

TESIS

**Perbandingan Algoritma *Machine Learning* untuk
Survivabilitas dan Biaya Pengobatan Pasien Kanker
Paru-paru di Taiwan**



**PALMY RAWINDA MELIALA
19916038**

PROGRAM STUDI TEKNIK INDUSTRI

PROGRAM MAGISTER

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2021

**PERBANDINGAN ALGORITMA *MACHINE*
LEARNINGS UNTUK SURVIVABILITAS DAN
BIAYA PENGOBATAN PASIEN KANKER PARU-
PARU DI TAIWAN**

**Tesis untuk memperoleh Gelar Magister pada Program
Studi Teknik Industri Program Magister
Fakultas Teknologi Industri
Universitas Islam
Indonesia**

**PALMY RAWINDA MELIALA
19916038**

**PROGRAM STUDI TEKNIK INDUSTRI
PROGRAM MAGISTER
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2021

PERNYATAAN KEASLIAN

Demi Allah, saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang setiap satunya telah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak kekayaan intelektual maka saya bersedia ijazah yang saya terima untuk ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 10 Desember 2021

A handwritten signature in black ink is written over a yellow 10,000 Rupiah Indonesian postage stamp. The stamp features the Garuda Pancasila emblem and the text 'METERAI TEMPEL' and '10000'. The serial number '8384AJX535266727' is visible at the bottom of the stamp.

Palmy Rawinda Meliala

19916038

LEMBAR PENGESAHAN PEMBIMBING

TUGAS AKHIR

**Perbandingan Algoritma *Machine Learning* untuk
Survivabilitas dan Biaya Pengobatan Pasien Kanker
Paru-paru di Taiwan**

Tesis telah disetujui pada tanggal

01 Desember 2021

Dosen Pembimbing 1

Dosen Pembimbing 2



(Dr. Drs. Imam Djati Widodo,

(Bambang Suratno, S.T., M.T.,

M.Eng.Sc.)

Ph.D.)

NIP. 935220102

NIP. 0529098401

الجامعة الإسلامية
Indonesian text in Arabic script

Mengetahui

Ketua Program Studi Teknik Industri Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Winda Nur Cahyo, S.T., M.T., Ph.D.)

NIP. 025200519

LEMBAR PENGESAHAN DOSEN PENGUJI

Perbandingan Algoritma *Machine Learning* untuk Survivabilitas dan Biaya Pengobatan Pasien Kanker Paru-paru di Taiwan

TUGAS AKHIR

Disusun Oleh:

Nama : Palmy Rawinda Meliala

NIM : 19 916 038

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Magister Teknik Industri Fakultas Teknologi Industri

Universitas Islam Indonesia

Yogyakarta, 28 Desember 2021

Tim Penguji

Dr. Drs.Imam Djati Widodo, M.Eng.Sc.

Ketua

Winda Nur Cahyo, S.T., M.T., Ph.D.

Anggota 1

Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D.

Anggota 2

Mengetahui

**Ketua Program Studi Teknik Industri Program Magister
Fakultas Teknologi Industri
Universitas Islam Indonesia**



(Winda Nur Cahyo, S.T., M.T., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Warrahmatullahi Wabarakatuh,

Alhamdulillah, segala puji dan syukur senantiasa penulis panjatkan atas kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul **“Perbandingan Algoritma *Machine Learning* Untuk Survivabilitas Pasien Kanker Paru-Paru Di Taiwan”** guna memenuhi salah satu prasyarat untuk memperoleh gelar sarjana Strata Dua (S2) pada program Magister Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Penulis menyadari kelemahan serta keterbatasan yang ada sehingga dalam menyelesaikan tesis ini penulis banyak mendapatkan bantuan, dukungan dan kesempatan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Hari Purnomo, Prof., Dr., Ir M.T. selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
2. Bapak Winda Nur Cahyo S.T., M.T., PhD. selaku Ketua Program Magister Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Dr. Drs.Imam Djati Widodo, M.Eng.Sc. dan Bambang Suratno, S.T., M.T., Ph.D. selaku dosen pembimbing yang senantiasa meluangkan waktunya untuk memberikan bimbingan.
4. Keluarga tercinta, Papa, Mama, kakak dan adek yang telah memberikan dukungan langsung maupun tidak langsung selama penyusunan tesis.
5. Seluruh responden dan pihak yang telah turut mendukung dan membantu dalam penyelesaian tesis ini yang tidak dapat disebutkan satu persatu oleh penulis.

Penulis menyadari bahwa laporan tesis ini masih jauh dari sempurna sehingga penulis mengharapkan kritik dan saran dari semua pembaca demi melengkapi kekurangan dalam laporan ini. Semoga laporan ini dapat bermanfaat bagi semua pihak yang berkepentingan. Aamiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, Desember 2021

Palmy Rawinda Meliala



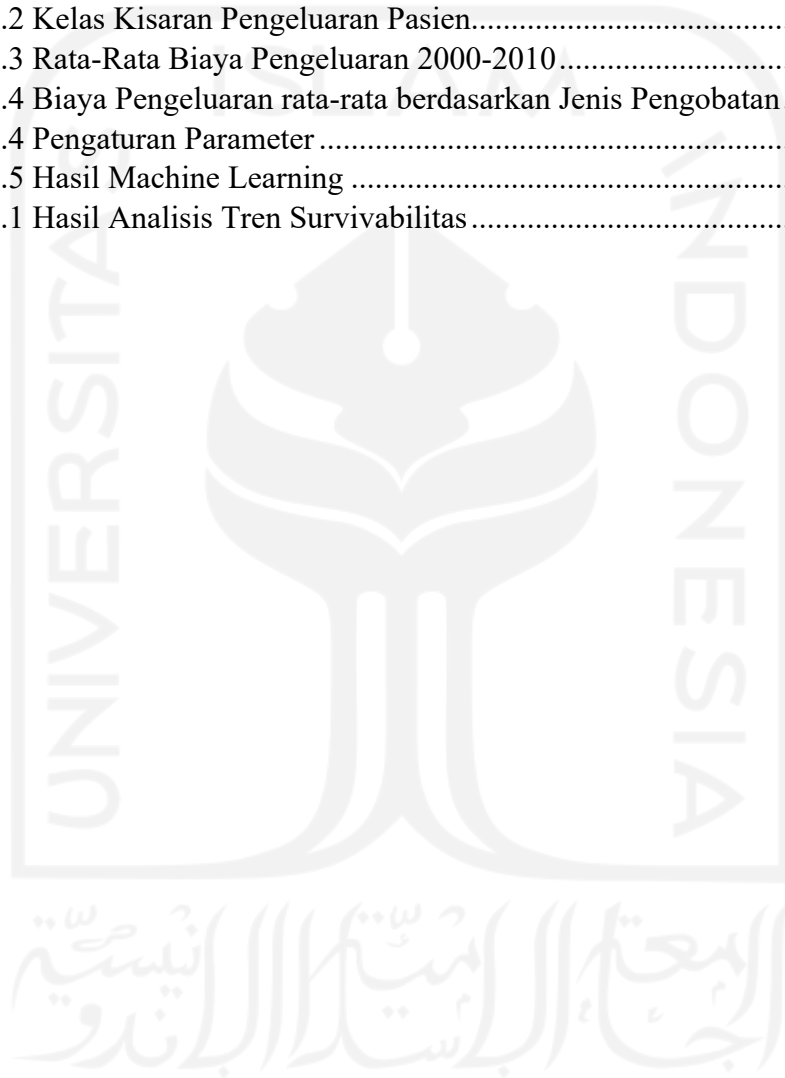
DAFTAR ISI

PERNYATAAN KEASLIAN	iv
LEMBAR PENGESAHAN PEMBIMBING	v
LEMBAR PENGESAHAN PENGUJI.....	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
ABSTRAK.....	xiii
BAB I PENDAHULUAN.....	2
1.1 Latar Belakang.....	2
1.2 Rumusan Masalah.....	4
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Kajian Induktif.....	7
2.2 Kajian Deduktif	13
2.1.1 Faktor Penyesuaian Resiko	13
2.1.2 Faktor Resiko untuk Kanker Paru-Paru	14
2.1.3 Kaplan Meier Estimation	15
2.1.4 Cox Proportionals Hazard.....	16
2.1.5 Algoritma Machine Learning	17
2.1.6 Pengukuran Performansi.....	20
BAB III METODE PENELITIAN	22
3.1 Sumber Data	22
3.2 Faktor-Faktor	23
3.3 Definisi Pemilihan Tahun	28

3.4	Prosedur Penelitian	30
BAB IV ANALISIS DATA.....		32
4.1	Estimasi Kaplan-Meier	32
4.2	Analisis Biaya Pengeluaran	34
4.3	Pengaturan Parameter	39
4.4	Hasil Eksperimen.....	41
BAB V PEMBAHASAN.....		43
5.1	Machine Learning	43
5.2	Analisis Survivabilitas	44
5.3	Analisis Biaya Pengeluaran	46
BAB VI SIMPULAN DAN SARAN		48
6.1	Kesimpulan	48
6.2	Saran	48
DAFTAR PUSTAKA.....		49
LAMPIRAN		55
Lampiran 1. Hasil Cox Proportional Hazard		55
Lampiran 2. Biaya Pengeluaran berdasarkan Biaya Pengobatan		63
Lampiran 3. Biaya Pengeluaran berdasarkan Tahun Kematian.....		66

DAFTAR TABEL

Tabel 2.1 Kajian Literatur.....	11
Tabel 3.1 Risk Factor Terpilih.....	24
Tabel 3.2 Variabel Utama Pemodelan	27
Tabel 3.3 Pengobatan Kanker Paru-Paru.....	29
Tabel 4.1 Probabilitas Rata-Rata Survivabilitas	34
Tabel 4.2 Kelas Kisaran Pengeluaran Pasien.....	35
Tabel 4.3 Rata-Rata Biaya Pengeluaran 2000-2010.....	35
Tabel 4.4 Biaya Pengeluaran rata-rata berdasarkan Jenis Pengobatan.....	37
Tabel 4.4 Pengaturan Parameter	39
Tabel 4.5 Hasil Machine Learning	41
Tabel 5.1 Hasil Analisis Tren Survivabilitas	45



DAFTAR GAMBAR

Gambar 3.1 Pemrosesan Pengambilan Data Sampling.....	23
Gambar 3.2 Cox PH Plot Keberlangsungan Hidup	26
Gambar 3.3 Cox PH Plot Biaya Pengeluaran	26
Gambar 3.4 Prosedur Penelitian	31
Gambar 4.1 Survival Curve Berdasarkan Jenis Kelamin	32
Gambar 4.2 Survival Curve Berdasarkan Lokasi Kanker.....	33
Gambar 4.3 Survival Curve Berdasarkan Jenis Pengobatan.....	33
Gambar 4.4 Proporsi Pengeluaran 2000-2003	36
Gambar 4.5 Proporsi Pengeluaran 2005-2010	36
Gambar 4.6 Rata-Rata Biaya Pengobatan Berdasarkan Jenis Perawatan	37
Gambar 4.7 Proporsi Pasien Berdasarkan Jenis Perawatan (2000-2003)	38
Gambar 4.8 Proporsi Pasien Berdasarkan Jenis Perawatan (2004-2010).....	38



ABSTRAK

Tingkat survivabilitas kanker paru-paru rendah di Taiwan selama tiga dekade terakhir, senilai sekitar lima belas persen untuk tingkat kelangsungan hidup lima tahun. Kanker paru-paru prognose pada tahap awal sebelum menjadi lebih parah adalah hal yang penting. Selain itu, biaya pengobatan untuk kanker paru-paru tinggi. Ini berada di tempat keempat di Taiwan untuk biaya medis paling mahal. Prediksi harapan hidup, terkait dengan biaya, diperlukan untuk meningkatkan tingkat kelangsungan hidup dan keseimbangan biaya dari pasien kanker paru-paru. Harapan hidup dan informasi biaya akan membantu rumah sakit atau dokter memberikan pengobatan efektif terbaik untuk pasien kanker paru-paru. Dalam penelitian ini, kami menyediakan algoritma perbandingan pembelajaran mesin yang dibantu dengan 10 *cross validation* dan *feature selection* untuk memprediksi kemampuan bertahan hidup dan pengeluaran medis pasien kanker paru-paru dari tahun 2000-2010. *Cox Proportional Hazard* digunakan untuk menentukan variabel prediktor penyakit kronis terbaik. Analisis tren akan dilakukan dengan Kaplan-Meier untuk analisis kelangsungan hidup dan mengetahui variabel dari pemanfaatan sebelumnya yang memiliki dampak langsung terhadap pengeluaran medis. Hasilnya menegaskan bahwa *Decision Tree* memiliki kinerja optimal untuk memprediksi kemampuan bertahan dan pengeluaran pasien kanker paru-paru untuk semua tahap, baik dalam pemilihan fitur atau tidak, dari akurasi, penarikan, dan skor F1 selalu menghasilkan di atas 90%.

Kata kunci: *Cox Proportional Hazard*; Kanker Paru-Paru; *Machine Learning*; Pengeluaran Medis; *Survivability*.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker terkenal menjadi penyebab utama kematian di setiap negara (WHO, 2018). Dari laporan 18,1 juta kasus kanker dan 9,6 juta kematian akibat kanker, kanker paru-paru berada di tempat pertama karena menjadi kanker yang paling banyak didiagnosis dengan 11,6% dari total kasus, dan penyebab utama kematian akibat kanker dengan 18,4% dari total kasus kematian yang ada di dunia (Bray et al., 2018).

Di Taiwan, kanker paru-paru juga merupakan penyebab utama kematian terbanyak yang terjadi di antara pria dan wanita (Wang *et al.*, 1998). Menurut laporan tahunan yang terdaftar sebagai kanker Taiwan, ada 12.462 kasus baru kanker paru-paru (12,1% dari total diagnosis kanker) dan 9.167 kematian (19,9% dari total kasus kematian) (Health, 2014). Tingkat survivabilitas kanker paru-paru di Taiwan termasuk rendah dari tiga dekade lalu, dengan persentase 15% untuk tingkat kelangsungan hidup selama lima tahun. Kanker paru-paru di Taiwan sering terjadi pada pasien yang memiliki komorbiditas. Komorbiditas memiliki peran yang besar dalam pengobatan awal dan efektivitas pengobatan perawatan pasien (Dutkowska & Antczak, 2016).

Dalam penelitian sebelumnya, pasien komorbiditas dengan kanker paru-paru telah menunjukkan respon yang kurang agresif dalam pengobatan yang diberikan (Asmis *et al.*, 2008; Janssen-Heijnen *et al.*, 2007). Jika pasien kanker paru-paru ini adalah seorang perokok, mereka sering memiliki komorbiditas seperti kardiovaskular (penyakit jantung iskemik atau hipertensi, arteriopati tungkai bawah, dll.) dan pernapasan (penyakit paru obstruktif kronik (PPOK), apnea tidur obstruktif, fibrosis interstisial biasa.). Sementara, komorbiditas yang tidak terkait dengan penggunaan tembakau adalah diabetes dan komplikasinya (insufisiensi ginjal, kerusakan kardiovaskular) (Antoni *et al.*, 2017). Oleh karena itu, menetapkan prognosis kanker paru-paru pada tahap awal merupakan bagian penting sebelum menjadi lebih parah dan tingkat kelangsungan hidup pasien

semakin rendah. Halabi & Owzar (2010) menyebutkan bahwa prognosis memainkan peran penting dalam penelitian klinis untuk manajemen pasien dan pengambilan keputusan. Prognosis dan prediksi yang lebih akurat pasti akan memberikan informasi yang bermanfaat bagi sektor medis. Informasi prognosis awal ini dapat membantu dalam merencanakan jenis pengobatan yang akan diberikan oleh pihak tenaga Kesehatan seperti dokter dan pihak Rumah Sakit itu sendiri, dalam meningkatkan tingkat survivabilitas kanker paru-paru yang ada di Taiwan.

National Health Insurance (NHI) adalah sistem asuransi kesehatan yang menjamin penduduk nasional terhadap biaya perawatan kesehatan Taiwan. Pada tahun 1995, Taiwan mengadopsi *National Health Insurance* (Wu *et al.*, 2010). Paket manfaat NHI mencakup semua layanan medis, obat-obatan, perawatan gigi, pengobatan Cina dan perawatan di rumah oleh perawat yang berkunjung (Hsiao *et al.*, 2019). Tetapi seiring waktu, orang lebih memperhatikan kesehatan setiap individu. Semakin banyak pasien, tentu semakin banyak biaya yang harus disiapkan NHI.

Kanker paru-paru ditempatkan di tempat keempat di Taiwan untuk biaya pengobatan yang paling mahal. Rata-rata yang tidak terhitung untuk biaya pengobatan dalam 10 tahun adalah NT \$ 448.371 (Wu *et al.*, 2010). Masalah biaya ini tidak hanya membebani pasien kanker paru-paru tetapi juga membebani sistem NHI. Terkadang, estimasi yang tidak akurat membuat NHI terbebani oleh masalah ekonomi mengenai biaya pengobatan langsung untuk pasien kanker paru-paru. Cara yang tepat untuk mengatasi biaya tak terduga adalah dengan mengendalikan biaya perawatan kesehatan yang dikeluarkan dari data historis.

Prediksi harapan hidup yang terkait dengan biaya diperlukan. Harapan hidup dan informasi biaya akan membantu rumah sakit atau dokter memberikan pengobatan yang paling efektif untuk pasien kanker paru-paru. Jika mortalitas dan morbiditas memiliki jumlah tertinggi, itu akan menjadi beban keuangan yang besar bagi pasien, keluarga, dan pemerintah. Dengan menggunakan *Machine Learning* ini harapannya biaya pengeluaran dapat membantu pihak asuransi dalam

melakukan *Budget Analysis* dari data historis yang ada sebelumnya, berdasarkan kategori range biaya yang dikeluarkan oleh pasien Kanker Paru-Paru di Taiwan.

Berdasarkan penelitian sebelumnya, peneliti menentukan *Machine Learning* yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM), *random forest*, *K-Nearest Neighbors* (KNN), *Naive Bayes Network* dan *Logistic Regression*. Algoritma ini berkinerja lebih baik daripada algoritma lain untuk memprediksi survivabilitas dan pengeluaran medis dari pasien Kanker paru-paru di Taiwan.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini sebagai berikut:

1. Apa algoritma *machine learning* yang memiliki performansi terbaik dalam memprediksi survivabilitas dan biaya pengeluaran yang diusulkan?
2. Apa rekomendasi yang bisa diberikan kepada sektor medis berdasarkan survivabilitas pasien?
3. Apa rekomendasi yang bisa diberikan kepada pihak NHI berdasarkan biaya pengeluaran pasien?

1.3 Batasan Permasalahan

Ruang lingkup dan batasan penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data sekunder dari database laboratorium.
2. Data yang dilakukan pengolahan hanya dari tahun 2000-2010.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini dijelaskan sebagai berikut:

1. Mengetahui hasil algoritma terbaik dalam memprediksi survivabilitas dan biaya pengeluaran dari pasien kanker paru-paru di Taiwan.
2. Mengetahui analisis tren dan memberikan rekomendasi kepada sektor medis berdasarkan survivabilitas pasien.
3. Mengetahui analisis tren dan memberikan rekomendasi kepada pihak NHI berdasarkan biaya pengeluaran pasien.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini dijelaskan sebagai berikut:

1. Bagi sektor medis dan asuransi dapat menggunakan pemodelan sebagai alat prediksi survivabilitas dan biaya pasien kanker paru-paru di Taiwan.
2. Bagi penelitian selanjutnya diharapkan dapat bermanfaat dan dapat menjadi referensi untuk perkembangan penelitian selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan tugas akhir disusun sebagai berikut:

BAB I PENDAHULUAN

Memberikan penjelasan terkait latar belakang permasalahan, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II KAJIAN LITERATUR

Menguraikan tentang konsep dan penjelasan metode yang digunakan. Selain itu terdapat hasil-hasil penelitian yang telah dilakukan sebelumnya oleh peneliti lain yang terkait dengan penelitian yang dilakukan.

BAB III METODE PENELITIAN

Menjelaskan terkait alur penelitian, teknik yang digunakan serta data yang akan dikaji.

BAB IV ANALISIS DATA

Menguraikan tentang data yang diperoleh yang ditampilkan dalam bentuk tabel serta cara menganalisa data tersebut. Pengolahan data termasuk analisis yang dilakukan terhadap data yang diperoleh.

BAB V PEMBAHASAN

Memaparkan pembahasan berupa implementasi dan penemuan dari algoritma yang diusulkan serta pembahasan hasil studi kasus beserta rekomendasi strategi pemasaran.

BAB VI SIMPULAN DAN SARAN

Menjelaskan tentang kesimpulan terhadap hasil penelitian yang telah dilakukan. Selain itu terdapat rekomendasi-rekomendasi terhadap penelitian lanjutan berdasarkan hasil yang telah dicapai.

DAFTAR PUSTAKA

LAMPIRAN



BAB II TINJAUAN PUSTAKA

Bab ini menyajikan penelitian sebelumnya yang terkait dengan penelitian ini. Metode-metode yang relevan juga dikaji untuk memahami konsep dan kelebihanannya guna memberikan kontribusi dari penelitian ini.

2.1 Kajian Induktif

Saat ini banyak penelitian yang telah dilakukan untuk mengatasi permasalahan pada kanker dengan menggunakan *Machine Learning*. Banyak metode yang dibandingkan satu sama lain dalam proses prediksi ataupun klasifikasi survivabilitas dan biaya pengeluaran dari penyakit Kanker.

Penelitian Maleki, *et al.*, (2020) melakukan pendeteksian klasifikasi kanker paru-paru di *stage* awal. Penelitian ini menggunakan *Decision Tree* dan *K-Nearest Neighbors* dengan menggunakan dan tanpa *feature selection*. Dari hasil perbandingan, penelitian ini menunjukkan bahwa KNN dengan menggunakan sebanyak 6 variabel dari *feature selection* memiliki akurasi pendeteksian yang sangat baik.

Penelitian yang dilakukan oleh K.R & N.C, (2018) menganalisis tingkat survivabilitas yang diprediksi untuk pasien kanker paru-paru berdasarkan faktor resiko yang ada. Metode yang digunakan adalah *Decision Tree C4.5*, *Naïve Bayes* dan *Support Vector Machine (SVM)*. Dari hasil perbandingan algoritma, *Decision Tree C4.5* memiliki akurasi yang paling baik diantara algoritma yang lain.

Xie, *et al.*, (2021) melakukan penelitian dengan tujuan untuk menemukan algoritma yang bisa melakukan prediksi tumor paru-paru berdasarkan *biomarker features*. Dalam penelitian digunakan algoritma *Support Vector Machine (SVM)*, *Random Forest*, *Neural Network*, *Naïve Bayes*, dan *Adaboost*. Dari hasil pengujian, algoritma yang memiliki algoritma paling baik adalah *Naïve Bayes*.

Pada penelitian yang dilakukan oleh Makond, *et al.*, (2015) dilakukan prediksi survivabilitas dan pasien kanker paru-paru yang memiliki *Brain Metastasis*. Algoritma yang digunakan dan akan dibandingkan pada penelitian ini adalah

Bayesian Network, *Naïve Bayes*, *Logistic Regression* dan *Support Vector Machine* (SVM). Algoritma ini juga dibandingkan berdasarkan penerapan *Synthetic Minority Over-sampling Technique* (SMOTE), yang digunakan sebagai teknik untuk mengeliminasi ketidakseimbangan data yang ada. Dari hasil pengujian didapatkan bahwa, *Bayesian Network* tanpa menggunakan SMOTE memiliki hasil akurasi yang lebih baik dibandingkan dengan metode yang lain.

Penelitian yang dilakukan Wang, *et al.*, (2019) bertujuan untuk melakukan diagnosa terhadap pasien kanker paru-paru dengan membandingkan *Fisher Model* dan *Support Vector Machine* (SVM). Dari hasil pengujian diketahui bahwa *Support Vector Machine* (SVM) memiliki performansi lebih baik daripada *Fisher Model*.

Ni, *et al.*, (2018) melakukan penelitian yang bertujuan untuk memprediksi kematian pasien kanker paru-paru berdasarkan 30 indikator dari kebiasaan manusia. Algoritma yang digunakan didalam penelitian ini adalah *Neural Network*, *Support Vector Machine* (SVM), dan *Linear Regression*. Hasil pengujian menunjukkan bahwa *Support Vector Machine* (SVM) memiliki performansi yang lebih baik dibandingkan metode yang lain.

Pada penelitian yang dilakukan oleh Bartholomai & Frieboes., (2021) bertujuan untuk model regresi terbaik untuk melakukan prediksi waktu survivabilitas pasien kanker paru-paru. Adapun algoritma yang dibandingkan adalah *Random Forest*, *General Linear Regression* (GL), *Gradient Boosted Machines* (GBM). Dari hasil pengujian menunjukkan *General Linear Regression* (GL) memiliki performansi yang lebih baik dibandingkan algoritma yang lain.

Penelitian oleh Wang, *et al.*, (2018) bertujuan untuk menentukan model prediksi stage awal pada pasien kanker paru-paru terbaik berdasarkan beberapa algoritma seperti *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree*, *Integrated Learning* (*Random Forest*, *Gradient Boost Decision Tree* (GBDT)). Dari hasil pengujian, diketahui bahwa *Gradient Boost Decision Tree* memiliki hasil performansi paling baik dibandingkan algoritma yang lain.

Faisal, *et al.*, (2020) melakukan penelitian yang berkaitan dengan evaluasi terhadap model prediksi *Machine Learning* untuk pendeteksian pasien kanker paru-paru di *stage* awal. Pada penelitian ini digunakan beberapa algoritma seperti

Support Vector Machine (SVM), C4.5 Decision tree, Multi-Layer Perceptron, Neural Network, Gradient Boosted Tree dan Naïve Bayes (NB). Dari hasil pengujian menunjukkan bahwa *Gradient Boosted Tree* memiliki hasil performansi yang paling baik dibandingkan algoritma lainnya.

Pada penelitian yang dilakukan oleh Nuhic & Kevric, (2020) menjelaskan penelitian mengenai klasifikasi tipologi kanker paru-paru menggunakan algoritma *Machine Learning*. Adapun algoritma yang digunakan pada penelitian ini adalah *Adaptive neuro-fuzzy inference system (ANFIS), Support Vector Machine (SVM), RUS Boosted Tree (RUSBoost) dan K-Nearest-Neighbor (K-NN)*. Dari hasil pengujian didapatkan *K-Nearest-Neighbor (K-NN)* sebagai algoritma dengan performansi terbaik.

Penelitian oleh Tapak, *et al.*, (2019) melakukan penelitian untuk mengkomparasi performansi dari beberapa jenis algoritma *Machine Learning* dalam melakukan prediksi survivabilitas dan metasis pada kanker otak. Adapun algoritma yang digunakan adalah *Naive Bayes (NB), Random Forest (RF), AdaBoost, Support Vector Machine (SVM), Least-square SVM (LSSVM), Adaboost, Adabag, Linear Discriminant dan Logistic Regression (LR)*. Dari penelitian tersebut didapatkan bahwa *Support Vector Machine (SVM)* merupakan algoritma terbaik untuk melakukan prediksi survivabilitas pasien kanker otak, sedangkan *Linear Discriminant* merupakan algoritma terbaik untuk mengetahui metastasis pada pasien kanker otak.

Amrane, *et al.*, (2020) melakukan penelitian dengan tujuan mengetahui algoritma mana yang memiliki performansi terbaik dalam mengklasifikasikan penyakit kanker otak. Adapun algoritma yang digunakan adalah *Naive Bayes (NB) dan K-Nearest Neighbor (KNN)*. Dari hasil pengujian didapatkan bahwa *Naïve Bayes* merupakan algoritma terbaik dalam melakukan klasifikasi penyakit kanker paru-paru.

Penelitian oleh Alabi, *et al.*, (2019) menjelaskan mengenai prediksi tingkat kambuh *stage* awal pada pasien kanker lidah dengan menggunakan algoritma *Machine Learning*. Ada 2 jenis algoritma yang digunakan yaitu *Artificial Neural Network (ANN) dan Logistic Regression*. Dari hasil pengujian didapatkan bahwa

Artificial Neural Network (ANN) memiliki performansi yang lebih baik daripada *Logistic Regression*.

Selanjutnya adalah penelitian yang dilakukan oleh Md. Milon Islam, *et al.*, (2017) bertujuan untuk membangun model klasifikasi yang bertujuan untuk memprediksi kanker otak. Algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN). Dari hasil pengujian ini diketahui bahwa *Support Vector Machine* (SVM) merupakan algoritma terbaik dalam melakukan prediksi.

Penelitian oleh Hiba Asri, *et al.*, (2016) menjelaskan mengenai prediksi resiko dan diagnosi menggunakan algoritma *Machine Learning* untuk pasien kanker otak. Adapun algoritma yang digunakan adalah *Support Vector Machine* (SVM), *Decision Tree* (C4.5), *Naïve Bayes* (NB) dan *K-Nearest Neighbors*. Dari hasil pengujian diketahui bahwa *Support Vector Machine* (SVM) merupakan algoritma terbaik untuk melakukan prediksi.

Penelitian oleh Yang, *et al.*, (2017) bertujuan untuk memprediksi biaya pengeluaran kesehatan pasien di amerika serikat berdasarkan data historis terdahulu. Adapun algoritma yang digunakan adalah *Linear regression* (LR), *regularized regression* (LASSO), *gradient boosting machine* (GBM), dan *recurrent neural networks* (RNN). Hasil pengujian menunjukkan bahwa algoritma *Recurrent Neural Networks* (RNN) memiliki performansi terbaik dalam memprediksi biaya pengeluaran dengan jenis data *time-series*.

Dari penelitian sebelumnya belum ditemukan penelitian yang menggabungkan tujuan untuk memprediksi survivabilitas dan biaya pengeluaran dengan menggunakan *Machine Learning*. Peneliti juga akan membandingkan algoritma yang memiliki hasil terbaik dari penelitian sebelumnya, untuk melakukan prediksi survivabilitas dan biaya pengeluaran pada pasien kanker paru-paru yang ada di Taiwan. Ringkasan penelitian-penelitian di atas apabila dikelompokkan berdasarkan kata kuncinya, dapat dilihat pada Tabel 2.1.

Tabel 2.1 Kajian Literatur

Penulis	Tujuan	Algoritma	Performansi	
Maleki, et al. (2020)	Melakukan klasifikasi dalam pendeteksian kanker paru-paru pada tahap awal	Decision Tree	95,2%	
		KNN tanpa <i>feature selection</i> (K=6)	99,8%	
		KNN tanpa <i>feature selection</i> (K=10)	95,2%	
		KNN dengan <i>feature selection</i>(K=6)	100%	
K.R & N.C, (2018)	Menganalisis Tingkat Survavibility pasien kanker paru-paru.	C4.5	Terbaik	
		Naïve Bayes	Rendah	
		Support Vector Machine (SVM)		
Xie, et al. (2021)	Untuk memprediksi tumor paru-paru berdasarkan fitur biomarkes metabolomik	Support Vector Machine (SVM)	94,70%	
		Random Forest	89,50%	
		Neural Network	94,70%	
		Naïve Bayes	100%	
		Adaboost	63,20%	
Makond, et al. (2015)	Memprediksi Survivabilitas Kanker Paru-paru dengan Metastasis Otak	Bayesian Network	SMOTE	Tanpa SMOTE
			88,30%	89,93%
		Naïve Bayes	87,45%	88,10%
		Logistic Regression	86,69%	86,91%
		Support Vector Machine (SVM)	87,90%	88,50%
Wang, et al. (2019)	Untuk mendiagnosis kanker paru-paru	Support Vector Machine (SVM)	Terbaik	
		Fisher Model	Rendah	
Ni, et al. (2018)	Untuk memprediksi kematian akibat kanker paru-paru	Linear regression	Rendah	
		Support Vector Machine (SVM)	Terbaik	
Bartholomai & Frieboes, (2021)	Untuk membuat model klasifikasi untuk memprediksi waktu bertahan hidup.	General Linear Regression (RMSE)	15,77	30,9092
		Random Forests (RMSE)	15,70	30,772
		Gradient Boosted Machines (GBM) -> (RMSE)	15,65	30,674
Wang, et al., (2018)	Untuk memprediksi kanker paru-paru berdasarkan faktor risiko	Support Vector Machine (SVM)	90%	
		K-Nearest Neighbors (KNN)	85%	
		Decision Tree	77%	
		Random Forest	88%	
		Gradient Boost Decision Tree (GBDT)	91%	
		Multi-layer Perceptron	78.33%	

Penulis	Tujuan	Algoritma	Performansi
Faisal, et al., (2020)	Untuk mengevaluasi terhadap model prediksi Machine Learning untuk pendeteksian pasien kanker paru-paru di stage awal	Naïve Bayes	85%
		Support Vector Machine (SVM)	79.17%
		Decision Tree	78.33%
		Gradient Boosted Tree	90%
		Nueral Network	71,67%
		Random Forest	79,17%
		Majority Voting (MLP+GBT+SVM)	88,57%
Nuhic & Kevric, (2020)	Untuk mengklasifikasikan kanker paru-paru berdasarkan penanda Biokimia	Adaptive neuro- fuzzy inference system (ANFIS)	86,63%
		Support vector machine (SVM)	94,00%
		RUS Boosted Tree (RUSBoost)	92,30%
		K-Nearest-Neighbor (K-NN)	94,60%
Tapak, et al., (2019)	Untuk membandingkan metode pembelajaran mesin untuk memprediksi waktu kelangsungan hidup kanker payudara dengan metastasis otak.	Naive Bayes (NB)	Rendah
		Random Forest (RF)	Rendah
		AdaBoost	Rendah
		Support Vector Machine (SVM)	Terbaik untuk survivabilitas
		Least-square SVM (LSSVM)	Rendah
		Adabag	Rendah
		Logistic Regression (LR)	Rendah
Linear Discriminant	Terbaik untuk metastasi otak		
Amrane, et al., (2020)	Untuk mengetahui metode mana yang dapat membangun pengklasifikasi yang tepat dan dapat diandalkan pada pasien kanker otak.	Naive Bayes (NB) classifier	97,51%
		K-nearest neighbor (KNN)	96,19%
Alabi, et al., (2019)	Untuk memprediksi kekambuhan pada kanker karsinoma sel skuamosa lidah oral tahap awal (OTSCC).	Artificial Neural Network (ANN)	92,70%
		Logistic Regression	86,50%
Md. Milon Islam, et al., (2017)	Untuk mebangun model klasifikasi yang bertujuan untuk memprediksi kanker otak	K-nearest neighbor (KNN)	98,57%
		Support Vector Machine (SVM)	97,14%
		Support Vector Machine (SVM)	97,13%

Penulis	Tujuan	Algoritma	Performansi
Hiba Asri, et al., (2016)	Untuk melakukan prediksi resiko dan diagnosi menggunakan algoritma Machine Learning untuk pasien kanker otak.	C4.5	95,12%
		Naive Bayes (NB)	95,28%
		K-nearest neighbor (KNN)	95,28%
Yang, et al. (2017)	Untuk memprediksi biaya pengeluaran kesehatan pasien di amerika serikat	Recurrent Neural Networks (RNN)	Terbaik
		Linear regression (LR)	Rendah
		Radient Boosting Machine (GBM)	Rendah
		Regularized Regression (LASSO)	Rendah

2.2 Kajian Deduktif

2.1.1 Faktor Penyesuaian Resiko

Faktor risiko dipilih untuk mengatasi efisiensi biaya pengobatan jangka panjang. Sangat penting untuk menentukan faktor risiko dan menetapkan model prediksi untuk memprediksi kemampuan bertahan hidup pasien dan pengeluaran medis berdasarkan faktor-faktor ini. Biaya pengobatan dapat dibagi menjadi biaya kronis, biaya akut, dan biaya kecelakaan secara umum (Layton, 2017). Enam kategori yang dapat membantu menemukan faktor risiko yang relevan adalah:

1. Faktor demografis meliputi jenis kelamin, usia, tingkat pendidikan dan area tempat tinggal. Faktor risiko ini dapat diperoleh dari statistik pemerintah dan catatan medis pasien. Namun, kurang efektif untuk memprediksi pengeluaran pribadi dengan menggunakan faktor demografis.
2. *Prior Utilization* meliputi catatan medis masa lalu dan pengeluaran medis dapat digunakan untuk memperkirakan informasi medis sebelumnya dan memprediksi pengeluaran di masa depan. Perubahan kebijakan pemerintah dan terapi pribadi akan mempengaruhi pengeluaran. Informasi tambahan tentang kasus-kasus ini dapat ditemukan di *National Healthcare Insurance Database*.
3. *Diagnostic-based factors* adalah faktor yang berkaitan dengan informasi rawat jalan dan rawat inap, faktor ini dapat membantu mengklasifikasikan pasien ke

dalam kelompok yang berbeda dan membantu memberikan rincian informasi penyakit. Kode ICD-9 digunakan untuk menemukan pasien dan catatan medis tertentu menggunakan *National Healthcare Insurance Database*. Faktor ini juga dapat digunakan untuk menentukan siapa yang memiliki penyakit kronis dari pasien. Kemudian, faktor yang diresepkan dapat membantu mengidentifikasi pasien yang perlu minum obat untuk waktu yang lama dan layanan medis yang sering yang sering dijalani oleh setiap pasien.

4. Faktor fisiologis meliputi status kesehatan pasien. Faktor ini mengandung tekanan darah, kolesterol, dan glukosa darah. Faktor fisiologis sangat mudah dalam membantu untuk memeriksa dan mencegah. Namun, itu tidak dapat diperoleh dari *National Healthcare Insurance Database*.
5. *Self-reported factor* adalah faktor yang berbeda dari penyedia asuransi. Faktor ini menggunakan kuesioner dan wawancara untuk mendapatkan informasi yang relevan. Faktor yang dilaporkan sendiri mengevaluasi kehidupan sehari-hari pasien, status fungsional, dan persyaratan pemanfaatan kesehatan.

2.1.2 Faktor Resiko untuk Kanker Paru-Paru

Karena fokus penelitian ini adalah pada pasien dengan kanker paru-paru, model ini akan dikembangkan berdasarkan faktor risiko kanker paru-paru. Nasser & Abu-Naser (2019) melakukan deteksi kanker paru-paru menggunakan Jaringan Saraf Tiruan. Variabel yang digunakan untuk diagnosis adalah gender; usia; riwayat merokok; dan gejala, seperti jari kuning; kecemasan; tekanan teman sebaya; penyakit kronis; kelelahan, alergi, mengi, alkohol, batuk, sesak napas, kesulitan menelan; dan nyeri dada. Dataset dikumpulkan dari situs web dunia data.

Lynch *et al.*, (2017) menggunakan pembelajaran mesin untuk memprediksi kelangsungan hidup pasien kanker paru-paru. Data diambil dari database SEER, yang berisi pasien yang didiagnosis dari 2004-2009. Model ini dibangun menggunakan 19 variabel yang terdiri dari kedua kontinu (numerik) dan kategoris (diskrit) jenis, yaitu usia, kelas, urutan radiasi dengan operasi, ukuran tumor, keterlibatan kelenjar getah bening, penyebaran tumor ke organ lain, radiasi, tahap, situs primer, indikator primer ganas pertama, nomor urut, kelenjar getah bening CS,

histologi recode-pengelompokan luas, prosedur pengangkatan / biopsi / aspirasi kelenjar getah bening regional, prosedur untuk mengangkat atau menghancurkan jaringan situs utama, tahap ringkasan SEER, pengukuran ukuran tumor dan waktu kelangsungan hidup. *Linear Regression, Decision Tree, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), dan custom ansamble* khusus digunakan untuk menentukan metode terbaik.

Hazra *et al.*, (2017) Mengembangkan model prediksi survivabilitas kanker paru-paru menggunakan dataset dari arsip pencitraan *NSCLC-Radiomics*, yang memiliki 422 pengamatan dan sepuluh variabel. Variabel yang digunakan adalah ID pasien, usia, stadium tumor klinis, tahap node klinis, tahap metastasis klinis, tahap keseluruhan, histologi, jenis kelamin, waktu bertahan hidup dan peristiwa status mati. Sebelum pemrosesan data, praproses data dan pemilihan fitur diterapkan. Studi sebelumnya di atas cenderung memiliki atribut atau variabel yang tepat yang digunakan untuk membangun model. Tetapi untuk komorbiditas, penelitian sebelumnya belum memiliki komorbiditas yang selalu hadir dalam kanker paru-paru. Komorbiditas yang digunakan biasanya berasal dari database yang ada, yang merupakan hasil dari catatan pasien.

2.1.3 Kaplan Meier Estimation

Metode non-parametrik digunakan untuk memperkirakan distribusi tingkat kelangsungan hidup. Metode yang digunakan di sini adalah tes *Kaplan-Meier*. Metode ini digunakan dengan menghitung pasien yang selamat dari penyakit atau kondisi tertentu setelah menerima intervensi atau pengobatan. Metode ini dapat dalam jangka waktu yang berbeda, tetapi umumnya, jika ada kehilangan anggota, proporsi kelangsungan hidup yang diperkirakan juga menurun. Oleh karena itu, kurva yang ditampilkan menggunakan tes *Kaplan-Meier* akan berbentuk seperti tangga menurun, yang mewakili setiap peserta yang hilang (Aakshi, 2017).

Untuk perhitungan *Kaplan-Meier*, variabel X mewakili variabel acak X_1, X_2, \dots, X_n , dan rata-rata $\mu = E(X)$, $E(X) < \infty$, $\sigma^2 \leq \infty$ dan fungsi distribusi $F(X) > 0$, fungsi kelangsungan hidup $\bar{F}(X) = 1 - F(X)$ dan distribusi sisa hidup rata-rata

$e(x)X$ untuk pengamatan yang disensor dihitung seperti di bawah ini: (Abdous & Berred, 2015).

$$e(x) = E(X - x | X > x) = \begin{cases} \int_x^{\infty} \frac{F(y)dy}{\bar{F}(x)} & \text{if } F(X) > 0 \\ 0 & \text{otherwise} \end{cases}$$

2.1.4 Cox Proportionals Hazard

Data kelangsungan hidup identik dengan memiliki lebih dari satu variabel. Model regresi dapat menganalisis hubungan satu set variabel dengan peristiwa dalam periode tertentu. Model regresi sesuai untuk melakukan analisis data kelangsungan hidup jika variabel independen lebih dari satu (Wuryandari *et al.*, 2018). *Cox Regression Proportional Hazard Model* adalah salah satu analisis kelangsungan hidup dalam model regresi. Bahaya proporsional adalah fungsi dari fungsi *Cox Regression Model (age-specific failure rate)*, yang dianggap sebagai fungsi dari variabel penjelasan. *Model Cox Proportional Hazard* dapat diperoleh dari koefisien regresi yang tidak diketahui dikalikan dengan fungsi waktu yang sewenang-wenang dan tidak diketahui (Cox, 1972).

Fungsi bahaya dapat diartikan sebagai risiko kematian pada satu waktu (t). $h(t)$ mewakili model *cox* dengan fungsi bahaya (Cox & Oakes, 1984).

$$h(t) = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \dots + b_px_p)$$

Waktu kelangsungan hidup mewakili (t) dalam model. Parameter vektor model diwakili oleh $b' = (b_1, b_2, \dots, b_p)$. Model hazard diwakili oleh $h(t)$ ditentukan oleh satu set p covariates $x' = (x_1, x_2, \dots, x_p)$. H_0 mewakili garis dasar bahaya. Ini menyesuaikan dengan nilai bahaya jika semua x' sama dengan nol (*quantity exp* (0) sama dengan 1). Tingkat bahaya $h_x(t)$ diperbolehkan untuk berubah dari waktu ke waktu, sedangkan h_r rasio bahaya tetap konstan; Ini adalah asumsi bahaya proporsional.

Jika HR lebih besar dari 1 ($b > 0$), maka risiko terjadinya akan meningkat untuk subjek dengan nilai kovariat x_2 dibandingkan dengan subjek dengan nilai kovariat x_1 , sedangkan HR lebih rendah dari 1 ($b < 0$), yang menjelaskan penurunan risiko. Dapat dilihat bahwa nilai bahaya di *Cox Proportional* adalah (Bellera *et al.*, 2010).

HR = 1: Tidak ada efek

HR < 1: Pengurangan bahaya

HR > 1: Peningkatan Bahaya.

2.1.5 Algoritma *Machine Learning*

Data mining adalah pendekatan pembelajaran statistik yang lebih efektif dan menggunakan fitur yang kuat untuk analisis data. Namun, ada beberapa masalah dengan kompleksitas pra-proses data dan pengetahuan dari para ahli. Kemajuan terbaru dalam teknologi pembelajaran mesin membantu memberikan paradigma baru yang efektif untuk mendapatkan model pembelajaran mesin pembelajaran *end-to-end* dari data yang kompleks (Mastoli *et al.*, 2019).

Machine Learning memiliki peran penting dalam bidang perawatan kesehatan, termasuk segmentasi dan pendaftaran gambar medis, fusi gambar multimodal, diagnosis berbantuan komputer, terapi yang dipandu gambar, anotasi gambar, dan pengambilan basis data gambar. Kegagalan bisa berakibat fatal (Khare *et al.*, 2017). Dengan mengungkapkan tren data yang sebelumnya tersembunyi, pembelajaran mesin menggunakan asumsi apriori distribusi data dan dapat menemukan pola dalam data yang dapat, pada gilirannya, digunakan untuk membuat hipotesis dan dalam hal tugas pengujian hipotesis yang ada dalam sistem kesehatan (Panch *et al.*, 2018). Meskipun model *Machine Learning* sulit untuk ditafsirkan, mereka dapat menggabungkan lebih banyak variabel. Mereka dapat digeneralisasi ke *array* yang jauh lebih luas dari jenis data, dan model juga dapat menghasilkan hasil dalam situasi yang lebih kompleks (Henglin *et al.*, 2017).

A. *Support Vector Machine*

Support Vector Machine (SVM) adalah metode machine learning yang biasanya digunakan dalam pengenalan pola dan masalah klasifikasi, terutama jika kumpulan data memiliki dua kelas. *Hyperplane* yang ideal yang memisahkan kelas adalah apa pengujian SVM adalah semua tentang. Fitur vektor atau pola input yang dimiliki diambil oleh pengklasifikasi dan menentukan kelasnya. Ini mengklasifikasikan data yang dapat dipisahkan secara linear, tetapi vektor fitur mungkin tidak dipisahkan secara linear secara

umum. Trik *kernel* digunakan untuk tes ini (Hussain *et al.*, 2011). Parameter untuk *Support Vector Machine* adalah *kernel*, *gamma* dan *c-value*. Fungsi *kernel* dapat diartikan untuk mengukur kesamaan antara satu sampel x_i dan sampel x_j untuk membantu pengklasifikasi SVM dalam memisahkan bahkan dengan batas-batas yang kompleks. *Gamma* (λ) adalah lebar *kernel* dan parameter tidak intuitif C (Amami *et al.*, 2013).

B. Random Forest

Random forest adalah salah satu algoritma machine learning yang paling populer. Random forest terdiri dari banyak pohon keputusan, dan output kelas yang dihasilkan adalah mode output kelas melalui pohon individu yang ada. Banyak dari pohon-pohon yang dihasilkan ini menciptakan hutan yang baik untuk masalah klasifikasi seperti ini dan tugas-tugas lain seperti regresi fungsional seperti yang dijelaskan di atas dengan menciptakan beberapa pohon ketika melatih dan mengeluarkan kelas atau memprediksi rata-rata pohon tertentu. Banyak pohon keputusan yang mendalam dilatih dalam kelompok terpisah dari dataset yang sama dan rata-rata dengan target pengurangan varians **Invalid source specified.**(Friedman *et al.*, 2001). Parameter tuning untuk hutan acak adalah kriteria, kedalaman maksimum, fitur maksimum dan jumlah estimator. Kriteria adalah nilai yang digunakan untuk memproses model. Fitur maksimum adalah split terbaik untuk model hutan acak. Estimator angka digunakan untuk membuat node daun dari node induk, dan kedalaman maksimum adalah jumlah batas instans di setiap node (Bentejaca *et al.*, 2006).

C. K-Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) adalah metode non-parametrik yang mudah namun fungsional yang digunakan untuk klasifikasi dan regresi. Model yang dihasilkan dihasilkan dari tes pelatihan dengan pertimbangan fungsi jarak dan fungsi pilihan kelas. Sebelum data baru diklasifikasikan, setiap data dibandingkan dengan data lain menggunakan ukuran kesamaan. K-Nearest Neighbor dari data kemudian dipertimbangkan, dan kelas yang paling umum di antara tetangga ditugaskan ke data untuk diklasifikasikan. Para tetangga

diberi bobot berdasarkan jarak. Parameter tuning untuk tetangga k-terdekat adalah algoritma, leaf_size, n_jobs, n_neighbors dan n_weight.

D. Naive Bayes Network

Naive Bayes adalah salah satu algoritma machine learning untuk klasifikasi biner. Efek nilai atribut pada kelas yang tidak bergantung pada nilai atribut lainnya adalah asumsi yang solid dalam menggunakan metode ini. Algoritma naive bayes adalah bentuk sederhana dari klasifikasi probabilistik berdasarkan teorema Bayes.

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

Di mana $P(B)$ adalah probabilitas posterior, $P(B | A)$ adalah probabilitas, $P(A)$ adalah probabilitas kelas sebelumnya, dan $P(B)$ adalah probabilitas prediktor sebelumnya. Keuntungan dari pengklasifikasi Naive Bayes adalah bahwa hal itu tidak tergantung pada data pelatihan. Hal ini dapat menggunakan sejumlah kecil data pelatihan untuk memperkirakan parameter yang diperlukan untuk klasifikasi. Penyetelan parameter naive bayes menghaluskan. Smoothing adalah parameter tuning kemungkinan maksimum untuk model menjadi lebih akurat.

E. Decision Tree

Model prediktif yang dapat mewakili model klasifikasi adalah decision tree (mungkin juga disebut Pohon Klasifikasi). Decision tree digunakan karena kesederhanaan dan transparansinya. Sebuah decision tree diwakili secara grafis sebagai struktur hierarkis. Sebuah decision tree sebagai klasifikasi dinyatakan sebagai pihak rekursif ruang input berdasarkan nilai atribut. Pemisahan file akan dipisahkan oleh setiap node internal yang memisahkan file menjadi dua atau lebih sub-spasi sesuai dengan fungsi tertentu dari nilai atribut input. Sebuah kelas mewakili nilai yang paling sesuai atau target, yang node daun akan sering mewakili. Contoh diklasifikasikan dengan melintasi pohon dari simpul akar ke daun sesuai dengan hasil node uji (Masha et al., 2012).

Parameter untuk decision tree sama seperti random forest, seperti kriteria, kedalaman maksimum, leaf_size dan C. Kriteria adalah nilai yang digunakan untuk memproses model. Estimator angka digunakan untuk membuat node daun dari node induk, dan kedalaman maksimum adalah jumlah batas instans di setiap node (Bentejaca et al., 2006). C adalah parameter yang tidak intuitif (Amami et al., 2013).

F. Regresi Logistik

Regresi logistik adalah teknik pembelajaran yang diawasi yang dipinjam oleh pembelajaran mesin dari bidang statistik. Output atau variabel target Regresi Logistik adalah variabel kategoris, tidak seperti Regresi Linier, dan merupakan algoritma klasifikasi biner yang mengkategorikan titik data ke salah satu kelas data. Persamaan umum Regresi Logistik adalah (Hosmer Jr et al., 2013):

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

Penyetelan parameter regresi logistik adalah nilai C, intersep, penalti dan tol.

2.1.6 Pengukuran Performansi

Akurasi dalam masalah klasifikasi dapat dihitung dari jumlah prediksi yang benar yang dibuat oleh model. Akurasi adalah ukuran yang baik ketika kelas variabel target dalam data hampir sama.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$$

di mana TP, TN dan FP, FN mewakili nilai Positif, Benar Negatif dan Positif Palsu dan Negatif Palsu, masing-masing.

Presisi adalah rasio True Positive terhadap jumlah True Positives dan False Positives. Contoh: Presisi mengukur proporsi pasien yang didiagnosis sebagai orang mati yang masih hidup. Prediksi optimis (Diprediksi meninggal karena pasien kanker paru-paru adalah TP dan FP) dan orang yang meninggal adalah TP.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Sensitivitas, juga dikenal sebagai recall. Sensitivitas mengukur tingkat pengamatan positif yang dilakukan dari apa pun yang secara akurat diprediksi sebagai positif. Penarikan adalah ukuran yang mewakili proporsi pasien yang meninggal yang didiagnosis oleh algoritma sebagai mati.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Akses presisi ke fraksi yang tepat dari kasus yang diklasifikasikan sebagai positif.

Rumus berikut menghitung presisi: F1-score adalah Harmonic Mean antara presisi dan recall. Kisaran untuk skor F1 adalah [0, 1]. Skor F1 menunjukkan seberapa akurat pengklasifikasi dan seberapa kuat pengklasifikasi pada saat yang sama.

$$\text{F - Score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

BAB III

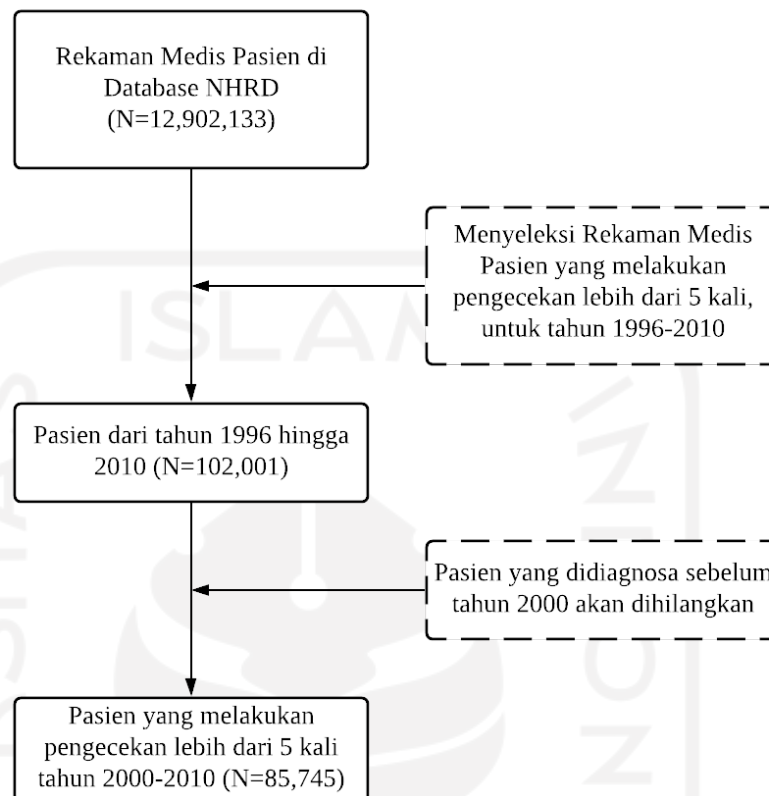
METODE PENELITIAN

3.1 Sumber Data

Dataset yang digunakan berasal dari National Health Insurance Research Database (NHIRD), yang mewakili 23 juta catatan medis di Taiwan yang berdasarkan jumlah pasien kanker paru-paru diidentifikasi dari tahun 1996 hingga 2010 *Ambulatory Care Expenditures by Visits* (CD) file, *Inpatient Expenditures by Admissions* (DD) file. Studi ini juga mempertimbangkan rincian *file Abulatory Care Orders* (OO) untuk menentukan jenis perawatan, Rincian perintah rawat inap (DO) dan *file Registry for Beneficiaries* (ID) untuk mendapatkan medis terperinci untuk melakukan penelitian. Penelitian ini digunakan pasien dari tahun 2000 hingga 2010. TRAN_CODE kolom file DD digunakan untuk membantu menentukan situasi kematian pasien yang ditandai dengan nomor 4 (meninggal), 9 (bunuh diri) dan A (rumah sakit kiri dalam kondisi kritis).

Penyakit ini diidentifikasi dan dibedakan oleh kode diagnostik ICD-9-CM. Ada tiga kode diagnostik ICD-9-CM pada file kunjungan rawat jalan yang membantu mendiagnosis penyakit untuk setiap kunjungan yang dapat diidentifikasi. Sementara itu, file CD dan DD menjelaskan penyediaan biaya medis untuk setiap kunjungan rawat inap dan rawat jalan, termasuk berbagai prosedur diagnostik dan perawatan yang ditentukan atau layanan medis lainnya. Berdasarkan informasi biaya medis ini, perkiraan biaya yang terkait dengan objek bunga dapat dihitung.

Pengolahan data merupakan langkah penting yang perlu dilakukan untuk mendapatkan hasil prediksi yang baik. Pemrosesan data untuk memilih *dataset* yang akan digunakan ditampilkan pada Gambar 3.1 Setelah menggunakan kode diagnostik ICD-9-CM untuk membedakan kanker paru-paru dan penyakit kronis, pilih pasien dalam *database* yang telah mengunjungi di atas lima kali dari tahun 1996 hingga 2010 memiliki 102.001 kasus. Kedua, pasien yang didiagnosis sebelum tahun 2000 akan dikecualikan. Jadi, data yang digunakan adalah 85.745 pasien.



Gambar 3.1 Pemrosesan Pengambilan Data Sampling

3.2 Faktor-Faktor

Variabel yang sesuai diperlukan dalam membuat prediksi. Dari tinjauan literatur sebelumnya, telah diperoleh beberapa variabel yang akan digunakan. Kondisi ini adalah perhatian utama terhadap komorbiditas yang ada atau penyakit kronis. Penyakit kronis adalah kontributor sebagian besar beban medis. Dua puluh tujuh penyakit kronis digunakan untuk mempertimbangkan variabel (Kementerian Kesehatan dan Kesejahteraan, 2015).

Dalam menganalisis data kelangsungan hidup, dua fungsi yang berkaitan dengan waktu dan perhatian khusus terhadap kelangsungan hidup dan fungsi bahaya. Fungsi kelangsungan hidup $S(t)$ didefinisikan sebagai probabilitas bertahan hidup setidaknya sampai waktu t . Fungsi bahaya $h(t)$ adalah probabilitas kematian bersyarat pada saat itu yang bertahan hingga saat itu. Peneliti menggunakan data

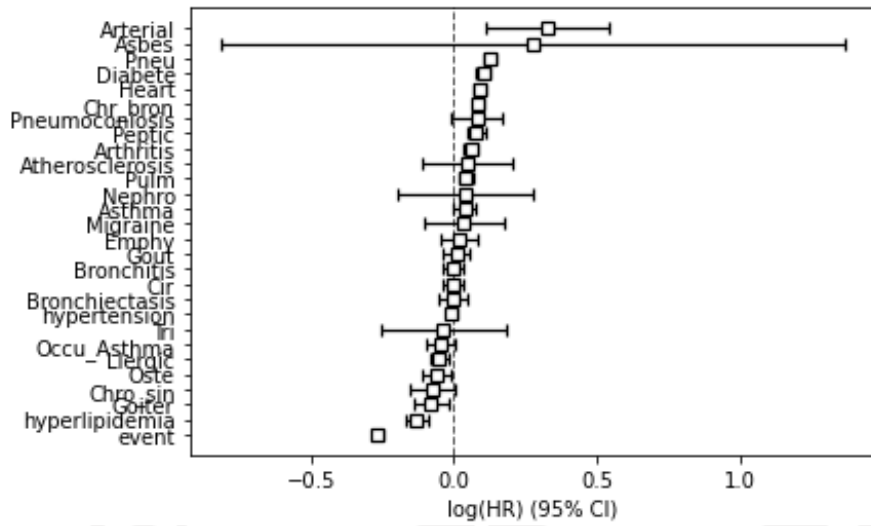
dari tahun (2000-2010). Dua puluh tujuh penyakit kronis telah diuji menggunakan *Cox Proportional Hazard* untuk menentukan faktor resiko yang paling berpengaruh dengan kemampuan bertahan hidup dari setiap pasien. pengeluaran medis dan kemampuan bertahan hidup. Metode ini bertujuan untuk menunjukkan bahwa tingkat risiko bervariasi dalam dimensi waktu. Parameter yang digunakan dalam menentukan penyakit kronis primer adalah nilai bahaya ($\exp(\text{coef})$) untuk pengeluaran dan survivabilitas.

Tabel 3.1 menunjukkan daftar penyakit kronis yang dianggap memiliki efek yang mungkin pada pasien kanker paru-paru berdasarkan kemampuan bertahan hidup dan pengeluaran medis. Variabel prediktor ini dipilih berdasarkan nilai rasio bahaya yang dihasilkan dalam Model Proporsional Cox, di mana $HR > 1$. Dari tabel, untuk tahap I pengeluaran, setiap variabel memiliki koararitas tinggi, dan tidak ada hasil spesifik yang dapat diperoleh. Tetapi untuk tahap II dan semua tahap, ada hasil spesifik untuk variabel prediktor yang memiliki rasio bahaya terhadap hasilnya, baik untuk survivabilitas atau pengeluaran. Gambar 3.2 dan 3.3 menunjukkan hasil rasio bahaya berdasarkan diagram plot. Misalkan posisi persegi dari variabel berada di sisi kanan. Dalam hal ini, itu berarti variabel memberikan koefisien positif dan memiliki dampak risiko pada kemampuan bertahan hidup atau pengeluaran.

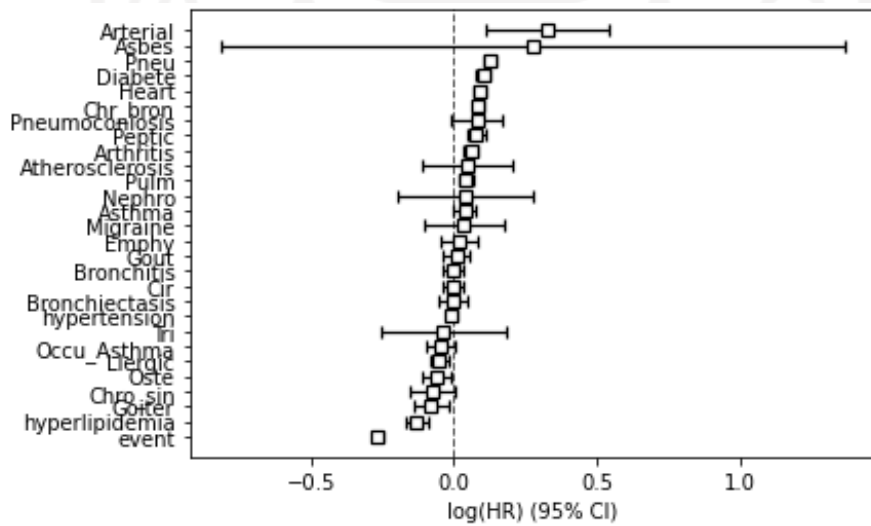
Tabel 3.1 Risk Factor Terpilih

Komorbiditas	(2000-2010)	
	Survivabilitas	Biaya Pengeluaran
Trigeminal neuralgia		
Heart disease	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Bronchitis		
Bronchiectasis		
Goiter		
Asbestosis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Cirrhosis		

Komorbiditas	(2000-2010)	
	Survivabilitas	Biaya Pengeluaran
Pneumonia	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Emphysema	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Pulmonary tuberculosis		<input checked="" type="checkbox"/>
Asthma	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Peptic ulcer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Osteoporosis		
hyperlipidemia		
hypertension		
Migraine		<input checked="" type="checkbox"/>
Arterial embolism and thrombosis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Atherosclerosis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Gout	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Nephrotic syndrome	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Allergic rhinitis		
Pneumoconiosis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Chronic_bronchitis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Chronic_sinusitis		
Diabetes mellitus	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Occurence_Asthma		
Arthritis	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Total Variable	14	16



Gambar 3.2 Cox PH Plot Keberlangsungan Hidup



Gambar 3.3 Cox PH Plot Biaya Pengeluaran

Dua belas faktor risiko utama tambahan secara signifikan mempengaruhi pasien kanker paru-paru untuk bertahan hidup dan pengeluaran medis, seperti yang ditunjukkan pada Tabel 3.2. Variabel penyakit kronis yang dipilih dan 10 variabel kunci yang ada akan digunakan nanti dalam model pembelajaran mesin.

Tabel 3.2 Variabel Utama Pemodelan

Faktor (X)	Deskripsi	Tipe data (YZ)
Usia (A)	Usia pasien ketika didiagnosis dengan kanker paru-paru	Continuous (Z)
Gender (G)	Laki-laki atau perempuan	Binary (Y)
Situs (S)	Delapan jenis lokasi kanker paru-paru, termasuk trakea, bronkus dan paru-paru (162), trakea (1620), bronkus (1622), lobus atas, bronkus atau paru-paru (1623), lobus tengah, bronkus atau paru-paru (1624), lobus bawah, bronkus atau paru-paru (1625), bagian lain dari bronkus atau paru-paru (1628), dan bronkus dan paru-paru, tidak ditentukan (1629)	Discrete (Y)
Pengobatan (T)	Pengobatan dibagi menjadi lima jenis: hanya menerima kemoterapi, menerima operasi, menerima kemoterapi sebelum operasi, menerima operasi sebelum kemoterapi, dan belum menerima operasi dan kemoterapi yang ditentukan.	Binary (Y)
Rekor kunjungan pertama (FV)	Catatan medis pasien kanker paru-paru pada tahun pertama.	Continuous (Z)
Rekor kunjungan terakhir (LV)	Catatan medis pasien kanker paru-paru pada tahun lalu.	Continuous (Z)
Pengeluaran rawat jalan (OE)	Pengeluaran Perawatan Rawat Jalan menurut file Kunjungan.	Continuous (Z)
Pengeluaran rawat inap (IE)	Pengeluaran Perawatan Rawat Jalan sesuai dengan file Penerimaan.	Continuous (Z)
Pengeluaran medis	Pengeluaran medis secara keseluruhan untuk tahun-tahun keterlacakan masa depan pribadi.	Continuous (Z)
Peristiwa	kondisi kehidupan pasien.	Binary

Faktor (X)	Deskripsi	Tipe data (YZ)
		(Y)

3.3 Definisi Pemilihan Tahun

Penggunaan obat untuk menghancurkan sel kanker disebut terapi sistemik. Tergantung pada terapi sistemik kanker paru-paru. Obat disortir ke dalam kemoterapi, terapi yang ditargetkan dan imunoterapi. Obat-obatan dalam kemoterapi berfokus pada bagaimana obat dapat menghancurkan sel-sel kanker, biasanya menjaga sel-sel kanker dari tumbuh, membelah, dan membuat lebih banyak sel. Obat-obatan ini akan digunakan untuk menghancurkan gen, protein, atau lingkungan jaringan spesifik kanker yang berkontribusi terhadap pertumbuhan kanker dan kelangsungan hidup untuk terapi yang ditargetkan. Dan untuk imunoterapi, obat-obatan ini dirancang untuk meningkatkan kekebalan tubuh pasien dari pengobatan (Kanker Paru - Sel Non-Kecil: Jenis Pengobatan, 2020).

Gefitinib secara signifikan menghasilkan reseptor faktor pertumbuhan epidermis (EGFR) dari kelangsungan hidup yang berkepanjangan, dibandingkan dengan Cisplatin dikombinasikan dengan Docetaxel pada kemoterapi NCLC. Hasil Gefitinib untuk kelangsungan hidup bebas perkembangan sebelumnya memiliki waktu kelangsungan hidup rata-rata 6,3 bulan, dan yang terakhir adalah 9,2 bulan (Mitsudomi, 2010). Pengobatan Erlotinib secara signifikan lebih lama daripada kemoterapi yang juga menargetkan EGFR (Zhou et al., 2011). Dan juga ditemukan dalam pengobatan kemoterapi NSCLC dan SCLC, Vinorelbine dengan terapi doublet berbasis Cisplatin dapat secara efektif meningkatkan tingkat kelangsungan hidup hingga 5,4% selama lima tahun (Douillard et al., 2010). Seperti yang dijelaskan sebelumnya, obat yang efektif untuk pasien kanker paru-paru muncul setelah tahun 2003. Karena itu, penelitian ini menggunakan data dari tahun 2000-2010, yang dianggap sudah ada obat yang mutakhir dibandingkan sebelum tahun 2000.

Tabel 3.3 Pengobatan Kanker Paru-Paru

Pengobatan	Perawatan yang Ditargetkan	Obat	Tahun Pendaftaran
Kemoterapi	Tumor Kanker Paru-paru	Ciplastin	1984
		Carboplatin	1991
		Pemetrexed (Alimta)	1991
		Gemcitabine (Gemzar)	1996
		Paclitaxel (Taxol)	1996
		Docetaxel (Taxotere)	1998
		Vinorelbine (Navelbine)	2005
Terapi bertarget	Inhibitor reseptor faktor pertumbuhan epidermis (EGFR)	Gefitinib (Iressa)	2003
		Erlotinib (Tarceva)	2006
		Afatinib (Gilotrif)	2013
		Dacomitinib (Vizimpro)	2018
		Osimertinib (Tagrisso)	2018
	Inhibitor anaplastic lymphoma kinase (ALK).	Crizotinib (Xalkori)	2011
		Alectinib (Alecensa)	2014
		Ceritinib (Zykadia)	2014
		Lorlatinib (Lorbrena)	2018
		Brigatinib (Alunbrig)	2020
	Obat-obatan yang menargetkan perubahan genetik ROS1.	Crizotinib (Xalkori)	2011
		Entrectinib (Rozlytrek)	2019
	Obat-obatan yang menargetkan fusi NTRK.	Larotrectinib (Vitrakvi)	2018
	Obat-obatan yang menargetkan mutasi BRAF V600E	Dabrafenib (Tafinlar)	2013
		tremetinib (Mekinist)	2013
	Capmatinib (Tabrecta)	2009	

Pengobatan	Perawatan yang Ditargetkan	Obat	Tahun Pendaftaran
	Obat-obatan yang menargetkan MET Exon 14 Skipping	Tepotinib (Tepmetko)	2021
	Obat-obatan yang menargetkan fusi RET.	Bevacizumab (Avastin, Mvasi)	2004
	Terapi anti-angiogenesis	Selpercatinib (LOXO-292)	2020
		Ramucirumab (Cyramza)	2020
Imunoterapi	Pertahanan alami tubuh	Pembrolizumab (Keytruda)	2015
		Durvalumab (Imfinzi)	2018
		Cemiplimab (Libtayo)	2018
		Durvalumab (Imfinzi)	2018
		Nivolumab (Opdivo)	2020

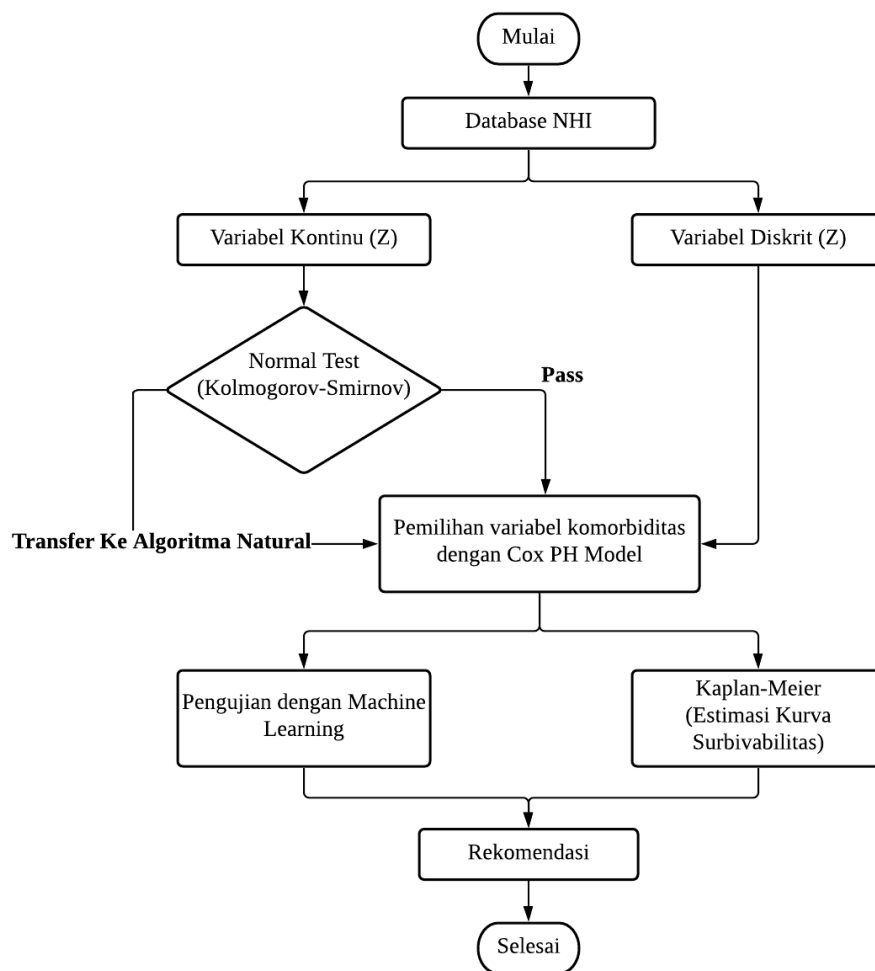
3.4 Prosedur Penelitian

Studi ini menggunakan *Machine Learning* untuk membangun model prediktif untuk kemampuan bertahan hidup yang ada. Berdasarkan prosedur Gambar 3.3, pertama, variabel yang dipilih diperlukan sebagai *database* untuk diekstraksi dari database. Kedua, menguji normalitas variabel kontinu. Variabel penyakit kronis akan dipilih dari data yang ada dengan menggunakan *Cox Proportional Hazard Model*. Setelah tahap data eksplorasi dilakukan, model akan dibangun menggunakan *Machine Learning*. Dan analisis survivabilitas akan dilakukan dengan metode *Kaplan-Meier* untuk memperkirakan kurva kelangsungan hidup.

Variabel target yang akan diprediksi adalah survivabilitas pasien kanker paru-paru berdasarkan data asuransi NHI yang ada. Metode klasifikasi yang digunakan adalah *Support Vector Machine*, *Random Forest*, *K-Nearest Neighbors*, *Naive Bayes*, *Decision Tree*, dan *Logistic Regression*.

Pada pengujian algoritma ini peneliti juga menggunakan *feature selection* untuk mengetahui variabel penting yang mempengaruhi biaya pengeluaran sekaligus tingkat survivabilitas yang ada setiap pasien kanker paru-paru. Harapannya informasi ini dapat mengurangi pengeluaran medis, meningkatkan penanganan dari institusi Kesehatan dan memberikan referensi penting dalam saran medis untuk pasien.

Untuk proses training dan testing, *ten-cross validation* akan digunakan untuk penerapan model algoritma. Adapun tujuan dari penggunaan *ten-cross validation* ini adalah untuk mengatasi permasalahan *overfitting* pada data yang ada serta bagaimana model dapat digeneralisais terhadap suatu set data yang independen.



Gambar 3.4 Prosedur Penelitian

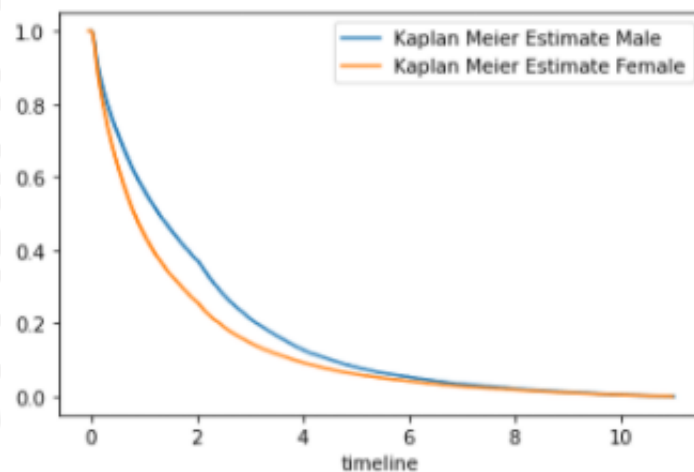
BAB IV

ANALISIS DATA

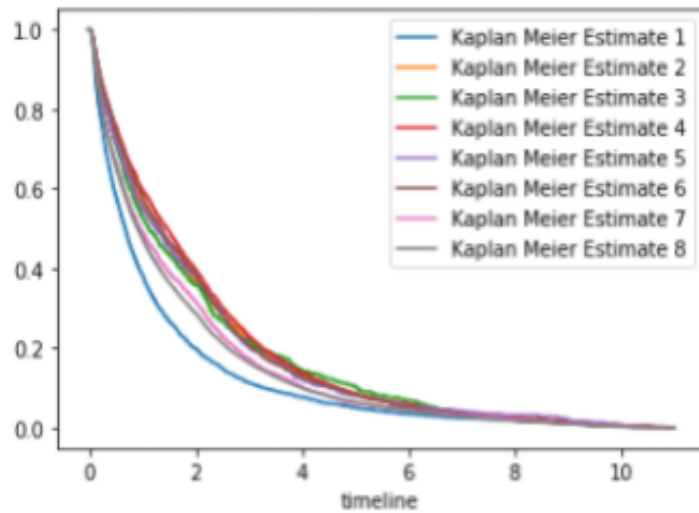
Bab ini menyajikan hasil eksperimen menggunakan algoritma yang diusulkan untuk mengevaluasi kinerjanya. Selain itu, terdapat hasil penerapan algoritma yang diusulkan terhadap data untuk segmentasi pelanggan.

4.1 Estimasi *Kaplan-Meier*

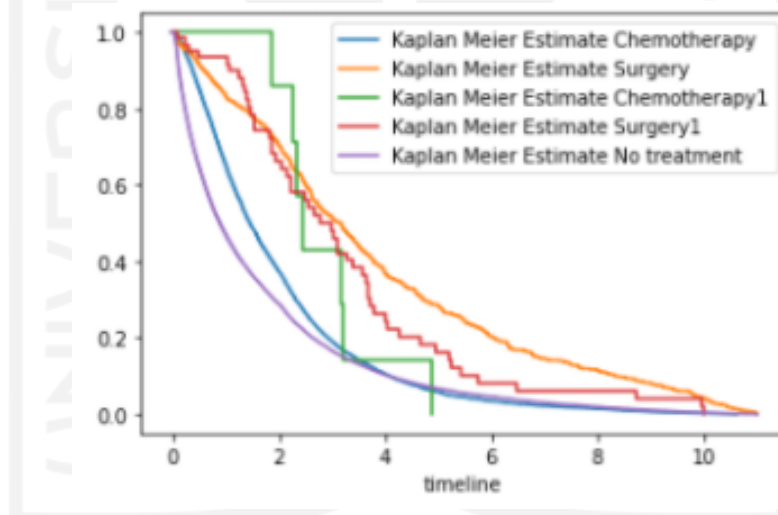
Untuk melengkapi analisis prediksi survivabilitas, maka dalam penelitian ini akan digunakan Estimasi *Kaplan-Meier*. *Kaplan-Meier* digunakan untuk melihat bagaimana kurva bertahan hidup. Analisis kelangsungan hidup pada penelitian ini akan dilakukan berdasarkan jenis kelamin pasien kanker paru-paru, lokasi kanker paru-paru pasien dan variabel prediktor sebelumnya, yaitu pengobatan, operasi dan kemoterapi.



Gambar 4.1 Survival Curve Berdasarkan Jenis Kelamin



Gambar 4.2 Survival Curve Berdasarkan Lokasi Kanker



Gambar 4.3 Survival Curve Berdasarkan Jenis Pengobatan

Dari hasil analisis menggunakan Kaplan-Meier, dapat diketahui bahwa tahun 2000-2010 memiliki kurva kelangsungan hidup lebih dari empat tahun. Berdasarkan jenis kelamin, laki-laki memiliki kemungkinan lebih tinggi untuk bertahan hidup, variabel No.3 pada jenis pengobatan atau menerima kemoterapi terlebih dahulu kemudian operasi memiliki kemungkinan tingkat survivabilitas yang lebih tinggi. Peneliti juga menemukan bahwa untuk lokasi kanker paru-paru No.5 atau bronkus atau lobus tengah, bronkus atau paru-paru memiliki dampak kemungkinan bertahan

hidup yang lebih tinggi daripada pasien yang memiliki lokasi kanker paru-paru pada bagian lain.

Tabel 4.1 Probabilitas Rata-Rata Survivabilitas

(2004-2010)		
Jenis Kelamin	Perempuan	0.178854
	Laki-laki	0.239499
Pengobatan	Kemoterapi	0.355339
	Bedah	0.567438
	Chemoterapy1	0.636364
	Operasi1	0.561781
	Tidak ada pengobatan	0.175643
Lokasi Kanker Paru-paru	Trakea, bronkus dan paru-paru	0.276837
	Trakea	0.35227
	Bronkus	0.483049
	Lobus atas, bronkus atau paru-paru	0.384181
	Lobus tengah, bronkus atau paru-paru	0.497958
	Lobus bawah, bronkus atau paru-paru	0.425166
	Bagian lain dari bronkus atau paru-paru	0.375552
	bronkus dan paru-paru, tidak ditentukan	0.183076

4.2 Analisis Biaya Pengeluaran

Untuk melengkapi analisis prediksi dan pemberian rekomendasi terhadap biaya pengeluaran, maka akan dilakukan analisis tren biaya medis. Analisis tren biaya

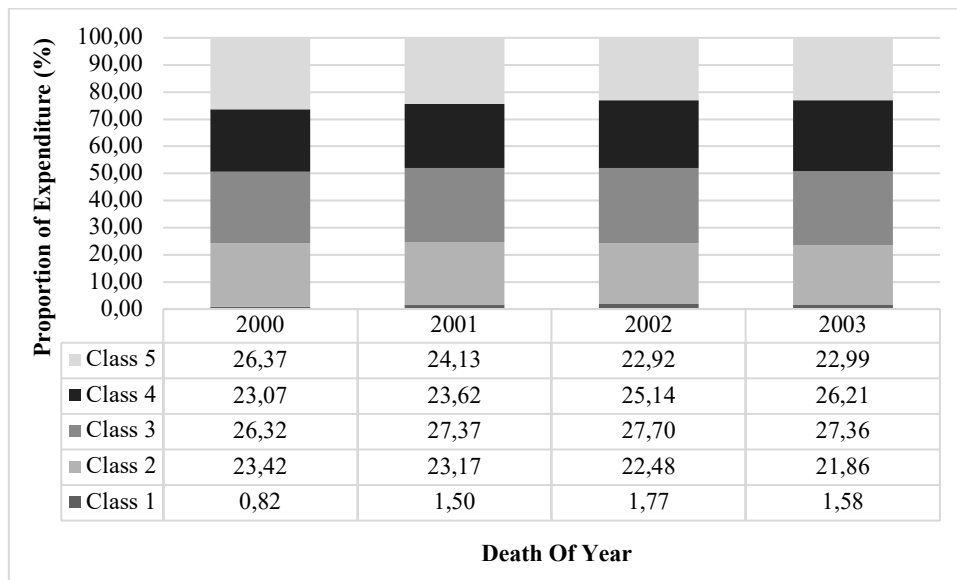
medis berdasarkan survivabilitas digunakan untuk melihat biaya rata-rata yang dikeluarkan oleh pasien kanker paru-paru berdasarkan kematian. Tujuan dari analisis tren di bagian ini adalah untuk melihat proporsi biaya yang paling signifikan yang dikeluarkan oleh pasien kanker paru-paru berdasarkan kelas biaya di setiap tahun. Pembagian kelas biaya ini juga digunakan dalam pengujian *Machine Learning*.

Tabel 4.2 Kelas Kisaran Pengeluaran Pasien

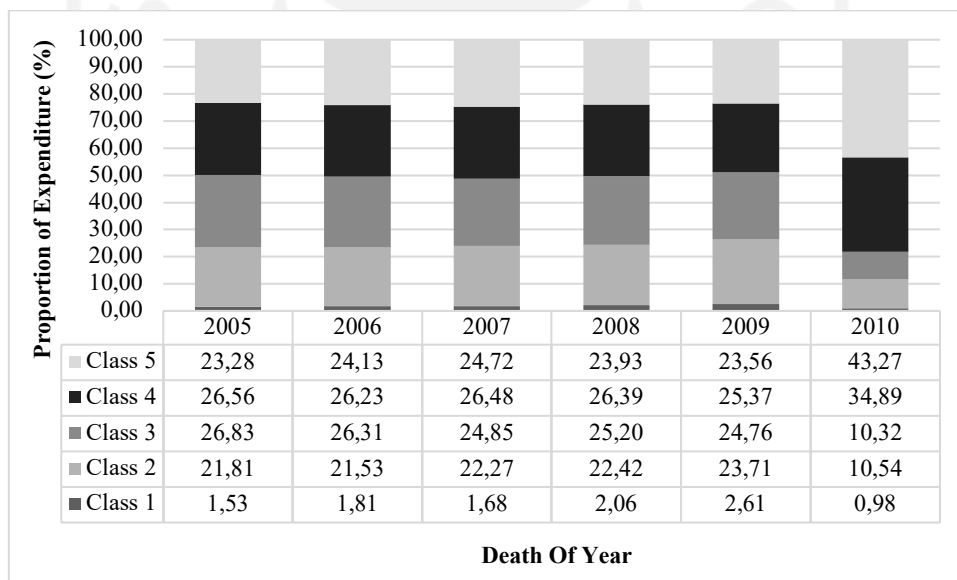
Kelas	Kisaran Pengeluaran (\$)
1	19964.2
2	19964.3-39591.4
3	39591.5-59218.6
4	59218.7-78845.8
5	>78845,9

Tabel 4.3 Rata-Rata Biaya Pengeluaran 2000-2010

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Kelas 1	14319 .8	13414 .7	13294 .4	13434 .6	13341 .7	12721 .3	13777 .2	13003 .3	13688 .8	12972 .6	12811. 1
Kelas 2	31037 .8	30849 .5	31322 .0	31213 .9	31036 .6	31018 .2	31100 .2	30799 .7	30975 .2	30631 .1	30936. 2
Kelas 3	49193 .9	49259 .7	49160 .4	49154 .6	49490 .4	49392 .4	49446 .2	49392 .2	49264 .5	49299 .3	49130. 4
Kelas 4	69033 .5	69058 .1	69085 .0	69046 .6	68979 .8	68783 .4	69145 .7	69108 .6	69093 .9	69199 .2	70124. 91
Kelas 5	88709 .0	88675 .7	88392 .3	88391 .4	88334 .8	88419 .9	88532 .8	88242 .7	88117 .2	88373 .4	88481. 61
Total Harga	59650 .3	58606 .6	58516 .0	58999 .5	59436 .7	59063 .1	59448 .4	59465 .8	58961 .9	58190 .0	71349. 8



Gambar 4.4 Proporsi Pengeluaran 2000-2003



Gambar 4.5 Proporsi Pengeluaran 2005-2010

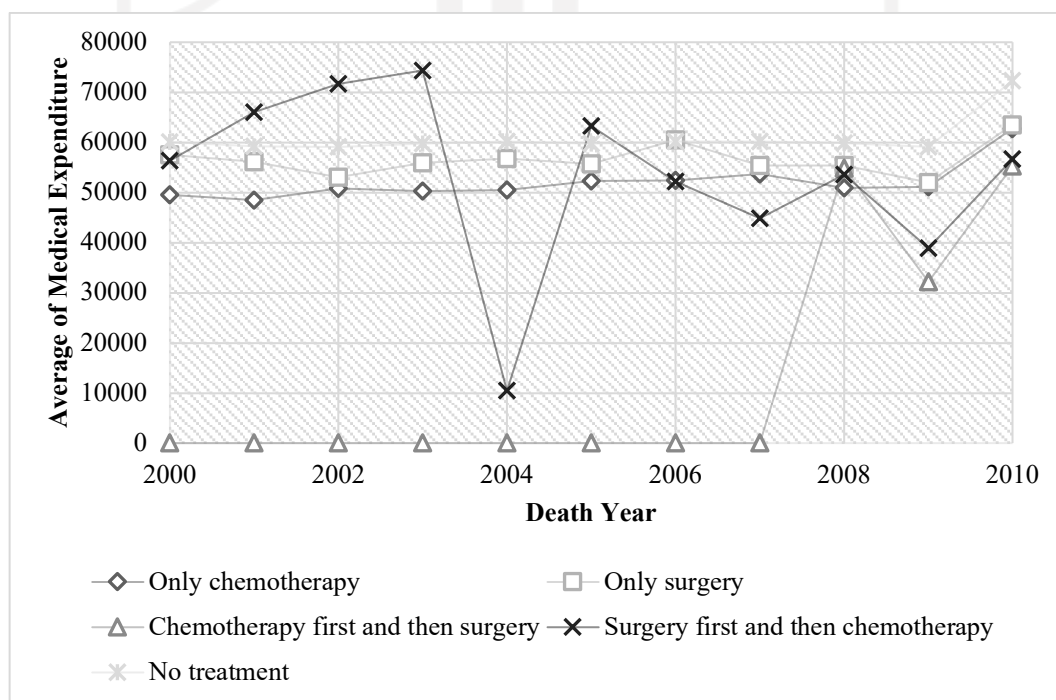
Berdasarkan Gambar 4.4 dan 4.5 dapat diketahui setiap kelas pengeluaran memiliki biaya rata-rata yang hampir sama dan tidak jauh berbeda. Tetapi untuk setiap tahun, dari tahun 2000-2010 memiliki proporsi kelas terbesar bervariasi. Pada tahun 2000, kelas 5 adalah biaya pengeluaran rata-rata tertinggi yang diperoleh pasien, yaitu 26,37% atau sebanyak 608 orang. Untuk tahun 2001, kelas 3 adalah proporsi tertinggi dengan 27,37% atau sebanyak 1328 orang. Dari total tahun, kelas 3 adalah

kelas dengan proporsi tertinggi setiap tahun di setiap tahap, yang berarti pasien kanker paru-paru membayar pengeluaran medis dalam kategori menengah pada umumnya.

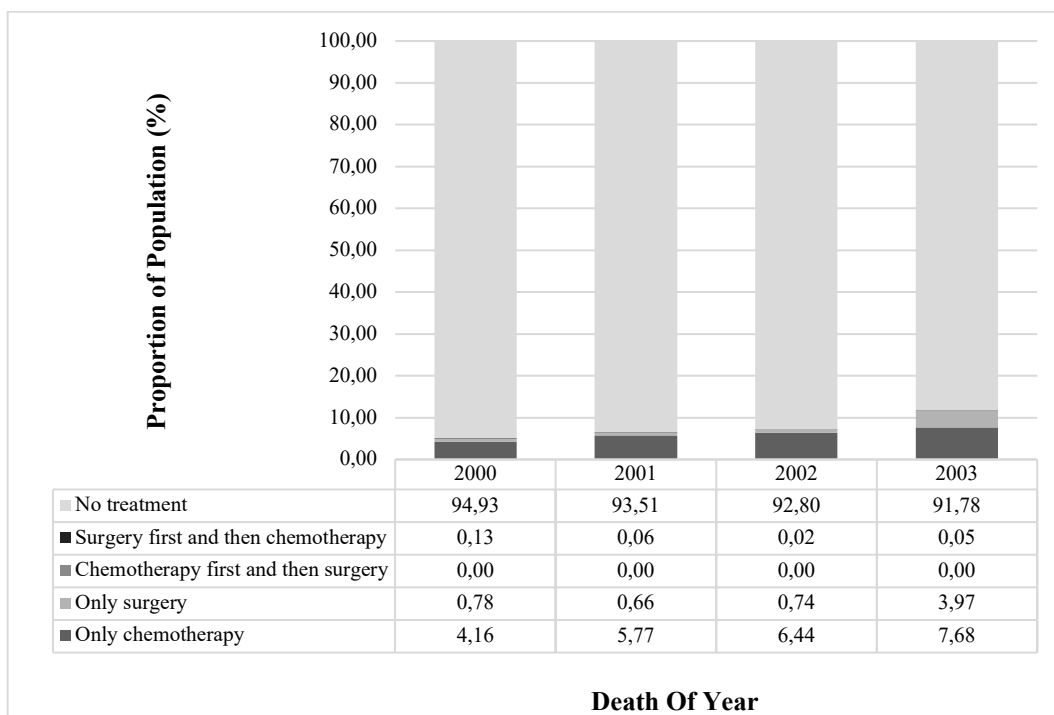
Biaya pengeluaran medis ini juga dapat dianalisis berdasarkan jenis pengobatan yang dilakukan oleh pasien kanker paru-paru. Setiap tahun, pengobatan untuk kanker paru-paru berdasarkan pengobatan khusus, operasi, dan kemoterapi memiliki arti yang paling signifikan untuk pembayaran NHI..

Tabel 4.4 Biaya Pengeluaran rata-rata berdasarkan Jenis Pengobatan

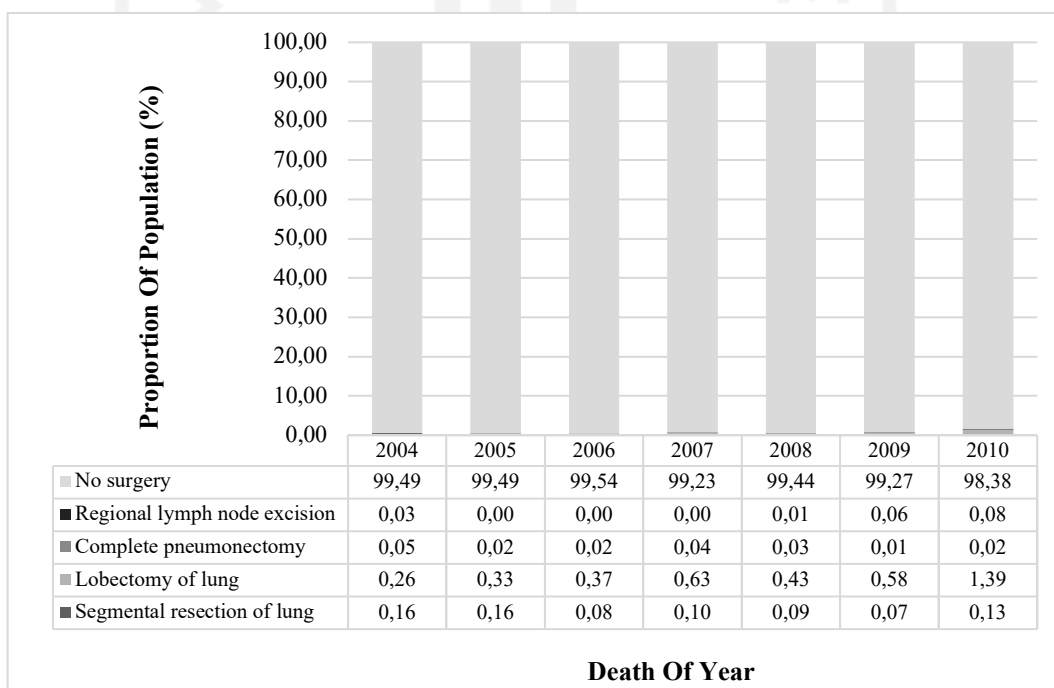
Jenis Pengobatan	2000-2010
Only chemotherapy	52074,8727
Only surgery	56536,5727
Chemotherapy first and then surgery	12942,0727
Surgery first and then chemotherapy	53495,1636
Tanpa Pengobatan	60892,8455



Gambar 4.6 Rata-Rata Biaya Pengobatan Berdasarkan Jenis Perawatan



Gambar 4.7 Proporsi Pasien Berdasarkan Jenis Perawatan (2000-2003)



Gambar 4.8 Proporsi Pasien Berdasarkan Jenis Perawatan (2004-2010)

Gambar 4.6 menunjukkan tren biaya pengeluaran berdasarkan data pasien kanker paru-paru pada tahun 2000-2010, yang berfokus pada Jenis pengobatan. Dapat dilihat bahwa dari tahun 2000 hingga 2007, tidak ada yang memilih untuk melakukan perawatan kemoterapi dan kemudian melakukan operasi. Pengeluaran rata-rata di setiap tahun berdasarkan jenis perawatan ditunjukkan pada tabel 4.4. Dapat diketahui bahwa pasien yang memilih tidak melakukan perawatan apa-apa merupakan jenis perawatan yang memiliki rata-rata biaya pengeluaran paling tinggi. Hal ini terjadi dikarenakan para pasien lebih memilih untuk tidak menjalani perawatan, yang dimana sesuai dengan gambar 4.7 dan 4.8 yang dimana proporsi pasien lebih dari 90% setiap tahunnya, pasien memilih untuk tidak menjalani perawatan.

4.3 Pengaturan Parameter

Sebelum memproses data dengan menggunakan *Machine Learning*, butuh adanya pengaturan parameter untuk mengurangi overfitting serta mendapatkan pengaturan parameter terbaik untuk memprediksi di masa depan. Untuk pengaturan parameter dalam *Machine Learning*, peneliti menggunakan *GridSearchCV*, yang digunakan untuk menemukan hiperparameter optimal untuk setiap algoritma *Machine Learning*, kecuali untuk *Recurrent Neural Network*. Hiperparameter *Recurrent Neural Network* dipilih berdasarkan akurasi hasil eksperimental, di mana para peneliti secara manual menentukan hasil eksperimen. Hiperparameter optimal untuk setiap algoritma adalah:

Tabel 4.5 Pengaturan Parameter

Algoritma		Survivabilitas		Pengeluaran	
		F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS
SVM	Kernel	RBF	RBF	RBF	RBF
	Gamma	0.001	0.01	0.01	0.001
	C	25	10	10	25
	Kriteria	Gini	Gini	Entropy	Entropy

Algoritma		Survivabilitas		Pengeluaran	
		F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS
Random Forest	Max_depth	4	4	8	8
	Max_features	Auto	Auto	Auto	Auto
	n_estimators	200	200	500	500
KNN	Algoritma	ball_tree	auto	Auto	Auto
	Leaf_size	1	1	1	1
	n_jobs	-1	-1	-2	-1
	n_neighbors	5	5	8	6
	Berat	Distance	Uniform	Uniform	Distance
Naïve Bayes	var_smoothing	0.005336	0.002310	6.579E-09	1.233E-06
Decision Tree	Kriteria	Gini	Gini	Gini	Gini
	max_depth	1	1	9	4
	Leaf_size	1	1	5	1
Logistic Regression	C	1	1	1	1
	Adang	1	1	1	1
	hukuman	12	12	12	12
	tol	0.001	0.001	0.0001	0.0001
RNN	Pengaktifan	Sigmoid	Sigmoid	Softmax	Softmax
	Lapisan padat	1	1	5	5
	Lapisan tersembunyi	4	4	4	4
	zaman	60	30	60	30
	batch_size	32	32	32	32
	Optimizer	Adam	Adam	Adam	Adam

Dari hasil tabel berikut dapat dilihat beberapa parameter yang membantu pengujian *Machine Learning*. Seperti pada bagian pengujian survivabilitas saat menggunakan F.Seleksi didapatkan hasil parameter terbaik untuk metode SVM adalah kernel “RBF”, dengan nilai gamma 0,001 dan nilai C sebesar 25.

4.4 Hasil Eksperimen

Dari hasil pengujian dengan menggunakan *Machine Learning*, dapat diketahui bahwa *Decision Tree* adalah algoritma optimal untuk memprediksi survivabilitas dan pengeluaran medis pada pasien kanker paru-paru yang meninggal pada tahun 2000-2010. Hasil ini dapat dilihat dari nilai survivabilitas akurasi 100% diikuti oleh nilai presisi prediksi, nilai recall dari rasio semua kasus positif yang diidentifikasi dengan benar dibagi dengan semua kasus positif dan skor F1 yang diperoleh dari rata-rata tertimbang presisi dan daya ingat juga sama dengan 100%. Untuk model biaya pengeluaran medis, *Decision Tree* tanpa *feature selection* memiliki akurasi tertinggi 99,9767% dibandingkan dengan algoritma lain, diikuti oleh presisi dengan 99,8319, recall dengan 99,9762% dan F1-skor dengan 99,9037% dan nilai waktu komputasi selama 0,26075.

Tabel 4.6 Hasil Machine Learning

Stage		Accuracy		Precision		Recall		F1-Score		Computational Time	
		F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS
SVM	Survival	80.9657	100	100	100	80.9657	100	89.4818	100	0.20445	0.265561
	Expenditure	79.4028	99.6968	66.2376	100	67.8162	99.6968	64.7102	99.6968	1238.39	708.008
Random Forest	Survival	100	100	100	100	100	100	100	100	0.237365	0.24634
	Expenditure	96.1161	99.9416	91.4690	99.6512	96.7166	99.9443	93.6969	99.7964	5.09735	3.86148
KNN	Survival	100	99.9883	100	99.9856	100	100	100	99.9928	1.8729914	3.46174
	Expenditure	85.7126	99.9416	89.2404	99.8006	85.7833	99.7979	86.9159	99.7992	1.61553	3.31598
Naïve Bayes	Survival	99.895	100	100	100	99.8705	100	99.9352	100	2.80944	0.14162
	Expenditure	73.2680	94.8098	62.6952	93.9282	64.0893	94.6658	62.5331	94.2755	2.94631	1.26185
Decision Tree	Survival	100	100	100	100	100	100	100	100	0.1615676	0.140624
	Expenditure	96.2211	99.9767	94.9425	99.8319	95.7105	99.9762	95.3094	99.9037	4.96647	0.26075

Stage		Accuracy		Precision		Recall		F1-Score		Computational Time	
		F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS	F. Seleksi	Tanpa FS
Logistic Regression	Survival	77.6766	77.6300	95.9377	95.8801	80.3183	80.3088	87.4359	87.4064	0.171834	0.126661
	Expenditure	67.9846	90.2962	55.6264	74.5525	54.6525	74.8903	54.8166	74.4762	4.08527	5.01446

Pada model yang menggunakan *Feature Selction* dapat diketahui beberapa variabel penting yang mempengaruhi survivabilitas dan biaya pengeluaran yang dikeluarkan oleh pasien kanker paru-paru di Taiwan. Adapun variabel yang ditemukan adalah:

1. Pada model experimental survivabilitas ditemukan bahwa *event* (Peristiwa), *gender* (jenis kelamin), *treatment* (jenis pengobatan), *lung_type* (lokasi kanker), *new_dead_year* (tahun kematian/kunjungan terakhir), *get_lung_year* (tahun kunjungan/ diagnosa pertama), umur dan penyakit kronis bawaan seperti *heart* (jantung), pneumonia, *emphasis*, diabetes, bronkitis kronis.
2. Pada model experimental untuk biaya pengeluaran ditemukan survivabilitas ditemukan bahwa pengeluaran medis, pengeluaran rawat jalan, pengeluaran rawat inap, *gender* (jenis kelamin), *treatment* (jenis pengobatan), *lung_type* (lokasi kanker), *new_dead_year* (tahun kematian/kunjungan terakhir), *get_lung_year* (tahun kunjungan/ diagnose pertama), umur dan penyakit kronis bawaan seperti pulmonaris, arthritis dan pneumonia.

BAB V PEMBAHASAN

Fokus dari penelitian ini adalah untuk meningkatkan kelangsungan hidup dan keseimbangan biaya pada pasien kanker paru-paru. Hal ini diperlukan untuk memprediksi harapan hidup yang terkait dengan biaya. Informasi tentang harapan hidup dan biaya akan membantu rumah sakit atau dokter memberikan pengobatan efektif terbaik untuk pasien kanker paru-paru. Jika tingkat kematian dan morbiditas memiliki tingkat tertinggi, itu akan menjadi beban keuangan yang besar bagi pasien, keluarga, dan pemerintah.

5.1 Machine Learning

Model yang dibuat untuk mendukung fokus penelitian ini diolah menggunakan *machine learning* dengan data yang terdapat pada tahun 2000-2010. Pemilihan tahun ini berdasarkan dari efektifitas jenis obat yang ditemukan setelah tahun 2000. Sebelum mengevaluasi kinerja pembelajaran mesin, *Cox Proportional Hazard Model* akan dilakukan untuk menentukan penyakit kronis yang akan diuji bersama dengan atribut utama lainnya. Penetapan ini akan dilakukan berdasarkan tingkat rasio bahaya terhadap hasil yang diprediksi, yaitu keberlangsungan hidup atau survivabilitas dari pasien dan biaya pengobatan yang dikeluarkan oleh pasien. Dari hasil penerapan *Cox Proportional Hazard*, Empat belas penyakit kronis akan digunakan untuk bertahan hidup, dan 16 penyakit kronis akan diggunakan pada biaya pengeluaran. Penyakit kronis digunakan sebagai tes atribut dengan atribut Umur (A), Jenis Kelamin (G), Lokasi kanker paru-paru (S), Kemoterapi (Chemo), Operasi (Sur), Jenis Pengobatan (T), *First-visit record* (FV), *Last Visit record* (LV), Pengeluaran Rawat Jalan (OE), pengeluaran rawat inap (IE), pengeluaran medis (ME).

Dalam pemrosesan *Machine Learning*, *Cross Validation* dan *Feature Selection* diggunakan untuk menghindari overfitting setiap algoritma Machine Learning. *Feature selection* juga membantu dalam mengidentifikasi variabel yang memiliki pengaruh penting dalam pengujian model eksperimental baik dari segi

survivabilitas dan biaya pengeluaran. Dari hasil pengujian terhadap data Kanker Paru-Paru pada tahun 2000-2010 dapat disimpulkan bahwa algoritma *Decision Tree* memiliki kinerja optimal yang dapat digunakan dalam melakukan pengujian survivabilitas dan pengeluaran medis pada pasien kanker paru-paru. *Decision Tree* menunjukkan hasilnya tidak pernah di bawah 90%, baik saat menggunakan pemilihan fitur atau tanpa pemilihan fitur. Hasil optimal di sini tidak berarti bahwa itu akan optimal jika digunakan dalam dataset lain. Peraturan parameter adalah salah satu cara untuk mengurangi *overfitting model* yang dijalankan dan mendapatkan hasil yang maksimal dalam pemodelan. Jenis dataset yang digunakan, baik dari data, nilai, atau variabel prediktor, juga akan mempengaruhi mendapatkan hasil pemodelan yang optimal. Jadi, lebih baik jika ada tes seperti ini untuk fokus pada pengaturan *parameter tuning* yang digunakan.

5.2 Analisis Survivabilitas

Analisis tren untuk kemungkinan pasien kanker paru-paru juga diperlukan. Analisis kelangsungan hidup berdasarkan faktor risiko juga membantu menentukan pengobatan, operasi, dan kemoterapi terbaik untuk memberi pasien umur panjang.

Dari hasil *Kaplan-Meier* sebelumnya, untuk analisis kelangsungan hidup. Dari hasil analisis menggunakan *Kaplan-Meier*, dapat diketahui bahwa tahun 2000-2010 memiliki kurva kelangsungan hidup lebih dari empat tahun. Berdasarkan jenis kelamin, laki-laki memiliki kemungkinan lebih tinggi untuk bertahan hidup, variabel No.3 pada jenis pengobatan atau menerima kemoterapi terlebih dahulu kemudian operasi memiliki kemungkinan yang lebih tinggi. Untuk lokasi kanker paru-paru No.5 atau bronkus atau lobus tengah, bronkus atau paru-paru memiliki dampak kemungkinan bertahan hidup yang lebih tinggi pada pasien. Untuk jenis pengobatan yang mungkin membantu pasien kanker paru-paru memiliki keberlangsungan hidup lebih lama adalah dengan melakukan bedah pada eksisi kelenjar getah bening regional dan rutin melakukan kemoterapi.

Tabel 5.1 Hasil Analisis Tren Survivabilitas

Tahun	Kaplan Meier					
	Survivabilitas	Jenis kelamin	Pengobatan	Lokasi Kanker Paru-paru	Bedah	Kemoterapi
2000-2010	> 4 Tahun	Laki-laki	No.3 atau menerima kemoterapi terlebih dahulu, kemudian operasi	Bronkus atau lobus tengah, bronkus atau paru-paru	Eksisi kelenjar getah bening regional.	Melakukan Kemoterapi

Hasil yang ditampilkan pada tabel 5.1 didapatkan berdasarkan nilai probabilitas dari hasil pengujian estimasi *Kaplan-Meier*. Pada pemberian rekomendasi survivabilitas peneliti juga melakukan analisis berdasarkan *Feature Selection* pada saat pengujian *Machine Learning*. Pada model experimental survivabilitas ditemukan bahwa *event* (Peristiwa), *gender* (jenis kelamin), *treatment* (jenis pengobatan), *lung_type* (lokasi kanker), *new_dead_year* (tahun kematian/kunjungan terakhir), *get_lung_year* (tahun kunjungan/ diagnosa pertama), umur dan penyakit kronis bawaan seperti *heart* (jantung), pneumonia, emphasis, diabetes, bronkitis kronis.

Hasil analisis yang didapatkan dari penelitian ini akan sangat menguntungkan bagi pihak medis. Pihak medis di Taiwan dapat dengan mudah melakukan perawalan atau prediksi survivabilitas dari seorang pasien berdasarkan berbagai variabel dari yang dihasilkan oleh *Feature Selection*. Pihak medis juga dapat mengusulkan jenis pengobatan no.3 yaitu menerima kemoterapi dan kemudian melakukan operasi, untuk bisa meningkatkan survivabilitas dari pasien kanker paru-paru di Taiwan. Analisis ini juga dapat menguntungkan dalam efisiensi waktu dan alokasi tenaga kerja. Dengan hasil prediksi ini harapannya pihak medis dapat dengan mudah melakukan perencanaan dalam proses pengobatan pasien kanker paru-paru di Taiwan. Rekomendasi yang dapat diberikan kepada sektor medis adalah dapat menggunakan *Decision Tree* dalam melakukan prediksi dan juga dapat mengusulkan jenis pengobatan no.3 yaitu menerima kemoterapi dan kemudian melakukan operasi kepada pasien kanker paru-paru.

5.3 Analisis Biaya Pengeluaran

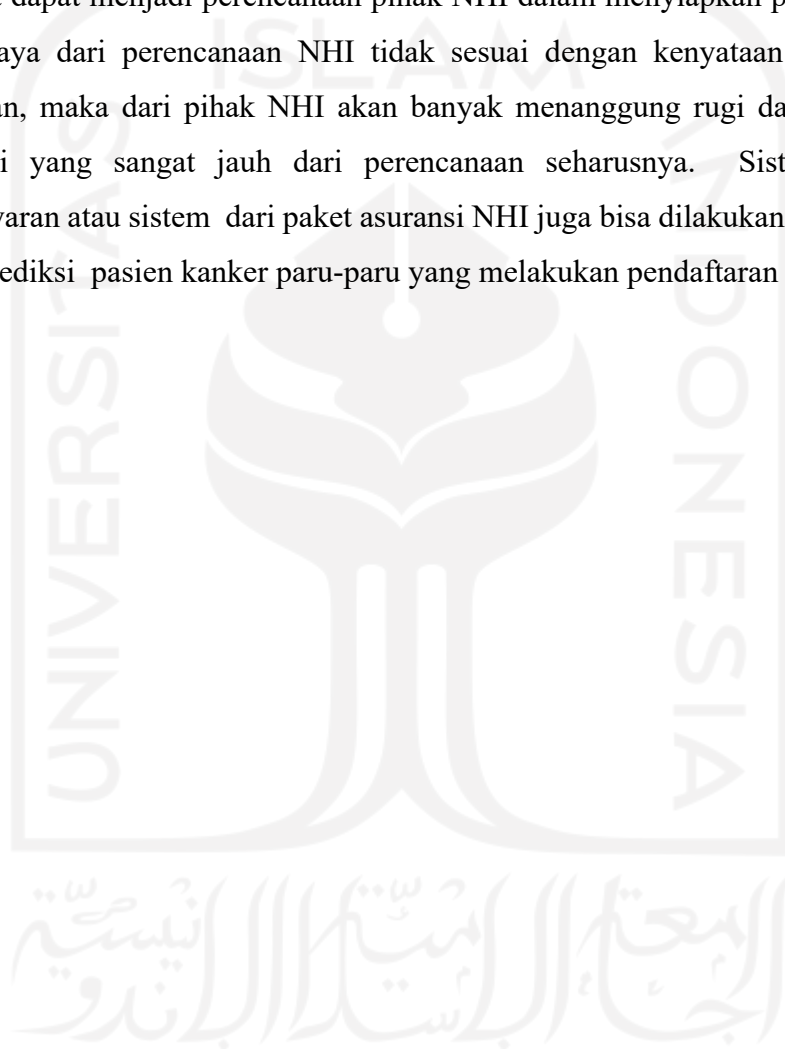
Karena biaya perawatan kesehatan terus meningkat, memahami tren biaya dalam perawatan kesehatan dan mengidentifikasi faktor-faktor yang berkontribusi terhadap peningkatan biaya perawatan akan diperlukan untuk merencanakan biaya perawatan kesehatan di masa depan dan memprioritaskan dan mengalokasikan sumber daya medis. Analisis tren biaya pengeluaran berdasarkan jenis pengobatan mengungkapkan peningkatan yang signifikan dalam proporsi pasien yang tidak menjalani perawatan apa pun. Peningkatan jumlah ini mungkin disebabkan oleh akumulasi data NHI tentang respons kanker paru-paru pasien yang memilih untuk tidak menjalani perawatan selama masa perawatan mereka. Hal ini juga mungkin dipengaruhi oleh pasien kanker paru-paru berdasarkan rekomendasi dokter, yang tidak tercantum pada data yang disediakan.

Proporsi biaya pengeluaran paling banyak setiap tahun berada pada kategori kelas 3 dengan kisaran biaya perawatan hingga kematian 39591,5-59218,6 di unit \$. Dapat dilihat bahwa untuk kelas, biaya yang harus dibayar oleh NHI masih dalam kategori menengah. Pada pemberian rekomendasi survivabilitas peneliti juga melakukan analisis berdasarkan *Feature Selection* pada saat pengujian *Machine Learning*. Pada model *experimental* untuk biaya pengeluaran ditemukan survivabilitas ditemukan bahwa pengeluaran medis, pengeluaran rawat jalan, pengeluaran rawat inap, *gender* (jenis kelamin), *treatment* (jenis pengobatan), *lung_type* (lokasi kanker), *new_dead_year* (tahun kematian/kunjungan terakhir), *get_lung_year* (tahun kunjungan/ diagnose pertama), umur dan penyakit kronis bawaan seperti pulmonaris, arthritis dan pneumonia.

Hasil yang diperoleh dari pengujian model prediksi dan analisis tren yang dilakukan dalam penelitian ini tidak hanya untuk membantu Asuransi Kesehatan Nasional. Tetapi data juga membantu perusahaan asuransi dalam memberikan penawaran untuk produk yang mereka miliki. Dalam asuransi jiwa, model prediktif menggunakan algoritma pembelajaran mesin memberikan hasil yang lebih mencolok dan tepat daripada metode tradisional untuk mendukung bisnis. Oleh karena itu, akan meningkatkan bisnis dengan memungkinkan layanan yang lebih

cepat kepada pelanggan dan perencanaan yang baik bagi perusahaan asuransi untuk meningkatkan kepuasan dan loyalitas.

Pihak NHI dapat memberikan rekomendasi paket asuransi *berdasarkan budget analysis* yang sekiranya akan dikeluarkan oleh asuransi untuk membiayai pengobatan dari pasien kanker paru-paru baru di Taiwan. Hasil analisis dan prediksi ini juga dapat menjadi perencanaan pihak NHI dalam menyiapkan paket asuransi. Jika biaya dari perencanaan NHI tidak sesuai dengan kenyataan yang ada di lapangan, maka dari pihak NHI akan banyak menanggung rugi dari nilai klaim asuransi yang sangat jauh dari perencanaan seharusnya. Sistem kenaikan pembayaran atau sistem dari paket asuransi NHI juga bisa dilakukan, berdasarkan hasil prediksi pasien kanker paru-paru yang melakukan pendaftaran asuransi.



BAB VI

SIMPULAN DAN SARAN

6.1 Kesimpulan

Hasil penelitian ini dapat disimpulkan sebagai berikut:

- (1) Hasil algoritma terbaik untuk melakukan prediksi pasien kanker paru-paru, baik dari sisi kebelangungan hidup dan biaya pengeluaran adalah menggunakan *Decision Tree*. Penggunaan *Decision Tree* baik dengan *feature selection* ataupun tidak menghasilkan hasil yang baik dibandingkan algoritma yang lain
- (2) Rekomendasi yang dapat diberikan kepada sektor medis adalah dapat menggunakan *Decision Tree* dalam melakukan prediksi survivabilitas dan juga dapat mengusulkan jenis pengobatan no.3 yaitu menerima kemoterapi dan kemudian melakukan operasi kepada pasien kanker paru-paru.
- (3) Rekomendasi yang dapat diberikan kepada pihak NHI adalah dapat menggunakan *Decision Tree* dalam melakukan prediksi biaya pengeluaran pasien baru kanker paru-paru di taiwan, dari hasil rekomendasi ini pihak asuransi dapat memberikan rekomendasi asuransi yang dapat diggunakan oleh pasien kanker paru-paru yang baru mendaftarkan asuransi.

6.2 Saran

Penelitian ini dapat ditingkatkan dengan beberapa saran sebagai berikut:

- (1) Penggunaan algoritma yang berbeda atau berfokus dengan *survival machine learning* yang mendalam untuk analisis survivabilitas, dengan menggunakan data yang berbeda.
- (2) Perbandingan metode *machine learning* tidak hanya dengan menggunakan satu jenis database, tetapi dengan menggunakan beberapa database.
- (3) Penambahan variabel lain untuk mengetahui karakteristik lebih detail seperti jenis obat yang diggunakan oleh pasien kanker paru-paru.

DAFTAR PUSTAKA

- Ahani, A. et al., 2019. Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, Volume 80, pp. 52-77.
- Alabi, R. O. et al., 2019. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool. *Virchows Archiv*, 475(4).
- Amjad, M. K. et al., 2018. Recent research trends in genetic algorithm based flexible job shop scheduling problems. *Mathematical Problems in Engineering*, pp. 1-32.
- Amrane, M., Oukid, S., gagaoua, I. & Ensari., T., 2020. Breast Cancer Classification Using Machine Learning. *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*.
- Ansari, A. & Riiasi, A., 2016. Customer clustering using a combination of fuzzy c-means and genetic algorithm. *International Journal of Business Management*, 11(7), pp. 59-66.
- Arunachalam, D. & Kumar, N., 2018. Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making. *Expert Systems With Applications*, Volume 111, pp. 11-34.
- Aryafar, A., Mikaeil, R., Haghshenas, S. S. & Haghshenas, S. S., 2018. Application of metaheuristic algorithms to optimal clustering of sawing machine vibration. *Measurement*, Volume 124, pp. 20-31.
- Askari, S., Montazerin, N. & Zarandi, M. H. F., 2017b. Generalized possibilistic fuzzy c-means with novel cluster validity indices for clustering noisy data. *Applied Soft Computing*, Volume 53, pp. 262-283.
- Askari, S., Montazerin, N., Zarandi, M. H. F. & Hakimi, E., 2017a. Generalized entropy based possibilistic fuzzy c-Means for clustering noisy data and its convergence proof. *Neurocomputing*, Volume 219, pp. 186-202.
- Bartholomai, J. A. & Frieboes., H. B., 2021. Lung cancer survival prediction via machine learning regression, classification, and statistical techniques. *Network Biology*, 11(2).
- Bie, R., Mehmood, R., Ruan, S. & Sun, Y., 2016. Adaptive fuzzy clustering by fast search and find of density peaks. *Pers Ubiquit Comput*, Volume 20, pp. 785-793.
- Borna, K. & Hashemi, V. H., 2014. An improved genetic algorithm with a local optimization strategy and an extra mutation level for solving traveling salesman problem. *International Journal of Computer Science, Engineering and Information Technology*, 4(4), pp. 47-53.

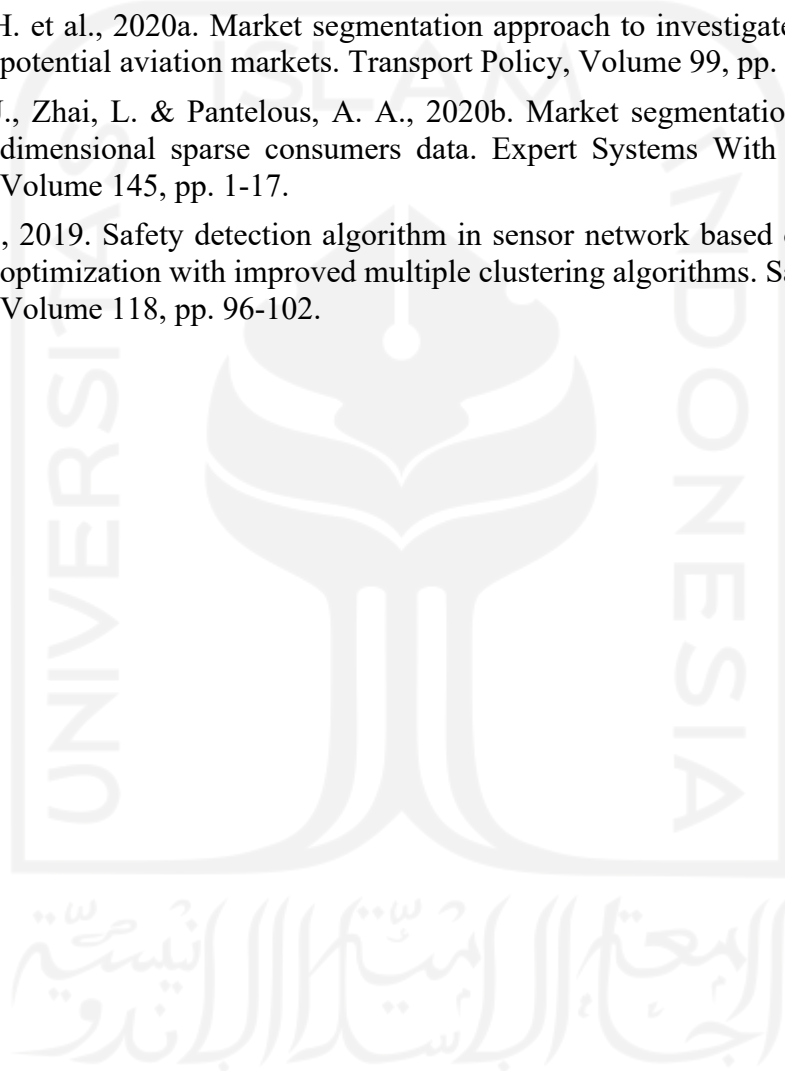
- Chehour, A. et al., 2017. A selection process for genetic algorithm using clustering analysis. *Algorithms*, 10(123), pp. 1-15.
- Chen, X., Li, D. & Li, H., 2016. An improved type-2 possibilistic fuzzy c-means clustering algorithm with application for MR image segmentation. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(11), pp. 363-370.
- Chip M. Lynch, B. A. J. D. F. et al., 2017. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International Journal of Medical Informatics*, Volume 108.
- Chou, J. S. & Nguyen, N. M., 2020. FBI inspired meta-optimization. *Applied Soft Computing*, Volume 93, pp. 1-28.
- Christy, A. J., Umamakeswari, A., Priyatharsini, L. & Neyaa, A., 2018. RFM ranking - an effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, pp. 1-7.
- Ding, Y. & Fu, X., 2016. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing*, Volume 188, pp. 233-238.
- Du, M., Ding, S., Xu, X. & Xue, Y., 2018. Density peaks clustering using geodesic distances. *International Journal Machine Learning & Cybernetics*, Volume 9, pp. 1335-1349.
- Faisal, M. I., Bashir, S., Khan, Z. S. & Khan., F. H., 2020. An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer. Pakistan, *International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST)*.
- Flores, K. G. & Garza, S. E., 2020. Density peaks clustering with gap-based automatic center detection. *Knowledge-Based Systems*, Volume 206, pp. 1-10.
- Franco, D. G. D. B. & Steiner, M. T. A., 2018. Clustering of solar energy facilities using a hybrid fuzzy c-means algorithm initialized by metaheuristics. *Journal of Cleaner Production*, Volume 191, pp. 445-457.
- Ghaemi, R., Sulaiman, N. B., Ibrahim, H. & Mustapha, N., 2011. A review: accuracy optimization in clustering ensembles using genetic algorithms. *Artificial Intelligence Review*, Volume 35, pp. 287-318.
- Ghaheri, A., Shoar, S., Naderan, M. & Hoseini, S. S., 2015. The applications of genetic algorithms in medicine. *Oman Medical Journal*, 30(6), pp. 406-416.
- Ghezelbash, R., Maghsoudi, A. & Carranza, E. J. M., 2020. Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GMKC) algorithm. *Computers & Geosciences*, Volume 134, pp. 1-11.
- Golmohammadi, A. M., Bani-Asadi, H., Zanjani, H. J. & Tikani, H., 2016. A genetic algorithm for preemptive scheduling of a single machine. *International Journal of Industrial Engineering Computations*, Volume 7, pp. 607-614.

- Haldurai, L., Madhubala, T. & Rajalakshmi, R., 2016. A study on genetic algorithm and its applications. *International Journal of Computer Sciences and Engineering*, 4(10), pp. 139-143.
- Hiba Asri, H. M., Moatassime, H. A. & Noel, T., 2016. Using Machine learning algorithms for breast cancer risk prediction and diagnosis. s.l., *Procedia in Computer Science*, pp. 1064-1069.
- Hou, J., Zhang, A. & Qi, N., 2020. Density peak clustering based on relative density relationship. *Pattern Recognition*, Volume 108, pp. 1-16.
- Islam, M. Z., Estivill-Castro, V., Rahman, M. A. & Bossomaier, T., 2018. Combining K-Means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering. *Expert Systems With Applications*, Volume 91, pp. 402-417.
- Jennings, P. C. et al., 2019. Genetic algorithms for computational materials discovery accelerated by machine learning. *Computational Materials*, Volume 5, pp. 1-6.
- K.R, P. & N.C, N., 2018. Lung Cancer Survivability Prediction based on performance using classification techniques of support vector machines, C4.5 and naïve bayes algorithms for healthcare analytics. India, *International Conference on Computational Intelligence and Data Science*, pp. 412-420.
- Kennedy, J. & Eberhart, R., 1995. Particle swarm optimization. Perth, IEEE, pp. 1942-1948.
- Khotimah, B. K., Irhamni, F. & Sundarwati, T., 2016. A genetic algorithm for optimized initial centers K-means clustering in SMEs. *Journal of Theoretical and Applied Information Technology*, 90(1), pp. 23-30.
- Kuo, R. J., Amornnikun, P. & Nguyen, T. P. Q., 2020. Metaheuristic-based possibilistic multivariate fuzzy weighted c-means algorithms for market segmentation. *Applied Soft Computing Journal*, Volume 96, pp. 1-14.
- Kuo, R. J., Lin, T. C., Zulvia, F. E. & Tsai, C. Y., 2018a. A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Applied Soft Computing*, Volume 67, pp. 299-308.
- Kuo, R. J., Potti, Y. & Zulvia, F. E., 2018b. Application of metaheuristic based fuzzy k-modes algorithm to supplier clustering. *Computers & Industrial Engineering*, Volume 120, pp. 298-307.
- Kuo, R. J., Zheng, Y. R. & Nguyen, T. P. Q., 2021. Metaheuristic-based possibilistic fuzzy k-modes algorithms for categorical data clustering. *Information Sciences*, Volume 557, pp. 1-15.
- Li, H., He, H. & Wen, Y., 2015. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. *Optik*, Volume 126, pp. 4817-4822.

- Liu, R., Wang, H. & Yu, X., 2018. Shared-nearest-neighbor-based clustering by fast search and find of density peaks. *Information Sciences*, Volume 450, pp. 200-226.
- Liu, X., Fan, J. & Chen, Z., 2019. Improved fuzzy c-means algorithm based on density peak. *International Journal of Machine Learning and Cybernetics*, Volume 11, pp. 545-552.
- Mai, D. N., Ngo, L. T., Trinh, L. H. & Hagra, H., 2021. A hybrid interval type-2 semi-supervised possibilistic fuzzy c-means clustering and particle swarm optimization for satellite image analysis. *Information Sciences*, Volume 548, pp. 398-422.
- Maina, S., 2015. The impact of market segmentation on the sales volume of a company's product or service. *European Journal of Business and Management*, 7(11), pp. 132-138.
- Makond, B., Wang, K.-J. & Wang, K.-M., 2015. Probabilistic modeling of short survivability in patients with brain metastasis from lung cancer. *Computer Methods and Programs in Biomedicine*, 119(3).
- Maleki, N., Zeinali, Y. & Niakib, S. T. A., 2020. A K-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. *Expert Systems With Application*, 164(5).
- Md. Milon Islam, H. I., Haque, M. R. & Hasan, M. K., 2017. Prediction of breast cancer using support vector machine and k-nearest neighbors. *IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*.
- Mirjalili, S., 2016. SCA: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems*, Volume 96, pp. 120-133.
- Moussa, D. A., Eissa, N. S., Abounaser, H. & Badr, A., 2018. Design of novel metaheuristic techniques for clustering. *IEEE Access*, Volume 6, pp. 77350-77358.
- Muller, H. & Hamm, U., 2014. Stability of market segmentation with cluster analysis - A methodological approach. *Food Quality and Preference*, Volume 34, pp. 70-78.
- Nayak, J., Naik, B., Behera, H. S. & Abraham, A., 2017. Hybrid chemical reaction based metaheuristic with fuzzy c-means algorithm for optimal cluster analysis. *Expert Systems With Applications*, Volume 79, pp. 282-295.
- Ni, D., Xiao, Z., Zhong, B. & Feng, X., 2018. Multiple human-behaviour indicators for predicting lung cancer mortality with support vector machine. *Scientific Reports*, 8(1).
- Nuhic, J. & Kevric, J., 2020. Lung cancer typology classification based on biochemical markers using machine learning techniques. *International Convention on Information, Communication and Electronic Technology (MIPRO)*.

- Pal, N. R., Pal, K., Keller, J. M. & Bezdek, J. C., 2005. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4), pp. 517-530.
- Rezaei, M., 2020. Improving a centroid-based clustering by using suitable centroids from another clustering. *Journal of Classification*, Volume 37, pp. 352-365.
- Rodriguez, A. & Laio, A., 2014. Clustering by fast search and find of density peaks. *Science*, 344(6191), pp. 1492-1496.
- Shabir, S. & Singla, R., 2016. A comparative study of genetic algorithm and the particle swarm optimization. *International Journal of Electrical Engineering*, 9(2), pp. 215-223.
- Silva, F. T., Silva, M. X. & Belchior, J. C., 2019. A new genetic algorithm approach applied to atomic and molecular cluster studies. *Frontiers in Chemistry*, Volume 7, pp. 1-21.
- Suthar, N., Rajput, I. J. & Gupta, V. K., 2013. A technical survey on DBSCAN clustering algorithm. *International Journal of Scientific & Engineering Research*, 4(5), pp. 1775-1781.
- Tapak, L. et al., 2019. Prediction of survival and metastasis in breast cancer patients using machine learning classifiers. *Clinical Epidemiology and Global Health*, 7(3), pp. 293-299.
- Tawhid, M. A. & Savsani, P., 2018. Discrete sine cosine algorithm (DSCA) with local search for solving traveling salesman problem. *Computer Engineering and Computer Science*, Volume 44, pp. 3669-3679.
- Tong, W., Liu, S. & Gao, X. Z., 2020. A density-peak-based clustering algorithm of automatically determining the number of clusters. *Neurocomputing*, pp. 1-12.
- Tripathi, S., Bhardwaj, A. & Poovammal, E., 2018. Approaches to clustering in customer segmentation. *International Journal of Engineering & Technology*, Volume 7, pp. 802-807.
- Wang, Q. et al., 2018. Explore machine learning for analysis and prediction of lung cancer related risk factors. s.l., *International Conference on Computer Science and Artificial Intelligence*, pp. 41-45.
- Wang, T., Ren, C., Luo, Y. & Tian, J., 2019. NS-DBSCAN: A density-based clustering algorithm in network space. *International Journal of Geo-Information*, 8(218), pp. 1-20.
- Wang, W. et al., 2019. Diagnostic value of plasma microRNAs for lung cancer using support vector machine model. *Journal of Cancer*, 10(21).
- Xie, Y. et al., 2021. Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational Oncology*, 14(1).
- Yang, C., Delcher, C., Shenkman, E. & Ranka., S., 2017. Machine learning approaches for predicting high cost high need patient expenditures in health care. *BioMedical Engineering OnLine*, Volume 113.

- Yu, H., Liu, Z. & Wang, G., 2014. An automatic method to determine the number of clusters using decision-theoretic rough set. *International Journal of Approximate Reasoning*, Volume 55, pp. 101-115.
- Zeebaree, D. Q., Haron, H., Abdulazeez, A. M. & Zeebaree, S. R. M., 2017. Combination of K-means clustering with genetic algorithm: A review. *International Journal of Applied Engineering Research*, 12(24), pp. 14238-14245.
- Zhou, H. et al., 2020a. Market segmentation approach to investigate existing and potential aviation markets. *Transport Policy*, Volume 99, pp. 120-135.
- Zhou, J., Zhai, L. & Pantelous, A. A., 2020b. Market segmentation using high-dimensional sparse consumers data. *Expert Systems With Applications*, Volume 145, pp. 1-17.
- Zhu, L., 2019. Safety detection algorithm in sensor network based on ant colony optimization with improved multiple clustering algorithms. *Safety Science*, Volume 118, pp. 96-102.



LAMPIRAN

Lampiran 1. Hasil Cox Proportional Hazard

1. Keberlangsungan Hidup

model	lifelines.CoxPHFitter
duration col	'time'
event col	'event'
penalizer	0.1
ll ratio	0
baseline estimation	breslow
number of observations	85745
number of events observed	69905
partial log-likelihood	-720162.84
time fit was run	2021-04-19 17:02:18 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
new_cost	0	1	0	0	0	1	1	-28.83	<0.005	604.81
Tri	-0.09	0.91	0.13	-0.34	0.16	0.71	1.17	-0.71	0.48	1.06
Heart	0.08	1.08	0.01	0.05	0.11	1.05	1.11	5.72	<0.005	26.48
Bronchitis	-0.05	0.95	0.02	-0.09	-0.01	0.91	0.99	-2.69	0.01	7.11
Bronchiectasis	0	1	0.03	-0.06	0.05	0.95	1.05	-0.03	0.97	0.04
Goiter	-0.21	0.81	0.04	-0.28	-0.14	0.75	0.87	-5.85	<0.005	27.64
Asbes	0.26	1.3	0.67	-1.05	1.58	0.35	4.84	0.39	0.7	0.52
Cir	-0.04	0.96	0.02	-0.08	0	0.93	1	-1.95	0.05	4.28
Pneu	0.11	1.12	0.01	0.09	0.14	1.09	1.15	8.76	<0.005	58.78
Emphy	0.09	1.09	0.04	0.02	0.16	1.02	1.17	2.46	0.01	6.17
Pulm	0.1	1.1	0.01	0.07	0.13	1.07	1.13	6.78	<0.005	36.23
Asthma	0.04	1.04	0.02	0	0.09	1	1.09	1.8	0.07	3.8
Peptic	0.08	1.08	0.02	0.05	0.11	1.05	1.12	4.62	<0.005	18
Oste	-0.08	0.92	0.03	-0.14	-0.03	0.87	0.97	-2.84	<0.005	7.77
hyperlipidemia	-0.25	0.78	0.02	-0.29	-0.2	0.75	0.82	-10.24	<0.005	79.34
hypertension	-0.04	0.96	0.01	-0.06	-0.02	0.94	0.98	-3.66	<0.005	11.98
Migraine	-0.02	0.98	0.08	-0.17	0.14	0.84	1.15	-0.21	0.84	0.26

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
Arterial	0.28	1.32	0.13	0.02	0.53	1.02	1.69	2.14	0.03	4.96
Atherosclerosis	0.07	1.08	0.09	-0.1	0.25	0.9	1.28	0.83	0.41	1.3
Gout	0.02	1.02	0.03	-0.03	0.07	0.97	1.07	0.64	0.52	0.93
Nephro	0.06	1.06	0.13	-0.2	0.32	0.82	1.38	0.45	0.65	0.61
Llergic	-0.11	0.9	0.02	-0.15	-0.07	0.86	0.93	-5.66	<0.005	26
Pneumoconiosis	0.08	1.08	0.05	-0.02	0.17	0.98	1.19	1.64	0.1	3.3
Chr_bron	0.11	1.11	0.01	0.09	0.13	1.09	1.14	9.81	<0.005	73.04
Chro_sin	-0.09	0.92	0.04	-0.17	0	0.84	1	-1.97	0.05	4.36
Diabete	0.1	1.11	0.01	0.07	0.13	1.08	1.13	7.78	<0.005	47.01
Occu_Asthma	-0.05	0.96	0.03	-0.1	0.01	0.9	1.01	-1.56	0.12	3.06
Arthritis	0.04	1.04	0.02	0	0.07	1	1.07	2.26	0.02	5.37

Concordance	0.56
Partial AIC	1440381.67
log-likelihood ratio test	1408.59 on 28 df
-log2(p) of ll-ratio test	925.61
Concordance	0.54

Partial AIC	1441190.07
log-likelihood ratio test	608.19 on 32 df
$-\log_2(p)$ of ll-ratio test	355.17



2. Biaya Pengeluaran

model	lifelines.CoxPHFitter
duration col	'time'
event col	'new_cost'
penalizer	0.1
l1 ratio	0
baseline estimation	breslow
number of observations	85745
number of events observed	85745
partial log-likelihood	-887561.68
time fit was run	2021-04-19 17:08:52 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
event	-0.27	0.76	0.01	-0.29	-0.25	0.75	0.78	-30.57	<0.005	679.4
Tri	-0.04	0.96	0.11	-0.25	0.18	0.78	1.2	-0.33	0.74	0.43

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
Heart	0.09	1.09	0.01	-0.07	0.11	1.07	1.12	7.18	<0.005	40.41
Bronchitis	0	1	0.02	-0.04	0.03	0.96	1.03	-0.16	0.88	0.19
Bronchiectasis	0	1	0.03	-0.05	0.05	0.95	1.05	-0.17	0.87	0.2
Goiter	-0.08	0.92	0.03	-0.14	-0.02	0.87	0.98	-2.67	0.01	7.05
Asbes	0.28	1.32	0.56	-0.82	1.37	0.44	3.94	0.5	0.62	0.69
Cir	0	1	0.02	-0.04	0.03	0.96	1.03	-0.24	0.81	0.3
Pneu	0.13	1.14	0.01	0.11	0.15	1.11	1.16	11.11	<0.005	92.92
Emphy	0.02	1.02	0.03	-0.05	0.08	0.95	1.09	0.53	0.6	0.74
Pulm	0.04	1.04	0.01	0.01	0.07	1.01	1.07	3.02	<0.005	8.62
Asthma	0.04	1.04	0.02	-0.01	0.08	0.99	1.08	1.71	0.09	3.53
Peptic	0.08	1.08	0.02	0.05	0.11	1.05	1.11	5.03	<0.005	20.98
Oste	-0.06	0.94	0.03	-0.11	-0.01	0.89	0.99	-2.25	0.02	5.37
hyperlipidemia	-0.13	0.88	0.02	-0.17	-0.09	0.84	0.91	-6.41	<0.005	32.67
hypertension	-0.01	0.99	0.01	-0.03	0.01	0.97	1.01	-0.89	0.37	1.42
Migraine	0.04	1.04	0.07	-0.1	0.17	0.9	1.19	0.51	0.61	0.71
Arterial	0.32	1.38	0.11	0.11	0.54	1.11	1.72	2.94	<0.005	8.26
Atherosclerosis	0.05	1.05	0.08	-0.11	0.21	0.89	1.23	0.58	0.56	0.84
Gout	0.01	1.01	0.02	-0.04	0.05	0.96	1.05	0.33	0.74	0.43

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
Nephro	0.04	1.04	0.12	-0.2	0.28	0.82	1.32	0.31	0.76	0.4
Llergic	-0.05	0.95	0.02	-0.09	-0.02	0.92	0.98	-3	<0.005	8.52
Pneumoconiosis	0.08	1.08	0.04	-0.01	0.17	0.99	1.18	1.79	0.07	3.76
Chr_bron	0.08	1.09	0.01	0.06	0.1	1.07	1.11	8.42	<0.005	54.58
Chro_sin	-0.08	0.92	0.04	-0.16	0	0.85	1	-1.95	0.05	4.28
Diabete	0.1	1.11	0.01	0.08	0.12	1.08	1.13	8.69	<0.005	57.9
Occu_Asthma	-0.05	0.95	0.03	-0.1	0.01	0.9	1.01	-1.7	0.09	3.49
Arthritis	0.06	1.06	0.01	0.03	0.09	1.03	1.09	4.09	<0.005	14.48

Concordance	0.52
Partial AIC	1775179.36
log-likelihood ratio test	1377.56 on 28 df
-log2(p) of ll-ratio test	903.65
Concordance	0.52
Partial AIC	1775180.94
log-likelihood ratio test	1383.99 on 32 df

$-\log_2(p)$ of ll-ratio test

897.03



Lampiran 2. Biaya Pengeluaran berdasarkan Biaya Pengobatan

1. Rata-rata biaya pengeluaran

Tahun		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total Price		59650. 5	58606. 6	58516. 0	58999. 5	59436. 7	59063. 1	59448. 4	59465. 8	58961. 9	58180. 0	71349. 8
Treatment	Only chemotherapy	49554. 9	48500. 2	50810. 0	50248. 5	50506. 0	52306. 6	52421. 0	53645. 9	50877. 2	51157. 1	62796. 2
	Only surgery	57528. 9	56107. 1	53037. 1	55955. 5	56829. 6	55680. 6	60479. 0	55350. 3	55383. 3	52019. 3	63531. 6
	Chemotherapy first and then surgery	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	54901. 0	32181. 0	55280. 8
	Surgery first and then chemotherapy	56334. 3	66095. 0	71687. 0	74364. 0	10428. 0	63239. 5	52215. 0	44909. 4	53569. 5	38940. 0	56665. 1
	No treatment	60115. 3	59243. 0	59091. 9	59739. 3	60225. 5	59752. 3	60173. 7	60159. 1	59870. 9	59144. 3	72306. 0

Tahun		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Surgery	Segmental resection of lung	62537. 7	56700. 3	39450. 4	57598. 8	57975. 4	63260. 4	59221. 0	65865. 4	49722. 0	62299. 0	61641. 9
	Lobectomy of lung	55317. 4	58240. 7	60273. 4	56041. 5	50356. 6	56339. 2	61207. 4	51062. 1	56841. 1	47580. 2	63249. 8
	Complete pneumonectomy	51626. 3	49997. 0	44683. 0	71249. 3	72729. 0	11406. 0	41023. 0	64265. 7	41532. 5	56493. 0	41259. 6
	Regional lymph node excision	45607. 5	57837. 0	35031. 0	64701. 0	38588. 5	0.0	0.0	0.0	80604. 0	59110. 0	66943. 2
	No surgery	59698. 1	58619. 7	58550. 9	59004. 2	59463. 2	59073. 1	59444. 9	59510. 8	58983. 6	58239. 0	71486. 4
Chemotherapy	Yes	49760. 3	48686. 7	50867. 0	50411. 4	50422. 7	52420. 3	52420. 6	53584. 9	50904. 5	51016. 8	62789. 6
	No	60094. 2	59921. 0	59044. 3	59719. 3	60206. 8	59733. 7	60175. 2	60120. 9	59847. 6	59092. 5	72300. 3

البيحة العامة
الاستاذة
البيحة العامة

2. Proporsi berdasarkan jenis pengobatan

Tahun		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total Price		2306	4852	5712	5742	6111	6292	6466	7144	7472	6673	26975
Treatment	Only chemotherapy	96	280	368	441	480	571	605	711	733	747	2278
	Only surgery	18	32	42	228	31	26	29	51	35	43	409
	Chemotherapy first and then surgery	0	0	0	0	0	0	0	0	1	3	6
	Surgery first and then chemotherapy	3	3	1	3	1	6	1	5	6	4	28
	No treatment	2189	4537	5301	5270	5599	5689	5831	6377	6697	5876	24254
Surgery	Segmental resection of lung	3	11	9	5	10	10	5	7	7	5	35
	Lobectomy of lung	15	19	30	21	16	21	24	45	32	39	375
	Complete pneumonectomy	3	4	4	3	3	1	1	3	2	1	5
	Regional lymph node excision	2	1	1	1	2	0	0	0	1	4	22
	No surgery	2283	4817	5668	5712	6080	6260	6436	7089	7430	6624	26538
Chemotherapy	Yes	99	283	369	441	481	577	606	716	740	754	2312
	No	2207	4569	5343	5271	5630	5715	5860	6428	6732	5919	24663

Lampiran 3. Biaya Pengeluaran berdasarkan Tahun Kematian

1. Rata-rata Biaya Pengeluaran

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Class 1	14319.7 9	13414.7 3	13294.4 1	13434.6 3	13341.6 6	12721.3 2	13777.2	13003.2 8	13688.8 1	12972.6 1	12811.1 3
Class 2	31037.8 4	30849.5 2	31322.0 4	31213.8 8	31036.6 4	31018.1 9	31100.1 9	30799.7 2	30975.1 8	30631.1 4	30936.1 6
Class 3	49193.9 5	49259.6 8	49160.3 5	49154.6	49490.3 9	49392.3 8	49446.2	49392.1 9	49264.4 9	49299.2 9	49130.3 9
Class 4	69033.5 3	69058.0 6	69085.0 1	69046.6 4	68979.7 6	68783.4 3	69145.6 6	69108.6 5	69093.8 7	69199.2 3	69115.5 6
Class 5	88708.9 6	88675.7	88392.2 8	88391.4 5	88334.8 3	88419.8 7	88532.7 7	88242.6 5	88117.2	88373.3 9	88552.2 6

2. Proporsi Biaya Pengeluaran Berdasarkan tahun kematian

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Class 1	0.82	1.50	1.77	1.58	1.62	1.53	1.81	1.68	2.06	2.61	0.98
Class 2	23.42	23.17	22.48	21.86	20.88	21.81	21.53	22.27	22.42	23.71	10.54

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Class 3	26.32	27.37	27.70	27.36	27.08	26.83	26.31	24.85	25.20	24.76	10.32
Class 4	23.07	23.62	25.14	26.21	26.72	26.56	26.23	26.48	26.39	25.37	34.89
Class 5	26.37	24.13	22.92	22.99	23.69	23.28	24.13	24.72	23.93	23.56	43.27



الجامعة الإسلامية
الاندونيسية

