

**TEKNIK *OVERSAMPLING* UNTUK MEMPREDIKSI STATUS
PASIEN BEDAH TORAKS MENGGUNAKAN METODE
*DECISION TREE C4.5***

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA**

Oleh :

TESYA RAHMAWATI

08011281722026



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA**

2022

HALAMAN PENGESAHAN

**TEKNIK *OVERSAMPLING* UNTUK MEMPREDIKSI STATUS PASIEN
BEDAH TORAKS MENGGUNAKAN METODE *DECISION TREE* C4.5**

SKRIPSI


**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Matematika Fakultas MIPA**

Oleh

**TESYA RAHMAWATI
NIM.08011281722026**

Indralaya, Januari 2022

Pembimbing Pembantu



**Endang Sri Kresnawati, M.Si
NIP. 197702082002122003**

Pembimbing Utama



**Dr. Yulia Resti, M.Si
NIP. 197307191997022001**

**Mengetahui,
Ketua Jurusan Matematika**



**Dra. Shgandi Yehdin, M.M
NIP. 195807271986031003**

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini :

Nama Mahasiswa : Tesya Rahmawati

NIM : 08011281722026

Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan srata (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain. Semua informasi yang dimuat dalam skripsi ini berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya

Indralaya, 31 Januari 2022
Penulis



Tesya Rahmawati
NIM.08011281722026

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai Civitas Akademik Universitas Sriwijaya, yang bertanda tangan di bawah ini :

Nama Mahasiswa : Tesya Rahmawati
NIM : 08011281722026
Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Sriwijaya “Teknik *Oversampling* Untuk Memprediksi Status Pasien Bedah Toraks Menggunakan Metode Decision Tree C4.5”. Dengan hak bebas royalti non-eksklusif ini Universitas Sriwijaya berhak menyimpan, mengalih, edit/memformatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir atau skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Indralaya, 31 Januari 2022
Penulis



Tesya Rahmawati
NIM.08011281722026

HALAMAN PERSEMBAHAN

MOTTO

“SEORANG PROFESIONAL BERMULA DARI AMATIR!!”

“Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu. Allah mengetahui, sedang kamu tidak mengetahui”

(QS. Al- Baqarah : 216)

Skripsi ini kupersembahkan untuk :

- ❖ ALLAH SWT**
- ❖ Rasulullah Muhammad SAW**
- ❖ Kedua Orangtuaku**
- ❖ Saudaraku dan Keluarga Besarku**
- ❖ Guru dan Dosenku**
- ❖ Sahabat-Sahabatku**
- ❖ Teman-Teman seperjuanganku**
- ❖ Almamaterku**

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur atas kehadiran Allah *Subhanahu wa ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Teknik *Oversampling* Untuk Memprediksi Status Pasien Bedah Toraks menggunakan Metode *Decision Tree* C4.5**” dengan baik. Shalawat serta salam semoga senantiasa selalu tercurahkan kepada Nabi Muhammad *Shallallahu 'alaihi wa sallam* beserta keluarga, sahabat dan seluruh pengikutnya hingga akhir zaman. Aamiin Yaa Rabbal 'Alamin.

Penulisan skripsi ini dimaksudkan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Jurusan Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan juga merupakan suatu sarana untuk menuangkan ilmu yang telah diperoleh selama mengikuti Pendidikan di perguruan tinggi.

Dengan penuh rasa hormat, cinta dan segala kerendahan hati, pertama penulis ingin mengucapkan terima kasih yang tak terhingga kepada kedua orang tua, yakni **Papa Syamsurizal** dan **Mama Leni Marlina** dengan segenap cinta, kasih sayang berlimpah, nasehat, dukungan moril maupun materi, didikan serta doa yang tak pernah berhenti untuk keberhasilan penulis. Keberhasilan dalam menyelesaikan skripsi ini juga tidak terlepas dari bantuan pembimbing dan semua pihak yang telah membantu baik secara langsung maupun tidak langsung. Penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada:

1. Bapak **Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya serta Pembimbing Akademik yang telah banyak memberikan bimbingan, arahan, saran dan kritikan kepada penulis selama penulis menempuh pendidikan di Jurusan Matematika FMIPA Universitas Sriwijaya.
2. Ibu **Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang membimbing dan memberikan arahan kepada penulis selama menempuh Pendidikan di Jurusan Matematika FMIPA Universitas Sriwijaya.
3. Ibu **Dr. Yulia Resti, M.Si** selaku Dosen Pembimbing Utama yang telah meluangkan waktu untuk memberikan bimbingan, arahan, ide pemikiran, kesabaran, saran dan kritikan, nasehat serta motivasi dalam proses penulis menyelesaikan skripsi.
4. Ibu **Endang Sri Kresnawati, M.Si** selaku Dosen Pembimbing Kedua yang bersedia meluangkan waktu, memberikan bimbingan, arahan, ide pemikiran, dorongan, saran dan kritikan kepada penulis selama proses pengerjaan skripsi.
5. Ibu **Dr. Yuli Andriani, M.Si** dan Ibu **Irmeilyana, M.Si** selaku Dosen Pembahas yang telah bersedia memberikan saran dan kritikan yang bermanfaat dalam perbaikan skripsi.
6. Ibu **Des Alwine Zayanti, M.Si** dan Ibu **Oki Dwipurwani, M.Si** selaku Ketua dan Sekretaris Pelaksana Seminar yang telah bersedia memandu

jalannya seminar dan memberikan saran serta kritik yang bermanfaat dalam perbaikan skripsi.

7. **Seluruh Dosen di Jurusan Matematika FMIPA Universitas Sriwijaya** atas ilmu yang diberikan.
8. Bapak **Irwansyah** dan Ibu **Hamidah** yang telah banyak membantu penulis dalam hal administrasi di Jurusan Matematika FMIPA UNSRI.
9. Saudara-saudara saya yang saya cintai **Abang Bibi, Tesyi, Egi, Syifa, Hafizh** dan juga saudara diperantauan **Yukpil** serta sepupu saya yang telah memberikan banyak doa dan dukungan dalam menyelesaikan studi.
10. Sahabat-sahabat tim skripsi **Olivia, Azizah, Mega, Muflhika, Kalim, Abu, Rendy, Agung, Yudha, Ridwan** dan **Shohif** yang telah membersamai penulis
11. Anggota kosan oren **Kak Pind, Kak Azizah, Kak Qoon, Mbak Evi, Nisa, Qoyin, Intan, Dila** dan **Alfia** yang telah menemani dan banyak membantu penulis selama masa studi.
12. Sahabat **OneHeart Ajeng, Filda, Ranil, Ayu, Weli, Enyta, Elsa, Feni, Calista**, dan sahabat **Kareot Kariah Ayu, Umik, Olip, Ririz, Elsa** serta teman-teman Angkatan 2017 yang telah banyak membantu dan menemani hari-hari penulis selama masa studi.
13. Sahabat **CircleSMA Fira, Dije, Jihan, Tesyi, Ajip, Deval, Dini, Aulia, Zikron, Edwin, Andreas, Gigih, Gema, Lany, Rahmi** dan **Randa** yang telah memberikan banyak dukungan dan doa.

14. Tim **PramukaSMP Dela, Riri, Gina, Laras, Suci, Naharin, Naufal, Fathur** dan **Qadri** yang telah memberikan banyak dukungan serta doa.
15. Rekan-rekan **BPH HIMASTIK 2018/2019** , **BPH KOSMIC** dan **PERMATO UNSRI** yang telah menemani dan banyak membantu penulis dalam kegiatan organisasi.
16. **Kakak tingkat 2014, 2015, dan 2016** yang telah memberikan ilmu dan pengalamannya serta **adik tingkat angkatan 2018 dan 2019** yang telah memberikan support dan doa.
17. **Semua Pihak** yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, Januari 2022

Tesya Rahmawati
NIM.08011281722026

**OVERSAMPLING TECHNIQUE FOR PREDICTING THE STATUS
OF PATIENTS WITH THORACIC SURGERY USING THE
DECISION TREE C4.5 METHOD**

By:

TESYA RAHMAWATI

08011281722026

ABSTRACT

Lung cancer is one of the diseases with the highest mortality rate in the world. The cause of lung cancer is closely related to smoking habits. There are several types of lung cancer treatment, one of which is surgery. Thoracic surgery is one of the most common operations performed on patients with lung cancer. One of the main problems in treating lung cancer patients is deciding whether to undergo thoracic surgery because of the high risk of death. Therefore, it is necessary to predict the survival of patients with lung cancer after thoracic surgery. This study uses secondary data obtained from UCI Machine Learning which has 16 variables with 470 data. In the data there is a class imbalance, the class balancing technique used is the Oversampling technique. The method used to predict the status of thoracic surgery patients is the Decision Tree C4.5 method. The results of this study obtained a value on Decision Tree C4.5 with an accuracy of 88.30%, recall of 47.37%, precision of 90%, F-Measure of 62.1% and G-Mean of 68.36%. Decision Tree C4.5 with Oversampling technique with 88.30% accuracy, 78.95% recall, 68.18% precision, 73.17 F-Measure and 84.6% G-Mean.

Keywords : Lung cancer, Thoracic surgery, Decision Tree C4.5, Oversampling

**TEKNIK *OVERSAMPLING* UNTUK MEMPREDIKSI STATUS
PASIEN BEDAH TORAKS MENGGUNAKAN METODE
*DECISION TREE C4.5***

Oleh:

TESYA RAHMAWATI

08011281722026

ABSTRAK

Kanker paru merupakan salah satu penyakit dengan angka kematian terbanyak di dunia. Penyebab kanker paru berhubungan erat dengan kebiasaan merokok. Terdapat beberapa jenis pengobatan kanker paru salah satunya dengan pembedahan. Bedah toraks merupakan salah satu operasi yang paling umum dilakukan pada pasien penderita kanker paru. Salah satu masalah utama dalam menangani penderita kanker paru adalah memutuskan apakah menjalani bedah toraks karena memiliki risiko meninggal yang tinggi. Oleh karena itu perlu dilakukan prediksi untuk kelangsungan hidup pasien penderita kanker paru pasca bedah toraks. Penelitian ini menggunakan data sekunder yang diperoleh dari *UCI Machine Learning* yang memiliki 16 variabel dengan 470 data. Pada data tersebut terdapat ketidakseimbangan kelas, teknik penyeimbangan kelas yang digunakan adalah teknik *Oversampling*. Metode yang digunakan untuk memprediksi status pasien bedah toraks adalah dengan metode *Decision Tree C4.5*. Hasil penelitian ini memperoleh nilai pada *Decision Tree C4.5* dengan *accuracy* sebesar 88,30%, *recall* sebesar 47,37%, *precision* sebesar 90%, *F-Measure* sebesar 62,1% dan *G-Mean* sebesar 68,36%. *Decision Tree C4.5* dengan teknik *Oversampling* dengan *accuracy* sebesar 88,30% , *recall* 78,95%, *precision* 68,18%, *F-Measure* sebesar 73,17 dan *G-Mean* sebesar 84,6%.

Kata Kunci : Kanker paru, Bedah toraks, *Decision Tree C4.5*, *Oversampling*

DAFTAR ISI

| | |
|-------------------------------------|-------------------------------------|
| SKRIPSI..... | i |
| HALAMAN PENGESAHAN..... | Error! Bookmark not defined. |
| HALAMAN PERSEMBAHAN | ii |
| KATA PENGANTAR..... | vi |
| ABSTRACT | x |
| ABSTRAK | xi |
| DAFTAR ISI..... | xii |
| DAFTAR TABEL | xv |
| DAFTAR GAMBAR..... | xvii |
| DAFTAR LAMPIRAN | xviii |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Batasan Masalah..... | 5 |
| 1.4 Tujuan..... | 5 |
| 1.5 Manfaat..... | 6 |
| BAB II TINJAUAN PUSTAKA..... | 7 |
| 2.1 Kanker Paru..... | 7 |

| | | |
|---|---|-----------|
| 2.2 | Bedah Toraks..... | 7 |
| 2.3 | <i>Statistical Learning</i> | 8 |
| 2.4 | Data Mining..... | 9 |
| 2.4.1 | <i>Machine Learning</i> | 9 |
| 2.5 | <i>Imbalanced Class</i> | 10 |
| 2.5.1 | Teknik <i>Resampling (Oversampling)</i> | 11 |
| 2.6 | Klasifikasi..... | 11 |
| 2.7 | <i>Decision Tree</i> | 12 |
| 2.7.1 | Algoritma C4.5..... | 12 |
| 2.8 | <i>Confusion Matrix</i> | 14 |
| BAB III METODOLOGI PENELITIAN..... | | 17 |
| 3.1 | Waktu dan Tempat | 17 |
| 3.1.1 | Waktu | 17 |
| 3.1.2 | Tempat..... | 17 |
| 3.2 | Alat dan Bahan | 17 |
| 3.3 | Metode Penelitian..... | 17 |
| BAB IV HASIL DAN PEMBAHASAN | | 20 |
| 4.1 | Deskripsi Data | 20 |
| 4.2 | Diskritisasi Data | 21 |
| 4.3 | Tabulasi Data..... | 21 |

| | |
|--|-------------------------------------|
| 4.4 Partisi Data | 23 |
| 4.5 Teknik <i>Resampling</i> untuk Penyeimbangan Kelas (<i>Balanced Class</i>) | 23 |
| 4.5.1 Tabulasi Data Latih Baru (<i>Oversampling</i>)..... | 24 |
| 4.6 Metode Decision Tree C4.5 Sebelum Penyeimbangan Kelas | 25 |
| 4.7 Metode Decision Tree C4.5 Setelah Penyeimbangan Kelas | 37 |
| 4.7.1 Teknik <i>Resampling</i> (<i>Oversampling</i>) | 37 |
| 4.8 Analisis Hasil | 50 |
| BAB V KESIMPULAN DAN SARAN | 52 |
| 5.1 Kesimpulan..... | 52 |
| 5.2 Saran..... | 52 |
| DAFTAR PUSTAKA | 54 |
| LAMPIRAN..... | Error! Bookmark not defined. |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2. 1 <i>Confusion matrix</i> dari dua kelas prediksi..... | 14 |
| Tabel 4. 1 Deskripsi data..... | 20 |
| Tabel 4. 2 Diskritisasi data kontinu | 21 |
| Tabel 4. 3 Tabulasi variabel respon (data asli) | 22 |
| Tabel 4. 4 Tabulasi variabel prediktor terhadap variabel respon (data asli) | 22 |
| Tabel 4. 5 Data latih | 23 |
| Tabel 4. 6 Data uji..... | 23 |
| Tabel 4. 7 Data latih baru (<i>Oversampling</i>)..... | 24 |
| Tabel 4. 8 Tabulasi variabel respon data latih baru (<i>Oversampling</i>) | 24 |
| Tabel 4. 9 Tabulasi variabel prediktor terhadap variabel respon data latih baru (<i>Oversampling</i>) | 24 |
| Tabel 4. 10 Lanjutan perhitungan <i>entropy</i> dan <i>gain root node</i> | 26 |
| Tabel 4. 11 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1 | 28 |
| Tabel 4. 12 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1.1..... | 30 |
| Tabel 4. 13 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1.1.1 | 32 |
| Tabel 4. 14 Prediksi metode <i>decision tree</i> | 34 |
| Tabel 4. 15 <i>Confusion matrix</i> | 35 |
| Tabel 4. 16 Lanjutan perhitungan <i>entropy</i> dan <i>gain root node</i> (<i>Oversampling</i>) .. | 38 |
| Tabel 4. 17 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1 | 40 |
| Tabel 4. 18 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1.2..... | 42 |
| Tabel 4. 19 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1.2.1 | 43 |
| Tabel 4. 20 Perhitungan <i>entropy</i> dan <i>gain node</i> 1.1.2.1.1 | 45 |

| | |
|--|----|
| Tabel 4. 21 Prediksi metode <i>decision tree</i> | 48 |
| Tabel 4. 22 <i>Confusion matrix</i> | 48 |
| Tabel 4.23 Perbandingan hasil | 50 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2. 1 Distribusi data sebelum dan setelah <i>oversampling</i> | 11 |
| Gambar 3. 1 Alur metodologi penelitian..... | 19 |
| Gambar 4. 1 Pohon keputusan <i>root node</i> | 28 |
| Gambar 4. 2 Pohon keputusan node 1.1 | 29 |
| Gambar 4. 3 Pohon keputusan 1.1.1 | 31 |
| Gambar 4. 4 Pohon keputusan 1.1.1.1 | 33 |
| Gambar 4. 5 Pohon keputusan root node | 39 |
| Gambar 4. 6 Pohon keputusan node 1.1 | 41 |
| Gambar 4. 7 Pohon keputusan 1.1.2 | 43 |
| Gambar 4. 8 Pohon keputusan 1.1.2.1 | 45 |
| Gambar 4. 9 Pohon keputusan 1.1.2.1.1 | 46 |

DAFTAR LAMPIRAN

Lampiran 1. Pohon Keputusan Decision Tree C4.5 Pada Data Asli..... **Error!**

Bookmark not defined.

Lampiran 2. Pohon Keputusan Decision Tree C4.5 Pada Data Latih Baru

(Oversampling)**Error! Bookmark not defined.**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker paru merupakan salah satu penyakit dengan angka kematian terbanyak di dunia. Kanker paru menempati urutan ke tiga baik di dunia maupun di Indonesia. Jumlah kasus kanker paru di Indonesia adalah sebesar 12,8 per 100.000 penduduk. Jenis kanker yang paling banyak menyerang masyarakat Indonesia berbeda pada setiap jenis kelamin. Kanker paru-paru dengan jumlah kasus sebesar 20,1 per 100.000 penduduk paling banyak menyerang laki-laki, sedangkan pada perempuan kanker yang paling banyak menyerang adalah kanker payudara dengan jumlah kasus 44,0 per 100.000 penduduk (World Health Organization (WHO), 2021).

Penyebab pasti dari kanker paru belum diketahui secara jelas. Paparan atau inhalasi berkepanjangan terhadap suatu zat yang bersifat karsinogenik merupakan faktor penyebab utama disamping adanya faktor lain seperti kekebalan tubuh, genetik, dan lain-lain. Dari beberapa kepustakaan telah dilaporkan bahwa etiologi kanker paru berhubungan erat dengan kebiasaan merokok (Joseph & Rotty, 2020).

Tingginya angka merokok di masyarakat menyebabkan kanker paru menjadi salah satu masalah kesehatan yang ada di Indonesia. Terdapat beberapa jenis pengobatan pada kanker paru diantaranya yaitu dengan pembedahan, radioterapi, kemoterapi, imunoterapi, hormonoterapi dan terapi gen (Indonesia, 2003). Penanganan penyakit kanker paru membutuhkan keahlian khusus dan kolaborasi

antara ahli paru, radiologi diagnostik, patologi anatomi, radiologi terapi, bedah toraks, rehabilitasi medik dan lain sebagainya (Sanjaya & Fitriyani, 2019).

Operasi toraks atau bedah toraks merupakan spesialisasi bedah yang difokuskan pada prosedur yang melibatkan bagian dada atau lebih dikenal sebagai toraks (Humedad, 2017 dalam Fitriyani, 2018). Bedah toraks merupakan salah satu operasi yang paling umum dilakukan pada pasien kanker paru. Salah satu masalah utama dalam menangani penderita kanker paru adalah memutuskan apakah menjalani bedah toraks karena memiliki risiko dan manfaat bagi pasien baik dalam jangka pendek maupun jangka panjang (Koklu et al., 2015).

Penelitian ini menggunakan data sekunder bedah toraks (*thoracic surgery*) yang diperoleh dari *UCI Machine Learning* yang terdiri atas 16 variabel prediktor dan 1 variabel respon. Permasalahan kelas yang tidak seimbang (*imbalanced class*) ditemui pada data ini dimana jumlah kelas hidup lebih banyak dibandingkan dengan kelas meninggal. Ketidakseimbangan kelas dapat menurunkan kinerja dari model prediksi dan akibat lain dari ketidakseimbangan kelas adalah dalam melakukan pengklasifikasian sebagian besar informasi yang diperoleh dengan benar dimiliki oleh kelas mayor tetapi memiliki kinerja yang buruk pada kelas lain, hal ini disebabkan oleh kelangkaan kelas minoritas (Cateni et al., 2014). Beberapa teknik untuk menyeimbangkan kelas yang tidak seimbang adalah dengan teknik *oversampling*, *undersampling*, dan *hybrid*. *Oversampling* bertujuan untuk meningkatkan jumlah sampel minor dengan melakukan *resampling* pada data asli sedangkan *undersampling* bertujuan untuk menghapus sampel kelas mayor pada data asli (Cateni et al., 2014).

Penelitian ini akan melakukan proses pengklasifikasian untuk prediksi status pasien penderita kanker paru pasca bedah toraks. Salah satu metode klasifikasi dalam data mining adalah *Decision Tree* C4.5. Pohon keputusan (*Decision Tree*) merupakan metode klasifikasi yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang dihubungkan dengan cabang-cabang dari simpul akar (*root node*) sampai ke simpul daun (*leaf node*). Variabel diuji pada *node-node* keputusan dengan setiap hasil yang mungkin menghasilkan sebuah cabang (*internal node*), dan setiap cabang akan diarahkan ke *internal node* lain atau *leaf node* yang akan menghasilkan sebuah keputusan (Larose & Larose, 2005).

Terdapat beberapa peneliti yang melakukan prediksi status pasien penderita kanker paru pasca bedah toraks dengan menggunakan berbagai metode klasifikasi diantaranya (Anshori & Riana, 2021) menggunakan metode *Neural Network*, *SMOTE*, dan *AdaBoost* memperoleh nilai *accuracy* sebesar 76,75%. (Sanjaya & Fitriyani, 2019) menggunakan seleksi fitur *Forward Selection* dan *K-Nearest Neighbor* memperoleh nilai *accuracy* terbesar adalah sebesar 85,74%, *recall* sebesar 8,57% dan *precision* sebesar 66,67%. (Sindhu et al., 2014 dalam Sanjaya and Fitriyani, 2019) menggunakan metode *Naïve Bayes* memperoleh nilai *accuracy* sebesar 82,34%, metode *J48 Decision Tree* memperoleh nilai *accuracy* sebesar 85,10% dan dengan metode *Decision Stump* memperoleh nilai *accuracy* sebesar 85,10%.

Terdapat beberapa peneliti yang mengimplementasikan metode *Decision Tree* C4.5 dalam penelitiannya. Swastina (2018) melakukan penelitian tentang penerapan algoritma C4.5 untuk penentuan jurusan mahasiswa yang

dilatarbelakangi oleh banyaknya kasus pemilihan jurusan yang tidak sesuai dengan kemampuan, kepribadian, minat dan bakat yang dapat mempengaruhi mahasiswa dalam mengikuti perkuliahan. Dari hasil eksperimen dan evaluasi yang telah dilakukan memperoleh nilai akurasi sebesar 82,64%.

Penelitian selanjutnya tentang penerapan algoritma C4.5 dalam pengklasifikasian data rekam medis berdasarkan ICD-10 yang dilatarbelakangi oleh catatan khusus pasien atau rekam medis seringkali menjadi data yang menumpuk dan tidak dilakukan penelusuran untuk menghasilkan pengetahuan yang berguna bagi rumah sakit. Tujuan dari penelitian ini adalah mengolah tumpukan data rekam medis untuk mengklasifikasikan jenis penyakit yang terjadi di rumah sakit berdasarkan kode penyakit internasional (ICD-10). Dari hasil penelitian ini memperoleh nilai akurasi sebesar 61,90% (Fiandra dkk., 2017).

Penelitian selanjutnya oleh Hermawanti (2012) yang meneliti tentang klasifikasi untuk diagnosis penyakit kanker payudara dengan algoritma C4.5 dan algoritma *naïve bayes* yang memperoleh akurasi sebesar 94.56% untuk algoritma C4.5 dan 94.1% untuk algoritma *naïve bayes*. Dari penelitian sebelumnya menunjukkan bahwa metode *Decision Tree* C4.5 cukup baik dalam melakukan pengklasifikasian. Oleh karena itu penulis melakukan penelitian serupa untuk memprediksi status pasien penderita kanker paru pasca bedah toraks.

1.2 Rumusan Masalah

1. Bagaimana memprediksi status pasien penderita kanker paru pasca bedah toraks menggunakan metode *Decision Tree* C4.5?

2. Bagaimana perbandingan tingkat ketepatan klasifikasi setelah diterapkan teknik *resampling* untuk mengatasi *imbalanced class* pada pasien penderita kanker paru pasca bedah toraks menggunakan metode *decision tree* C4.5

1.3 Batasan Masalah

1. Data *Thoracic Surgery Dataset* dari *Wroclaw Thoracic Surgery Centre* yang menjelaskan tentang data pasien penderita kanker paru yang melakukan operasi bedah toraks tahun 2009-2014. Data terdiri dari 470 data dan 17 variabel diantaranya adalah *diagnosis*, *FVC*, *FEV1*, *performance status*, *pain*, *haemoptysis*, *dyspnoea*, *cough*, *weakness*, *tumor size*, *diabetes mellitus*, *MI_6mo*, *PAD*, *smoking*, *asthma*, *age* dan *Death_1yr*.
2. Data dipartisi menjadi 80% data latih dan 20% data uji.
3. Teknik Resampling yang digunakan adalah *Oversampling*.
4. Tingkat ketepatan klasifikasi dibatasi oleh nilai *accuracy*, *recall*, *precision*, *F-Measure* dan *G-Mean*.

1.4 Tujuan

1. Memprediksi status pasien penderita kanker paru pasca bedah toraks menggunakan metode *Decision Tree* C4.5
2. Menghitung tingkat ketepatan dalam memprediksi status pasien penderita kanker paru pasca bedah toraks setelah diterapkan teknik *resampling* untuk mengatasi *imbalanced class* menggunakan metode *Decision Tree* C4.5

1.5 Manfaat

1. Membantu tenaga medis terkait informasi dalam membuat keputusan untuk memprediksi pasien penderita kanker paru pasca operasi bedah toraks.
2. Sebagai bahan referensi untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- Achmad, B. D. M., & Slamet, F. (2012). Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree. *Jurnal IPTEK*, 16(1), 18–19.
- Ahmad, A. (2017). Mengenal Artificial Intelligence , Machine Learning , Neural Network , dan Deep Learning. *Jurnal Teknologi Indonesia*.
- Ahmed, H., Abd-El Ghany, S. F., Youn, E. M. G., Omran, N. F., & Ali, A. A. (2019). Stroke Prediction Using Distributed Machine Learning Based on Apache Spark. *International Journal of Advanced Science and Technology*, 28(15), 89–97.
- Anshori, I. F., & Riana, D. (2021). *Prediksi Harapan Hidup Pasien Kanker Paru-Paru Pasca Operasi Bedah Thoraks Menggunakan Boosted Neural Network Dan Smote*. 6(1), 9–15.
- Cateni, S., Colla, V., & Vannucci, M. (2014). A method for resampling imbalanced datasets in binary classification tasks for real-world problems. *Neurocomputing*, 135, 32–41. <https://doi.org/10.1016/j.neucom.2013.05.059>
- Fiandra, Y. A., Defit, S., & Yuhandri. (2017). Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(2), 82–89.
- Fitriani, R. D., Yasin, H., & Tarno. (2021). Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20.
- Fitriyani. (2018). Metode Bagging Untuk Imbalance Class Pada Bedah Toraks Menggunakan Naïve Bayes. *Jurnal Kajian Ilmiah*, 18(3), 270–282.
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). *The Elements of Statistical*

Learning (2nd Ed) (2nd Editio). Springer.

Heranova, O. (2019). Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 443–450. <https://doi.org/10.29207/resti.v3i3.1275>

Hermawanti, L. (2012). Penerapan algoritma klasifikasi c4.5 untuk diagnosis penyakit kanker payudara. 7(1), 57–64.

Indonesia, P. D. P. (2003). *Pedoman Diagnosis & Penatalaksanaan Penyakit Kanker Paru Di Indonesia*. Website Resmi Perhimpunan Dokter Paru Indonesia.

Jian, C., Gao, J., & Ao, Y. (2016). Author ' s Accepted Manuscript To appear in : Neurocomputing. *Neurocomputing*.

Joseph, J., & Rotty, L. W. A. (2020). Kanker Paru: Laporan Kasus. *Medical Scope Journal*, 2(1), 17–25. <https://doi.org/10.35790/msj.2.1.2020.31108>

Kasih, P. (2019). Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara. *Innovation In Research Of Informatics (INNOVATICS)*, 1(2), 63–69.

Koklu, M., Kahramanli, H., & Allahverdi, N. (2015). Applications of Rule Based Classification Techniques for Thoracic Surgery. *Joint International Conference 2015, November*, 1991–1998.

Larose, D. T., & Larose, C. D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*.

Mujiasih, S. (2011). Pemanfatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi Dan Geofisika*, 12(2), 189–195. <https://doi.org/10.31172/jmg.v12i2.100>

Prachuabsupakij, W., & Doungpaisan, P. (2016). Matching Preprocessing Methods

for Improving the Prediction of Student's Graduation. *IEEE International Conference on Computer and Communications*, 33–37.

Puruhito. (2013). *Ilmu Bedah Toraks Primer, Kardiak, dan Vaskular*. Airlangga University Press.

Putra, P. P., & Chan, A. S. (2018). *Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4 . 5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)*.

Ren, F., Cao, P., Li, W., Zhao, D., & Zaiane, O. (2017). Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm. *Computerized Medical Imaging and Graphics*, 55, 54–67. <https://doi.org/10.1016/j.compmedimag.2016.07.011>

Sanjaya, R., & Fitriyani. (2019). Prediksi Bedah Toraks Menggunakan Seleksi Fitur Forward Selection dan K-Nearest Neighbor. (*JEPIN*) *Jurnal Edukasi Dan Penelitian Informatika*, 5(3), 316–320. <https://doi.org/https://dx.doi.org/10.26418/jp.v5i3.35324>

Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote Dan K-Nearest Neighbor. *Isd*, 3(1), 44–49.

Suherman, Purnamasari, M., & Hastuti, F. D. (2021). *Klasifikasi Siswa Berdasarkan Mata Pelajaran Lintas Minat Menggunakan Metode Decision Tree C4.5*. 8(2), 141–149.

Swastina, L. (2018). Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa. *Gema Aktualita*, 2(1), 93–98.

Worker Health Protection Program (WHPP). (2013). *Understanding Your Breathing Test Results*. <http://www.worker-health.org/breathingtestresults.html>

World Health Organization (WHO). (2021). *Global Cancer Observatory (GCO)*. Data Source GLOBOCAN 2020. <https://gco.iarc.fr/>

- Yunus, M. (2017). Analisis mutasi gen epidermal growth factor receptor pada sampel sediaan hapusan sitologi dan plasma darah pasien kanker paru. In *Tesis Program Studi Bioteknologi Institut Pertanian Bogor*.
- Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Big Data Research Finding the Best Classification Threshold in Imbalanced Classification ☆. *Big Data Research, 1*, 1–7. <https://doi.org/10.1016/j.bdr.2015.12.001>