUJUAN PEMBIMBING . . . . . **iii**
PENGESAHAN TIM PENGUJI SKRIPSI . . . . . **iv**
LEMBAR PERSETUJUAN PUBLIKASI . . . . . **v**
ABSTRAK . . . . . **vi**
ABSTRACT . . . . . **vii**
DAFTAR ISI **1**
DAFTAR TABEL **4**
DAFTAR GAMBAR **5**
I PENDAHULUAN **6**
1.1. Latar Belakang Masalah 6
1.2. Rumusan Masalah 11
1.3. Tujuan Penelitian 11
1.4. Manfaat Penelitian 11
1.5. Batasan Masalah 12
1.6. Sistematika Penulisan 12
II TINJAUAN PUSTAKA **14**
2.1. Kanker Payudara 14
2.2. Faktor-faktor yang mempengaruhi rekurensi pasien kanker payudara 18
2.2.1. Usia 18
2.2.2. Jenis Pengobatan yang Dijalani 19
2.2.3. Status Pernikahan 21
2.2.4. Riwayat Kanker Payudara Dalam Keluarga 21
2.2.5. Ukuran Tumor 22
2.2.6. Stadium Klinis 22
2.3. Klasifikasi 23

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.4. *Decision Tree* 24
2.5. *Random Forest* 25
2.6. Ukuran Tingkat Kepentingan 30
2.7. *Confussion Matrix* 30
2.8. Integrasi dalam Al-qur'an 34
III METODE PENELITIAN 36
3.1. Jenis Penelitian 36
3.2. Sumber Data dan Variabel Penelitian 36
3.3. Teknik Analisis Data 37
IV HASIL DAN PEMBAHASAN 41
4.1. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Usia Pasien 41
4.2. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Jenis Pengobatan Yang Dijalani 42
4.3. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Status Pernikahan 43
4.4. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Riwayat Kanker Dalam Keluarga 44
4.5. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Ukuran Tumor 45
4.6. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Stadium Klinik 45
4.7. Hasil Analisis Menggunakan Metode *Random Forest* 46
4.7.1. Penentuan Data Latih *(Training)* dan Data Uji *(Testing)* . . . 46
4.7.2. Penentuan *Mtry* (Peubah Penjelas) dan *Ntree* (Jumlah Pohon) Terbaik 47
4.7.3. Analisis Model 49
4.7.4. Pengujian Akurasi atau *Confussion Matrix* 61
V PENUTUP 63
5.1. Kesimpulan 63
5.2. Saran 64
DAFTAR PUSTAKA 65

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

A SKRIP PROGRAM R 69

B Data Pasien Kanker Payudara Periode Januari 2017 - Desember 2018 72

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

2.1 Rekurensi Pasien Berdasarkan Stadium Klinik 22

2.2 Evaluasi Klasifikasi pada Akurasi 31

3.1 Variabel Independen 37

4.1 Data Latih (*Training*) dan Data Uji (*Testing*) 47

4.2 Pengujian nilai *error* OOB 48

4.3 Pengujian nilai *error* OOB Dengan Nilai *ntree* Berbeda 49

4.4 Data Pasien Kanker Payudara 51

4.5 Data Pasien Kanker Payudara 54

4.6 Data Pasien Kanker Payudara 56

4.7 Karakteristik Pasien Rekuren Kanker Payudara di Rumah Sakit F Surabaya 60

4.8 Karakteristik Pasien Tidak Rekuren Kanker Payudara di Rumah Sakit F Surabaya 61

4.9 Hasil Prediksi Data Uji (*Testing*) 61

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

2.1 Contoh *Classification (Decision Tree)* 24

2.2 Contoh *Random Forest* 27

3.1 Diagram Alur Penelitian 40

4.1 Distribusi Usia Pasien Kanker Payudara Tahun 2017-2018 41

4.2 Distribusi Jenis Pengobatan Pasien Kanker Payudara Tahun 2017-2018 42

4.3 Distribusi Status Pernikahan Pasien Kanker Payudara Tahun 2017-2018 43

4.4 Distribusi Riwayat Kanker Dalam Keluarga Pasien Kanker Payudara Tahun 2017-2018 44

4.5 Distribusi Ukuran Tumor Pasien Kanker Payudara Tahun 2017-2018 45

4.6 Distribusi Stadium Klinik Pasien Kanker Payudara Tahun 2017-2018 46

4.7 Ukuran Tingkat Kepentingan 50

4.8 *Tree* Ke-1 Dengan *Root* Pertama 54

4.9 *Tree* Dengan *Root* Kedua 56

4.10 *Tree* Yang Terbentuk Berdasarkan Perhitungan Dengan Data Sampel n=5 58

4.11 *Decision Tree* Klasifikasi Optimum 58

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Indonesia mengalami perubahan dan juga menghadapi beban masalah yang besar. Hal tersebut ditandai dengan adanya peristiwa penyakit yang menular serta tidak menular secara serentak dalam bermasyarakat. Perubahan tersebut ditandai dengan adanya pergeseran pola penyakit dengan pola dari kematian dalam masyarakat, yaitu berkurangnya angka peristiwa penyakit yang menular tertentu dan bertambahnya jenis penyakit tidak menular (Gusti dkk , 2015). Kanker merupakan salah satu jenis penyakit tidak menular yang angka kejadiannya memiliki kecenderungan meningkat pada setiap tahunnya. Data Riset Kesehatan Dasar (Riskesdas) pada tahun 2007 menyebutkan bahwa kanker menempati urutan ke tujuh sebagai penyebab kematian terbesar di Indonesia setelah stroke, TB, hipertensi, cedera, perinatal dan DM. Perkiraan terdapat 100 penderita kanker baru setiap 100.000 penduduk per tahunnya (Muhammad dkk , 2018).

Kanker payudara adalah tumor ganas yang berasal dari sel-sel yang berada pada payudara, penyakit ini berupa neoplasma ganas yang berasal dari parenkim, pertumbuhan jaringan payudara abnormal yang tidak mengikut jaringan disekitarnya tumbuh infiltratif dan destruktif, serta dapat bermetastase (Desiyani , 2018). Kanker payudara merupakan salah satu jenis kanker terbanyak di Indonesia. Kanker payudara dapat terjadi pada pria maupun wanita, hanya saja prevalensi pada wanita jauh lebih tinggi. Menurut Organisasi Kesehatan Dunia

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

(WHO), 8-9% wanita akan mengalami kanker kanker payudara. Hal ini membuat kanker payudara menjadi jenis kanker yang paling banyak wanita di seluruh dunia (Muhammad dkk , 2018).

Payudara merupakan aset perempuan yang sangat berharga. Kelainan pada organ ini pastilah merupakan mimpi buruk bagi perempuan. Percaya diri lenyap dan tidak jarang mempengaruhi hubungan dengan pasangan. Jika seorang wanita menemukan benjolan di payudaranya, pertama-tama tentu akan timbul perasaan khawatir dan selanjutnya disikapi dengan berbeda-beda. Setiap benjolan di payudara tentu menimbulkan banyak kekhawatiran, diantaranya kemungkinan adanya kanker, operasi, efek samping radiasi dan kemoterapi, sampai kematian (Muhammad dkk , 2018).

Data Riset Kesehatan Dasar (Riskesdas) pada tahun 2007 menyebutkan bahwa kanker payudara menempati urutan pertama pada pasien rawat inap di seluruh rumah sakit di Indonesia (16,85%), disusul kanker leher rahim (11,78%), kanker hati dan saluran empedu intrahepatik (9,69%), leukimia (7,42%) dan Limfoma non hodgkin (6,69%) (Cici dkk , 2017). Angka nasional kanker adalah 4,3% per 1000 penduduk dengan angka kejadian yang lebih tinggi perempuan dari pada laki-laki, yaitu sebesar 5,7% per 1000 penduduk pada perempuan dan 2,9% per 1000 penduduk pada laki-laki.

Sampai sekarang, penyebab kanker payudara belum diketahui secara pasti, tetapi ada beberapa faktor resiko yang diduga menyebabkan seorang wanita menjadi lebih mungkin menderita kanker payudara yaitu usia, pernah menderita kanker payudara, status pernikahan, riwayat keluarga yang menderita kanker payudara, faktor genetika dan hormonal, pernah menderita penyakit payudara non-kanker, jenis pengobatan, ukuran tumor, dan stadium klinik. Hampir semua kanker payudara dapat diketahui dengan mengidentifikasi faktor yang ada, dengan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

mengetahui faktor tersebut akan memudahkan untuk mengidentifikasi apakah wanita tersebut tergolong risiko tinggi atau tidak, sehingga kanker payudara dapat dideteksi secara dini (Melda , 2016).

Rekurensi pasien kanker payudara adalah pertumbuhan kembali neoplasma ganas ke payudara setelah dibedah atau di operasi. Tanda awal rekurensi yaitu berupa terjadinya perubahan pada jaringan parenkim seperti terdapat perubahan kulit yang menebal di sekitar payudara, keluar cairan berwarna dari puting. Terdapat 2 faktor yang mempengaruhi terjadinya rekurensi yaitu faktor pembedahan dan faktor lingkungan. Faktor lingkungan yang berkaitan dengan rekurensi antara lain umur, jenis kelamin (wanita lebih rentan), belum pernah hamil dll (Emy , 2018).

Teknik bedah atau penanganan penyakit ini bervariasi, masing-masing teknik bedah memiliki kelebihan dan kekurangan. teknik bedah idealnya simpel, cepat, tingkat komplikasi rendah dan dapat diterima. Teknik operasi memiliki hubungan yang relatif tinggi dengan kejadian rekurensi. 50% rekurensi terjadi pada hari ke 120 setelah operasi. 97% rekurensi terjadi selama 1 tahun pasca bedah (Trisnadewi , 2018).

Pasien rekuren sering didapatkan, telah dilaporkan rata-rata mencapai 24% - 89% tergantung pada teknik operasi yang digunakan untuk operasi kanker payudara. Rekurensi paska bedah kanker payudara di Indonesia cukup besar dan bervariasi sekitar 35% - 52%. Penelitian di RS. dr. Soetomo Surabaya kejadian rekurensi sebesar 52% dalam studi lain di RSCM Jakarta ditemukan rekurensi sebesar 61,1% dengan prevalensi tertinggi pada penderita dibawah umur 50 tahun (Rokhana dkk , 2017).

Tingginya kasus rekurensi kanker payudara disebabkan karena penderita kanker payudara sering tidak menyadari atau merasakan dengan jelas gejala-gejala

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

kanker payudara dan kurangnya kesadaran untuk melakukan pencegahan sehingga sebagian besar pasien terdeteksi kanker ketika telah memasuki stadium lanjut. Apabila kanker payudara terdeteksi pada stadium lanjut mengakibatkan pengobatan menjadi lebih mahal, lebih sulit dan hasil pengobatannya tidak maksimal bahkan cenderung mempercepat kematian (Rifqy dkk , 2016).

Pengklasifikasian merupakan salah satu metode statistika dengan cara mengelompokkan atau mengklasifikasikan suatu data yang disusun secara matematis. Masalah klasifikasi sering dijumpai pada kehidupan nyata, salah satu masalah rekurensi pada pasien kanker payudara. Ada penyelesaian masalah klasifikasi yang perlu diperhatikan dalam memilih metode klasifikasi yang tepat. Terdapat berbagai macam metode klasifikasi, yaitu CART *(Classification And Regression Trees)*, CHAID *(Chi-Square Automatic Interaction Detection)*, *Random Forest*, SVM *(Support Vector Machine)* dan metode klasifikasi lainnya (Tutus dkk , 2017). Analisis yang dapat digunakan dalam metode klasifikasi ini adalah metode *Random Forest*, karena metode ini bisa mengatasi data *imbalanced* atau kosong, meningkatkan klasifikasi yg akurat sehingga metode tersebut bisa menghasilkan *tree* yang banyak (Trisnadewi , 2018).

Random Forest merupakan salah satu algoritma pohon keputusan dari klasifikasi dengan tingkat akurasi yang baik (Devi , 2015). *Random Forest* merupakan sebuah metode *ensemble* yang terdiri dari beberapa pohon keputusan sebagai *classifier*. Kelas yang dihasilkan dari proses klasifikasi ini diambil dari kelas terbanyak yang dihasilkan oleh pohon-pohon keputusan yang ada pada *random forest*. Dengan melakukan voting pada pohon-pohon keputusan yang tersedia membuat akurasi dari *Random Forest* meningkat (Breiman dkk , 1993).

Ada beberapa penelitian yang sudah melakukan penelitian menggunakan metode *Random Forest* antara lain Nidhomiddin dan Otok (2015) yang meneliti

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Klasifikasi Penderita HIV/AIDS Menggunakan Metode *Random Forest* dan *Multivariate Adaptive Regression Spline* (MARS) *binary response*. Hasil yang diperoleh adalah mengetahui variabel yang paling berpengaruh untuk menentukan status HIV/AIDS yaitu usia, jenis pekerjaan, pernah ditahan karena kasus NAPZA, status nikah dan selalu pakai jarum steril. Akurasi steril didapatkan ketika menggunakan metode *Random Forest* diperoleh akurasi terbaik yaitu 97,80%.

Penelitian sejenis juga yang dilakukan oleh Nugroho dan Emiliyawati (2017) dalam penelitiannya mereka melakukan Klasifikasi Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode *Random Forest*. Diperoleh hasil bahwa variabel yang mempengaruhi tingkat Penerimaan konsumen terdiri dari harga pembelian, biaya perawatan, jumlah pintu, kapasitas penumpang, ukuran bagasi, taksiran keselamatan penumpang. Didapatkan akurasi sebesar 82,40%. Budi Adnyana (2016) melakukan penelitian tentang Prediksi Lama Studi Mahasiswa Dengan Metode *Random Forest* dengan memperoleh nilai *accuracy* sebesar 83,54%.

Berdasarkan hasil penelitian Yuanita, dkk (2017) yang meneliti Klasifikasi PT. Gudang Garam Menggunakan Metode *Random Forest* tahun 2010. Hasil penelitiannya mendapatkan nilai *accuracy* sebesar 94,30%. Tutus Praningki, dkk (2017) yang meneliti tentang Penggunaan Metode *Random Forest* dan SVM *(Support Vector Machine)* di R Studi Kasus Klasifikasi Kanker Payudara. Diperoleh nilai akurasi terbaik untuk *Random Forest* sebesar 85,04%, sedangkan nilai akurasi dari SVM *(Support Vector Machine)* sebesar 70,02%.

Melihat betapa pentingnya untuk mengidentifikasi penyakit kanker payudara membuat banyak peneliti yang melakukan penelitian tersebut. Oleh karena itu, penulis tertarik melakukan penelitian tentang klasifikasi pasien kanker payudara yang rekuren dan tidak rekuren dengan judul **"Klasifikasi Rekurensi**

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Pasien Kanker Payudara Menggunakan Metode *Random Forest*".

1.2. Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana karakteristik pohon klasifikasi rekurensi pasien kanker payudara dengan menggunakan *Random Forest*?
2. Bagaimana nilai akurasi dari proses klasifikasi metode *Random Forest*?

1.3. Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah, maka tujuan dalam penelitian ini adalah :

1. Untuk mengetahui karakteristik pohon klasifikasi rekurensi pasien kanker payudara dengan menggunakan *Random Forest*.
2. Untuk mengetahui nilai akurasi dari proses klasifikasi metode *Random Forest*.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Bagi Penulis

   Penelitian ini diharapkan dapat menambah pengetahuan dan wawasan lebih mengenai *Random Forest* yang dapat digunakan untuk klasifikasi rekurensi pada pasien kanker payudara dengan melakukan pengolahan.

2. Bagi pihak lain

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

a. Diharapkan dapat memberikan informasi dan menambah wawasan atau pengetahuan bagi pembacanya dan sebagai bahan referensi apabila mengadakan penelitian yang sama seperti dilaporan ini.

b. Diharapkan dapat mempermudah untuk mengklasifikasi pasien rekurensi pada penyakit kanker payudara sehingga pasien dapat segera ditangani sesuai dengan klasifikasi rekurensi.

1.5. Batasan Masalah

Berdasarkan masalah yang telah diuraikan, didapatkan beberapa batasan penelitian agar penelitian ini tidak melebar dan lebih terarah. Batasan-batasan tesebut sebagai berikut :

1. Data yang digunakan dalam penelitian ini diperoleh dari data rekam medis pasien kanker payudara yang berobat di rumah sakit Periode Januari 2017-Desember 2018.

2. Variabel dari data yang digunakan dalam penelitian ini adalah usia, jenis pengobatan yang dijalani oleh pasien, status pernikahan, riwayat kanker dalam keluarga, ukuran tumor, dan stadium klinik.

3. Data dibagi menjadi dua yaitu data *training* 70% dan data *testing* 30%.

1.6. Sistematika Penulisan

Adapun sistematika penyusunan yang dilakukan yaitu, sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, dan sistematika penyusunan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

penelitian.

BAB II : TINJAUAN PUSTAKA

Bab ini berisi penjelasan tentang teori yang digunakan dalam mendukung penyelesaian penelitian yaitu Penggunaan Metode *Random Forest* untuk klasifikasi rekurensi pasien kanker payudara.

BAB III : METODE PENELITIAN

Bab ini berisi tentang variabel penelitian, jenis data, pengumpulan data yang digunakan dalam penelitian ini, serta metode penyelesaian yang dilakukan dalam penelitian.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari penelitian mengenai klasifikasi rekurensi pasien kanker payudara dengan menggunakan metode *Random Forest* dari hasil software R.

BAB V : PENUTUP

Bab ini berisi tentang kesimpulan yaitu rangkuman keseluruhan isi yang sudah dibahas dan saran yaitu saran perluasan, pengembangan, pendalaman dan pengkajian ulang.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II

TINJAUAN PUSTAKA

2.1. Kanker Payudara

Kanker adalah penyakit umum, satu dari tiga orang bisa terdiagnosa salah satu jenis kanker semasa hidupnya. Kanker payudara adalah tumor ganas yang tumbuh di dalam jaringan payudara. Kanker payudara mempunyai andil terbesar dalam kematian wanita di Indonesia karena kanker-kanker maligna. Insidensi karsinoma payudara di kebanyakan negara berkembang 11,5% tiap tahun, sehingga mulai tahun 2006 kira-kira terdapat 20.000 penderita baru tiap tahun (Fourina dkk , 2016).

Setiap wanita Indonesia, selama hidupnya mendapat kanker payudara. Kematian karena karsinoma payudara berkat perbaikan diagnostik dan terapi, meskipun insidennya meningkat, tetap tidak berubah. Tetapi untuk wanita pada umur 35 sampai dengan 50 tahun kanker payudara merupakan penyebab kematian yang penting untuk diwaspadai. Terobosan terakhir dalam penelitian sekarang, risiko pada wanita yang penyakitnya meningkat dapat diidentifikasi dengan kemungkinan penyembuhan dengan pasti. Diagnostik dini dengan skrining mamografik membantu pengenalan penyakit ini pada stadium dini. Perkembangan dalam kemoterapi dan radioterapi, kebanyakan dalam kombinasi dengan pembedahan, merupakan terapi baru dengan kemungkinan penyembuhan yang lebih besar (Ida , 2019).

Gejala umum kanker payudara menurut Suryaningsih dan Sukaca (2009)

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

adalah adanya benjolan pada payudara yang dapat diraba dan biasanya semakin mengeras, tidak beraturan, serta terkadang menimbulkan nyeri. Gejala lain yang tampak, misalnya perubahan bentuk dan ukuran, kerutan pada kulit payudara sehingga tampak menyerupai kulit jeruk, adanya cairan tidak normal berupa nanah, darah, cairan encer, atau air susu pada ibu tidak hamil atau tidak sedang menyusui yang keluar dari puting susu. Gejala kanker payudara umumnya juga tampak dari adanya pembengkakan di salah satu payudara, tarikan pada puting susu atau puting susu terasa gatal, serta nyeri. Pada kanker payudara stadium lanjut, dapat menimbulkan nyeri tulang, pembengkakan lengan, ulserasi kulit, atau penurunan berat badan (Gusti dkk , 2015). Adapun jenis-jenis kanker payudara yaitu :

a. Jenis kanker payudara noninvasif atau in-situ (Tidak Ganas)

   Kanker yang terjadi pada kantung (tube) susu yaitu penghubung antara alveolus (kelenjar yang memproduksi susu) dan puting payudara. *Ductal Carcinoma In Situ* (DCIS) merupakan bentuk kanker payudara non-invasif yang paling umum terjadi sedangkan *Lobular Carcinoma In Situ* (LCIS) lebih jarang terjadi (Reza , 2016).

   1. *Ductal Carcinoma In Situ* (DCIS)

      Karsinoma duktal berasal dari sel-sel yang melapisi saluran yang menuju ke puting susu sekitar 90% kanker payudara merupakan karsinoma duktal. Kanker ini bisa terjadi sebelum maupun sesudah masa menopause. Kanker ini tampak sebagai bintik-bintik kecil dari endapan kalsium (mikrokalsifikasi). Kanker ini biasanya terbatas pada daerah tententu di payudara dan bisa diangkat secara keseluruhan melalui pembedahan. Sekitar 25 sampai dengan 35% penderita karsinoma duktal akan menderita kanker invasif (biasanya pada payudara yang sama).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. *Lobular Carcinoma In Situ* (LCIS)

   Karsinoma lobular dikenal juga dengan nama neoplasia lobular. Sel karsinoma ini sebenarnya tidak termasuk kanker, tetapi terlihat seperti sel kanker yang tumbuh di *lobulus* payudara (jaringan yang memproduksi susu).

b. Jenis kanker payudara invasif (Ganas)

   Kanker yang telah menyebar keluar bagian kantung susu dan menyerang jaringan sekitarnya bahkan dapat menyebabkan penyebaran (metastase) keanggota tubuh lainnya seperti kelenjar lempa dan lainnya melalui peredaran darah (Putu , 2016).

1. *Mucinous Carcinoma*

   Kanker payudara yang termasuk dalam kategori ini sekitar 3%. Tipe kanker ini merupakan bentuk lain dari *ductal cancer*.

2. *Medullary Carcinoma*

   Jenis kanker ini terjadi sekiitar 5% dari seluruh kejadian kanker payudara dan merupakan satu jenis kanker *invasif* yang membentuk satu batas yang tidak lazim antara jaringan tumor dan jaringan normal.

Gejala dan pertumbuhan kanker payudara tidak mudah dideteksi karena awal pertumbuhan sel kanker payudara tidak dapat diketahui dengan gejala umumnya baru diketahui setelah stadium kanker berkembang agak lanjut, karena tahap dini biasanya tidak menimbulkan keluhan. Penderita merasa sehat, tidak merasa nyeri, dan tidak mengganggu aktivitas. Gejala-gejala kanker payudara yang tidak disadari dan tidak dirasakan pada stadium dini menyebabkan banyak penderita yang berobat dalam kondisi kanker stadium lanjut. Hal tersebut akan mempersulit penyembuhan dan semakin kecil peluang untuk disembuhkan. Tanda yang mungkin muncul pada stadium dini adalah teraba benjolan kecil di payudara yang tidak terasa nyeri (Jackie , 2008).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Gejala yang dapat diamati atau dirasakan oleh orang yang terkena penyakit kanker payudara ini antara lain adanya semacam benjolan yang tumbuh pada payudara, yang lama kelamaan bisa menimbulkan rasa nyeridan mendenyut-denyut (Melda , 2016). Secara umum, tanda dan gejala kanker payudara yaitu :

1. Benjolan atau massa di ketiak.
2. Perubahan ukuran atau bentuk payudara.
3. Keluar cairan yang abnormal dari puting susu (biasanya berdarah atau berwarna kuning sampai hijau, mungkin juga bernanah).
4. Perubahan pada warna atau tekstur kulit pada payudara, puting susu maupun *areola* (daerah berwarna coklat tua di sekeliling puting susu) payudara tampak kemerahan.
5. Kulit di sekitar puting susu bersisik.
6. Puting susu tertarik ke dalam atau terasa gatal.
7. Nyeri payudara atau pembengkakan salah satu payudara.

Gejala kanker payudara pada pria sama seperti kanker payudara yang dialami wanita, mulanya hanya benjolan. Umumnya benjolan hanya dialami di satu payudara, dan bila diraba terasa keras. Bila stadium kanker sudah lanjut, ada perubahan pada puting dan daerah hitam di sekitar puting. Kulit putingnya bertambah merah mengerut, tertarik ke dalam ataupun puting mengeluarkan cairan (Fourina dkk , 2016).

Stadium penyakit kanker adalah suatu keadaan hasil penilaian dokter saat mendiagnosis suatu penyakit kanker yang diderita pasiennya, sudah sejauh manakah tingkat penyebaran kanker tersebut baik ke organ atau jaringan sekitar

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

atau penyebaran ketempat lain. Staging (Penentuan stadium kanker) penentuan stadium kanker penting sebagai panduan pengobatan, *follow-up* dan menentukan *prognosis* (Krisan , 2017).

Sistem *Tumor Node Metastasis* (TNM) merupakan singkatan dari "T"yaitu *Tumor Size* atau ukuran tumor, "N"yaitu *node* atau kelenjar getah bening regional dan "M"yaitu *metastasis* atau penyebaran jauh. Ketiga faktor T, N dan M dinilai baik secara klinis sebelum dilakukan operasi, juga sesudah operasi dan dilakukan pemeriksaan histopatologi.

2.2. Faktor-faktor yang mempengaruhi rekurensi pasien kanker payudara

Faktor-faktor yang mempengaruhi rekurensi pasien kanker payudara antara lain :

2.2.1. Usia

Usia merupakan faktor yang penting dalam menentukan prevalensi kanker payudara. *Worldwide Cancer* melaporkan di Inggris antara tahun 2009 dan 2011, sekitar 80% dari kasus kanker payudara didiagnosis pada perempuan berusia > 50 tahun, dan sekitar seperempat (24%) didiagnosis pada perempuan berusia 75 dan lebih. Kejadian kanker payudara meningkat seiring bertambahnya usia. Semakin tua usia seorang wanita, semakin tinggi resiko untuk menderita kanker payudara. Menurut Mulyani dan Rinawati (2013) kategori usia yang paling beresiko terkena kanker payudara yaitu pada usia 50-69 tahun. Resiko terkena kanker payudara akan lebih meningkat pada wanita yang mengalami menopause terlambat yaitu pada usia lebih dari 55 tahun (Rifqy dkk , 2016).

Kanker payudara merupakan penyakit dengan massa laten panjang yaitu sekitar 10-15 tahun. Maka lebih banyak ditemukan pada wanita yang berusia di

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

atas 50 tahun. Hal ini dikarenakan semakin bertambahnya usia seorang wanita semakin lemah pula sistem imunitas tubuh, sehingga semakin tua lebih mudah untuk terkena kanker payudara (Jalu , 2019).

2.2.2. Jenis Pengobatan yang Dijalani

Pengobatan kanker payudara tergantung tipe dan stadium yang dialami penderita. Pengobatan bertujuan untuk memusnahkan kanker atau membatasi perkembangan serta menghilangkan gejala-gejalanya. Macam-macam pengobatan kanker payudara, yaitu (Devi , 2015) :

a. Pembedahan atau Operasi

Ada 3 jenis pembedahan atau biasa disebut *mastectomy* yaitu :

1. *Radical mastectomy*, merupakan operasi pengangkatan sebagian dari payudara (*lumpectomy*) dan operasi ini selalu diikuti dengan pemberian radioterapi. *Lumpectomy* ini biasanya direkomendasikan pada pasien yang besar tumornya kurang dari 2 cm dan letaknya di sebelah payudara.
2. *Total mastectomy*, merupakan operasi pengangkatan seluruh payudara saja bukan kelenjar di ketiak.
3. *Modified radical mastectomy* merupakan operasi pengangkatan seluruh payudara, jaringan payudara di tulang dada, tulang selangka, dan tulang iga serta benjolan di sekitar ketiak.

b. Terapi Radiasi

Terapi radiasi ini dilakukan dengan sinar-x dengan intensitas tinggi untuk membunuh kanker yang tidak terangkat saat pembedahan. Terapi radiasi ini bertujuan untuk menyembuhkan atau mengecilkan kanker pada stadium

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dini. Radiasi dalam pengobatan kanker disebut *ionizing radiation*.

Bentuk lain dari radiasi, diantaranya adalah gelombang radio, gelombang *micro* atau gelombang cahaya yang disebut *non-ionizing*. Sedangkan jenis ini tidak mempunyai energi yang besar dan tidak bisa mengionize sel. Terapi radiasi biasanya diberikan setiap hari, lima hari dalam seminggu, selama 6-7 minggu berturut-turut tergantung ukuran, lokasi, jenis kanker, kesehatan penderita secara umum, dan pengobatan lain yang diberikan.

c. Kemoterapi

Kemoterapi adalah suatu proses pemberian obat-obatan anti kanker dapat secara oral (diminum) dan intravenous (diinfuskan). Kemoterapi umumnya diberikan pada siklus yang berlangsung tiga hingga empat minggu dan diberikan enam hingga delapan kali siklus. Tes darah diambil sebelum berlangsungnya setiap siklus kemoterapi guna memeriksa jumlah darah bagi kemungkinan anemia atau tanda-tanda menurunnya penolakan terhadap daya tahan tubuh, fungsi ginjal maupun fungsi hati.

Para pasien boleh jadi dianjurkan untuk menjalani pengobatan anti rasa mual atau steroid sebelum dimulainya kemoterapi. Obat-obat biasanya diberikan melalui infuse intravena (diteteskan melalui vena) yang berlangsung beberapa menit hingga beberapa jam. Mereka biasanya diberikan berdasarkan kejadian di hari itu (pasien tidak harus perlu semalam di rumah sakit). Efek dari kemoterapi adalah pasien mengalami mual dan muntah serta rambut rontok karena pengaruh obat-obatan yang diberikan pada saat kemoterapi.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.3. Status Pernikahan

Pada wanita yang menikah akan terjadi aktivitas reproduksi pada saat kehamilan atau laktasi hormon. Hormon yang berperan besar adalah estrogen dan progesteron. Wanita yang belum menikah resiko terkena kanker payudara lebih besar dibandingkan dengan wanita yang sudah menikah, hal ini dikarenakan wanita yang sudah menikah kemungkinan akan memiliki anak dan menyusui, sedangkan wanita yang tidak atau belum menikah tidak menyusui sehingga laktasi hormonnya mengumpul yang lama-kelamaan bisa menyebabkan terjadinya kanker payudara (Haslinda dkk , 2017).

2.2.4. Riwayat Kanker Payudara Dalam Keluarga

Jika ibu, saudara perempuan, adik, kakak memiliki kanker payudara (terutama sebelum umur 40 tahun), maka resiko terkena kanker payudara lebih tinggi. Resiko dapat berlipat ganda jika ada lebih dari satu anggota keluarga inti yang terkena kanker payudara dan semakin mudah ada anggota keluarga yang terkena kanker maka akan semakin besar penyakit tersebut besifat keturunan. Wanita dengan riwayat keluarga ada yang menderita kanker payudara pada ibu, saudara perempuan maupun adik atau kakak resikonya 2 hingga 3 kali lebih tinggi.

Adanya mutasi pada beberapa gen yang berperan penting dalam pembentukan kanker payudara, gen yang dimaksud adalah beberapa gen yang bersifat onkogen dan gen yang bersifat mensupresitumor. Gen-gen kanker payudara menunjukkan bahwa wanita mempunyai kesempatan 80% untuk terkena kanker payudara dan 50% kemungkinan mewariskan gen ini. Jika seorang wanita memiliki salah satu dari gen tersebut, maka resiko kemungkinan menderita kanker payudara sangat besar (Rini , 2018).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.5. Ukuran Tumor

Pasien kanker payudara pada tahap pertama ditemukan tumor dengan ukuran diameter lebih dari 2,9 cm diperkirakan sekitar 48%. Ukuran diameter tumor yang lebih besar yaitu sekitar 4 cm terjadi pada 28% pasien dan ukuran tumor yang kurang dari 1 cm hanya sekitar 5% terjadi pada pasien kanker payudara. Pada pengamatan yang telah dilakukan Helmuth, tingkat pasien rekurensi yang terjadi dengan masing-masing ukuran tumor adalah 40%, 20%, dan 18%. Hal tersebut dapat disimpulkan bahwa semakin kecil ukuran tumor pertama ketika didiagnosis, maka tingkat pasien rekurensi akan semakin sedikit.

2.2.6. Stadium Klinis

Stadium klinis dari kanker payudara merupakan indikator terbaik untuk menentukan tingkat penyebaran dan potensi metastasis penyakit ini. Angka pasien rekurensi pada penderita kanker payudara yang telah menjalani pengobatan yang sesuia mendekati :

Tabel 2.1 Rekurensi Pasien Berdasarkan Stadium Klinik

Rekurensi Pasien Kanker Payudara	Stadium
7%	0
36%	I
66%	II
88%	III
95%	IV

Pasien yang mengalami metastasis lama setelah didiagnosa awal atau yang mengalami metastasis ke tulang atau jaringan lunak memiliki tingkat penyebaran

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang lebih baik.

2.3. Klasifikasi

Klasifikasi adalah proses untuk mencari model atau fungsi yang menggambarkan dan membedakan kelas-kelas atau konsep data untuk mengklasifikasikan target class ke dalam kategori yang dipilih. Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia. Salah satu tahapannya yaitu data *training* dan data *testing*.

Data *training* digunakan oleh algoritma untuk membentuk sebuah model klasifikasi. Model ini merupakan representasi pengetahuan yang akan digunakan untuk mengukur sejauh mana kasifikasi berhasil melakukan prediksi dengan benar. Karena itu, data yang ada pada data *testing* seharusnya tidak boleh ada pada data *training* sehingga dapat diketahui apakah model klasifikasi dapat melakukan klasifikasi dengan baik.

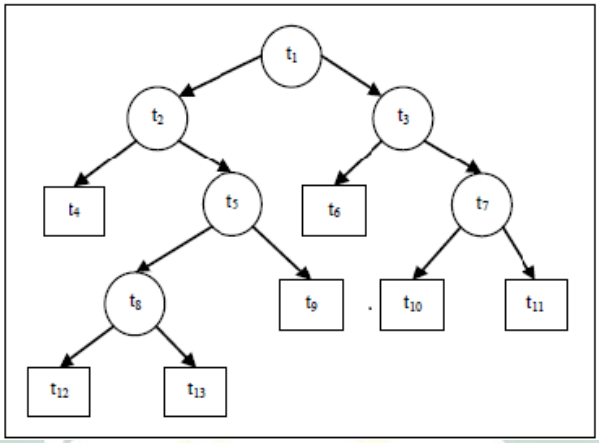
Proporsi antara data *training* dan data *testing* tidak mengikat tetapi agar variasi dalam model tidak telalu besar maka disarankan data *training* lebih besar dibandingkan data *testing*. Oleh karena itu, penelitian yang menghasilkan keakuratan model klasifikasi optimum dengan proporsi 70:30 untuk data *training* dan data *testing* (Untari , 2017).

Salah satu contohnya adalah *decision tree* yaitu sebuah *flowchart* yang menyerupai struktur pohon, dimana tiap node menunjukkan sebuah hasil dari test yang mempresentasikan kelas-kelas atau distribusi kelas (Gusti dkk , 2015). Dalam klasifikasi, terdapat target variabel kategori, misalnya pendapat tinggi, sedang dan rendah. Ada banyak metode untuk membangun klasifikasi seperti *naive-bayesian, Support Vector Machine* (SVM), *random forest, neighbor classification* dan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sebagainya.

2.4. *Decision Tree*



Gambar 2.1 Contoh *Classification (Decision Tree)*

Decision Tree atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplor data dan menemukan hubungan yang tersembunyi. Menurut Saputra (2014) *decison tree* merupakan metode yang paling efisien untuk menyaring sesuatu lewat pohon keputusan apakah suatu data lolos atau tidak terhadap saringan dengan proses yang cukup cepat dengan tahapan dalam membuat sebuah pohon keputusan sebagai berikut :

a. Menyiapkan data *training* yang sudah dikelompokkan ke dalam kelas tertentu.

b. Menentukan akar dan atribut, dimana akar akan diambil dari atribut yang terpilih yaitu dengan cara menghitung nilai gain dari masing-masing atribut. Nilai gain yang tertinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung nilai *entropy*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Untuk menghitung nilai *entropy* menggunakan rumus:

$$entropy(y) = -p_1 log_2 p_1 - p_2 log_2 p_2 \cdots - -p_n log_n p_n \tag{2.1}$$

dimana :

y = himpunan kasus

$-p_1, p_1, \ldots, p_n$ = proporsi kelas 1, kelas 2, . . . , dan kelas n dalam *output*.

2.5. *Random Forest*

Random Forest diusulkan oleh Tin Kam Ho pada tahun 1995 dengan penelitiannya yang berjudul *Random Decision Forest*, yang kemudian dikembangkan Leo Breiman dan kemudian dipatenkan pada tahun 2001 (Breiman dkk , 1993). *Random Forest* (RF) adalah suatu algoritma yang digunakan pada klasifikasi data jumlah yang besar. Klasifikasi *random forest* dilakukan melalui penggabungan pohon *(tree)* dengan melakukan *training* pada sampel data yang dimiliki.

Penggunaan pohon *(tree)* yang semakin banyak akan mempengaruhi akurasi yang akan didapatkan menjadi lebih baik. Penentuan klasifikasi dengan *random forest* diambil berdasarkan hasil voting dari *tree* yang terbentuk. Pemenang dari *tree* yang terbentuk ditentukan dengan vote terbanyak. Pembangunan pohon *(tree)* pada *random forest* sampai mencapai ukuran maksimum dari pohon data (Desiyani , 2018). Akan tetapi, pembangunan pohon *random forest* tidak dilakukan pemangkasan *(prunning)* yang merupakan sebuah metode untuk mengurangi kompleksitas ruang. Pembangunan dilakukan dengan penerapan metode *random forest feature selection* untuk meminimalisir kesalahan. Pembentukan pohon *(tree)* dengan sampel data menggunakan variabel yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

diambil secara acak dan menjalankan klasifikasi pada semua *tree* yang terbentuk (Agung dkk , 2017).

Random Forest adalah pengembangan dari metode *Classification and Regression Tree* (CART), yakni dengan menerapkan metode *bootstrap aggregating* (bagging) dan *random feature selection*. CART yaitu metode eksplorasi data yang didasarkan pada teknik pohon keputusan dimana pohon klasifikasi dihasilkan saat peubah respon berupa data kategorik sedangkan pohon regresi dihasilkan saat peubah respons berupa data numerik (Rokhana dkk , 2017). Pembangunan pohon klasifikasi CART meliputi tiga hal, yaitu :

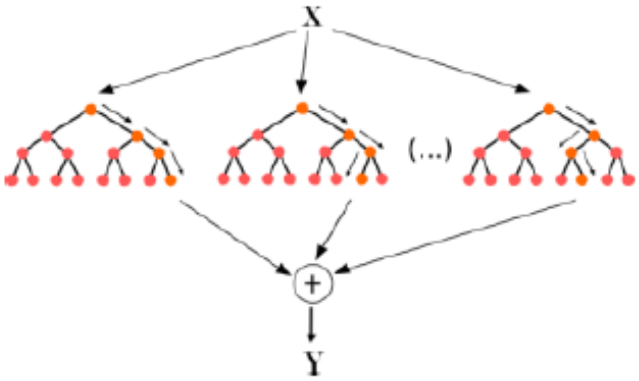
1. Pemilihan pemilah
2. Penentuan simpul terminal
3. Penandaan label kelas

Bootstrap aggregarting (Bagging) adalah teknik yang dapat digunakan untuk membentuk sampel *bootstrap* dimana setiap pohon keputusan dibangun dengan menggunakan sampel *bootstrap* dari data kandidat atribut untuk dibagi di setiap node yang berasal dari himpunan atribut acak dari hasil data yang dipilih dan disimpan (Nidhomuddin dkk , 2015). Pada setiap kali pembentukan *tree*, peubah penjelas yang digunakan untuk melakukan pemisahan bukanlah seluruh peubah yang terlibat, namun hanya sebagian saja hasil pemilihan secara acak. Proses ini menghasilkan kumpulan *tree* tunggal dengan ukuran dan bentuk berbeda-beda. *Bagging* merupakan metode *ensemble* yang bertujuan meningkatkan akurasi pengklasifikasi dengan menghubungkan pengklasifikasi tunggal dan hasilnya lebih baik dari pada random sampling (Tutus dkk , 2017).

Random forest melalui proses pengacakan yang tidak hanya dilakukan pada

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

data sampel saja melainkan juga pada pengambilan variabel bebas sehingga pohon klasifikasi yang dibangkitkan akan memiliki ukuran dan bentuk yang berbeda-beda. *Random forest* adalah klasifikasi yang terdiri dari beberapa pohon keputusan yang dibangun dengan menggunakan vektor acak (Fourina dkk , 2016).



Gambar 2.2 Contoh *Random Forest*

Menurut Dewo (2011) pada saat pemilihan klasifikasi secara keseluruhan, pohon-pohon yang buruk akan membuat prediksi yang merupakan prediktor yang baik akan muncul sebagai jawaban, agar ukuran kepentingan semakin stabil disarankan untuk menggunakan banyak pohon ketika penelitian, untuk mempertimbangkan ukuran kepentingan dan dihadapkan pada variabel bebas yang banyak. Tedapat ukuran kepentingan pada random forest, yaitu *Mean Decrease Accuracy* (MDA) dan *Mean Decrease Gini* (MDG).

Pembentukan *tree* pada *Random forest* dilakukan dengan cara melakukan *training* pada sampel. Variabel yang digunakan untuk *split* diambil secara acak dan klasifikasi dijalankan setelah semua *tree* terbentuk. Penentuan klasifikasi pada *Random forest* ini diambil berdasarkan *vote* dari masing-masing *tree* dan *vote* terbanyak yang menjadi pemenang (Krisan , 2017). Pada pembentukan *Random forest* menggunakan nilai *Gini Index* untuk menentukan *split* yang akan dijadikan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

root/node.

$$Gini(t) = 1 - \sum [p(j|t)]^2 \tag{2.2}$$

Keterangan :

$p(j|t)$ = Frekuensi relatif kelas j pada *node* t

Langkah berikutnya menghitung nilai Gini $_{split}$.

$$Gini_{split} = \sum_{i=1}^{k} \frac{n_i}{n} Gini(t) \tag{2.3}$$

dimana :

n_i = jumlah *record* pada partisi i

n = jumlah *record* pada *node p*.

Adapun perhitungan *Gini Index* mempengaruhi dari setiap *node* teratas dan *node* pemisah. Apabila perhitungan *Gini Index* sudah memiliki hasil akhir = 0, maka perhitungan akan berhenti. Namun apabila perhitungan *Gini Index* masih memiliki hasil akhir berupa angka, maka perhitungan akan tetap berlanjut (Haslinda dkk , 2017).

Untuk mendapatkan hasil yang optimal, metode *random forest* harus menentukan m jumlah variabel penjelas *(feature)* yang diambil secara acak dan juga untuk menentukan berapa banyak k pohon yang akan dibentuk. Nilai k yang disarankan untuk diggunakan pada metode *bagging* yang telah di uji coba adalah $k = 50$ dimana nillai k tersebut memberikan hasil yang memuaskan untuk klasifikasi. Sedangkan apabila nilai k yang diberikan adalah ≥ 100, nilai tersebut cenderung menghasilkan tingkat klasifikasi yang rendah (Jalu , 2019).

Hasil respons dari suatu observasi diprediksi dengan cara menggabungkan *(Aggregating)* hasil prediksi k pohon. Dalam melakukan klasifikasi, hasil akhir dari suatu klasifikasi yang terbentuk dari beberapa pohon tersebut kemudian

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dipilih berdasarkan *Majority Vote* (suara terbanyak). Untuk menduga adanya kesalahan dalam klasifikasi *random forest*, dapat dilakukan dengan perolehan *error OOB*. Data OOB atau *Out of bag* adalah data yang tidak termuat ketika melakukan pembuatan *bootstrap* (Breiman dkk , 1993). Adapun perolehan error OOB dilakukan dengan cara sebagai berikut :

1. Melakukan prediksi terhadap setiap data OOB pada pohon yang terbentuk.
2. Secara rata-rata, setiap observasi dari kelompok data asli akan menjadi data OOB dari banyak pohon yang telah dihasilkan. Jika a adalah sebuah pengamat dari kelompok data asli, maka hasil prediksi pada *random forest* terhadap a adalah gabungan dari hasil prediksi setiap kali a menjadi data OOB.
3. Error OOB dihitung dari proporsi data yang telah menjadi klasifikasi hasil prediksi *random forest* dari seluruh observasi kelompok data asli.

Sedangkan menurut (Breiman dkk , 1993), terdapat 3 cara untuk mendapatkan nilai *mtry* untuk mengamati *error* OOB yaitu :

$$\textit{Mtry}\ ke\ 1 = \frac{1}{2} \mid \sqrt{p} \mid \tag{2.4}$$

$$\textit{Mtry}\ ke\ 2 = \mid \sqrt{p} \mid \tag{2.5}$$

$$\textit{Mtry}\ ke\ 3 = 2\times \mid \sqrt{p} \mid \tag{2.6}$$

dimana p adalah total variabel.

Untuk melakukan pengamatan lebih teliti terhadap error OOB ketika k kecil, dan kemudian memilih m yang menghasilkan error OOB paling kecil. Apabila *random forest* akan dilakukan dengan menggunakan *variable importance*, maka

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

untuk menggunakan banyak pohon, misalnya 1000 pohon atau lebih. Jika variabel penjelas yang dianalisis sangat banyak, nilai tersebut dapat menjadi lebih besar agar *variable importance* yang akan dihasilkan semakin stabil (Breiman dkk , 1993).

2.6. Ukuran Tingkat Kepentingan

Salah satu ukuran kepentingan variabel bebas yang dihasilkan oleh *random forest* adalah *Mean Decrease Accuracy* (MDA) dan *Mean Decrease Gini* (MDG). MDA digunakan untuk menampilkan besar tambahan observasi yang mengalami misklasifikasi jika satu persatu variabel bebas tidak diikutsertakan dalam pengujian. Menurut Sandri dan Zuccolotto (2006) *Mean Decrease Gini* (MDG) merupakan salah satu ukuran tingkat kepentingan peubah penjelas yang dihasilkan oleh metode *random forest*. Misalkan terdapat *p* peubah penjelas dengan h = 1,2,3,...,p. Maka MDG mengukur tingkat kepentingan peubah penjelas X_h dengan cara :

$$MDG_h = \frac{1}{k}\sum_t [d(h,t)I(h,t)] \tag{2.7}$$

dimana :

d(h,t) : besar penurunan indeks Gini untuk peubah penjelas X_h

I(h,t) : memilih simpul t

k : banyaknya pohon dalam random forest (ukuran random forest).

2.7. *Confussion Matrix*

Untuk mengukur kineja klasifikasi digunakan *confusion matrix* yang memberikan keputusan yang diperoleh dalam pelatihan dan pengujian. *Confussion Matrix* adalah sebuah tabel untuk mendeskripsikan performa dari sebuah model klasifikasi, digunakan untuk mengukur ketika melakukan analisis terhadap sebuah

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pengklasifikasian. *Confussion matrix* mengandung informasi tentang aktual dan prediksi klasifikasi yang dilakukan oleh sistem, hasil dari sebuah sistem sering dievaluasi dengan menggunakan pengujian *confussion matrix* (Breiman dkk , 1993).

Menurut Wang Yao (2013),*confussion matrix* kelas bisa menggunakan matriks *recall, Precision* dan *Accuracy*. Tiga variabel pengukur tersebut merupakan evaluasi prediktor yang lebih komprehensif dalam konteks ketidakseimbangan dan dapat mengevaluasi kinerja pada saat data yang tidak seimbang. Misalkan terdapat *confussion matrix* 2×2 seperti matriks dibawah ini :

Tabel 2.2 Evaluasi Klasifikasi pada Akurasi

Kelas	Kelas hasil klasifikasi	
	Yes	No
Yes	TP	FP
No	FN	TN
Total	P'	N'

Keterangan :

a. TN *(true negative)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi negatif.

b. FP *(false positive)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi positif.

c. FN *(false negative)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi negatif.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

d. TP *(true positive)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi positif.

1. *Specificity*

   Specificity merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$Specificity = \frac{TN}{TN + FP} \tag{2.8}$$

   dimana :

   TN *(true negative)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi negatif.

   FP *(false positive)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi positif.

2. *Recall*

   Recall merupakan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Rumus dari *recall* adalah sebagai berikut :

$$Recall = \frac{TP}{TP + FN} \tag{2.9}$$

   dimana :

   TP *(true positive)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi positif.

   FN *(false negative)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi negatif.

3. *Precision*

   Precision merupakan jumlah data kategori positif yang diklasifikasi

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

secara benar dibagi dengan total data yang di klasifikasi positif. Rumus dari *precision* adalah sebagai berikut :

$$Precision = \frac{TP}{TP + FP} \tag{2.10}$$

dimana :

TP *(true positive)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi positif.

FP *(false positive)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi positif.

4. *Accuracy*

Accuracy smerupakan seberapa akurat sistem dapat mengklasifikasi data secara benar. Maka rumus yang digunakan adalah sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{2.11}$$

dimana :

TN *(true negative)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi negatif.

FP *(false positive)* adalah jumlah data dengan nilai sebenarnya negatif dan hasil prediksi positif.

FN *(false negative)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi negatif.

TP *(true positive)* adalah jumlah data dengan nilai sebenarnya positif dan hasil prediksi positif.

Dapat didefinisikan seperti rumus berikut :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$Accuracy = \frac{\sum data\ yang\ diklasifikasikan\ dengan\ tepat}{\sum data\ testing} \times 100\%$$

2.8. Integrasi dalam Al-qur'an

Agama Islam sangat menganjurkan bagi setiap umatnya untuk terus berusaha dan bersabar dalam menghadapi cobaan yang diberikan Allah SWT yang berupa suatu penyakit. Di dalam Al-qur'an telah dijelaskan bahwa dibalik terkena penyakit pasti ada hikmahnya, seperti yang tertera pada sebuah hadits riwayat Imam Muslim yang menuturkan sabda Rasulullah *shallallahu 'alaihi wa sallam* yang berbunyi :

مَا مِنْ مُسْلِمٍ يُصِيبُهُ أَذًى مِنْ مَرَضٍ، فَمَا سِوَاهُ إِلَّا حَطَّ اللهُ بِهِ سَيِّئَاتِهِ، كَمَا تَحُطُّ الشَّجَرَةُ وَرَقَهَا

Artinya : "Tidaklah seorang muslim terkena suatu penyakit dan lainnya kecuali karena Allah menggugurkan kejelekan-kejelekan sebagaimana sebuah pohon menggugurkan daunnya."

Imam Nawawi memberikan penjelaskan bahwa didalam hadits ada pelajaran bahwa kesalahan-kesalahan akan dilebur dengan berbagai penyakit didunia meskipun hanya sedikit kesusahannya. Imam Muslim juga meriwayatkan dari 'Atho bin Abi Robah yang mengatakan :

امْرَأَةٌ مِنْ أَهْلِ الْجَنَّةِ؟ قُلْتُ: بَلَى، قَالَ: هَذِهِ الْمَرْأَةُ السَّوْدَاءُ، أَتَتِ النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَتْ: إِنِّي أُصْرَعُ وَإِنِّي أَتَكَشَّفُ، فَادْعُ اللهَ لِي، قَالَ: وَإِنْ شِئْتِ دَعَوْتُ اللهَ أَنْ يُعَافِيَكِ قَالَتْ: أَصْبِرُ، قَالَتْ: فَإِنِّي أَتَكَشَّفُ فَادْعُ اللهَ أَنْ لَا أَتَكَشَّفَ فَدَعَا لَهَا

Artinya : "Ibnu Abas berkata kepadaku, 'Mau kah kau kuperlihatkan seorang perempuan ahli surga?' Aku menjawab, 'Ya.' Ia berkata, "Perempuan hitam

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ini telah datang kepada Nabi dan berkata, ‘Sesungguhnya aku mengidap penyakit ayan dan auratku sering tersingkap karenanya. Maka berdo’alah kepada Allah untuk kesembuhanku”Nabi bersabda, “Kalau kau mau bersabar bagimu surga. Dan bila kau mau aku mau mendo’akanmu agar Allah menyembuhkanmu.”Perempuan itu berkata “Aku mau bersabar saja.”Ia berkata lagi, “Auratku sering terungkap, maka mohonlah kepada Allah agar auratku tak terungkap lagi.”Maka Nabi mendo’akannya.”Oleh karena itu, sebagai manusia yang bisa dilakukan dengan bersabar, berdo’a dan berusaha menjalani pengobatan untuk kesembuhan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian yang dilakukan merupakan jenis penelitian kuantitatif yaitu dengan mengambil data yang diperlukan yang kemudian dianalisis. Penelitian menggunakan pendekatan kuantitatif menekankan analisisnya pada data numeri (angka) yang diolah.

3.2. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut merupakan data rekam medis rekurensi pasien kanker payudara yang rumah sakit periode Januari 2017 - Desember 2018. Variabel dependen yang digunakan adalah status rekurensi kanker payudara. Rekurensi yang dimaksud adalah kembalinya pasien ke rumah sakit karena penyakit yang sama, terdiri dari :

Y = 0, jika tidak rekuren

Y = 1, jika rekuren

Sedangkan variabel independen yang digunakan sebagai berikut :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 3.1 Variabel Independen

Variabel	Nama Variabel	Kategori	Skala
X_1	Usia Pasien	-	Rasio
X_2	Jenis Pengobatan yang dijalani Pasien	0 = Operasi 1 = Terapi radiasi 2 = Kemoterapi	Nominal
X_3	Status Pernikahan	0 = Belum menikah 1 = Menikah	Nominal
X_4	Riwayat Kanker dalam Keluarga	0 = Tidak ada riwayat 1 = Ada riwayat	Nominal
X_5	Ukuran Tumor	0 = $\leq 5cm$ 1 = $> 5cm$	Nominal
X_6	Stadium klinik	0 = stadium 0 1 = stadium I 2 = stadium II 3 = stadium III 4 = stadium IV	Ordinal

3.3. Teknik Analisis Data

Teknik analisis data merupakan langkah-langkah penyelesaian permasalahan hingga selesai. Teknik analisis data dalam penelitian ini sebagai berikut :

1. Langkah awal yang dilakukan adalah mencari dan membaca referensi berbentuk paper yang berkaitan dengan rekurensi pasien kanker payudara

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

menggunakan metode *random forest* di internet maupun di perpustakaan.

2. Selanjutnya adalah pengumpulan data. Data yang dikumpulkan diperoleh dari bagian rekam medis periode Januari 2017 - Desember 2018 yang meliputi usia, jenis pengobatan yang dijalani, status pernikahan, riwayat kanker dalam keluarga, ukuran tumor dan stadium klinik. Kemudian dimasukkan atau melakukan input data ke *Microsoft excel*.

3. Data yang sudah dimasukkan ke *Microsoft excel* akan diolah untuk mengetahui mana yang menyebabkan rekurensi pasien menggunakan metode *random forest* dengan bantuan aplikasi R untuk mendapatkan model atau *tree* yang baik.

4. Input data atau salin data dari Micrsoft Excel ke aplikasi R, lalu masuk ke proses pembagian data. Data dibagi menjadi 2 bagian yaitu data *training* dan data *testing*.

5. Data *training* yaitu data latih yang digunakan, pada penelitian ini yang digunakan adalah 70% dari total data.

6. Data *testing* yaitu data yang digunakan untuk menguji tingkat akurasi metode klasifikasi, pada penelitian ini yang digunakan adalah 30%.

7. Setelah pembagian data, maka langkah selanjutnya menentukan berapa jumlah *tree* yang akan dibuat untuk kebutuhan *training*. Kemudian dilakukan *bagging* atau pengambilan data sampel secara *random* untuk membuat *tree* dengan memasukkan data tersebut kedalam perhitungan $Gini(t) = 1 - \sum[p(j|t)]^2$ untuk menemukan *node* teratas dan *node* pemisah dari masing-masing *tree* menggunakan rumus $Gini_{split} = \sum_{i=1}^{k} \frac{n_i}{n} Gini(t)$.

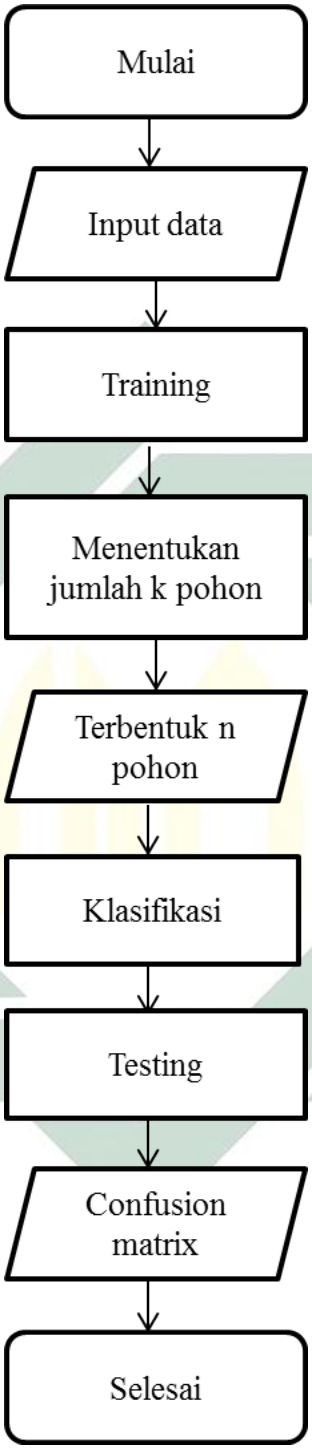
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Maka terbentuklah aturan keputusan yang merupakan hasil model yang akan dijadikan pada perhitungan sistem.

8. Kemudian menghitung tingkat *accuracy, specificity, precision,* dan *recall* menggunakan data *testing*, maka akan menghasilkan tabel *confusion matrix*.

9. Di tahap terakhir ini adalah menarik kesimpulan dari hasil yang sudah didapatkan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 3.1 Diagram Alur Penelitian

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

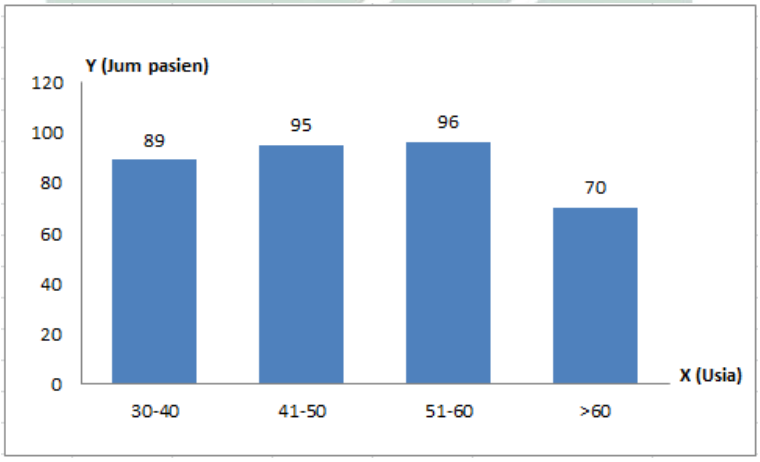
BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas hasil penelitian yang meliputi deskripsi data dan pembahasan hasil penelitian yang didapat dari hasil deskriptif faktor-faktor yang mempengaruhi rekurensi pasien kanker payudara yang terdiri atas usia pasien (X_1), jenis pengobatan yang dijalani pasien (X_2), status pernikahan (X_3), riwayat kanker dalam keluarga (X_4), ukuran tumor (X_5), dan stadium klinik (X_6).

4.1. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Usia Pasien

Terdapat empat kategori usia pasien yang berobat di Rumah Sakit F Surabaya, diantaranya yaitu pasien dengan usia 30-40 tahun, 41-50 tahun, 51-60, dan >61 tahun. Berikut adalah distribusi usia pasien pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.1**.



Gambar 4.1 Distribusi Usia Pasien Kanker Payudara Tahun 2017-2018

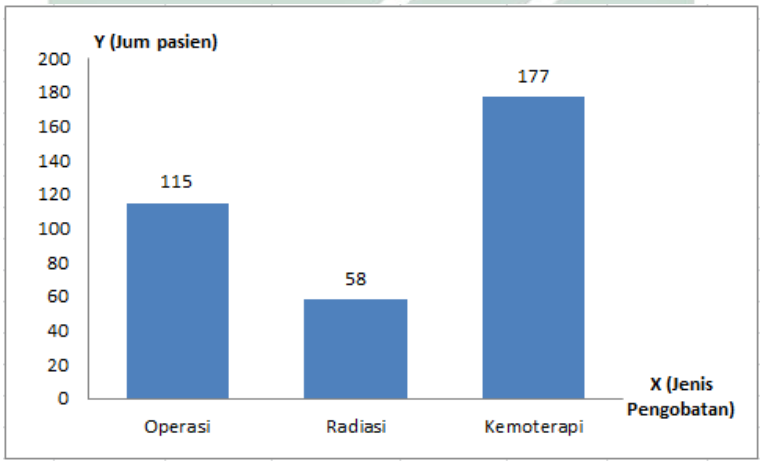
Gambar 4.1 menujukkan hasil distribusi usia pasien yang terdiagnosa

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

kanker payudara. Pada sumbu x menyatakan kelompok usia pasien ketika awal terjangkit dan sumbu y menyatakan banyaknya pasien. Berdasarkan **Gambar 4.1** peringkat pertama pasien yang terdiagnosa kanker payudara terjadi pada usia 51-60 tahun, yaitu sebesar 96 pasien. Selanjutnya yang kedua terjadi pada pasien yang berusia 41-50 tahun sebesar 95 pasien. Peringkat yang ketiga kanker payudara terjadi pada pasien dengan usia 30-40 tahun sebesar 89 pasien. Peringkat yang terakhir kanker payudara terjadi pada pasien dengan usia >60 tahun sebesar 70 pasien.

4.2. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Jenis Pengobatan Yang Dijalani

Terdapat tiga kategori jenis pengobatan yang dijalani oleh pasien kanker payudara yang berobat di Rumah Sakit F Surabaya, diantaranya yaitu pasien yang melakukan pengobatan dengan operasi atau pembedahan, terapi radiasi, dan kemoterapi. Berikut adalah distribusi jenis pengobatan pasien kanker payudara pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.2**.



Gambar 4.2 Distribusi Jenis Pengobatan Pasien Kanker Payudara Tahun 2017-2018

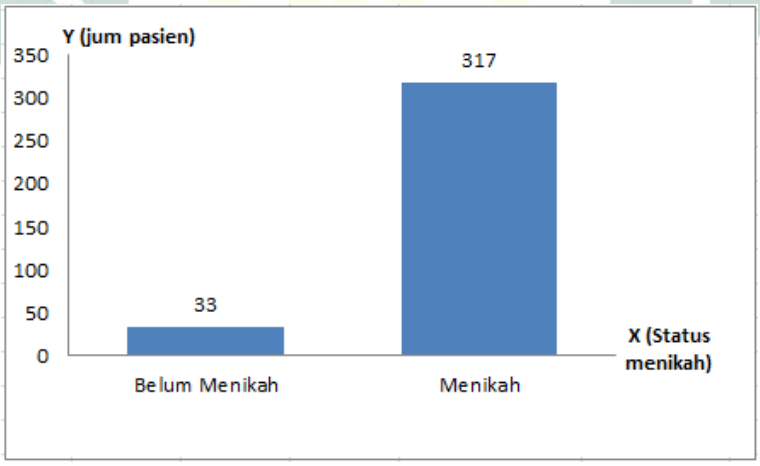
Gambar 4.2 menunjukkan hasil distribusi jenis pengobatan pasien kanker

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

payudara. Pada sumbu x menujukkan kelompok jenis pengobatan pasien dan sumbu y menyatakan jumlah pasien. Dari 350 pasien yang diteliti, pasien yang paling banyak melakukan kemoterapi yaitu sebesar 177 pasien. Selanjutnya urutan kedua adalah operasi yaitu sebesar 115 pasien. Selanjutnya yang terakhir adalah radiasi yaitu sebesar 58 pasien.

4.3. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Status Pernikahan

Terdapat dua kategori berdasarkan status pernikahan pada pasien kanker payudara yang berobat di Rumah Sakit F Surabaya, diantaranya yaitu pasien yang belum menikah dan menikah. Berikut adalah distribusi status penikahan pasien kanker payudara pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.3**



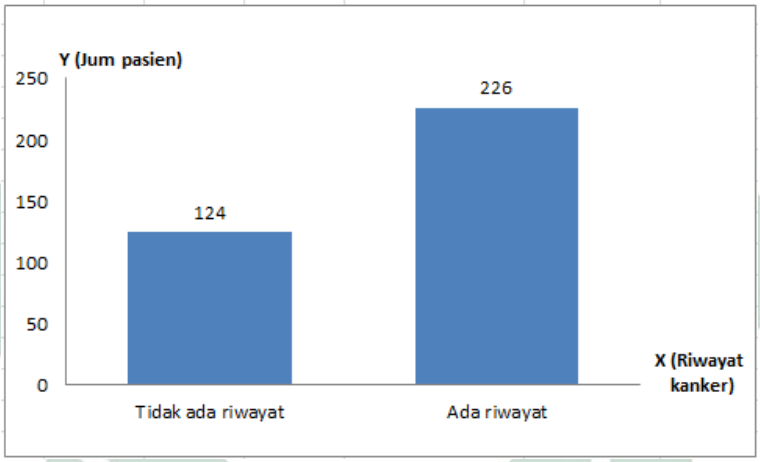
Gambar 4.3 Distribusi Status Pernikahan Pasien Kanker Payudara Tahun 2017-2018

Gambar 4.3 menunjukkan hasil distribusi status penikahan pasien kanker payudara. Pada sumbu x menyatakan kelompok status penikahan pasien dan sumbu y menyatakan banyaknya pasien. Peringkat pertama terjadi pada pasien yang sudah menikah yaitu sebesar 317 pasien. Dilanjutkan pasien kanker payudara yang belum menikah yaitu sebesar 33 pasien.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.4. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Riwayat Kanker Dalam Keluarga

Terdapat dua kategori riwayat kanker dalam keluarga pada pasien kanker payudara yang berobat di Rumah Sakit F Surabaya yaitu pasien yang tidak ada riwayat kanker dalam keluarga dan ada riwayat kanker dalam keluarga. Berikut adalah distribusi riwayat kanker dalam keluarga pasien kanker payudara pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.4**.



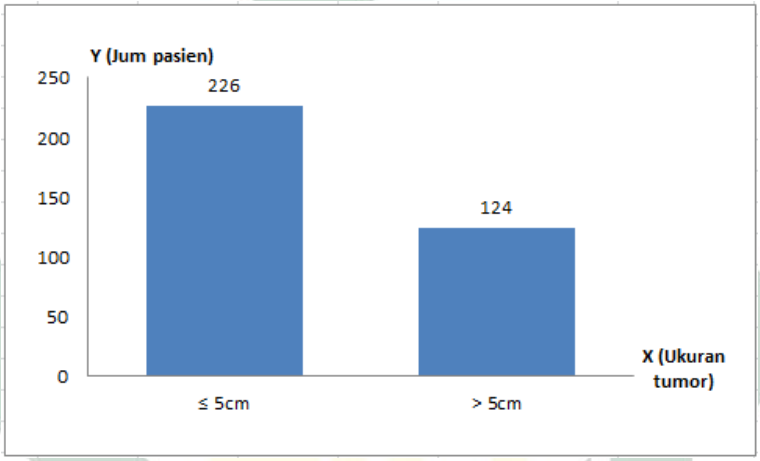
Gambar 4.4 Distribusi Riwayat Kanker Dalam Keluarga Pasien Kanker Payudara Tahun 2017-2018

Gambar 4.4 menunjukkan hasil distribusi riwayat kanker dalam keluarga pasien kanker payudara. Pada sumbu x menyatakan kelompok riwayat kanker dalam keluarga pasien dan sumbu y menyatakan jumlah pasien. Pasien yang terdiagnosa kanker payudara terbanyak adalah pasien yang mempunyai riwayat penyakit kanker dalam keluaga sebesar 226 pasien. Kemudian diikuti pasien yang terdiagnosa kanker payudara yang tidak mempunyai riwayat penyakit dalam keluarga sebesar 124 pasien.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.5. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Ukuran Tumor

Terdapat dua kategori ukuran tumor pada pasien kanker payudara yang berobat di Rumah Sakit F Surabaya, diantaranya adalah pasien dengan ukuran tumor $\leqslant$ 5 cm dan $>$ 5 cm berikut adalah distribusi ukuran tumor pasien kanker payudara pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.5**.



Gambar 4.5 Distribusi Ukuran Tumor Pasien Kanker Payudara Tahun 2017-2018

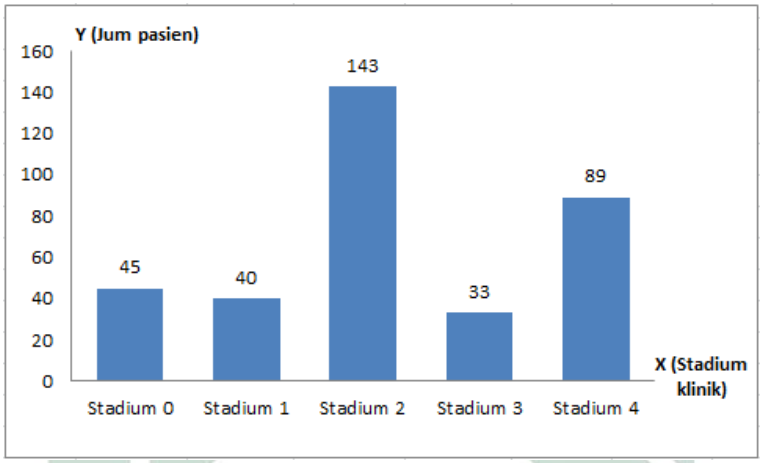
Gambar 4.5 menunjukkan hasil distribusi ukuran tumor pasien yang terdiagnosis kanker payudara. Pada sumbu x menyatakan kelompok ukuran tumor pasien dan sumbu y menyatakan jumlah pasien. Peringkat pertama pasien yang terdiagnosa kanker payudara memiliki ukuran tumor sebesar $\leqslant$ 5cm dengan 226 pasien. Peringkat kedua adalah pasien yang terdiagnosa kanker payudara memiliki ukuran tumor sebesar $>$ 5cm dengan 124 pasien.

4.6. Karakteristik Pasien Kanker Payudara Berdasarkan Faktor Stadium Klinik

Terdapat empat kategori stadium klinik pada pasien kanker payudara yang berobat di Rumah Sakit F Surabaya, diantaranya adalah pasien dengan stadium 0,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1, 2, 3, dan 4. Berikut adalah distribusi stadium klinik pasien kanker payudara pada tahun 2017-2018 yang diilustrasikan pada **Gambar 4.6**.



Gambar 4.6 Distribusi Stadium Klinik Pasien Kanker Payudara Tahun 2017-2018

Gambar 4.6 menunjukkan hasil distribusi stadium klinik pasien yang terdiagnosis kanker payudara. Pada sumbu x menyatakan kelompok ukuran stadium klinik dan sumbu y menyatakan jumlah pasien. Dari data 350 pasien yang diteliti, pasien yang memiliki stadium tertinggi yakni stadium 2 sebesar 143 pasien. Selanjutnya yang terendah adalah stadium 3 yaitu 33 pasien.

4.7. Hasil Analisis Menggunakan Metode *Random Forest*

4.7.1. Penentuan Data Latih *(Training)* dan Data Uji *(Testing)*

Data yang akan diolah untuk melakukan analisis klasifikasi terlebih dahulu dibagi menjadi 2 yaitu data *training* dan data *testing*. Data *training* digunakan untuk melatih program melakukan analisis klasifikasi dan data *testing* digunakan untuk menguji apakah hasil analisis dari data *training* sebelumnya memiliki tingkat akurasi yang besar atau tidak, dengan kata lain adalah untuk mengetahui seberapa besar tingkat ketepatan metode dalam melakukan klasifikasi data.

Pembagian data *training* dan data *testing* dapat dilihat dari **Tabel 4.1**.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.1 Data Latih (*Training*) dan Data Uji (*Testing*)

Keterangan	Data *Training*	Data *Testing*	Total
Jumlah	245	105	350
Persentase	70%	30%	100%

Berdasarkan **Tabel 4.1**, total data yang digunakan pada penelitian ini adalah 350, dengan komposisi 70% untuk data *training* atau latih yaitu sebesar 245 data, dan sisanya 105 untuk data *testing* atau uji coba. Pembagian data *training* dan *testing* tersebut secara *random* atau acak dengan bantuan software.

4.7.2. Penentuan *Mtry* (Peubah Penjelas) dan *Ntree* (Jumlah Pohon) Terbaik

Setelah menentukan data latih atau *training* dan data uji atau *testing*, langkah selanjutnya adalah menentukan nilai *Mtry* atau peubah penjelas dan *Ntree* atau jumlah pohon agar mendapatkan model yang optimal dan nilai *error Out Of Bag* (OOB) yang bernilai kecil. Penentuan *mtry* agar didapatkan nilai yang optimal digunakan dengan tiga cara, yaitu :

$$\begin{aligned} \textit{Mtry}\ ke\ 1 &= \frac{\sqrt{total\ variabel}}{2} \\ &= \frac{\sqrt{6}}{2} \\ &= 1,22 \approx 1 \end{aligned}$$

$$\begin{aligned} \textit{Mtry}\ ke\ 2 &= \sqrt{total\ variabel} \\ &= \sqrt{6} \\ &= 2,44 \approx 2 \end{aligned}$$

$$\begin{aligned} \textit{Mtry}\ ke\ 3 &= \sqrt{total\ variabel} \times 2 \\ &= \sqrt{6} \times 2 \end{aligned}$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$= 2,44 \times 2$

$= 4,88 \approx 5$

Berdasarkan perhitungan, diperoleh nilai masing-masing *mtry* yaitu 1, 2, dan 5. Selanjutnya masing-masing nilai *mtry* tersebut diujikan menggunakan default jumlah pohon (*ntree*) sebanyak 500, maka didapatkan nilai *error Out Of Bag* (OOB) sebagai berikut:

Tabel 4.2 Pengujian nilai *error* OOB

Mtry	*Error Out Of Bag* (OOB)
1	39,34%
2	29,50%
5	24,18%

Berdasarkan **Tabel 4.2**, diperoleh nilai *error Out Of Bag* (OOB) terendah yaitu 24,18% pada saat nilai *mtry* sebesar 5, sedangkan nilai *error Out Of Bag* (OOB) tertinggi pada saat nilai *mtry* sebesar 1. *Error* OOB terendah yang terpilih menjadi nilai *mtry* yang optimum. Breiman (2001) menyebutkan bahwa *error Out Of Bag* (OOB) bergantung pada korelasi antarpohon dan kekuatan masing-masing pohon dalam *random forest*. Peningkatan korelasi meningkatkan *error Out Of Bag* (OOB), sedangkan peningkatan kekuatan pohon menurunkan *error Out Of Bag* (OOB). Saat melakukan *random forest*, ukuran contoh peubah penjelas *mtry* yang digunakan sangat mempengaruhi korelasi dan kekuatan tersebut. Meningkatkan *mtry* akan meningkatkan keduanya, begitu juga sebaliknya. Penggunaan *mtry* yang tepatt akan menghasilkan *random forest* dengan korelasi antar pohon cukup kecil namun kekuatan setiap pohon cukup besar. Hal tersebut ditunjukkan dengan diperolehnya *random forest* dengan *error OOB* bernilai kecil (Breiman dkk ,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1993).

Selanjutnya mencari *ntree* optimum dengan memasukkan nilai *mtry* sebesar 5. Nilai *ntree* yang akan dicoba ada 5 yaitu 25, 50, 100, 500, dan 1000. Hasil dari nilai *error* OOB yang didapat adalah sebagai berikut:

Tabel 4.3 Pengujian nilai *error* OOB Dengan Nilai *ntree* Berbeda

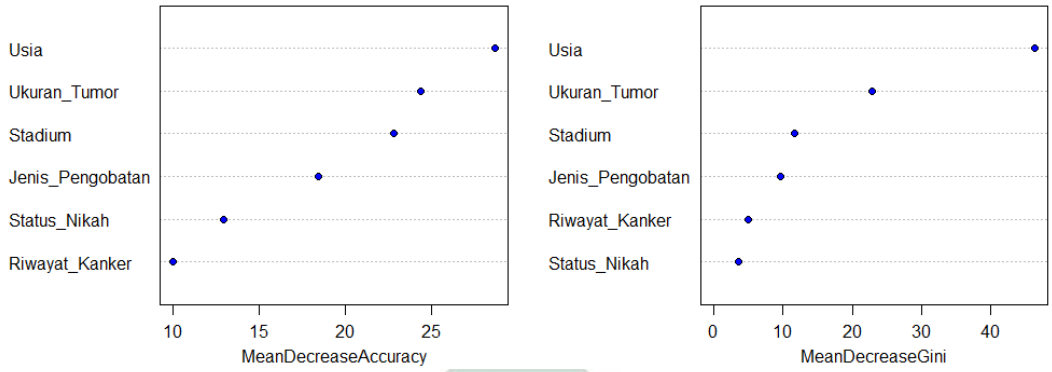
Ntree	*Error* OOB
25	41,90%
50	42,55%
100	53,30%
500	48,57%
1000	42,85%

Berdasarkan **Tabel 4.3**, diperoleh nilai *Ntree* optimum atau terbaik yaitu sebesar 25 dengan menghasilkan nilai *error* terkecil yaitu sebesar 41,90%. Setelah mendapatkan nilai *Mtry* dan *Ntree* optimum, maka nilai tersebut digunakan untuk menentukan prediksi *random forest* untuk *training* dan kemudian akan diuji tingkat akurasinya terhadap data *testing*.

4.7.3. Analisis Model

Selanjutnya dilakukan pengukuran tingkat kepentingan tiap variabel bebas yang diteliti.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.7 Ukuran Tingkat Kepentingan

Metode *random forest* menghasilkan ukuran tingkat kepentingan pada masing-masing variabel bebas yaitu *Mean Decrease Accuracy* (MDA) dan *Mean Decrease Gini* (MDG) seperti **Gambar 4.7**. Dimana *Mean Decrease Accuracy* (MDA) yang bertujuan untuk mengetahui seberapa besar observasi mengalami misklasifikasi jika salah satu variabel bebas tidak diikutsertakan pada pengujian, sedangkan *Mean Decrease Gini* (MDG) digunakan untuk mengetahui kestabilan variabel bebas dari X_1 sampai X_6. Tingkat kepentingan disusun berdasarkan peringkat.

Pada ukuran tingkat kepentingan *Mean Decrease Accuracy* (MDA), dua variabel yang tertinggi yaitu variabel X_1 (usia) dan variabel X_5 (ukuran tumor) yang berarti variabel tersebut memiliki peranan penting sebagai variabel bebas yang mempengaruhi variabel *dependent* (terikat) pada metode tersebut. Sedangkan untuk dua peringkat terakhir yaitu variabel X_3 (status pernikahan) dan variabel X_4 (riwayat kanker dalam keluarga) yang berarti bahwa variabel tersebut menandakan tingkat kepentingannya terhadap variabel rekuren dan tidak rekuren atau variabel *dependent* kecil.

Pada ukuran tingkat kepentingan *Mean Decrease Gini* (MDG), dimana variabel bebas yang memiliki tingkat kepentingan tertinggi adalah X_1 (usia),

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dilanjutkan X_5 (ukuran tumor) dan seterusnya, hingga variabel yang terakhir yaitu X_3 (status pernikahan). Jadi dapat disimpulkan baik dari MDA dan MDG, bahwa dua variabel terbawah dalam ukuran tingkat kepentingan adalah variabel X_4 (riwayat kanker dalam keluarga) dan variabel X_3 (status pernikahan).

Berikut sampel perhitungan untuk membentuk *decision tree* dengan melakukan penarikan data secara acak untuk n = 5.

Tabel 4.4 Data Pasien Kanker Payudara

No.	Status Pernikahan	Riwayat Kanker	Rekurensi
1.	belum menikah	ada riwayat	tidak rekuren
2.	menikah	tidak ada riwayat	rekuren
3.	menikah	ada riwayat	tidak rekuren
4.	menikah	ada riwayat	tidak rekuren
5.	belum menikah	tidak ada riwayat	rekuren

1. Rekurensi

$$P(tidak\ rekuren) = \frac{3}{5}$$

$$P(rekuren) = \frac{2}{5}$$

$$I_{GINI}(rekurensi) = 1 - \sum[p(tidak\ rekuren|rekuren)]^2$$

$$= 1 - [(\frac{3}{5})^2 + (\frac{2}{5})^2]$$

$$= 0{,}48$$

2. Status pernikahan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$\text{P}(belum\ menikah) = \frac{2}{5}$$

$$\text{P}(tidak\ rekuren|belum\ menikah) = \frac{1}{2}$$

$$\text{P}(rekuren|belum\ menikah) = \frac{1}{2}$$

$$\text{P}(menikah) = \frac{3}{5}$$

$$\text{P}(tidak\ rekuren|menikah) = \frac{2}{3}$$

$$\text{P}(rekuren|menikah) = \frac{1}{3}$$

$$I_{GINI}(belum\ menikah) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|belum\ menikah))^2 + (\text{P}(rekuren|menikah))^2]$$

$$= 1 - [(\tfrac{1}{2})^2 + (\tfrac{1}{2})^2]$$

$$= 0{,}5$$

$$I_{GINI}(menikah) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|menikah))^2 + (\text{P}(rekuren|menikah))^2]$$

$$= 1 - [(\tfrac{2}{3})^2 + (\tfrac{1}{3})^2]$$

$$= 0{,}44$$

$$GINIsplit(status\ pernikahan) = I_{GINI}(S) - \sum_{i=1}^{k} \frac{n_i}{n} GINI(t)$$

$$= I_{GINI}(rekuren) - \sum \text{P}(status\ pernikahan) * I_{GINI}(status\ pernikahan)$$

$$= 0{,}48 - [(\tfrac{2}{5} * 0,5) + (\tfrac{3}{5} * 0,44)]$$

$$= 0{,}48 - [0{,}464]$$

$$= 0{,}016$$

3. Riwayat kanker dalam keluarga

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$P(adariwayat) = \frac{3}{5}$$

$$P(tidak\ rekuren|ada\ riwayat) = 3$$

$$P(rekuren|ada\ riwayat) = 0$$

$$P(tidak\ ada\ riwayat) = \frac{2}{5}$$

$$P(tidak\ rekuren|tidak\ ada\ riwayat) = 0$$

$$P(rekuren|tidak\ ada\ riwayat) = 2$$

$$I_{GINI}(ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(P(tidak\ rekuren|ada\ riwayat))^2 + (P(rekuren|ada\ riwayat))^2]$$

$$= 1 - [(\frac{3}{3})^2 + 0]$$

$$= 0$$

$$I_{GINI}(tidak\ ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(P(tidak\ rekuren|tidak\ ada\ riwayat))^2 + (P(rekuren|tidak\ ada\ riwayat))^2]$$

$$= 1 - [(0 + (\frac{2}{3})^2]$$

$$= 0$$

$$GINIsplit(status\ pernikahan) = I_{GINI}(S) - \sum_{i=1}^{k} \frac{n_i}{n} GINI(t)$$

$$= I_{GINI}(rekuren) - \sum P(status\ pernikahan) * I_{GINI}(status\ pernikahan)$$

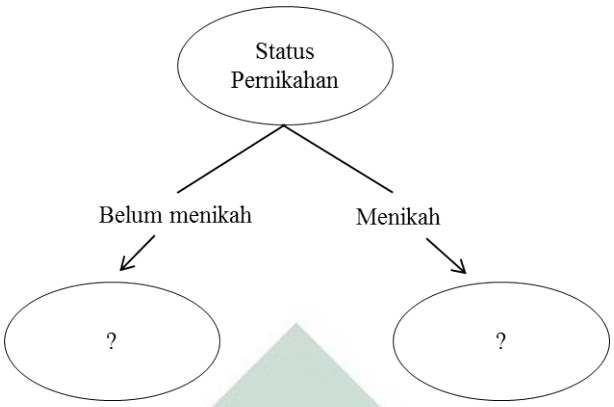
$$= 0{,}48 - [(\frac{3}{5} * 0) + (\frac{2}{5} * 0)]$$

$$= 0{,}48 - [0]$$

$$= 0{,}48$$

Dari variabel status pernikahan, riwayat kanker dalam keluarga, nilai $Gini_{split}$ yang paling kecil menjadi *root* untuk *tree* pertama. Akhirnya yang menjadi *root* pertama adalah status pernikahan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.8 *Tree* Ke-1 Dengan *Root* Pertama

Berikutnya mengambil barisan data di status pernikahan yang belum menikah :

Tabel 4.5 Data Pasien Kanker Payudara

No.	Status Pernikahan	Riwayat Kanker	Rekurensi
1.	belum menikah	ada riwayat	tidak rekuren
2.	belum menikah	tidak ada riwayat	rekuren

1. Rekurensi

$$P(tidak\ rekuren) = \frac{1}{2}$$

$$P(rekuren) = \frac{1}{2}$$

$$I_{GINI}(rekurensi) = 1 - \sum[p(tidak\ rekuren|rekuren)]^2$$

$$= 1 - [(\tfrac{1}{2})^2 + (\tfrac{1}{2})^2]$$

$$= 0{,}5$$

2. Status pernikahan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$\text{P}(belum\ menikah) = \frac{1}{2}$$

$$\text{P}(tidak\ rekuren|belum\ menikah) = 1$$

$$\text{P}(rekuren|belum\ menikah) = 0$$

$$\text{P}(menikah) = \frac{1}{2}$$

$$\text{P}(tidak\ rekuren|menikah) = 0$$

$$\text{P}(rekuren|belum\ menikah) = 1$$

$$I_{GINI}(ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|ada\ riwayat))^2 + (\text{P}(rekuren|ada\ riwayat))^2]$$

$$= 1 - [(\tfrac{1}{1})^2 + 0]$$

$$= 0$$

$$I_{GINI}(tidak\ ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|tidak\ ada\ riwayat))^2 + (\text{P}(rekuren|tidak\ ada\ riwayat))^2]$$

$$= 1 - [(0 + (\tfrac{1}{1})^2]$$

$$= 0$$

$$GINIsplit(status\ pernikahan) = I_{GINI}(S) - \sum_{i=1}^{k} \frac{n_i}{n} GINI(t)$$

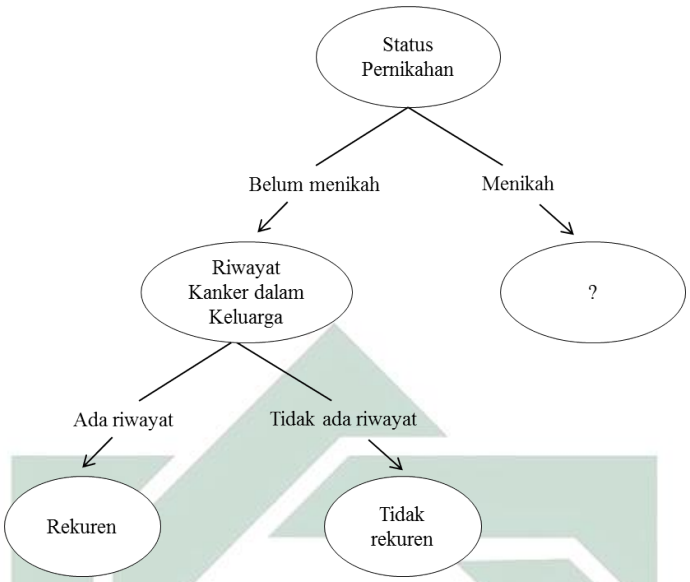
$$= I_{GINI}(rekuren) - \sum \text{P}(status\ pernikahan) * I_{GINI}(status\ pernikahan)$$

$$= 0{,}5 - [(\tfrac{1}{2} * 0) + (\tfrac{1}{2} * 0)]$$

$$= 0{,}5$$

Akibatnya diperoleh sebagai berikut :

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.9 *Tree* Dengan *Root* Kedua

Berikutnya mengambil barisan data di status pernikahan yang menikah :

Tabel 4.6 Data Pasien Kanker Payudara

No.	Status Pernikahan	Riwayat Kanker	Rekurensi
1.	menikah	tidak ada riwayat	rekuren
2.	menikah	ada riwayat	tidak rekuren
3.	menikah	ada riwayat	tidak rekuren

1. Rekurensi

P($tidak\ rekuren$) $= \frac{2}{3}$

P($rekuren$) $= \frac{1}{3}$

$I_{GINI}(rekurensi) = 1 - \sum[p(tidak\ rekuren|rekuren)]^2$

$= 1 - [(\frac{2}{3})^2 + (\frac{1}{3})^2]$

$= 0{,}44$

2. Riwayat kanker dalam keluarga

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$\text{P}(ada\ riwayat) = \frac{2}{3}$$

$$\text{P}(rekuren|ada\ riwayat) = 0$$

$$\text{P}(tidak\ rekuren|ada\ riwayat) = 2$$

$$\text{P}(tidak\ ada\ riwayat) = \frac{1}{3}$$

$$\text{P}(rekuren|tidak\ ada\ riwayat) = 1$$

$$\text{P}(tidak\ rekuren|tidak\ ada\ riwayat) = 0$$

$$I_{GINI}(ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|ada\ riwayat))^2 +$$

$$(\text{P}(rekuren|ada\ riwayat))^2]$$

$$= 1 - [(\tfrac{2}{2})^2 + 0]$$

$$= 0$$

$$I_{GINI}(tidak\ ada\ riwayat) = 1 - \sum[p(j|t)]^2$$

$$= 1 - [(\text{P}(tidak\ rekuren|tidak\ ada\ riwayat))^2 +$$

$$= (\text{P}(rekuren|tidak\ ada\ riwayat))^2]$$

$$= 1 - [(0 + (\tfrac{1}{1})^2]$$

$$= 0$$

$$GINIsplit(status\ pernikahan) = I_{GINI}(S) - \sum_{i=1}^{k} \frac{n_i}{n} GINI(t)$$

$$= I_{GINI}(rekurensi) - \sum P(status\ pernikahan) *$$

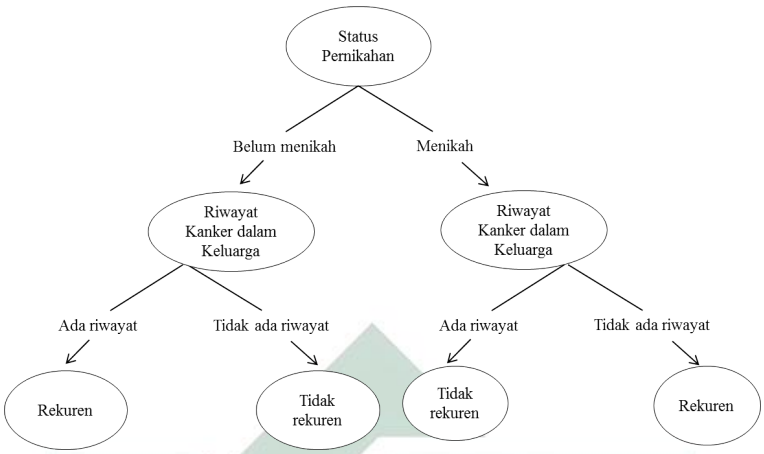
$$I_{GINI}(status\ pernikahan)$$

$$= 0{,}44 - [(\tfrac{2}{3} * 0) + (\tfrac{1}{3} * 0)]$$

$$= 0{,}44 - [0]$$

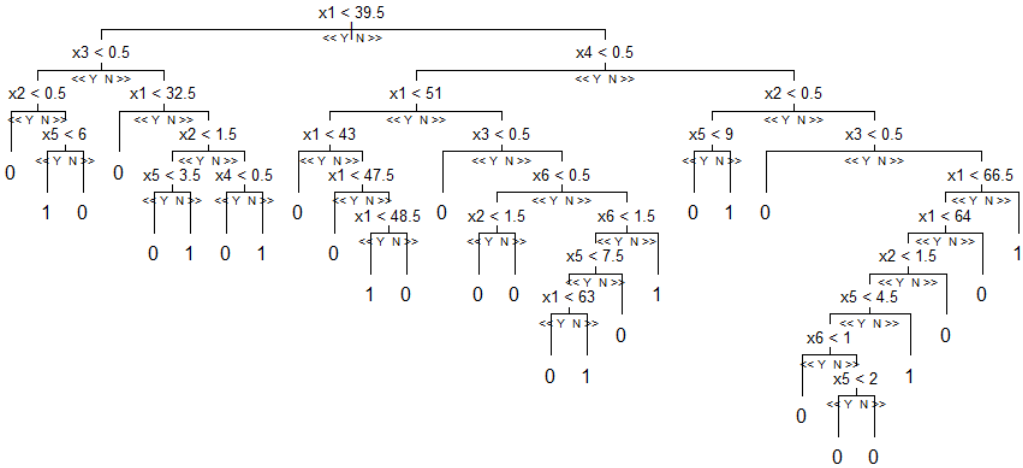
$$= 0{,}44$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.10 *Tree* Yang Terbentuk Berdasarkan Perhitungan Dengan Data Sampel n=5

Lebih lanjut, hasil model menggunakan metode *random forest* dari 350 data yang dibentuk dengan bantuan *software* adalah sebagai berikut :



Gambar 4.11 *Decision Tree* Klasifikasi Optimum

Decision Tree pada **Gambar 4.10**, menujukkan bahwa variabel usia (X_1) oleh pasien kanker payudara merupakan variabel pemilah utama dan paling menentukan klasifikasi pasien kanker payudara. Selanjutnya variabel yang juga menentukan klasifikasi pasien kanker payudara adalah status pernikahan (X_3), riwayat kanker dalam keluarga (X_4), jenis pengobatan (X_2), ukuran tumor jenis

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

pengobatan (X_5) dan stadium klinik (X_6). Karakteristik simpul yang sudah mengklasifikasi data menjadi salah satu dari kedua status rekurensi yaitu pada pasien kanker payudara yang diprediksi sebagai pasien yang rekuren adalah (X_1) usia kurang dari 39,5 tahun. Terjadi pada pasien yang belum menikah (X_3) status pernikahan, menjalani pengobatan (X_2) terapi radiasi, dan memiliki ukuran tumor (X_5) $< 6cm$.

Pada pasien kanker payudara yang diprediksi sebagai pasien yang tidak rekuren adalah, (X_1) usia berkisar 32,5 sampai 39,5 tahun. Terjadi pada pasien yang menjalani pengobatan (X_2) terapi radiasi, memiliki ukuran tumor (X_5) $< 3,5cm$, dan tidak memiliki riwayat kanker dalam keluarga (X_4).

Pada pasien kanker payudara yang diprediksi sebagai pasien yang rekuren adalah (X_1) usia 39,5 sampai 51 tahun dan tidak memiliki riwayat kanker dalam keluarga (X_4). Terjadi pada pasien yang belum menikah (X_3), menjalani pengobatan (X_2) terapi radiasi, memiliki (X_6) stadium 0 sampai 1, dan ukuran tumor (X_5) $< 7,5cm$.

Pada pasien kanker payudara yang diprediksi sebagai pasien yang rekuren adalah (X_1) usia kurang dari 39,5 tahun. Terjadi pada pasien yang tidak memiliki riwayat kanker dalam keluarga (X_4), menjalani pengobatan (X_2) operasi dan memiliki ukuran tumor (X_5) $< 9cm$.

Pada pasien kanker payudara yang diprediksi sebagai pasien yang tidak rekuren adalah X_1 (usia) berkisar 39,5 sampai 66,5 tahun. Terjadi pada pasien yang tidak memiliki riwayat kanker dalam keluarga (X_4), menjalani pengobatan (X_2) operasi, belum menikah (X_3), stadium klinik (X_1) yakni stadium 1 dan memiliki ukuran tumor (X_5) $< 2cm$ sampai $< 4,5cm$.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.7 Karakteristik Pasien Rekuren Kanker Payudara di Rumah Sakit F Surabaya

No.	Rekuren
1.	Pasien yang (X_1) berusia kurang dari 39,5 tahun, belum menikah (X_3) status pernikahan, menjalani pengobatan (X_2) terapi radiasi, dan memiliki ukuran tumor (X_5) $< 6cm$
2.	Pasien yang (X_1) usia 39,5 sampai 51 tahun, tidak memiliki riwayat kanker dalam keluarga (X_4), belum menikah (X_3), menjalani pengobatan (X_2) terapi radiasi, memiliki (X_6) stadium 0 sampai 1 dan ukuran tumor (X_5) $< 7,5cm$.
3.	Pasien yang (X_1) usia kurang dari 39,5 tahun, tidak memiliki riwayat kanker dalam keluarga (X_4), menjalani pengobatan (X_2) operasi dan memiliki ukuran tumor (X_5) $< 9cm$.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.8 Karakteristik Pasien Tidak Rekuren Kanker Payudara di Rumah Sakit F Surabaya

No.	Tidak Rekuren
1.	Pasien yang (X_1) usia berkisar 32,5 sampai 39,5 tahun, menjalani pengobatan (X_2) terapi radiasi, memiliki ukuran tumor (X_5) $< 3,5cm$ dan tidak memiliki riwayat kanker dalam keluarga (X_4).
2.	Pasien yang X_1 (usia) berkisar 39,5 sampai 66,5 tahun, tidak memiliki riwayat kanker dalam keluarga (X_4), menjalani pengobatan (X_2) operasi, belum menikah (X_3), stadium klinik (X_1) yakni stadium 1 dan memiliki ukuran tumor (X_5) $< 2cm$ sampai $4,5cm$

4.7.4. Pengujian Akurasi atau *Confussion Matrix*

Hasil pengujian akurasi menggunakan data uji (*testing*) adalah sebagai berikut :

Tabel 4.9 Hasil Prediksi Data Uji (*Testing*)

Observasi	Tidak *rekuren*	*Rekuren*	total
Tidak *rekuren*	52	40	92
Rekuren	42	36	78
total	94	76	170

Berdasarkan **Tabel 4.8**, diperoleh banyaknya pasien kanker payudara yang *rekuren* yaitu sebanyak 76, hasil prediksi benar untuk pasien kanker payudara yang *rekuren* adalah sebanyak 40 dan prediksi salah sebanyak 36. Untuk pasien kanker

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

payudara yang tidak *rekuren* dari prediksi data uji (*testing*) adalah 94, hasil prediksi benar untuk pasien kanker payudara yang tidak *rekuren* adalah sebanyak 42 dan prediksi salah adalah sebanyak 52. Berikut adalah hasil perhitungan *specificity, recall, precision* dan *accuracy*.

Specificity $= \frac{36}{78}$

$= 0{,}4615 \times 100\% = 46{,}15\%$

Recall $= \frac{52}{92}$

$= 0{,}5652 \times 100\% = 56{,}52\%$

Precision $= \frac{52}{94}$

$= 0{,}5531 \times 100\% = 55{,}31\%$

Akurasi $= \frac{\sum(prediksi\ benar)}{\sum(semua\ prediksi)}$

$= \frac{52+36}{170}$

$= 0{,}5176 \times 100\% = 51{,}76\%$

Berdasarkan hasil diatas, menunjukkan nilai *specificity* atau pengukuran proporsi true negative (pasien kanker payudara tidak rekuren) yang diidentifikasikan dengan benar yang memiliki nilai sebesar 46,15%. Sedangkan nilai *recall* atau pengukuran proporsi *true positive* (rekuren) yang diidentifikasi dengan benar sebesar 56,52%.

Hasil akurasi data *testing* sebesar 0,5176 atau 51,76% menunjukkan metode *random forest* untuk melakukan prediksi tepat terhadap data *testing* dan prediksi salah sebesar 48,24% atau bisa dikatakan dengan prediksi yang cukup. Hal tersebut dikarenakan *tree* yang terbentuk kurang banyak jadi semakin banyak *tree* maka semakin bagus akurasinya serta data yang di peroleh jangka waktunya kurang panjang, minimal 5 tahun (Gunn , 1998).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V

PENUTUP

Pada bab ini akan diberikan simpulan dan saran-saran yang dapat diambil berdasarkan materi-materi yang telah dibahas pada bab-bab sebelumnya.

5.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan diperoleh kesimpulan sebagai berikut :

1. Karakteristik pasien kanker payudara berdasarkan status rekurensi kanker payudara diperoleh bahwa pasien yang berusia 39,5 sampai 51 tahun, tidak memiliki riwayat kanker dalam keluarga, menjalani pengobatan terapi radiasi dan operasi, belum menikah, dan memiliki ukuran tumor $6cm$ sampai $9cm$. Pada pasien yang tidak rekuren adalah pasien yang berusia berkisar 32,5 sampai 66,5 tahun, tidak memiliki riwayat kanker dalam keluarga, belum menikah, menjalani pengobatan terapi radiasi dan operasi, memiliki stadium klinik 1 dan memiliki ukuran tumor $< 2cm$ sampai $4,5cm$.

2. Dengan menggunakan data latih *(training)* sebesar 70% dan data uji *(testing)* sebesar 30% maka mendapatkan nilai akurasi sebesar 51,76% atau bisa dikatakan cukup.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.2. Saran

Setelah membahas dan mengimplementasikan metode *random forest*, penulis ingin menyampaikan beberapa saran.

1. Pada penelitian selanjutnya diharapkan agar menggunakan metode pengembangan *Random Forest* lainnya, misalnya *Balanced Random Forest* untuk meningkatkan hasil klasifikasi yang lebih bagus.
2. Untuk meningkatkan akurasi dan kualitas hasil klasifikasi, sebaiknya pada penelitian selanjutnya menggunakan data yang lebih banyak lagi serta menambah variabel yang mendukung.
3. Sebaiknya menggunakan waktu penelitian lebih panjang minimum 5 tahun karena penyakit ini bukan penyakit bisa dsembuhkan dalam jangka pendek.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

Agung Waluyo, Moch. Abdul Mukid, Triastuti Wuryandari. 2017, *Perbandingan Klasifikasi Nasabah Kredit Menggunakan Regresi Logistik Biner dan Random Forest (RF)*. Media Statistika. Vol. 7, No. 2.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1993, *Classification and Regression Trees (CART)*. New York: Chapman And Hall.

Cici Priyatin, Elisa Ulfiana, Sri Sumarni. 2017, *Faktor Risiko Yang Berpengaruh Terhadap Kejadian Kanker Payudara Di RSUP DR. Kariadi Semarang*. Jurnal Kebidanan. Vol. 2 No. 5. Semarang.

Desiyani Nani. 2018, *Analisis Faktor-faktor Yang Berhubungan Dengan Kejadian Kanker Payudara Di Rumah Sakit Pertamina Cilacap*. Jurnal Keperawatan Soedirman (The Soedirman Journal of Nursing). Volume 4, No. 2. Purwokerto.

Devi Nur Oktaviana. 2015, *Faktor-faktor Risiko Kanker Payudara pada Pasien Kanker Payudara Wanita di Rumah Sakit Kanker Dharmais Jakarta*. Bogor : Institut Pertanian Bogor.

Emy Rianti. 2018, *Faktor Rekurensi Kanker Payudara Pada Pasien Wanita di Rumah Sakit Kanker Dharmais Jakarta*. Jurnal kanker (Journal of Cancer).

Fourina Ayu Novianti, Santi Wulan Purnami. 2016, *Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi*. Jurnal Sains Dan Seni ITS. Vol. 1, No. 1. Surabaya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Gunn Steve. 1998, *Classification and Regression* . Southampion: University of Sounthaton.

Gusti Ayu Triara Dewi, Lucia Yovita Hendrati. 1992, *Analisis Risiko Kanker Payudara Berdasarkan Riwayat Pemakaian Kontrasepsi Hormonal Dan Usia Menarche"*. Jurnal Berkala Epidemiologi. Volume 3, No. 1. Surabaya.

Haslinda, Erna Kadrianti, Suarnianti. 2017, *Faktor Risiko Kejadian Kanker Payudara di RSUP Magetan*. Jurnal Kesehatan.

Ida Rahmawati. 2019, *Metode CART, Random Forest (FR) dan CHAID Untuk Pengklasifikasian Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember*. Jurnal Matematika. Volume 4, No 7. Jember.

Iin Yulianti, Henry Setyawan, Dwi Sutiningsih. 2016, *Faktor-faktor Risiko Kanker Payudara (Studi Kasus Pada Rumah Sakit Ken Saras Semarang)*. Jurnal Kesehatan Masyarakat. Volume 4, Nomor 4. Diponegoro.

Jackie, Lincoln-wilensky. 2008, *Kanker Payudara Diagnosis dan Solusinya*. Jakarta: Prestasi Pustaka.

Jalu Nurcahyo. 2019, *Awas Bahaya Kanker Rahim dan Kanker Payudara*. Jakarta: Tim Wahana.

Krisan Aprian Widagdo. 2017, *Pembentukan Pohon Klasifikasi Biner Dengan Algoritma CART (Classification and Regression Trees) dan Random Forest (FR) (Studi Kasus Penyakiit Diabetes Suku Pima Indian)*. Jurnal Sains. Volume 5, No. 2. Jakarta: Prestasi Pustaka.

Melda Yanti. 2016, *Faktor Resiko Kanker Payudara Pada Wanita Di Poliklinik*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

RSUP DR. M. Djamil Padang. Fakultas Kesehatan Masyarakat Universitas Andalas.

Muhammad Husni, Siti Romadoni, Desi Rukiyati. 2018, *Hubungan Dukungan Keluarga Dengan Kualitas Hidup Pasien Kanker Payudara Di Instalasi Rawat Inap Bedah RSUP Dr. Mohammad Hoesin Palembang Tahun 2018*. Jurnal Keperawatan Sriwijaya. Volume 2, Nomor 2. Palembang.

Nidhomuddin dkk. 2015, *Random Forest dan Multivariate Adaptive Regression Spline (MARS) Binary Response Untuk Penderita HIV/AIDS di Surabaya*. Jurnal Statistik. Vol.1 No.3 Hal 49-57. Surabaya.

Putu Okky Martha K. 2016, *Hubungan Antara Pengetahuan Kanker Payudara dan Perilaku SADARI (Pemeriksaan Payudara Sendiri) pada Wanita Dewasa Dini*. Jurnal Psiklogi. Yogyakarta.

Rahmi Islamiana Heri. 2017, *Karakteristik Pasien Kanker Payudara di RS Wahidin Sudirohusodo Periode Juni-November 2017*. Jurnal Kedokteran. Volume 2, No. 5. Makassar.

Reza Fitryesta R. 2016, *Pengaruh Penyuluhan Pemeriksaan Payudara Sendiri (SADARI) Dengan Penggunaan Media Video Terhadap Pengetahuan dan Sikap Deteksi Dini Kanker Payudara pada Siswi SMA Negri 1 Sumbawa*. Jurnal Pendidikan Kebidanan. Volume 2, No 3. Surabaya.

Rifqy Marwah Akhsanti, Widyanti Rahayu, Vera Maya Santi. 2016, *Klasifikasi Diagnosis Penyakit Kanker Payudara dengan Pendekatan Regresi Logistik Biner, Metode MARS dan Random Forest (FR)*. Jurnal Kesehatan. Jakarta.

Rini Mayasari. 2018, *Hubungan Antara Umur dan Status Perkawinan dengan*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Kejadian Kanker Payudara di Instalasi Bedah Rumah Sakit Umum Pusat Dr. Mohammad Hoesin Palembang Tahun 2018. Jurnal Keperawatan. Palembang.

Rokhana Dwi Bekti, Noviana Pratiwi, Maria Titah Jatipaningrum, Dina Auliana. 2017, *Analisis Pengaruh Lokasi dan Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan CART*. Media Statistika 10(2) 2017: 119-130.

Tutus Praningki, Indra Budi. 2017, *Sistem Prediksi Penyakit Kanker payudara Menggunakan CART, Random Forest (FR), Naive Bayes, dan k-NN*. Citec Journal, Vol. 4, No. 2. Yogyakarta.

Trisnadewi. 2018, *Faktor Kanker Payudara Pada Wanita di RSUP Sanglah Denpasar*. *Laporan Hasil Penelitian*. hal 181-188.

Untari, D. 2017, *Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-aktif Menggunakan Metode Decision Tree*. Jurnal Teknik Informatika. Volume 3, No. 5.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id