

SKRIPSI

**ESTIMASI PELUANG *SURVIVAL* MENGGUNAKAN METODE
KAPLAN-MEIER DAN MODEL *SURVIVAL TREE* SERTA
ANALISIS KOVARIAT MENGGUNAKAN MODEL REGRESI
COX PADA DATA PASIEN KANKER OTAK**



ADAM FADHLURRAHMAN PUTRA SUNTIYATNO

NPM: 6162001198

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2025**

FINAL PROJECT

ESTIMATION OF SURVIVAL PROBABILITY USING THE
KAPLAN-MEIER METHOD AND SURVIVAL TREE MODEL
ALONG WITH COVARIATE ANALYSIS USING THE COX
REGRESSION MODEL ON BRAIN CANCER PATIENT DATA



ADAM FADHLURRAHMAN PUTRA SUNTIYATNO

NPM: 6162001198

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
PARAHYANGAN CATHOLIC UNIVERSITY
2025

LEMBAR PENGESAHAN

ESTIMASI PELUANG *SURVIVAL* MENGGUNAKAN METODE KAPLAN-MEIER DAN MODEL *SURVIVAL TREE* SERTA ANALISIS KOVARIAT MENGGUNAKAN MODEL REGRESI COX PADA DATA PASIEN KANKER OTAK

Adam Fadhlurrahman Putra Suntiayatno

NPM: 6162001198

Telah lulus ujian skripsi pada 10 Januari 2025 dengan penguji:
Iwan Sugiarto, M.Si. dan Dr. Ferry Jaya Permana

Bandung, 31 Januari 2025

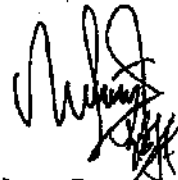
Menyetujui,

Pembimbing 1



Farah Kristiani, Ph.D.

Pembimbing 2



Robyn Irawan, M.Sc.

Mengetahui,

Ketua Program Studi



Jonathan Hoseana, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

**ESTIMASI PELUANG *SURVIVAL* MENGGUNAKAN METODE
KAPLAN-MEIER DAN MODEL *SURVIVAL TREE* SERTA ANALISIS
KOVARIAT MENGGUNAKAN MODEL REGRESI COX PADA DATA
PASIEN KANKER OTAK**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 31 Januari 2025



Adam Fadhlurrahman Putra Suntiyatno
NPM: 6162001198

ABSTRAK

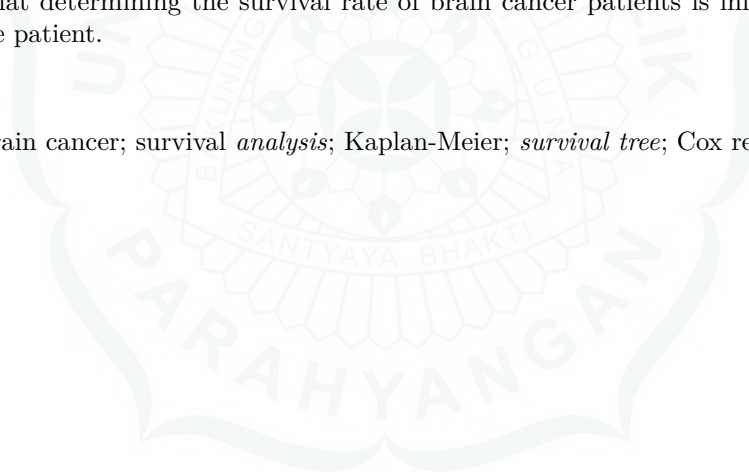
Penelitian ini memiliki dua fokus, yaitu mengestimasi peluang *survival* pasien kanker otak menggunakan metode Kaplan-Meier dan model *survival tree* dan analisis variabel-variabel yang memengaruhi tingkat *survival* pasien kanker otak dengan menggunakan model regresi Cox. Dengan tingkat kematian yang disebabkan oleh kanker otak tinggi secara global, penelitian ini bertujuan untuk memprediksi peluang *survival* pasien serta mengidentifikasi pengaruh variabel dalam memperkirakan tingkat *survival*. Metode Kaplan-Meier digunakan untuk menganalisis variabel kategorik, sementara model *survival tree* dan model regresi Cox digunakan untuk analisis lebih lanjut terhadap data *survival* pasien kanker otak. Data yang digunakan mencakup variabel seperti *Gender*, *Diagnosis*, *Lokasi Tumor*, *Karnofsky Index*, *Ukuran Tumor* dan *Metode Terapi*. Hasil analisis menunjukkan bahwa metode Kaplan-Meier mampu memvisualisasikan kurva *survival* berdasarkan variabel tertentu, sedangkan model *survival tree* mengidentifikasi pembagian kelompok berdasarkan variabel yang berpengaruh terhadap tingkat *survival*. Model regresi Cox menunjukkan variabel-variabel yang berpengaruh signifikan yaitu *Diagnosis*, *Karnofsky Index*, dan *Ukuran Tumor*. Ketiga pendekatan secara konsisten menunjukkan bahwa penentuan tingkat *survival* pasien kanker otak dipengaruhi oleh hasil *Diagnosis* pada pasien tersebut.

Kata-kata kunci: kanker otak; analisis *survival*; Kaplan-Meier; *survival tree*; regresi Cox.

ABSTRACT

This study has two focuses, namely estimating the survival probability of brain cancer patients using Kaplan-Meier Estimator and survival tree model as well as analyzing the variables influencing the survival rates of brain cancer patients using Cox regression model. With the high global mortality rate caused by brain cancer, this research aims to predict patients' survival rates and identify the impact of variables in estimating survival. The Kaplan-Meier method is employed to analyze categorical variables, while the survival tree model and Cox regression model are used for an in-depth analysis of the survival data of brain cancer patients. The data includes variables such as Gender, Diagnosis, Lokasi Tumor, Karnofsky Index, Ukuran Tumor, and Metode Terapi. The results show that the Kaplan-Meier method effectively visualizes survival curves based on specific variables, while the survival tree model identifies subgroup divisions influenced by variables affecting survival rates. The Cox regression model highlights significant variables such as Diagnosis, Karnofsky Index, and Ukuran Tumor. All three approaches consistently demonstrate that determining the survival rate of brain cancer patients is influenced by the diagnosis of the patient.

Keywords: brain cancer; survival *analysis*; Kaplan-Meier; *survival tree*; Cox regression.



Life's too short to fold fitted sheets, embrace the chaos.



KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa, penulis dapat menyelesaikan skripsi dengan judul "**Estimasi Peluang *Survival* Menggunakan Metode Kaplan-Meier dan Model *Survival Tree* Serta Analisis Kovariat Menggunakan Model Regresi Cox pada Data Pasien Kanker Otak**". Skripsi ini disusun sebagai salah satu persyaratan untuk menyelesaikan studi Strata-1, Program Studi Matematika, Fakultas Sains, Universitas Katolik Parahyangan. Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu penulis ingin mengucapkan terima kasih yang setulusnya kepada:

1. Ibu dan nenek yang senantiasa memberikan dukungan dan doa dalam segala hal hingga proses penyusunan skripsi.
2. Ibu Farah Kristiani, Ph.D. dan Bapak Robyn Irawan, M.Sc. selaku dosen pembimbing yang telah memberikan ilmu, dorongan, arahan, serta saran sehingga skripsi ini dapat diselesaikan dengan baik.
3. Bapak Iwan Sugiarto, M.Si. dan Dr. Ferry Jaya Permana selaku dosen penguji yang telah memberikan masukan dan saran yang berharga untuk menyempurnakan skripsi ini.
4. Bapak Dr. Andreas Parama Wijaya selaku koordinator kelas skripsi yang senantiasa membimbing proses penyusunan skripsi sehingga berjalan dengan baik.
5. Seluruh dosen Fakultas Sains khususnya dosen Program Studi Matematika yang senantiasa memberikan ilmu dan pengetahuan baru selama masa perkuliahan.
6. Seluruh staf Fakultas Sains yang senantiasa membantu penulis dalam masa perkuliahan.
7. Aifa Aini Soewandi, yang kehadirannya sangat berarti dalam memberikan dukungan dan semangat dalam menyelesaikan perkuliahan dan juga skripsi.
8. Teman-teman dari "WahyuDeen Skwad": Dimas, Jalali, Jawad, Pey, Dzaki, Iban, Veka, Amir yang telah menjadi teman semenjak SMP hingga saat ini.
9. Teman-teman "12 Murid Adam": Aifa, Alfonsa, Andika, Bella, Devina, Dwinanda, Elsy, Gabriella, Kores, Monika, Nurul, Sabilla, Septia, dan Revina, yang telah menjadi teman semenjak perkuliahan *offline* hingga akhir masa perkuliahan.
10. Johannes Dimas Paramasatya dan orang tua yang telah memberikan penulis dukungan semenjak SD.
11. Teman-teman matematika UNPAR angkatan 2019, 2020, 2021, dan semua pihak yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis dengan terbuka menerima kritik dan saran yang membangun untuk memperbaiki dan menyempurnakan skripsi ini. Akhir kata, semoga skripsi ini dapat dikembangkan menjadi karya yang lebih baik.

Bandung, 24 Januari 2025

Penulis



DAFTAR ISI

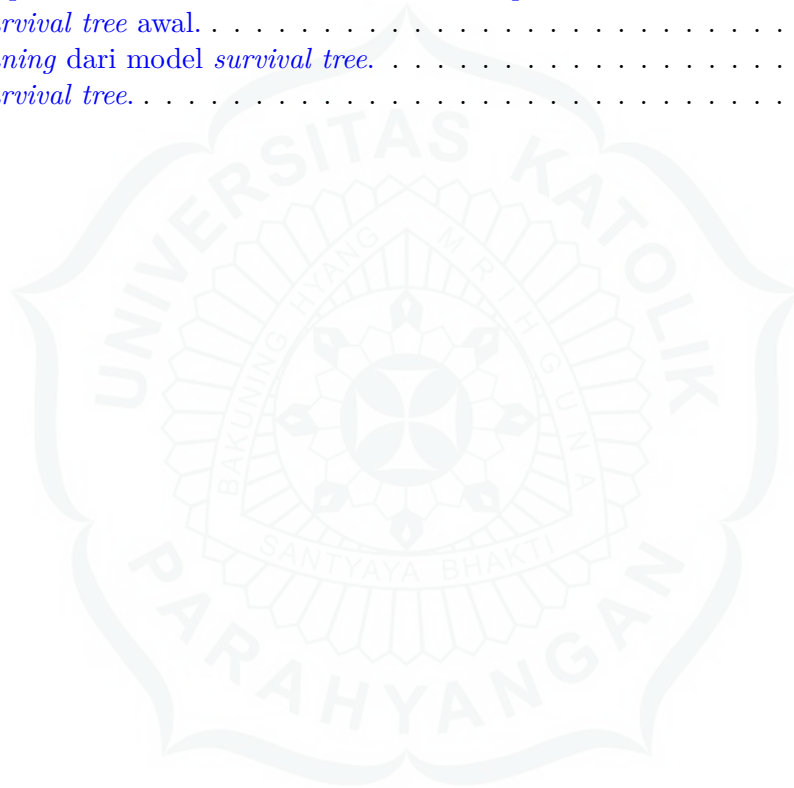
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR NOTASI	xiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 <i>State of the Art</i>	2
1.5 Sistematika Pembahasan	3
2 LANDASAN TEORI	4
2.1 Kanker Otak	4
2.2 Fungsi Survival	4
2.3 Fungsi Distribusi Kumulatif	5
2.4 Fungsi Kepadatan Peluang	6
2.5 Fungsi Laju Kegagalan	6
2.6 Penyensoran Data	7
2.7 Analisis <i>Survival</i>	7
2.8 Distribusi Eksponensial	8
2.9 Metode Iterasi Newton-Raphson	8
3 METODE ESTIMASI PELUANG <i>Survival</i> PADA PASIEN KANKER OTAK	10
3.1 Estimator Kaplan-Meier	10
3.2 <i>Survival Tree</i>	13
3.2.1 <i>Exponential Log-Likelihood Loss</i>	13
3.2.2 <i>Pruning</i>	15
3.3 Model Regresi <i>Cox Proportional Hazard</i> (CPH)	16
3.3.1 Estimasi <i>Likelihood</i> Untuk Model CPH	17
3.3.2 Asumsi <i>Proportional Hazard</i>	19
3.3.3 Penentuan Model Terbaik	20
3.3.4 Uji Signifikansi Model	20
4 ANALISIS PADA KEMATIAN PASIEN KANKER OTAK	22
4.1 Data	22
4.2 Aplikasi Metode Estimator Kaplan-Meier pada Data Pasien Kanker Otak	24
4.3 Aplikasi Model <i>Survival Tree</i> pada Data Pasien Kanker Otak	28
4.4 Aplikasi Model Regresi Cox pada Data Pasien Kanker Otak	31

4.5 Analisis Hasil	35
5 KESIMPULAN DAN SARAN	38
5.1 Kesimpulan	38
5.2 Saran	38
DAFTAR REFERENSI	40
A DATA	42



DAFTAR GAMBAR

2.1	Ilustrasi variabel waktu dan <i>event</i> (kematian).	8
3.1	Kurva Kaplan-Meier untuk contoh data.	12
4.1	Kurva Kaplan-Meier untuk variabel <i>Gender</i> .	25
4.2	Kurva Kaplan-Meier untuk variabel <i>Diagnosis</i> .	26
4.3	Kurva Kaplan-Meier untuk variabel Lokasi Tumor.	27
4.4	Kurva Kaplan-Meier untuk variabel Metode Terapi.	27
4.5	Model <i>survival tree</i> awal.	29
4.6	Hasil <i>pruning</i> dari model <i>survival tree</i> .	30
4.7	Model <i>survival tree</i> .	30



DAFTAR TABEL

3.1	Asumsi data pasien kanker otak yang telah diamati selama 10 bulan	11
3.2	Hasil perhitungan metode estimator Kaplan-Meier terhadap data asumsi	12
3.3	Asumsi data pasien kanker otak	15
3.4	Hasil perhitungan terhadap variabel <i>Gender</i>	15
3.5	Hasil perhitungan terhadap variabel Umur	15
4.1	Tabel Data Pasien Kanker Otak	22
4.2	Tabel data terhadap variabel <i>Gender</i>	24
4.3	Hasil perhitungan metode Kaplan-Meier untuk variabel <i>Gender</i>	24
4.4	Nilai <i>deviance</i> yang diperoleh dari model	29
4.5	Pemisahan variabel <i>dummy</i> pada data pasien kanker otak	32
4.6	Model CPH menggunakan empat kovariat dengan interaksi pada data pasien kanker otak	33
4.7	Nilai AIC untuk model-model CPH	33
4.8	Hasil uji parsial model CPH	34
4.9	Hasil uji asumsi PH model pada data pasien kanker otak	34
4.10	Interpretasi hasil model CPH	35
A.1	Data pasien kanker otak	43

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan penyakit dengan tingkat kematian tertinggi kedua di dunia menurut *World Health Organization* (WHO).¹ Kanker otak, salah satu bentuk kanker mematikan, dapat bersifat jinak atau ganas. Tumor jinak umumnya tidak menyebar dan dapat diangkat melalui pembedahan, sementara tumor ganas, seperti glioma, bersifat agresif dan sering menyerang jaringan otak di sekitarnya. Glioma berasal dari jaringan otak selain sel saraf dan pembuluh darah, menyebabkan berbagai gangguan neurologis. Sebaliknya, meningioma tumbuh dari membran yang melapisi otak dan biasanya bersifat jinak dengan pertumbuhan lebih lambat [1].

Pada tahun 2022, tingkat insiden kanker otak di seluruh dunia yang telah disesuaikan berdasarkan populasi standar dunia, tercatat sebesar 3,5 per 100.000 populasi. Secara spesifik, tingkat insiden pada laki-laki mencapai 3,9 per 100.000, sementara pada perempuan sedikit lebih rendah, yaitu 3,1 per 100.000. Angka ini merepresentasikan sekitar 173.699 kasus baru kanker otak pada laki-laki dan 148.032 kasus pada perempuan, menghasilkan total estimasi 321.731 individu yang didiagnosis dengan tumor ganas primer otak secara global pada tahun tersebut [2].

Berdasarkan data di atas, tingkat insiden kanker otak menunjukkan urgensi dalam memahami prognosis pasien secara lebih mendalam. Maka dari itu, diperlukan pemahaman mengenai seberapa lama pasien kanker otak dapat bertahan setelah diagnosis atau perawatan tertentu menggunakan analisis *survival*. Mengingat sifat agresif tumor ganas seperti glioma, terutama *High-Grade* glioma, tingkat kelangsungan hidup pasien sangat bervariasi dan bergantung pada banyak faktor, termasuk jenis tumor, stadium, serta efektivitas pengobatan [1]. Analisis *survival* memerlukan data berisi kovariat yang berpotensi menjadi pengaruh kematian pasien kanker otak yang akan digunakan untuk memprediksi tingkat *survival* pasien kanker otak di masa yang akan datang. Data ini meliputi durasi waktu yang diperlukan dari awal waktu pengamatan hingga akhir waktu pengamatan, yang dikenal sebagai data *survival* [3]. Kemudian, data *survival* tersebut dapat digunakan untuk menganalisis tingkat *survival* pasien kanker otak di masa yang akan datang.

Terdapat beberapa metode dan model yang dapat digunakan untuk menganalisis data *survival*, antara lain: metode estimator Kaplan-Meier, model *survival tree*, model regresi Cox, dan model Weibull. Menurut Sami [4], metode estimator Kaplan-Meier merupakan metode analisis data *survival* non-parametrik, kemudian model *Survival Tree* dan model regresi Cox merupakan model semi-parametrik, sementara model Weibull merupakan model analisis data *survival* parametrik. Model analisis data *survival* non-parametrik dan semi-parametrik lebih fleksibel dibandingkan model

¹<https://www.who.int/news-room/fact-sheets/detail/cancer>, diakses pada 29 November 2024

parametrik karena tidak memerlukan asumsi tentang distribusi waktu *survival* sehingga pada skripsi ini tidak dibahas model Weibull.

Pada skripsi ini, akan digunakan metode Kaplan-Meier untuk mengestimasi peluang *survival* pada masing-masing variabel yang diduga memengaruhi peluang *survival* pasien kanker otak. Model *survival tree* juga diterapkan untuk menganalisis pola dan peluang *survival* pasien secara bertahap berdasarkan kovariat yang memiliki nilai *deviance* yang mengecil secara terurut. Selanjutnya, model regresi Cox digunakan untuk menganalisis dan menentukan kovariat-kovariat yang signifikan terhadap tingkat *survival* pasien kanker otak.

1.2 Rumusan Masalah

Berikut masalah-masalah yang akan dibahas dalam skripsi ini.

1. Bagaimana metode Kaplan-Meier dan model *survival tree* digunakan untuk mengestimasi peluang *survival* pasien kanker otak?
2. Bagaimana pengaruh dari kovariat-kovariat data pasien kanker otak terhadap tingkat *survival* pasien berdasarkan dari karakteristik pasien menurut model Cox *Proportional Hazard*?

1.3 Tujuan

Tujuan penulisan skripsi ini adalah

1. Menguraikan dan menjelaskan penggunaan metode Kaplan-Meier dan model *survival tree* untuk mengestimasi peluang *survival* pasien kanker otak.
2. Menganalisis pengaruh kovariat-kovariat pada data pasien kanker otak terhadap tingkat *survival* pasien berdasarkan karakteristik pasien menggunakan model Cox *Proportional Hazard*.

1.4 State of the Art

Pengembangan jurnal acuan tulisan Sami [4] dan Pahlevi [5] adalah

1. Jurnal acuan berdasarkan Sami [4] memberikan contoh analisis *survival* pada data kasus COVID-19 di Kur-distan menggunakan estimator Nelson-Aalen dan estimator Kaplan-Meier. Dalam skripsi ini, digunakan metode estimator Kaplan-Meier untuk analisis *survival* pada data pasien kanker otak. Pengembangan dalam skripsi ini adalah menganalisis variabel bebas kategorik menggunakan metode estimator Kaplan-Meier untuk melihat pengaruh variabel bebas tersebut dalam menentukan peluang *survival* pasien kanker otak.
2. Jurnal acuan berdasarkan tulisan Pahlevi [5] memberikan contoh analisis *survival* menggunakan model regresi Cox. Dalam jurnal tersebut digunakan delapan kovariat pada data sekunder pasien penderita *stroke* hemoragik *unspecified*. Dalam skripsi ini, model regresi Cox digunakan dalam analisis pengaruh enam kovariat terhadap pasien kanker otak dan mengidentifikasi kovariat yang berpengaruh signifikan dalam tingkat *survival* pasien kanker otak.

3. Jurnal tulisan Wang [6], memberikan masukan dan memperkenalkan aplikasi model pembelajaran mesin dalam data *survival*, salah satunya model *survival tree*. Dalam jurnal tersebut diberikan deskripsi secara singkat mengenai ide dari model *survival tree* dan bagaimana algoritma model *survival tree* bekerja. Dalam skripsi ini, digunakan model *survival tree* untuk menganalisis peluang *survival* pasien kanker otak dengan mempertimbangkan banyak kovariat secara bersamaan.

1.5 Sistematika Pembahasan

Berikut adalah penjelasan setiap bagian dalam skripsi ini.

BAB 2: Landasan Teori

Dalam bab ini, dijelaskan teori-teori dasar terkait analisis *survival*, metode dan model yang digunakan, dan penyakit kanker otak secara umum.

BAB 3: Metode Estimasi Peluang *Survival* Pada Pasien Kanker Otak

Dalam bab ini dijelaskan metode estimator Kaplan-Meier, model *survival tree*, dan model regresi Cox yang digunakan untuk mengestimasi peluang *survival* dan menganalisis kovariat-kovariat yang dominan memengaruhi kematian pasien kanker otak.

BAB 4: Analisis pada Kematian Pasien Kanker Otak

Dalam bab ini, dijelaskan data dan analisis *survival* menggunakan metode Kaplan-Meier, model *survival tree*, dan model regresi Cox.

BAB 5: Kesimpulan dan Saran

Dalam bab ini, dijelaskan kesimpulan dan saran berdasarkan hasil yang diperoleh dari skripsi ini.

BAB 2

LANDASAN TEORI

Pada bab ini dijelaskan notasi mengenai fungsi *survival*, serta konsep-konsep dasar lainnya yang relevan untuk memahami analisis *survival* yang diambil dari buku Dickson dan London [7, 8]. Pembahasan mencakup fungsi *survival*, distribusi kumulatif, kepadatan peluang, laju kegagalan, serta penyensoran. Selain itu, dijelaskan juga dasar-dasar metode distribusi eksponensial, metode iterasi Newton-Raphson, hingga pengetahuan dasar kanker otak.

2.1 Kanker Otak

Pada umumnya, kanker otak terjadi ketika sel-sel di dalam otak tumbuh secara tidak normal dan membentuk sebuah tumor. Kanker otak pada umumnya dibedakan antara tumor primer dan metastasis, dengan tumor primer merupakan tumor yang berasal dari otak, dan metastasis merupakan tumor yang menyebar ke otak dari bagian tubuh lain. Beberapa tipe yang umum ditemukan dalam kanker otak di antaranya adalah glioma dan meningioma. Glioma merupakan kategori tumor yang berasal dari sel glial di otak, yang terbagi menjadi beberapa subkategori, di antaranya *High-Grade* Glioma yang tumbuh cepat, dan *Low-Grade* Glioma yang cenderung tumbuh lambat. Lebih lanjut, meningioma merupakan tumor yang berkembang di lapisan otak dan sumsum tulang belakang tetapi cenderung lebih jinak dibandingkan glioma [9].

Beberapa pendekatan yang digunakan dalam data skripsi ini adalah radioterapi stereotaktik (SRT) dan *radiosurgery* stereotaktik (SRS). Kedua teknik ini menggunakan radiasi yang diarahkan untuk menargetkan sel-sel kanker tanpa merusak jaringan di sekitarnya. SRT merupakan prosedur yang dilakukan secara bertahap, di mana dosis radiasi diberikan dalam beberapa sesi untuk untuk memaksimalkan perlambatan dan pembunuhan sel kanker dan meminimalisasi efek samping. Sementara itu, SRS merupakan prosedur yang lebih intensif yang seringkali dilakukan dalam satu kali sesi dengan dosis yang tinggi [10].

2.2 Fungsi Survival

Fungsi survival atau *Survival Distribution Function (SDF)* merupakan probabilitas kegagalan (*failure*) yang akan terjadi setelah waktu t . Dalam fungsi survival, digunakan variabel acak T yang didefinisikan sebagai waktu kegagalan dan $S_x(t)$ menyatakan peluang atau probabilitas dari seseorang yang saat ini berusia x tahun, yang diketahui hidup pada saat $t = 0$ dan dapat bertahan

hidup hingga waktu t . Fungsi *survival* dapat ditulis sebagai

$$S_x(t) = Pr(T > t). \quad (2.1)$$

Fungsi survival memiliki 3 sifat yaitu

1. $S_x(0) = 1$;
2. $S_x(t)$ merupakan fungsi monoton tak naik;
3. $\lim_{t \rightarrow \infty} S_x(t) = 0$.

Sifat pertama dapat diinterpretasikan sebagai probabilitas seorang individu yang menjadi subjek analisis dipastikan masih hidup saat waktu pengamatan dimulai. Sifat kedua dapat diinterpretasikan sebagai probabilitas seorang individu monoton tak naik karena semakin lama waktu pengamatan, maka probabilitas *survival* suatu individu akan menurun. Sifat ketiga dapat diinterpretasikan sebagai probabilitas suatu individu bernilai 0 dalam waktu pengamatan yang cukup besar.

2.3 Fungsi Distribusi Kumulatif

Fungsi distribusi kumulatif atau *Cumulative Distribution Function* (CDF) merupakan probabilitas kegagalan dari peubah acak T yang tidak melebihi dari waktu t . Fungsi distribusi kumulatif untuk seseorang yang saat ini berusia x tahun, dapat dinotasikan dengan $F_x(t)$ dan dapat dituliskan sebagai

$$F_x(t) = Pr(T \leq t). \quad (2.2)$$

Dengan menggunakan persamaan (2.1) dan (2.2), dapat dituliskan

$$S_x(t) + F_x(t) = Pr(T > t) + Pr(T \leq t) \quad (2.3)$$

$$= 1. \quad (2.4)$$

Dengan memperhatikan persamaan (2.3), didapatkan hubungan antara fungsi *survival* dengan fungsi distribusi kumulatif yang dapat dituliskan sebagai

$$F_x(t) = 1 - S_x(t). \quad (2.5)$$

Dengan mensubstitusikan sifat 1 dari fungsi *survival* pada fungsi distribusi kumulatif, maka didapatkan

$$\begin{aligned} F_x(0) &= 1 - S_x(0) \\ &= 1 - 1 \\ &= 0, \end{aligned}$$

dan dengan mensubstitusikan sifat 3 dari fungsi *survival* akan didapatkan

$$\lim_{\Delta t \rightarrow \infty} F_x(t) = \frac{F_x(t + \Delta t) - F_x(t)}{\Delta t}$$

$$= 1.$$

2.4 Fungsi Kepadatan Peluang

Fungsi kepadatan peluang atau *Probability Density Function* (PDF) merupakan peluang kepadatan kegagalan pada waktu t . Fungsi kepadatan peluang seseorang yang saat ini berusia x tahun, mengalami kegagalan pada waktu t , dengan syarat individu tersebut hidup pada saat $t = 0$ dinotasikan dengan $f_x(t)$. Berlainan dengan $F_x(t)$ dan $S_x(t)$, yang merupakan peluang yang berkaitan dengan suatu interval waktu, $f_x(t)$ memperhatikan titik waktu tertentu. Fungsi kepadatan peluang dapat dituliskan sebagai

$$f_x(t) = \lim_{\Delta t \rightarrow 0} \frac{F_x(t + \Delta t) - F_x(t)}{\Delta t}. \quad (2.6)$$

Fungsi kepadatan peluang didefinisikan sebagai turunan dari $F_x(t)$, sehingga

$$f_x(t) = \frac{d}{dt} F_x(t). \quad (2.7)$$

Dengan mensubstitusikan persamaan (2.5) pada $f_x(t)$, akan diperoleh

$$f_x(t) = -\frac{d}{dt} S_x(t). \quad (2.8)$$

Lebih lanjut, dengan memperhatikan persamaan (2.2), fungsi distribusi kumulatif dapat dituliskan sebagai

$$F_x(t) = \int_0^t f_x(y) dy, \quad (2.9)$$

dan dengan memperhatikan persamaan (2.1), fungsi *survival* dapat dituliskan sebagai

$$S_x(t) = \int_t^\infty f_x(y) dy. \quad (2.10)$$

Kemudian, dengan menggunakan persamaan (2.3), (2.9), dan (2.10), dapat ditulis

$$\int_0^\infty f_x(y) dy = 1. \quad (2.11)$$

2.5 Fungsi Laju Kegagalan

Menurut London [8], fungsi laju kegagalan atau *hazard rate function* dinotasikan sebagai $h(t)$ merupakan peluang terjadinya kegagalan pada interval waktu yang sangat kecil. Fungsi laju kegagalan dapat dituliskan sebagai

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\frac{Pr(t < T < t + \Delta t \wedge T \geq t)}{Pr(T \geq t)}}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T < t + \Delta t \wedge T \geq t)}{Pr(T \geq t) \cdot \Delta t} \end{aligned}$$

$$\begin{aligned}
&= \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T < t + \Delta t)}{Pr(T \geq t) \cdot \delta t} \\
&= \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T < t + \Delta t) - Pr(T \leq t)}{Pr(T \geq t) \cdot \delta t}.
\end{aligned}$$

Dengan menggunakan persamaan (2.1) dan (2.2), diperoleh

$$\begin{aligned}
h_x(t) &= \lim_{\Delta t \rightarrow 0} \frac{F_x(t + \delta t) - F_x(t)}{S_x(t) \cdot \delta t} \\
&= \frac{1}{S_x(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{F_x(t + \delta t) - F_x(t)}{\delta t},
\end{aligned}$$

lebih lanjut dengan memanfaatkan persamaan (2.6) diperoleh

$$h_x(t) = \frac{f_x(t)}{S_x(t)}, \quad (2.12)$$

yang merupakan fungsi laju kegagalan atau *hazard rate* subjek pada waktu x .

2.6 Penyensoran Data

Penyensoran data adalah praktik dalam analisis *survival* karena membantu memperjelas kegunaan suatu data. Penyensoran data dilakukan ketika terdapat informasi yang tidak lengkap dalam data, yaitu ketika tidak diketahui kapan *event* terjadi. *Event* dalam konteks ini berarti sebagai kematian individu yang disebabkan oleh kanker otak. Berdasarkan Frees [11], terdapat tiga jenis penyensoran data.

1. Data tersensor kanan

Data tersensor kanan terjadi ketika subjek yang diamati hidup dari awal waktu pengamatan sampai akhir waktu pengamatan.

2. Data tersensor kiri

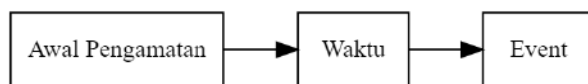
Data tersensor kiri terjadi jika *event* yang diamati sudah terjadi sebelum waktu pengamatan.

3. Data tersensor secara interval

Data tersensor interval terjadi ketika *event* yang diamati terjadi di dalam interval waktu pengamatan, tetapi tidak diketahui waktu tepat *event* terjadi.

2.7 Analisis *Survival*

Analisis *survival* merupakan serangkaian prosedur statistik untuk menganalisis data *survival*, di mana variabel hasil yang dicari adalah waktu hingga terjadinya suatu *event* (kematian). Waktu dalam konteks ini dapat berupa tahun, bulan, minggu, atau mulai dari awal pengamatan terhadap individu sampai terjadinya suatu peristiwa. Variabel waktu dapat dipresentasikan pada Gambar 2.1. Dalam analisis *survival*, umumnya variabel waktu disebut sebagai waktu *survival*. Variabel waktu menunjukkan berapa lama individu telah *survive* (bertahan hidup) selama periode pengamatan. Peristiwa yang menjadi perhatian utama disebut sebagai *event* (kematian). Tujuan analisis *survival* di antaranya



Gambar 2.1: Ilustrasi variabel waktu dan *event* (kematian).

1. melakukan estimasi dan interpretasi fungsi *survival* atau fungsi laju kegagalan;
2. membandingkan dua atau lebih fungsi *survival* atau fungsi laju kegagalan;
3. melihat hubungan antar variabel independen dengan waktu *survival*

2.8 Distribusi Eksponensial

Berdasarkan Ross [12], distribusi eksponensial merupakan salah satu distribusi probabilitas kontinu yang sering digunakan untuk memodelkan waktu antara kejadian yang bersifat acak dalam proses kontinu. Perhatikan bahwa fungsi kepadatan peluang pada distribusi eksponensial ditulis sebagai

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0; \\ 0, & x < 0. \end{cases} \quad (2.13)$$

Kemudian dengan merujuk persamaan (2.9), fungsi distribusi kumulatif dari distribusi eksponensial dapat ditulis sebagai

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \lambda > 0; \\ 0, & x < 0, \end{cases} \quad (2.14)$$

dan dengan memanfaatkan persamaan (2.5), fungsi *survival* dari distribusi eksponensial dapat ditulis sebagai

$$S(x) = e^{-\lambda x}. \quad (2.15)$$

Menurut Ross [12], distribusi eksponensial memiliki sifat *memoryless property* yang berarti bahwa peluang suatu kejadian akan berlangsung tidak bergantung pada waktu yang telah berlalu.

2.9 Metode Iterasi Newton-Raphson

Menurut Dickson [7], metode iterasi Newton-Raphson merupakan metode yang digunakan untuk memperoleh nilai estimasi akar-akar dari suatu sistem persamaan. Misalkan, terdapat sistem persamaan dengan n buah variabel dan n buah persamaan sehingga dapat ditulis sebagai

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0; \\ f_2(x_1, x_2, \dots, x_n) = 0; \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = 0, \end{cases}$$

di mana solusi dari persamaan tersebut adalah suatu nilai (x_1, x_2, \dots, x_n) yang membuat seluruh persamaan bernilai nol. Penyelesaian metode iterasi Newton-Raphson diperoleh dengan menggunakan

ekspansi deret Taylor yang dapat ditulis sebagai

$$f_{k,i+1} = f_{k,i} + (x_{1,i+1} - x_{1,i}) \frac{\partial f_{k,i}}{\partial x_1} + (x_{2,i+1} - x_{2,i}) \frac{\partial f_{k,i}}{\partial x_2} + \cdots + (x_{n,i+1} - x_{n,i}) \frac{\partial f_{k,i}}{\partial x_n}. \quad (2.16)$$

Dalam skripsi ini, metode Newton-Raphson digunakan untuk memaksimalkan fungsi *likelihood* dalam mengestimasi parameter β dengan $\beta = (\beta_1, \beta_2, \dots, \beta_n)$. Adapun langkah-langkah yang perlu dilakukan untuk mengestimasi parameter β menggunakan metode iterasi Newton-Raphson:

1. pilih nilai $\beta_{(0)}$ sebagai nilai estimasi awal dari parameter β ,
2. definisikan sebuah vektor baris $\mathbf{U}(\beta)$ yaitu

$$\mathbf{U}(\beta) = \left[\frac{\partial \ell(\beta)}{\partial \beta_1}, \frac{\partial \ell(\beta)}{\partial \beta_2}, \dots, \frac{\partial \ell(\beta)}{\partial \beta_p} \right] \quad (2.17)$$

di mana $\frac{\partial \ell(\beta)}{\partial \beta_k}$ adalah turunan parsial *log-likelihood* terhadap setiap parameter β_k untuk $k \in \{1, 2, \dots, p\}$,

3. definisikan matriks $I(\beta)$ yang memiliki ukuran $p \times p$ dengan elemen ke (i, j) adalah

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j}$$

Dengan matriks $I(\beta)$ merupakan turunan parsial kedua dari fungsi *partial log-likelihood* terhadap setiap parameter pada vektor β ,

4. aproksimasi parameter β dengan iterasi, di mana iterasi pada langkah ke- $i + 1$ dinyatakan sebagai

$$\beta_{i+1} = \beta_i - I^{-1}(\beta_c) U(\beta_c), c \in \{0, 1, 2, \dots\},$$

5. menghentikan proses iterasi untuk aproksimasi parameter β ketika perbedaan antara nilai β_{i+1} dan β_i memenuhi kriteria akurasi tertentu, demikian hasil estimasi parameter β yang dilambangkan $\hat{\beta}$ adalah β_{i+1} .

BAB 3

METODE ESTIMASI PELUANG *SURVIVAL* PADA PASIEN KANKER OTAK

Pada bab ini, dijelaskan metode dan model *survival* yang digunakan untuk mengestimasi peluang *survival* dari pasien kanker otak. Metode pertama yang digunakan yaitu metode estimator Kaplan-Meier yang merupakan sebuah metode non-parametrik untuk menganalisis fungsi *survival* dari sebuah populasi berdasarkan masing-masing kovariat. Lebih lanjut, digunakan dua model yang melibatkan beberapa kovariat sekaligus yaitu model *survival tree* dan model regresi Cox yang keduanya merupakan model semi-parametrik. Model *survival tree* digunakan untuk mengidentifikasi pola antar kovariat yang membagi data ke dalam kelompok, sedangkan model regresi Cox digunakan untuk menganalisis kovariat yang menjadi pengaruh dominan terjadinya kematian pada pasien kanker otak.

3.1 Estimator Kaplan-Meier

Berdasarkan Frees [11], estimator Kaplan-Meier atau *Product Limit Estimator* adalah metode statistik non-parametrik yang digunakan untuk mengestimasi peluang *survival* dari suatu populasi. Pada dasarnya, estimator Kaplan-Meier bekerja dengan cara mengumpulkan data waktu *survival* dari populasi, terutama dalam studi medis dan ilmiah lainnya. Dalam skripsi ini, estimator Kaplan-Meier akan digunakan pada data pasien kanker otak.

Misalkan $d_1 < d_2 < \dots < d_K$ menyatakan titik waktu kematian (*failure*) pasien-pasien kanker otak. Perhatikan bahwa peluang pasien bertahan lebih lama dari titik waktu d_k (*survive*) dapat ditulis sebagai

$$Pr(T > d_k) = Pr(T > d_k | T > d_{k-1})Pr(T > d_{k-1}) + Pr(T > d_k | T \leq d_{k-1})Pr(T \leq d_{k-1}), \quad (3.1)$$

dengan T merupakan peubah acak waktu kematian pasien. Perhatikan bahwa $d_{k-1} < d_k$, hal ini menyatakan bahwa $Pr(T > d_k | T \leq d_{k-1}) = 0$ dikarenakan tidak mungkin untuk pasien *survive* sampai waktu ke d_k jika pasien tersebut tidak *survive* sampai waktu ke d_{k-1} . Oleh karena itu, persamaan (3.1) dapat disederhanakan dan dinotasikan sebagai

$$\begin{aligned} S(d_k) &= Pr(T > d_k) \\ &= Pr(T > d_k | T > d_{k-1})Pr(T > d_{k-1}) \\ &= Pr(T > d_k | T > d_{k-1})S(d_{k-1}). \end{aligned} \quad (3.2)$$

Dari persamaan (3.2) perhatikan bahwa

$$S(d_{k-1}) = Pr(T > d_{k-1} | T > d_{k-2}) Pr(T > d_{k-2}),$$

sehingga persamaan (3.2) dapat ditulis sebagai

$$S(d_k) = Pr(T > d_k | T > d_{k-1}) \times \cdots \times Pr(T > d_2 | T > d_1) Pr(T > d_1). \quad (3.3)$$

Peluang *survive* pasien pada titik waktu d_j dapat diestimasi dengan rumus

$$\hat{Pr}(T > d_j | T > d_{j-1}) = \frac{r_j - q_j}{r_j}, \quad (3.4)$$

di mana q_j menyatakan banyaknya pasien yang mengalami kematian antara waktu ke- $(j - 1)$ dan waktu ke- j dan r_j merupakan banyaknya pasien yang *survive* pada waktu ke- j . Dengan merujuk pada persamaan (3.3) dan (3.4), estimasi peluang *survival* pasien kanker otak dengan metode estimator Kaplan-Meier dapat ditulis sebagai

$$\hat{S}(d_k) = \prod_{j=1}^k \frac{r_j - q_j}{r_j}. \quad (3.5)$$

Setelah memahami konsep estimator Kaplan-Meier, berikut akan diberikan contoh penerapan pada data waktu bertahan dari 10 pasien kanker otak yang disajikan pada Tabel 3.1. Berdasarkan persamaan (3.5), akan dihitung estimasi peluang suatu pasien tidak mengalami kematian (*survive*). Dari data tersebut, dapat dilihat bahwa pada titik untuk $k = 1, 4, 6, 8$ terdapat pasien yang meninggal karena kanker otak. Maka dari itu dapat dihitung peluang *survive* dari pasien kanker otak pada keempat titik waktu itu. Untuk titik waktu lainnya, yaitu $k = 2, 3, 5, 9$, dan 10, peluang *survival* tidak mengalami perubahan dengan satu titik waktu sebelumnya karena tidak terdapat pasien yang meninggal. Hasil dari perhitungan peluang *survival* pasien kanker otak dapat dilihat pada Tabel 3.2.

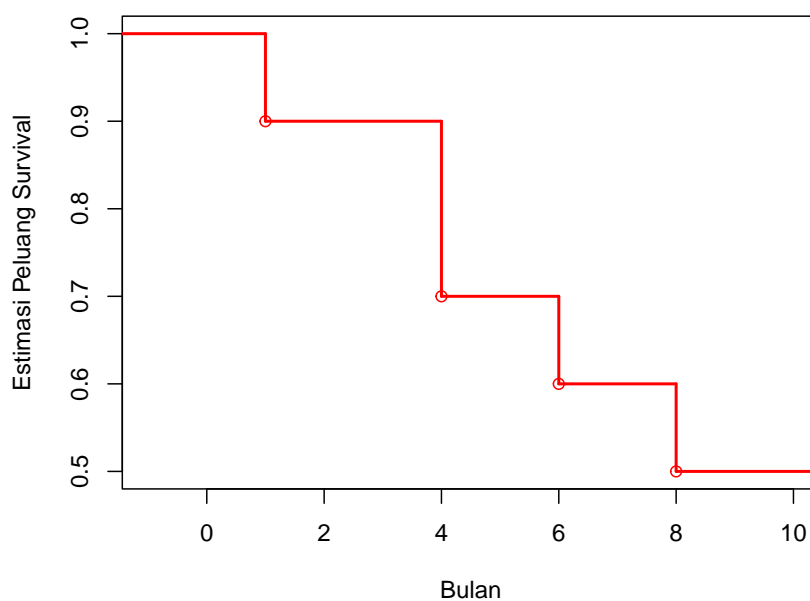
Tabel 3.1: Asumsi data pasien kanker otak yang telah diamati selama 10 bulan

Titik waktu (k)	Jumlah dalam risiko (r_j)	Jumlah kematian (q_j)
1	10	1
2	9	0
3	9	0
4	9	2
5	7	0
6	7	1
7	6	0
8	6	1
9	5	0
10	5	0

Tabel 3.2: Hasil perhitungan metode estimator Kaplan-Meier terhadap data asumsi

k	r_k	q_k	$\hat{S}(d_k)$
1	10	1	$\hat{S}(d_k) = \prod_1^1 \frac{r_k - q_k}{r_k}$ $= \frac{10 - 1}{10}$ $= 0,9$
4	9	2	$\hat{S}(d_k) = \prod_1^4 \frac{r_k - q_k}{r_k}$ $= \frac{10 - 1}{10} \cdot 1 \cdot 1 \cdot \frac{9 - 2}{9}$ $= 0,7$
6	7	1	$\hat{S}(d_k) = \prod_1^6 \frac{r_k - q_k}{r_k}$ $= \frac{10 - 1}{10} \cdot 1 \cdot 1 \cdot \frac{9 - 2}{9} \cdot 1 \cdot \frac{7 - 1}{7}$ $= 0,6$
8	6	1	$\hat{S}(d_k) = \prod_1^8 \frac{r_k - q_k}{r_k}$ $= \frac{10 - 1}{10} \cdot 1 \cdot 1 \cdot \frac{9 - 2}{9} \cdot 1 \cdot \frac{7 - 1}{7} \cdot 1 \cdot \frac{6 - 1}{6}$ $= 0,5$

Dari hasil perhitungan pada Tabel 3.2, dapat dibentuk kurva Kaplan-Meier yang dapat dilihat pada Gambar 3.1.



Gambar 3.1: Kurva Kaplan-Meier untuk contoh data.

Dalam skripsi ini, perhitungan peluang *survival* pasien kanker otak dengan metode estimator

Kaplan-Meier akan dilakukan dengan menggunakan perangkat lunak RStudio. *Package* yang digunakan adalah *survival* dengan perintah *survfit*. Pada RStudio, perintah *survfit* digunakan untuk membentuk kurva peluang *survival*. Fungsi yang digunakan untuk formula tersebut adalah $\text{Surv}(\text{time}, \text{status}) \sim \text{kovariat}$ dengan *time* berisi durasi waktu *survival* untuk data tersensor kanan. Kemudian, *status* berisikan indikator apakah data tersensor atau *event*, dan *kovariat* berisi kovariat yang memiliki potensi pengaruh terhadap kematian pasien kanker otak. Metode ini akan diterapkan menggunakan data pasien kanker otak yang dikumpulkan berdasarkan waktu *survival* yang dijelaskan pada subbab 4.1.

3.2 Survival Tree

Survival tree merupakan model analisis statistik menggunakan pembelajaran mesin yang digunakan untuk menganalisis data *survival* [13]. Model *survival tree* digunakan untuk menangani data yang tersensor dan bertujuan untuk memprediksi kapan suatu peristiwa terjadi dan menganalisis pengaruh utama peristiwa berdasarkan kovariat yang diamati [14]. Lebih lanjut, dalam setiap simpul model *survival tree*, digunakan metode estimator Kaplan-Meier untuk membentuk kurva *survival* pada setiap simpul tersebut.

3.2.1 Exponential Log-Likelihood Loss

Langkah awal dalam pembentukan model *survival tree* yaitu memilih kovariat pertama sebagai simpul akar atau akar pertama, kemudian memilih kovariat lain sebagai daun-daunnya. Pemilihan kovariat ini dilakukan dengan menggunakan pendekatan *exponential log-likelihood loss* [14], di mana setiap pemilihan kovariat untuk suatu simpul berdasarkan pada nilai penyimpangan (*deviance*) dari semua kovariat-kovariat yang dilibatkan. Terlebih dahulu, diasumsikan δ_i sebagai indikator pensensoran dengan kondisi sebagai berikut

$$\delta_i = \begin{cases} 1, & \text{jika data tidak tersensor;} \\ 0, & \text{jika data tersensor,} \end{cases} \quad (3.6)$$

untuk $i \in \{1, 2, \dots, n\}$ dengan n adalah banyaknya data. Lebih lanjut, didefinisikan T_i adalah waktu tersensor pasien dan diasumsikan $T_i \sim \exp(\lambda)$, sehingga fungsi kepadatan peluang untuk T_i otak dapat ditulis sebagai

$$f(T_i | \delta_i) = \begin{cases} \lambda e^{-\lambda T_i}, & \text{jika } \delta_i = 1 ; \\ e^{-\lambda T_i}, & \text{jika } \delta_i = 0, \end{cases} \quad (3.7)$$

dengan λ merupakan laju kematian per unit waktu untuk pasien kanker otak. Perhatikan bahwa persamaan (3.7) dapat ditulis sebagai

$$f(T_i | \delta_i) = (\lambda e^{-\lambda T_i})^{\delta_i} \cdot [e^{-\lambda T_i}]^{1-\delta_i}.$$

Dengan asumsi bahwa setiap pasien tidak memiliki keterkaitan antara satu dengan yang lain

maka fungsi *likelihood* untuk kematian pasien kanker otak ke- i dapat ditulis sebagai

$$L(\lambda) = \prod_{i=1}^n (\lambda e^{-\lambda T_i})^{\delta_i} \cdot [e^{-\lambda T_i}]^{1-\delta_i}. \quad (3.8)$$

Dengan memanfaatkan persamaan tersebut, fungsi *exponential log-likelihood* dapat ditulis sebagai

$$\begin{aligned} l(\lambda) &= \ln L(\lambda) \\ &= \ln \prod_{i=1}^n (\lambda e^{-\lambda T_i})^{\delta_i} \cdot (e^{-\lambda T_i})^{1-\delta_i} \\ &= \sum_{i=1}^n \delta_i \ln(\lambda) - \lambda \sum_{i=1}^n T_i, \end{aligned} \quad (3.9)$$

sehingga dapat dicari estimasi dari λ , yaitu $\hat{\lambda}$ dengan menyelesaikan persamaan

$$\begin{aligned} 0 &= \frac{d \ln L(\lambda)}{d\lambda} \\ &= \frac{d}{d\lambda} \left(\sum_{i=1}^n \delta_i \ln(\lambda) - \lambda \sum_{i=1}^n T_i \right) \\ &= \sum_{i=1}^n \frac{\delta_i}{\lambda} - \sum_{i=1}^n T_i. \end{aligned} \quad (3.10)$$

Berdasarkan persamaan (3.10), estimasi dari λ dapat dicari dengan persamaan

$$\hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n T_i}. \quad (3.11)$$

Lebih lanjut, untuk menentukan kovariat yang paling cocok untuk dipakai pada simpul pertama, dihitung nilai dari persamaan (3.9) dengan nilai λ yang dapat diestimasi menggunakan persamaan (3.11). Kemudian dihitung *deviance* dari persamaan (3.9) dengan rumus

$$D_j = -2 \sum_{k=1}^K n_{j_k} \ln(\lambda_{j_k}), \quad (3.12)$$

dengan j merupakan indikator suatu variabel bebas, k merupakan kategori ke k dalam variabel bebas ke- j , dan n_{j_k} menyatakan banyaknya observasi pada variabel bebas ke- j kategori ke k . Pemilihan kovariat untuk menjadi simpul akar atau simpul batang dalam *splitting* dipilih dengan menghitung nilai *deviance* dari masing-masing kovariat. Kovariat dengan nilai *deviance* yang paling kecil dari kumpulan kovariat-kovariat yang dilibatkan, akan menjadi pilihan untuk simpul akar atau simpul batang [15]. Pada Tabel 3.3, dapat dilihat contoh data pasien kanker otak dengan beberapa kovariat yang akan digunakan untuk diterapkan metode *exponential log-likelihood loss* dalam pemilihan kovariat pertama sebagai simpul akar pada model *survival tree*.

Tabel 3.3: Asumsi data pasien kanker otak

ID	Umur	Gender	Waktu Survive	Status
1	55	Male	10	1
2	60	Female	8	1
3	65	Male	12	0
4	70	Female	15	1

Dengan Status sebagai indikator *event* ($1 = event$ terjadi dan $0 = data$ tersensor). Berdasarkan data dari Tabel 3.3, akan dihitung nilai *deviance* dari Umur dan *Gender* untuk menentukan variabel mana yang lebih cocok digunakan sebagai simpul akar dari model. Hasil dari perhitungan nilai *deviance* untuk *Gender* dan Umur dapat dilihat pada Tabel 3.4 dan Tabel 3.5.

Tabel 3.4: Hasil perhitungan terhadap variabel *Gender*

λ_{Male}	λ_{Female}	D_{Gender}
$\lambda_{Male} = \frac{\sum_{i=1}^4 \delta_i}{\sum_{i=1}^4 T_i}$	$\lambda_{Female} = \frac{\sum_{i=1}^4 \delta_i}{\sum_{i=1}^4 T_i}$	$D_{Gender} = -2(n_{jM} \ln(\lambda_M) + n_{jF} \ln(\lambda_F))$
$= \frac{1}{22}$	$= \frac{2}{23}$	$= -2 \cdot (2 \ln 0,0454 + 2 \ln 0,0869)$
$= 0,0454$	$= 0,0869$	$= 31,3513$

Kemudian untuk menghitung nilai *deviance* dari Umur, dikarenakan variabel tersebut merupakan variabel numerik, perlu mengikuti langkah sebagai berikut:

1. Urutkan nilai Umur dari nilai terkecil. Dalam kasus ini urutan tersebut adalah [55, 60, 65, 70];
2. Menghitung rata-rata untuk setiap dua baris di mana nilai setiap rata-rata merupakan potensi pemisahan. Dalam hal ini, nilai rata-rata untuk setiap dua baris pada variabel Umur adalah [57,5; 62,5; 67,5];
3. Mencari nilai D terbaik untuk setiap titik pemisahan yang telah ditentukan.

Karena proses cukup panjang, maka akan diberikan contoh untuk salah satu titik pemisahan saja. Hasil perhitungan diperlihatkan pada Tabel 3.5. Berdasarkan hasil perhitungan di atas, dapat dilihat bahwa nilai *deviance* dari Umur ($\leq 62,5$ dan $> 62,5$) lebih kecil dibandingkan dengan nilai *deviance* dari *Gender*, maka Umur akan dipilih sebagai simpul akar dari model.

Tabel 3.5: Hasil perhitungan terhadap variabel Umur

$\lambda_{Umur \leq 62,5}$	$\lambda_{Umur > 62,5}$	D_{Umur}
$\lambda_{Umur \leq 62,5} = \frac{\sum_{i=1}^4 \delta_i}{\sum_{i=1}^4 T_i}$	$\lambda_{Umur > 62,5} = \frac{\sum_{i=1}^4 \delta_i}{\sum_{i=1}^4 T_i}$	$D_{Umur} = -2(n_{j \leq 62,5} \ln(\lambda_{\leq 62,5}) + n_{j > 62,5} \ln(\lambda_{> 62,5}))$
$= \frac{2}{18}$	$= \frac{1}{27}$	$= -2 \cdot (2 \ln 0,1111 + 2 \ln 0,0370)$
$= 0,1111$	$= 0,0370$	$= 21,9766$

3.2.2 Pruning

Berdasarkan Stolfo [16], *pruning* adalah suatu metode pemangkasan yang digunakan dalam model *survival tree* yang bertujuan untuk menghasilkan pohon yang optimal. Proses *pruning* dimulai

dengan membangun model *survival tree* yang besar yang dinotasikan sebagai P_0 . Dari P_0 , dilakukan *pruning* secara bertahap dimulai dari simpul paling bawah untuk membentuk serangkaian pohon model *survival tree* yang lebih kecil, yaitu pohon P_i , di mana $i = \{1, 2, \dots, N\}$. Perhatikan bahwa P_N merupakan pohon terkecil yang hanya terdiri dari satu akar. Proses *pruning* dilakukan dengan meminimalkan nilai *cost-complexity*, yaitu kombinasi antara kesalahan klasifikasi dan kompleksitas pohon. Lebih lanjut, nilai *cost-complexity* dapat dihitung dengan rumus

$$R_a(P_i) = R(P_i) + \alpha C(P_i), \quad (3.13)$$

di mana $R(P_i)$ merupakan tingkat kesalahan klasifikasi pada pohon P_i , $C(P_i)$ merupakan ukuran kompleksitas pohon, dan α adalah bobot penalti terhadap ukuran pohon di mana nilai $\alpha \geq 0$.

Semakin besar nilai α , semakin besar penalti terhadap ukuran pohon. Dengan demikian, pohon yang dihasilkan akan memiliki simpul lebih sedikit dikarenakan algoritma akan memprioritaskan pengurangan ukuran pohon dibandingkan meminimalkan kesalahan klasifikasi. Sebaliknya jika nilai α kecil, pohon yang dihasilkan cenderung lebih kompleks karena penalti terhadap ukuran pohon kecil. Metode *pruning* secara iteratif menghasilkan serangkaian pohon yang terurut berdasarkan ukuran dan kompleksitasnya. Pohon terbaik dipilih berdasarkan nilai minimum dari $R_a(P_i)$, yang mengoptimalkan keseimbangan antara akurasi prediksi dan kesederhanaan model.

Dalam skripsi ini, model akan dibuat menggunakan perangkat lunak RStudio. *Package* yang digunakan adalah `Survival` seperti yang sudah dibahas pada metode estimator Kaplan-Meier dan `rpart` dengan perintah `rpart`. Perintah `rpart` pada RStudio digunakan untuk membentuk model *survival tree*. Pada penggunaan `rpart`, formula yang digunakan sama dengan metode Kaplan-Meier, yaitu `Surv(time,event) ~ kovariat` [17]. Adapun langkah-langkah yang lebih rinci dari pengerjaan yang dapat dilihat secara langsung di Bab 4.

3.3 Model Regresi Cox Proportional Hazard (CPH)

Model Regresi Cox merupakan model *survival* semi-parametrik yang digunakan untuk menilai pengaruh dari banyak kovariat secara bersamaan. Model Cox *Proportional Hazard* (CPH) merupakan salah satu model regresi yang digunakan untuk analisis dan mengolah data *survival*. Model ini berguna untuk mengevaluasi akibat dari variabel-variabel secara bersamaan terhadap suatu *event* (kematian). Dalam skripsi ini, model CPH digunakan untuk menganalisis kovariat yang menjadi pengaruh terjadinya kematian pada pasien kanker otak. Berdasarkan Kleinbaum dan Klein [18], model CPH dapat ditulis sebagai

$$\begin{aligned} h(t | \mathbf{X}) &= h_0(t) \exp(\boldsymbol{\beta} \mathbf{X}) \\ &= h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p), \end{aligned} \quad (3.14)$$

dengan $h(t | \mathbf{X})$ menyatakan fungsi tingkat kematian pasien kanker otak pada waktu t dengan memperhitungkan kovariat \mathbf{X} , di mana $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$ menyatakan kovariat yang memengaruhi kematian pasien kanker otak, $h_0(t)$ menyatakan nilai fungsi dasar tingkat kematian pada saat t dan $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ merupakan parameter yang menyatakan koefisien dari kovariat.

Fungsi tingkat kematian dari pasien kanker otak, yang dinotasikan sebagai $h_0(t)$, merupakan

fungsi tingkat kematian saat kovariat $X_1 = X_2 = \dots = X_p = 0$. Hal ini dapat dibuktikan ketika dilakukan substitusi terhadap persamaan (3.14), sehingga

$$\begin{aligned} h(t | \mathbf{X} = \mathbf{0}) &= h_0(t) \exp(\beta_1 \cdot 0 + \beta_2 \cdot 0 + \dots + \beta_p \cdot 0) \\ &= h_0(t). \end{aligned}$$

Lebih lanjut, model CPH dapat diperluas dengan menambahkan efek interaksi antar kovariat. Efek interaksi digunakan ketika diduga fungsi tingkat kematian dan suatu kovariat dipengaruhi oleh kovariat lainnya. Model CPH dengan interaksi dapat ditulis sebagai

$$h(t | \mathbf{X}) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \beta_{ij} X_i X_j) \quad (3.15)$$

dengan β_{ij} merupakan parameter yang menyatakan koefisien dari kovariat X_i yang berinteraksi dengan kovariat X_j .

3.3.1 Estimasi *Likelihood* Untuk Model CPH

Berdasarkan Kleinbaum dan Klein [18], dalam mengestimasi parameter β digunakan *partial likelihood estimation*. Metode *partial likelihood estimation* menghitung peluang bersyarat dari pasien kanker otak yang diamati, yang terpilih dari himpunan pasien kanker otak kemungkinan mengalami kematian. Dengan kata lain, metode ini digunakan untuk mengestimasi data yang tidak tersensor saja. Terlebih dahulu, diasumsikan δ_i sebagai indikator terjadinya penyensoran dengan kondisi sebagai berikut

$$\delta_i = \begin{cases} 1, & \text{jika data tidak tersensor;} \\ 0, & \text{jika data tersensor,} \end{cases} \quad (3.16)$$

untuk $i \in \{1, 2, \dots, n\}$ dengan n sebagai banyaknya data. Lebih lanjut, fungsi *likelihood* untuk pasien ke- i dapat ditulis dengan

$$L_i(\beta) = \begin{cases} f(T_i | \mathbf{X}_i), & \text{jika } \delta_i = 1 ; \\ S(T_i | \mathbf{X}_i), & \text{jika } \delta_i = 0, \end{cases} \quad (3.17)$$

dengan T_i merupakan waktu kematian pasien ke- i dan $\mathbf{X}_i = (X_{1i}, X_{2i}, \dots, X_{pi})$ merupakan kovariat yang memengaruhi kematian pasien ke- i dengan jumlah kovariat maksimum sebanyak p . Lebih lanjut, fungsi *likelihood* pada persamaan (3.17) dapat ditulis sebagai

$$L_i(\beta_i | \delta_i) = [f(T_i | \mathbf{X}_i)]^{\delta_i} \cdot [S(T_i | \mathbf{X}_i)]^{1-\delta_i}.$$

Dengan asumsi bahwa setiap pasien tidak memiliki keterkaitan antara satu dengan yang lain, fungsi *likelihood* untuk kematian pasien kanker otak dapat ditulis sebagai

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n L(\beta | \delta_i) \\ &= \prod_{i=1}^n [f(T_i | \mathbf{X}_i)]^{\delta_i} \cdot [S(T_i | \mathbf{X}_i)]^{1-\delta_i}. \end{aligned}$$

Perhatikan bahwa persamaan di atas dapat ditulis ulang sebagai

$$L(\beta) = \prod_{i=1}^n [h(T_i|\mathbf{X}_i) \cdot S(T_i|\mathbf{X}_i)]^{\delta_i} \cdot [S(T_i|\mathbf{X}_i)]^{1-\delta_i} \quad (3.18)$$

$$= \prod_{i=1}^n [h(T_i|\mathbf{X}_i)]^{\delta_i} \cdot [S(T_i|\mathbf{X}_i)]. \quad (3.19)$$

Misalkan $R(T_i)$ merupakan himpunan pasien kanker otak yang berpotensi untuk mengalami *event* pada waktu T_i . Dengan manipulasi aljabar, persamaan (3.18) dapat diubah menjadi

$$L(\beta) = \prod_{i=1}^n \left[\frac{h(T_i|\mathbf{X}_i)}{\sum_{j \in R(T_i)} h(T_i|\mathbf{X}_j)} \right]^{\delta_i} \cdot \left(\left[\sum_{j \in R(T_i)} h(T_i|\mathbf{X}_j) \right]^{\delta_i} \cdot S(T_i|\mathbf{X}_i) \right). \quad (3.20)$$

Berdasarkan persamaan (3.20), jelas bahwa unsur pengali pertama dan kedua memiliki informasi tentang β dan fungsi laju dasar kematian pasien kanker otak. Perhatikan bahwa unsur pengali pertama memiliki fungsi laju dasar kematian pasien kanker otak sebagai pembilang dan penyebutnya, sehingga dapat membagi satu sama lain. Di sisi lain, unsur pengali kedua tidak berbentuk pecahan sehingga masih memiliki fungsi laju dasar kematian pasien kanker otak. Karena fungsi laju dasar kematian pasien kanker otak tidak diketahui, maka digunakan metode *partial likelihood estimation*, sehingga hanya unsur pengali pertama yang diperhatikan dan unsur pengali kedua akan diabaikan.

Dengan demikian, fungsi *partial likelihood* untuk pasien kanker otak dapat ditulis sebagai

$$L(\beta) = \prod_{i=1}^n \left[\frac{h(T_i|\mathbf{X}_i)}{\sum_{j \in R(T_i)} h(T_i|\mathbf{X}_j)} \right]^{\delta_i}. \quad (3.21)$$

Lebih lanjut, dengan menggunakan persamaan (3.14), persamaan (3.21) dapat ditulis ulang menjadi

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \left[\frac{h_0(T_i) \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})}{\sum_{j \in R(T_i)} h_0(T_i) \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj})} \right]^{\delta_i} \\ &= \prod_{i=1}^n \left[\frac{\exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj})} \right]^{\delta_i}. \end{aligned} \quad (3.22)$$

Dengan memanfaatkan persamaan (3.22), fungsi *partial log likelihood* dapat ditulis sebagai

$$\begin{aligned} l(\beta) &= \ln L(\beta) \\ &= \ln \left(\prod_{i=1}^n \left[\frac{\exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj})} \right]^{\delta_i} \right) \\ &= \sum_{i=1}^n \delta_i \left(\ln \left[\frac{\exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj})} \right] \right) \\ &= \sum_{i=1}^n \delta_i \left(\ln [\exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})] \right. \\ &\quad \left. - \ln \left[\sum_{j \in R(T_i)} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj}) \right] \right) \end{aligned}$$

$$= \sum_{i=1}^n \delta_i (\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}) - \ln \left[\sum_{j \in R(T_i)} \exp(\beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj}) \right]. \quad (3.23)$$

Perhatikan bahwa persamaan (3.23) dapat disederhanakan menjadi

$$l(\beta) = \sum_{i=1}^n \delta_i \left(\beta \mathbf{X}_i - \ln \left[\sum_{j \in R(T_i)} \exp(\beta \mathbf{X}_j) \right] \right), \quad (3.24)$$

sehingga nilai estimasi dari β dapat dicari dengan memaksimalkan $l\beta$ yang dapat diselesaikan dengan persamaan

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta_a} &= 0 \\ &= \frac{\partial \sum_{i=1}^n \delta_i \left(\beta \mathbf{X}_i - \ln \left[\sum_{j \in R(T_i)} \exp(\beta \mathbf{X}_j) \right] \right)}{\partial \beta_a} \\ &= \sum_{i=1}^n \delta_i \left(\mathbf{X}_i - \frac{\sum_{j \in R(T_i)} X_j \exp \beta \mathbf{X}_j}{\sum_{j \in R(T_i)} \exp(\beta \mathbf{X}_j)} \right). \end{aligned} \quad (3.25)$$

Perhatikan bahwa estimasi parameter pada model CPH dengan efek interaksi menggunakan cara yang serupa dengan model CPH tanpa interaksi. Lebih lanjut, persamaan (3.25) dapat diselesaikan dengan menggunakan metode iterasi Newton-Raphson untuk memperoleh nilai dari β_a , dengan $a \in \{1, 2, \dots, p\}$. Menurut Pahlevi [5], jika nilai $\beta_a > 0$ atau $\exp(\beta_a) > 1$, maka efek dari kovariat akan semakin besar terhadap tingkat *survival* pasien kanker otak untuk mengalami kematian. Sebaliknya, jika $\beta_a < 0$ atau $\exp(\beta_a) < 1$ maka efek dari kovariat semakin kecil terhadap tingkat *survival* pasien kanker otak untuk mengalami kematian.

3.3.2 Asumsi *Proportional Hazard*

Setelah membangun model CPH, langkah selanjutnya adalah menguji asumsi *proportional hazard* (PH). Berdasarkan Collet [19], uji ini bertujuan untuk memastikan bahwa model yang dibangun sesuai untuk dianalisis menggunakan pendekatan CPH. Asumsi PH mengindikasikan bahwa rasio *hazard* antara kelompok tetap konstan sepanjang waktu. Jika asumsi ini tidak terpenuhi, hasil interpretasi dari model kemungkinan besar tidak akurat ataupun bias.

Dalam skripsi ini, pendekatan utama dalam menguji asumsi PH yang digunakan yaitu menggunakan Schoenfeld *residuals test*. Pendekatan *Schoenfeld residuals test* melibatkan analisis statistik menggunakan *p-value*. Nilai *p-value* dari Schoenfeld *residuals test* akan memberikan indikasi tentang signifikansi statistik dari uji tersebut yang akan menentukan apakah asumsi PH terpenuhi atau tidak. Adapun langkah-langkah pengujian dengan Schoenfeld *residuals test* sebagai berikut

1. membangun model CPH dan menghitung nilai Schoenfeld *residuals test* untuk setiap kovariat;
2. mengurutkan data berdasarkan waktu kejadian (*event*);
3. menguji korelasi antara Schoenfeld *residuals test* dan urutan waktu kejadian tersebut.

Lebih lanjut, rumusan hipotesis dalam melakukan uji asumsi PH adalah

H_0 : Kovariat bersifat tidak signifikan;

H_1 : Kovariat bersifat signifikan.

Di mana ada keputusan yang dibuat berdasarkan nilai p -value sebagai berikut

1. jika p -value $< 0,05$, maka H_0 ditolak, artinya asumsi PH tidak terpenuhi;
2. jika p -value $> 0,05$ maka H_0 diterima, yang menunjukkan bahwa asumsi PH terpenuhi.

Asumsi PH dianggap terpenuhi apabila hasil analisis menunjukkan bahwa kovariat tidak memiliki signifikansi statistik. Sebaliknya, apabila sebuah kovariat bersifat signifikan secara statistik, maka asumsi PH tidak terpenuhi. Jika asumsi PH tidak terpenuhi, alternatif yang dapat dilakukan adalah menggunakan model *Stratified Cox* (SC), yang digunakan untuk mengelompokan data berdasarkan kovariat yang melanggar asumsi PH.

3.3.3 Penentuan Model Terbaik

Dalam analisis CPH, pemilihan model terbaik dilakukan dengan mempertimbangkan berbagai kombinasi kovariat untuk membentuk beberapa model. Kovariat yang digunakan pada model dipilih berdasarkan pengaruhnya yang signifikan terhadap *event* yang diamati. Proses pemilihan model terbaik menggunakan *Akaike Information Criterion* (AIC). AIC merupakan nilai yang mempertimbangkan keseimbangan antara kecocokan model dan kompleksitasnya, sehingga model dengan nilai AIC terkecil dianggap sebagai model terbaik. Berdasarkan Collet [19], nilai AIC dapat dihitung berdasarkan rumus

$$AIC = -2 \ln l(\beta) + 2P,$$

dengan $l(\beta)$ merupakan nilai *log-likelihood* untuk parameter β dan P merupakan jumlah parameter dalam model. Dalam skripsi ini, nilai AIC dihitung menggunakan perangkat lunak RStudio dengan perintah AIC.

3.3.4 Uji Signifikansi Model

Setelah model CPH dibentuk, berikutnya perlu untuk menguji signifikansi model untuk menentukan apakah model yang dihasilkan signifikan secara statistik atau tidak. Dalam skripsi ini, akan digunakan dua uji signifikansi model, yaitu uji serentak dan uji parsial [20].

Uji Serentak

Uji serentak merupakan uji statistik yang digunakan untuk mengevaluasi signifikansi model secara keseluruhan dengan mempertimbangkan semua kovariat secara bersamaan. Uji ini dilakukan dengan memanfaatkan *likelihood ratio test*. Bentuk uji statisik *likelihood ratio test* yang digunakan adalah

$$LR = -2 [\ln(L_0) - \ln(L_p)],$$

di mana L_0 merupakan fungsi likelihood tanpa parameter β , dan L_p sebagai fungsi *likelihood* yang mempertimbangkan B_i , dengan $i \in \{1, 2, \dots, p\}$ Adapun hipotesis H_0 dan H_1 yang diasumsikan

sebagai:

H_0 : Model tidak signifikan;

H_1 : Model signifikan.

Keputusan diambil berdasarkan nilai *p-value* sebagai berikut

1. jika $p\text{-value} \leq \alpha$, $\alpha = 0,05$ maka H_0 ditolak, yang berarti model signifikan secara keseluruhan,
2. jika $p\text{-value} > \alpha$, $\alpha = 0,05$, maka H_0 gagal ditolak, yang berarti model tidak signifikan secara keseluruhan.

Uji Parsial

Uji parsial atau sering disebut sebagai Uji Wald merupakan uji statistik yang digunakan untuk mengevaluasi signifikansi masing-masing kovariat dalam model. Tujuan dari uji ini adalah menentukan apakah setiap kovariat secara individual memiliki pengaruh signifikan terhadap kejadian yang diamati. Berikut merupakan bentuk statistik Uji Wald

$$Z^2 = \frac{\beta_i}{SE(\beta_i)}.$$

Adapun hipotesis H_0 dan H_1 dari uji parsial yang ditulis sebagai

H_0 : Kovariat tidak signifikan;

H_1 : Kovariat signifikan.

Keputusan diambil berdasarkan nilai *p-value* berupa

1. jika $p\text{-value} \leq \alpha$, $\alpha = 0,05$ maka H_0 ditolak, yang berarti kovariat signifikan;
2. jika $p\text{-value} > \alpha$, $\alpha = 0,05$ maka H_0 gagal ditolak, yang berarti kovariat tidak signifikan.

Dalam skripsi ini, nilai *p-value* dari uji serentak dan uji parsial diperoleh dengan menggunakan perangkat lunak RStudio. *Package* dan perintah yang digunakan disesuaikan dengan model CPH.

BAB 4

ANALISIS PADA KEMATIAN PASIEN KANKER OTAK

Pada bab ini, dilakukan analisis terhadap data *survival* pasien kanker otak untuk mengestimasi peluang *survival* dan pengaruh kovariat-kovariat terhadap tingkat *survival* pasien kanker otak. Pembahasan dimulai dengan deskripsi data yang digunakan, termasuk kovariat-kovariat yang menjadi potensi pengaruh kematian pasien kanker otak seperti *Gender*, *Diagnosis*, *Lokasi Tumor*, *Karnofsky Index*, *Ukuran Tumor*, dan *Metode Terapi*. Selanjutnya, diterapkan metode estimator Kaplan-Meier dan model *survival tree* untuk estimasi peluang *survival*, kemudian model regresi Cox diterapkan untuk analisis hubungan kovariat dengan tingkat *survival* pasien kanker otak. Setiap metode akan dijelaskan langkah-langkah penerapannya, hasil yang diperoleh, serta interpretasinya pada estimasi peluang *survival* dan pengaruh masing-masing kovariat pada tingkat *survival* pasien kanker otak.

4.1 Data

Pada penelitian ini, subjek penelitian yang digunakan adalah data *survival* pasien kanker otak dari suatu studi. Kumpulan data ini dirancang untuk mendukung penelitian lebih lanjut dalam bidang onkologi dan statistik medis, serta menyediakan dasar untuk analisis data yang dapat digunakan dalam penelitian akademik dan pengembangan model prediktif. Kumpulan data ini memuat 88 data *survival* pasien kanker otak yang dilampirkan pada Lampiran A, sebagian kecil dari data dapat dilihat dalam Tabel 4.1.

Tabel 4.1: Tabel Data Pasien Kanker Otak

No	Variabel yang dipertimbangkan						status	time	j	r_j	q_j
	<i>sex</i>	diagnosis	loc	ki	gtv	stereo					
1	Male	HG glioma	Supratentorial	70	33,69	SRT	1	0,07	1	87	1
2	Male	LG glioma	Infratentorial	90	30,41	SRT	0	1,18			
3	Male	Meningioma	Supratentorial	70	0,97	SRT	1	1,41	2	86	1
4	Female	Meningioma	Infratentorial	70	2,94	SRS	0	1,54			
5	Female	Meningioma	Supratentorial	80	1,57	SRT	0	2,03	3	86	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
88	Male	HG glioma	Supratentorial	90	10,8	SRT	0	82,56	83	53	0

Lebih lanjut akan dijelaskan variabel-variabel tentang pasien yang telah didiagnosis menderita kanker otak serta notasi yang dipakai pada Tabel 4.1.

1. *Sex (Gender)*

Variabel ini terdiri dari dua kategori yang menyatakan jenis kelamin pasien yang di antaranya terdapat 45 pasien perempuan dan 43 pasien laki-laki.

2. **Diagnosis**

Variabel diagnosis terdiri dari empat kategori yang menyatakan diagnosis kanker yang diderita oleh pasien. Diagnosis kanker pada variabel ini dibagi menjadi meningioma, *low-grade* (LG) glioma, *high-grade* (HG) glioma, dan diagnosis kanker lainnya.

3. **Loc (Lokasi Tumor)**

Variabel ini menyatakan lokasi tumor pasien yang dibagi menjadi dua kategori, yaitu *infratentorial* (terletak pada otak kecil atau batang otak) dan *supratentorial* (terletak pada bagian lobus otak).

4. **KI (Karnofsky Index)**

Karnofsky index merupakan skala yang digunakan untuk menilai tingkat fungsi dan kemampuan pasien dalam melakukan aktivitas sehari-hari dengan skor 0 (meninggal) sampai 100 (fungsi normal tanpa gejala penyakit).

5. **GTV (Ukuran Tumor)**

Gross tumor volume menyatakan volume tumor pasien, baik di dalam maupun di luar bagian tubuh dalam satuan cm^3 .

6. **Stereo (Metode Terapi)**

Variabel ini menggambarkan metode stereotaktik yang terdiri dari dua kategori, yaitu *radiosurgery* stereotaktik (SRS) atau radioterapi stereotaktik fraksinasi (SRT), seperti yang ditampilkan pada Tabel 4.1. Variabel ini mencatat apakah pasien menerima salah satu dari kedua metode tersebut.

7. **Status**

Status merupakan variabel yang menyimpan data sensor pasien kanker otak, di mana 1 menyatakan pasien telah mengalami kematian, dan 0 menyatakan pasien *survive* atau data tersensor.

8. **Time**

Variabel *time* menyimpan data waktu pasien mengalami kematian ketika nilai variabel status = 1, dan waktu pasien *survive* atau tersensor pada saat variabel status = 0.

9. **Notasi j, r_j , dan q_j**

Seperti dengan yang sudah dibahas pada Bab 3.1, notasi j menyatakan titik waktu dalam bulan, dengan nilai $j \in \{1, 2, \dots, k\}$. Lebih lanjut, r_j merupakan banyaknya pasien kanker otak yang *survive* pada waktu ke j , q_j merupakan banyaknya pasien kanker otak yang mengalami kematian pada waktu ke j .

4.2 Aplikasi Metode Estimator Kaplan-Meier pada Data Pasien Kanker Otak

Analisis awal dilakukan dengan melihat efek dari masing-masing variabel terhadap terjadinya kematian pasien kanker otak menggunakan metode estimator Kaplan-Meier. Berikut akan ditunjukkan contoh perhitungan estimasi peluang *survival* Kaplan-Meier yang mempertimbangkan variabel *Gender* secara manual menggunakan tiga data awal yang diurutkan dengan waktu. Informasi dapat dilihat pada Tabel 4.2.

Tabel 4.2: Tabel data terhadap variabel *Gender*

No	<i>Gender</i>	status	time (bulan)	k	r_k	r_k Male	r_k Female	q_k	q_k	Male	q_k	Female
1	Male	1	0,07	1	87	42	45	1	1	1	0	0
2	Male	0	1,18									
3	Male	1	1,41	2	86	41	45	1	1	1	0	0
4	Female	0	1,54									

Dalam perhitungan estimasi peluang *survival* terhadap variabel *Gender*, akan digunakan formula yang sudah dibahas pada persamaan (3.5). Hasil dari perhitungan terhadap dua titik waktu pertama dapat dilihat pada Tabel 4.3.

Tabel 4.3: Hasil perhitungan metode Kaplan-Meier untuk variabel *Gender*

k	$S(d_k)$ Male	$S(d_k)$ Female
1	$S(d_1) = \prod_{j=1}^1 \frac{r_k - q_k}{r_k}$ $= \frac{42 - 41}{42}$ $= 0,976$	$S(d_1) = \prod_{j=1}^1 \frac{r_k - q_k}{r_k}$ $= \frac{45 - 0}{45}$ $= 1$
2	$S(d_2) = \prod_{j=1}^2 \frac{r_k - q_k}{r_k}$ $= \frac{42 - 41}{42} \cdot \frac{41 - 1}{41}$ $= 0,975$	$S(d_2) = \prod_{j=1}^2 \frac{r_k - q_k}{r_k}$ $= \frac{45 - 0}{45} \cdot \frac{45 - 0}{45}$ $= 1$

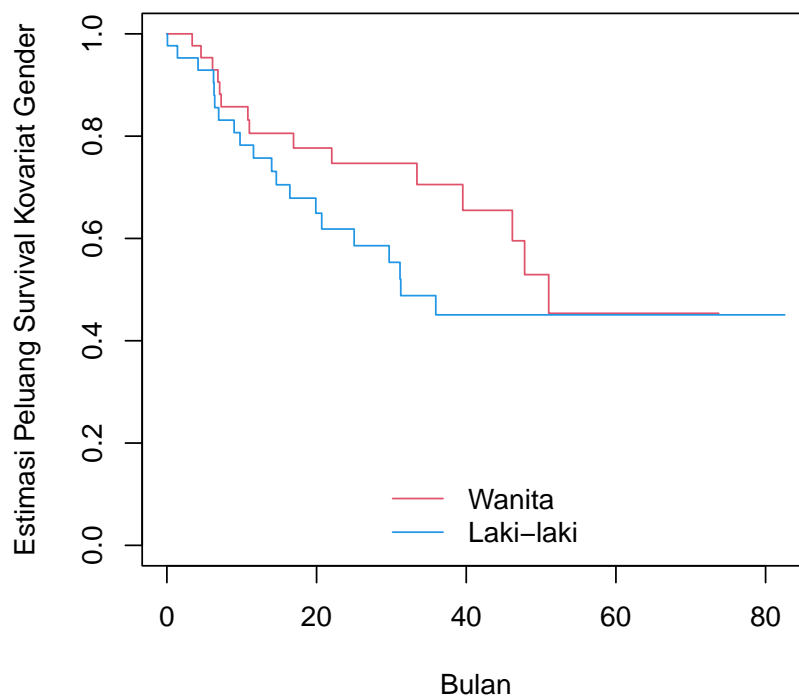
Karena data yang cukup banyak, untuk mempermudah perhitungan, maka sesuai dengan yang dibahas pada Bab 3.1, akan digunakan perangkat lunak RStudio. Berikut merupakan langkah-langkah pengerjaan metode Kaplan-Meier menggunakan RStudio:

1. memanggil *package survival*;
2. mendefinisikan *BrainCancer* sebagai variabel yang menyimpan data *survival* dari 88 pasien kanker otak yang telah dijelaskan pada subbab 4.1;
3. membentuk variabel yang berisi objek dari analisis *survival*;
4. menuliskan formula *survival* metode estimator Kaplan-Meier sesuai dengan variabel yang ingin dilihat seberapa besar pengaruhnya terhadap kematian pasien kanker otak, misalnya `survfit(Surv(time, status) ~ variabel)`. Dalam hal ini, variabel yang akan dilihat adalah *Gender*, *Diagnosis*, *Lokasi Tumor*, dan *Metode Terapi*. Variabel tersebut mengacu

pada variabel yang dijelaskan pada subbab 4.1 yaitu *Gender*, *diagnosis*, *loc*, dan *stereo*, secara berurutan;

5. membentuk kurva Kaplan-Meier dari hasil yang diperoleh dengan menggunakan perintah `plot`;
6. menganalisis apakah masing-masing variabel yang diamati memiliki pengaruh dominan terhadap kematian pasien kanker otak;
7. menentukan variabel yang memiliki pengaruh dominan terhadap terjadinya kematian pasien kanker otak.

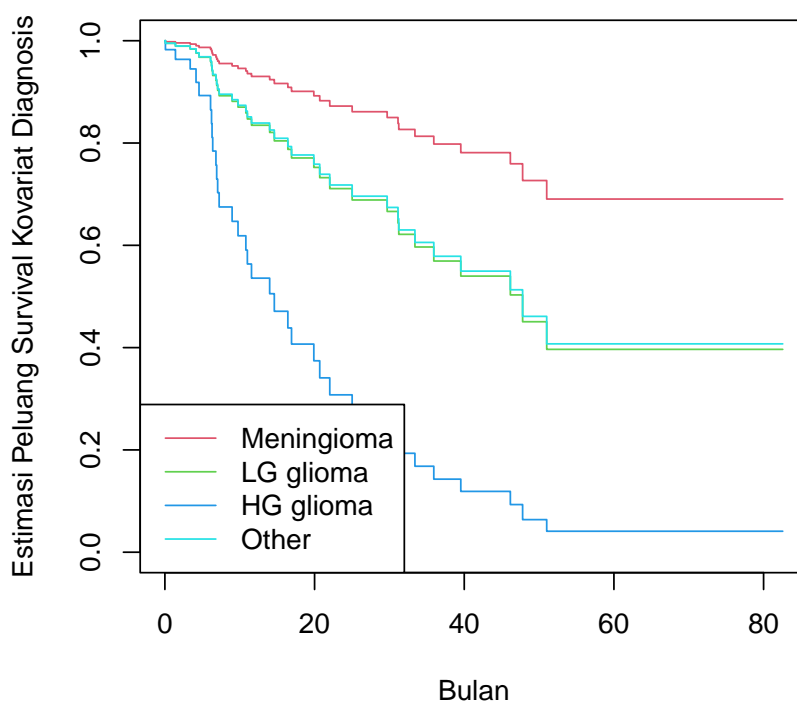
Dengan mengikuti langkah-langkah pengerjaan di atas, akan dibentuk kurva Kaplan-Meier untuk masing-masing variabel yang terdapat pada data pasien kanker otak. Berikut merupakan hasil kurva Kaplan-Meier untuk pasien kanker otak.



Gambar 4.1: Kurva Kaplan-Meier untuk variabel *Gender*.

Variabel memiliki pengaruh dalam kematian pasien kanker otak dilihat dari kurva Kaplan-Meier yang semakin menurun saat durasi waktu pengamatan pasien semakin besar. Kurva pada Gambar 4.1 menunjukkan perbandingan peluang *survival* pasien berdasarkan variabel *Gender*. Dapat dilihat bahwa pasien perempuan memiliki peluang *survival* yang lebih tinggi dibandingkan pasien laki-laki, meskipun pada akhirnya tidak terlihat perbedaan saat pasien telah diamati selama 50 bulan atau lebih. Hal ini menunjukkan kemungkinan adanya faktor respons yang berbeda berdasarkan jenis kelamin.

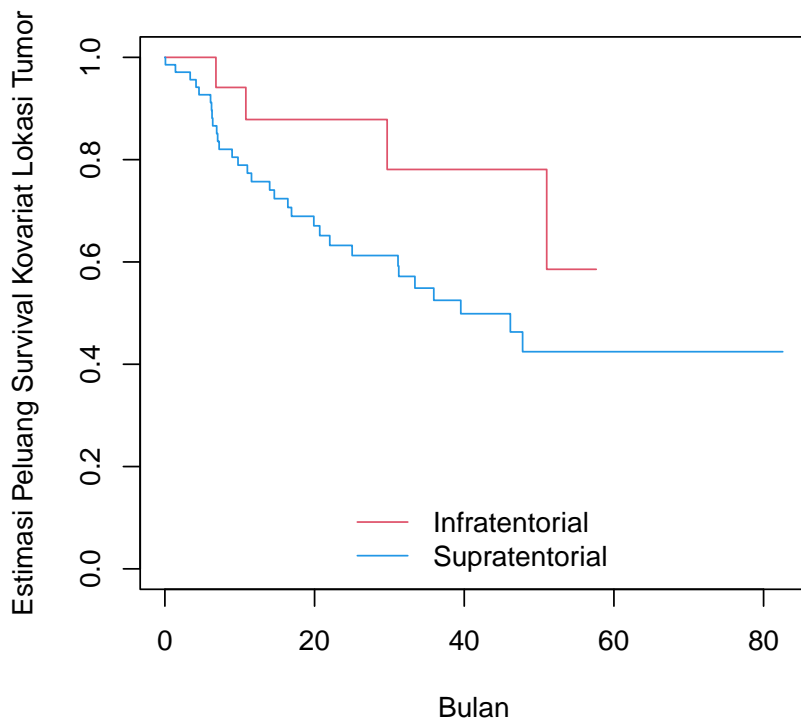
Gambar 4.2 menunjukkan bahwa peluang *survival* pasien kanker otak dengan jenis diagnosis meningioma memiliki probabilitas bertahan hidup paling tinggi dibandingkan dengan jenis diagnosis lainnya, sedangkan pasien dengan diagnosis HG glioma menunjukkan peluang *survival* yang paling rendah. Peluang *survival* pasien dengan diagnosis LG glioma dan *other* berada di antara meningioma dan HG glioma. Dengan memperhatikan kurva dari masing-masing kategori diagnosis pada Gambar 4.2, dapat disimpulkan pasien dengan diagnosis HG glioma memiliki peluang *survival* yang sangat kecil, LG glioma dan *Other* memiliki peluang *survival* yang serupa, sementara Meningioma memiliki peluang *survival* yang paling tinggi dibandingkan dengan jenis tumor lainnya.



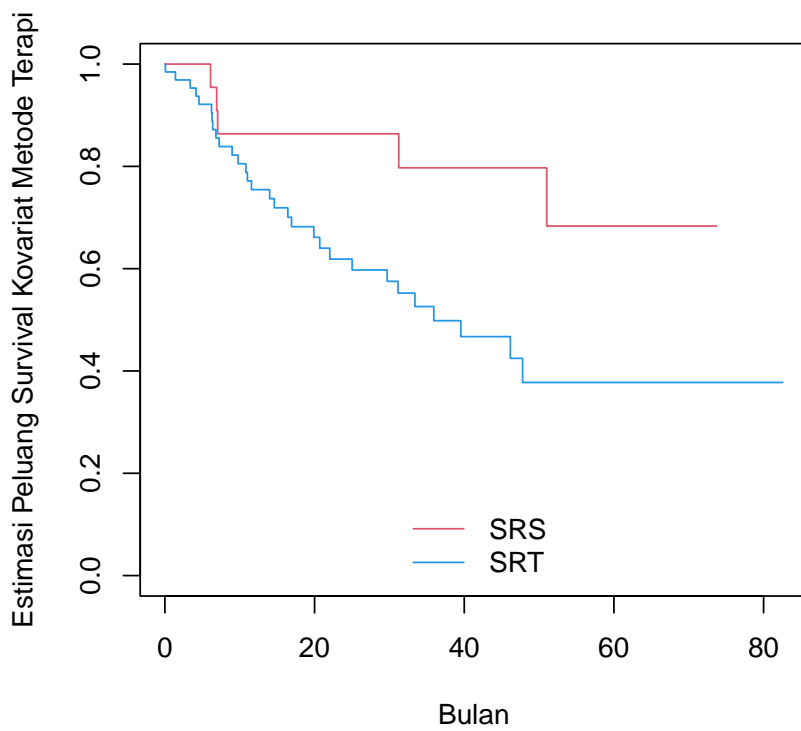
Gambar 4.2: Kurva Kaplan-Meier untuk variabel *Diagnosis*.

Kurva pada Gambar 4.3 menunjukkan perbandingan peluang *survival* berdasarkan lokasi tumor pasien kanker otak. Dapat dilihat bahwa pasien dengan kategori lokasi tumor *infratentorial* menunjukkan peluang *survival* lebih tinggi dibandingkan dengan pasien dengan kategori lokasi tumor *supratentorial* dari awal pengamatan. Dalam hal ini, dapat disimpulkan bahwa pasien dengan lokasi tumor *infratentorial* memiliki peluang *survival* yang lebih baik secara keseluruhan dibandingkan dengan pasien yang memiliki lokasi tumor pada bagian *supratentorial*.

Pada Gambar 4.4 diperlihatkan kurva yang membandingkan peluang *survival* pasien kanker otak berdasarkan metode terapi SRS dan metode terapi SRT. Dapat diperoleh informasi juga bahwa pasien yang menerima metode terapi SRS menunjukkan peluang *survival* yang lebih tinggi dibandingkan pasien yang menerima metode terapi SRT dari awal pengamatan. Hal ini mengindikasikan bahwa pasien yang menggunakan metode terapi SRS memiliki peluang *survival* lebih besar secara keseluruhan dibandingkan dengan pasien yang menggunakan metode terapi SRT.



Gambar 4.3: Kurva Kaplan-Meier untuk variabel Lokasi Tumor.



Gambar 4.4: Kurva Kaplan-Meier untuk variabel Metode Terapi.

4.3 Aplikasi Model *Survival Tree* pada Data Pasien Kanker Otak

Analisis lebih lanjut akan dimodelkan peluang *survival* pasien kanker otak menggunakan model *survival tree*. Model *survival tree* memungkinkan untuk mendiagnosis peluang *survival* pasien kanker otak secara bertahap. Pembangunan model dilakukan dengan menggunakan pendekatan *exponential log-likelihood loss* yang sudah dibahas pada bagian 3.2.1. Adapun langkah-langkah pembentukan model *survival tree* menggunakan perangkat lunak RStudio:

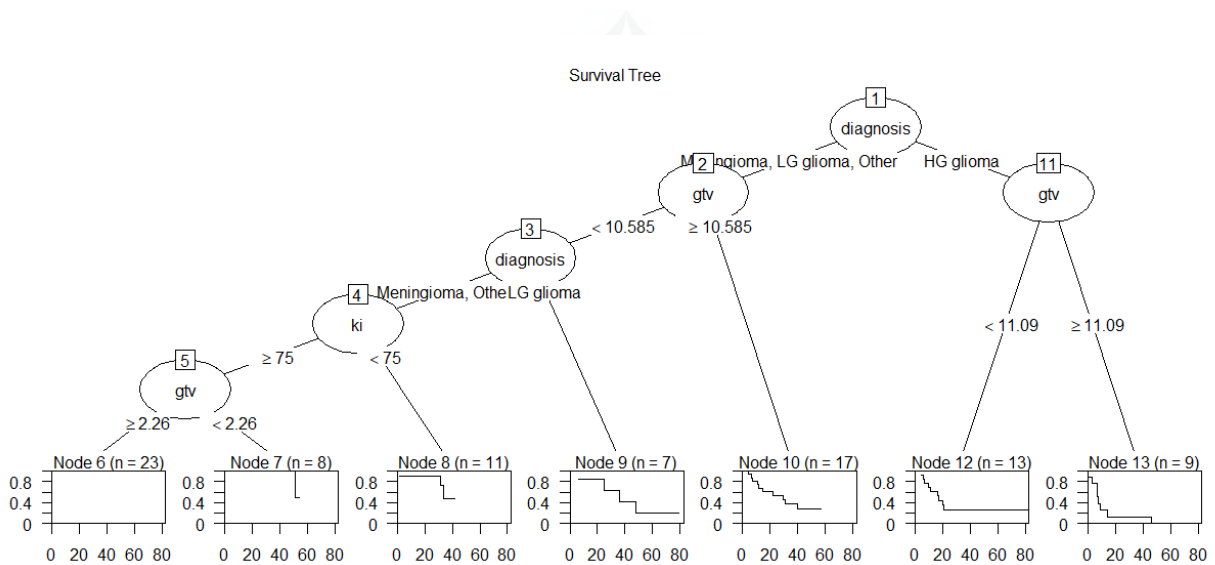
1. memanggil package `ISLR2`, `survival`, `rpart`, dan `rpart.plot` dengan menuliskan perintah `library(ISLR2)`, `library(survival)`, `library(rpart)`, dan `library(rpart.plot)`;
2. mendefinisikan `BrainCancer` sebagai variabel yang menyimpan data 88 pasien kanker otak;
3. menggunakan fungsi `Surv()` untuk membentuk kurva objek *survival* dari data `BrainCancer` dengan variabel `time` dan `status`. Objek *survival* ini berguna untuk membagi status pasien yang telah mengalami kegagalan dan pasien tersensor;
4. mencari nilai *deviance* terbesar dari masing-masing kovariat untuk menjadi kandidat simpul akar model *survival tree*. Pada masing-masing sisi simpul hasil *splitting* simpul kiri dan kanan akar, diterapkan pencarian kovariat lagi untuk dilakukan *splitting* kembali. Langkah ini diulang hingga *splitting* tidak dapat dilakukan lagi, di mana jika banyaknya observasi pada masing-masing simpul kurang dari 10;
5. membentuk model *survival tree* dengan menggunakan kovariat-kovariat `Gender`, `Diagnosis`, `Lokasi Tumor`, `Karnofsky Index`, `Ukuran Tumor`, dan `Metode Terapi`;
6. melakukan *pruning* atau pemangkasan sebagaimana yang telah dibahas pada bagian 3.2.2 terhadap model *survival tree* dengan nilai α terbaik. Dalam perangkat lunak RStudio, α pada persamaan (3.13) dinotasikan sebagai `cp`, dan perintah yang digunakan adalah `plotcp`;
7. Nilai `cp` pada langkah 6 dipilih nilai `cp = 0` untuk menunjukkan struktur *survival tree* yang paling kompleks. Dengan `plotcp` dapat diperoleh ragam nilai-nilai `cp` beserta nilai *error*. Selanjutnya, model terbaik dipilih berdasarkan nilai *error* terkecil dari `plotcp`.
8. memvisualisasikan hasil dari model *survival tree* yang telah dipangkas dengan perintah `rpart.plot`;
9. membuat interpretasi dari model *survival tree* yang telah dipangkas, berdasarkan kovariat penting, yaitu kovariat yang muncul dalam model *survival tree* beserta simpul daun dari struktur model *survival tree* yang terbentuk.

Dengan merujuk pada langkah-langkah pembentukan model *survival tree* di atas, akan dibentuk model dengan simpul akar berupa kovariat dengan nilai *deviance* yang tertinggi. Dengan cara yang sudah dibahas pada bagian 3.2.1 dan bantuan perangkat lunak RStudio, diperoleh nilai *deviance* untuk setiap kovariat untuk penentuan akar yang dapat dilihat pada Tabel 4.4 dan Gambar 4.5. Hal ini mengindikasikan bahwa kovariat `Diagnosis` pada keseluruhan pasien kanker otak memiliki pengaruh paling besar terhadap waktu *survival*. Dalam gambar tersebut, ada 6 *splitting* yang

dilakukan berdasarkan nilai *deviance* terbaik, dan *splitting* tidak dilakukan lagi pada simpul **5** dan **11** sehingga simpul tersebut merupakan simpul daun. Simpul-simpul tersebut menjadi simpul daun karena banyaknya observasi dalam simpul tersebut kurang dari 10 observasi.

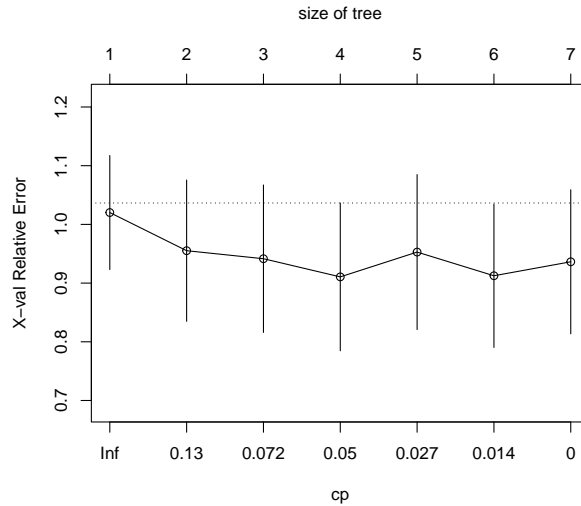
Tabel 4.4: Nilai *deviance* yang diperoleh dari model

Simpul	Kovariat	Nilai Deviance
1	Diagnosis	114,240961
2	Ukuran Tumor	66,895855
3	Diagnosis	37,499114
4	<i>Karnofsky Index</i>	24,270788
5	Ukuran Tumor	19,891534
6	Ukuran Tumor	17,775835



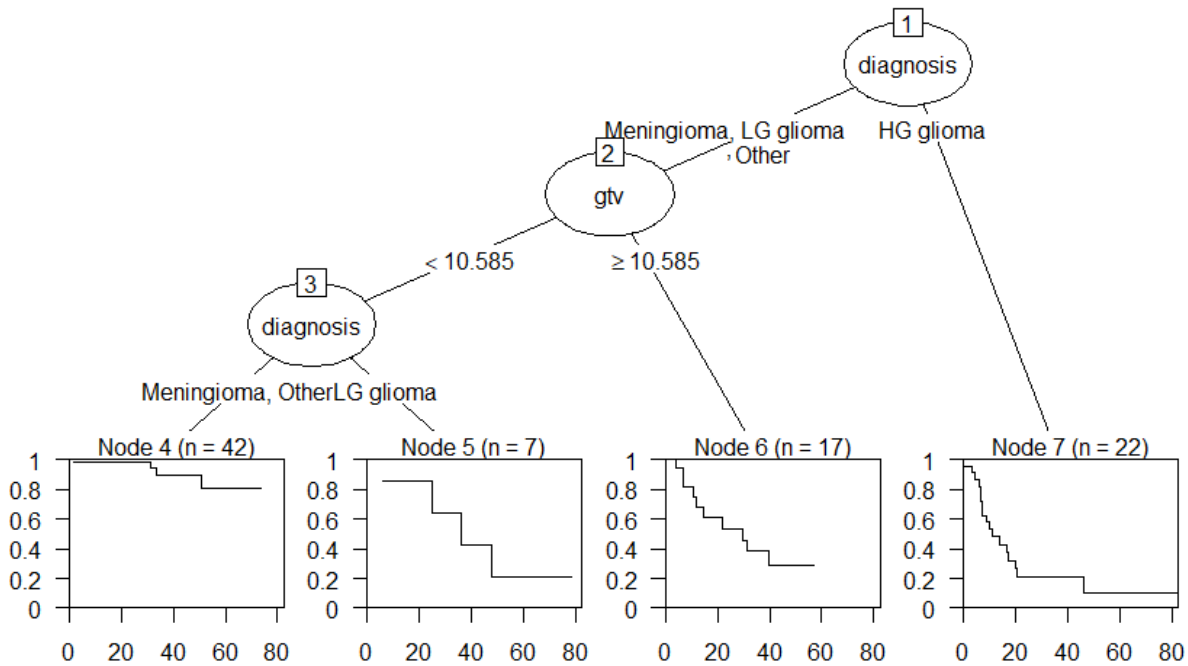
Gambar 4.5: Model *survival tree* awal.

Perhatikan bahwa nilai *cp* pada model *survival tree* Gambar 4.5 adalah 0. Hal ini menyebabkan model *survival tree* kompleks dan fit terhadap data. Oleh karena itu, perlu diterapkan *pruning* terhadap model *survival tree* awal sehingga diperoleh model yang optimal. *Pruning* model *survival tree* dibantu dengan perangkat lunak RStudio menggunakan perintah `plotcp`. Hasil dapat dilihat pada Gambar 4.6, di mana akan dipilih banyaknya simpul daun dengan nilai *error* terkecil sebagai hasil *pruning* dari model *survival tree* dengan *cp* = 0,05.



Gambar 4.6: Hasil *pruning* dari model *survival tree*.

Berdasarkan gambar tersebut, dapat dilihat bahwa nilai *error* paling kecil saat model terdiri dari 4 daun. Lebih lanjut diperoleh informasi dari gambar tersebut bahwa nilai *cp* atau α untuk persamaan (3.13) yang optimal adalah 0,05, sehingga akan dibentuk model *survival tree* baru yang terdiri dari empat daun. Hasil model *survival tree* yang dibentuk terhadap data pasien kanker otak setelah diterapkan *pruning* dapat dilihat pada Gambar 4.7. Model tersebut terdiri dari tiga simpul dan empat daun yang masing-masing menggambarkan pemisahan berdasarkan kovariat yang berpengaruh pada waktu *survival* pasien.



Gambar 4.7: Model *survival tree*.

Berdasarkan model *survival tree* yang dapat dilihat pada Gambar 4.7, struktur pohon mencer-

minkan pola dan hierarki pengaruh kovariat terhadap peluang *survival* pasien. Berdasarkan Gambar tersebut, dengan memperhatikan kurva *survival* sebelah kanan bawah dengan **Node 7**, diperoleh informasi bahwa diagnosis kategori HG glioma memiliki peluang *survival* kanker otak yang sangat signifikan dibandingkan diagnosis lainnya sehingga diagnosis kategori HG glioma dipisahkan dengan diagnosis kategori lainnya. Sementara, pada pasien dengan diagnosis meningioma, LG glioma, dan *Other* belum cukup jelas dalam menjelaskan peluang *survival* sehingga dilakukan *splitting* kembali. Pembagian selanjutnya dilakukan berdasarkan kovariat Ukuran Tumor yang dinotasikan dengan *gtv* pada batas *splitting* 10,585. Hal ini menunjukkan bahwa kovariat Ukuran Tumor memiliki pengaruh cukup besar terhadap peluang *survival* pasien dengan diagnosis meningioma, LG glioma, dan *others*. Dapat dilihat kurva *survival* pada **Node 6** bahwa pasien dengan ukuran tumor besar ($\geq 10,585$) terlihat memiliki peluang *survival* yang cenderung jelas sehingga tidak diperlukan *splitting* lagi.

Selanjutnya, dilakukan *splitting* lagi pada pasien dengan ukuran tumor kecil ($< 10,585$) dibagi kembali berdasarkan pasien yang terdiagnosa Meningioma, *Other*, dan LG glioma, di mana pasien dengan diagnosis meningioma memiliki tingkat *survival* yang lebih tinggi dibandingkan dengan diagnosis LG glioma dan *others*. Kurva *survival* di setiap simpul menggambarkan perbedaan peluang *survival* antar kelompok dengan peluang *survival* terbaik terdapat pada pasien dengan diagnosis meningioma dan Ukuran Tumor kecil. Sedangkan, peluang *survival* terburuk terdapat pada pasien dengan diagnosis HG glioma, terlepas dari ukuran tumor. Hal ini mengindikasikan bahwa kombinasi kovariat Diagnosis dan Ukuran Tumor menjadi faktor utama dalam menentukan peluang *survival* pasien kanker otak. Misalkan ada pasien kanker otak dengan jenis tumor LG glioma, ukuran tumor $8,88\text{cm}^3$, maka diestimasi peluang *survival* pasien kanker otak mengikuti model *survival tree* pada **Node 5**. Diestimasikan pasien memiliki peluang *survive* sekitar 90% pada 2 tahun pertama waktu pengamatan. Setelah sekitar 4 tahun, peluang *survive* pasien diestimasi sekitar 20%.

4.4 Aplikasi Model Regresi Cox pada Data Pasien Kanker Otak

Sebelum dibentuk model regresi Cox untuk data pasien kanker otak, diperlukan variabel *dummy* untuk mengubah variabel yang memiliki struktur kualitatif menjadi kuantitatif. Dalam penelitian ini, kovariat yang memiliki struktur kualitatif adalah *Gender*, *Diagnosis*, *Lokasi Tumor*, *Metode Terapi*. Variabel *dummy* akan bernilai 1 jika sesuai dengan kondisi yang tertera, dan bernilai 0 jika tidak sesuai. Misalkan terdapat n kategori dalam suatu kovariat kualitatif, maka akan dibuat $n - 1$ variabel *dummy*. Tabel 4.5 menunjukkan pembuatan variabel *dummy* untuk data pasien kanker otak.

Tabel 4.5: Pemisahan variabel *dummy* pada data pasien kanker otak

Nomor	Kovariat	Variabel <i>dummy</i>	Penggolongan	Nilai
1	<i>Gender</i>	<i>sex</i> ₁	Jenis kelamin laki-laki	1
2			Lainnya	0
3	<i>Diagnosis</i>	<i>diagnosis</i> ₁	Pasien terdiagnosa kanker otak LG Glioma	1
4			Diagnosa lainnya	0
5		<i>diagnosis</i> ₂	Pasien terdiagnosa kanker otak HG Glioma	1
6			Diagnosa lainnya	0
7		<i>diagnosis</i> ₃	Pasien terdiagnosa kanker otak selain LG Glioma, HG Glioma, Meningioma	1
8			Diagnosa lainnya	0
9	<i>Lokasi Tumor</i>	<i>loc</i> ₁	Supratentorial	1
10			Lainnya	0
11	<i>Metode Terapi</i>	<i>stereo</i> ₁	SRT	1
12			Lainnya	0

Setelah dibentuk variabel *dummy*, akan dibuat model regresi Cox untuk data pasien kanker otak. Sesuai dengan yang dijelaskan pada Bab 3.3, model dibuat dengan menggunakan perangkat lunak RStudio. Adapun langkah-langkah pembuatan model regresi Cox dengan menggunakan RStudio sebagai berikut:

1. memanggil *package* '`survival`';
2. mendefinisikan `fit.cox` yang menyimpan data *survival* pasien kanker otak dengan membaca tabel data menggunakan RStudio;
3. membuat variabel baru untuk menyimpan hasil analisis dari model CPH yang telah diestimasi;
4. menuliskan formula *survival* sesuai dengan kovariat yang akan digunakan untuk membangun model CPH. Sebagai contoh, formula *survival* ditulis sebagai '`Surv(time, status) ~ kovariat1 + kovariat2 + kovariat3`';
5. mementuk beberapa model CPH dengan kombinasi kovariat yang memiliki pengaruh kuat dan memilih model terbaik berdasarkan nilai AIC;
6. melakukan uji serentak dan uji parsial pada model CPH terbaik yang diperoleh;
7. menguji asumsi PH pada model CPH menggunakan perintah '`cox.zph`'; apabila asumsi PH terpenuhi, membuat interpretasi model yang paling sesuai.

Dengan merujuk pada langkah-langkah tersebut, dibentuk variabel '`fit.all`' berisi model CPH menggunakan empat kovariat yang terdapat pada data. Formula *survival* yang digunakan pada perangkat lunak RStudio dapat ditulis sebagai '`Surv(time, status) ~ sex * diagnosis + ki + gtv`'. Hasil model CPH pada data pasien kanker otak dengan empat kovariat dapat dilihat pada Tabel 4.6.

Tabel 4.6: Model CPH menggunakan empat kovariat dengan interaksi pada data pasien kanker otak

α	Kovariat	Variable <i>dummy</i>	β_α	$\exp(\beta_\alpha)$
1	<i>Gender</i>	sex_1	1,37080	3,93850
2		$diagnosis_1$	2,43012	11,36026
3	Diagnosis	$diagnosis_2$	2,86159	17,48931
4		$diagnosis_3$	2,14071	8,50547
5	<i>Karnofsky Index</i>	-	-0,05027	0,95097
6	Ukuran Tumor	-	0,05120	1,05253
7		$sex_1 * diagnosis_1$	-2,37276	0,09322
8	Interaksi: <i>Gender</i> dengan Diagnosis	$sex_1 * diagnosis_2$	-1,21360	0,29713
9		$sex_1 * diagnosis_3$	-2,68997	0,06788

Lebih lanjut, dengan memanfaatkan nilai β_a pada Tabel 4.6 dan persamaan (3.14), diperoleh model CPH

$$\begin{aligned}
 h(t | \mathbf{X}) = & h_0(t) \exp(1,3708 \cdot sex_{(1)} + 2,4301 \cdot diagnosis_{(1)} + 2,8616 \cdot diagnosis_{(2)} \\
 & + 2,1407 \cdot diagnosis_{(3)} - 0,0503 \cdot ki + 0,0512 \cdot gtv - 2,3728 \cdot (sex_{(1)} \times diagnosis_{(1)}) \\
 & - 1,2136 \cdot (sex_{(1)} \times diagnosis_{(2)}) - 2,6900 \cdot (sex_{(1)} \times diagnosis_{(3)})
 \end{aligned} \quad (4.1)$$

Sebagai pembanding untuk model CPH yang diperoleh, dibentuk model-model CPH lain dengan menggunakan kombinasi kovariat lainnya. Untuk membandingkan model CPH mana yang lebih baik, digunakan AIC pada persamaan (3.3.3) untuk membandingkan nilai-nilai evaluasi model. Nilai AIC pada masing-masing model dapat dilihat pada Tabel 4.7. Karena nilai AIC terkecil diperoleh dari model CPH yang terdiri dari empat kovariat, maka model CPH yang digunakan merupakan model CPH pada persamaan (4.1).

Tabel 4.7: Nilai AIC untuk model-model CPH

Model	Kovariat	Banyak Parameter	Nilai AIC
1	<i>Gender, Diagnosis, Karnofsky Index, Ukuran Tumor</i>	9	245,5934
2	<i>Gender, Diagnosis, Lokasi Tumor, Karnofsky Index, Ukuran Tumor, Metode Terapi</i>	11	249,032
3	<i>Gender, Diagnosis, Lokasi Tumor, Karnofsky Index, Ukuran Tumor, Metode Terapi</i>	6	249,4955

Lebih lanjut, akan dilakukan uji signifikansi model pada model CPH yang paling baik seperti yang sudah dibahas di subsubbab 3.3.4. Hasil kedua uji dapat dilihat sebagai berikut.

1. Uji Serentak

Pada uji serentak, diasumsikan hipotesis H_0 dan H_1 sebagai berikut

$$H_0 : \text{Model bersifat tidak signifikan};$$

$$H_1 : \text{Model bersifat signifikan}.$$

Dengan menggunakan perangkat lunak RStudio, diperoleh *p-value* dengan nilai 3×10^{-7} . Karena nilai *p-value* lebih kecil dari nilai α yang digunakan, yaitu 0,05, maka H_0 ditolak, sehingga model bersifat signifikan.

2. Uji Parsial

Pada uji parsial, diasumsikan hipotesis H_0 dan H_1 sebagai berikut

H_0 : Kovariat bersifat tidak signifikan;

H_1 : Kovariat bersifat signifikan.

Dengan menggunakan perangkat lunak RStudio untuk melakukan uji parsial, diperoleh nilai p -value dan interpretasinya yang dapat dilihat pada Tabel 4.8

Tabel 4.8: Hasil uji parsial model CPH

No	Kovariat	Variabel <i>dummy</i>	<i>p</i> -value	Keputusan	Interpretasi
1	<i>Gender</i>	sex_1	0,0561	H0 tidak ditolak	sex_1 tidak signifikan
2	Diagnosis	$diagnosis_1$	0,0092	H0 ditolak	$diagnosis_1$ signifikan
		$diagnosis_2$	0,000051	H0 ditolak	$diagnosis_2$ signifikan
		$diagnosis_3$	0,0110	H0 ditolak	$diagnosis_3$ signifikan
3	<i>Karnofsky Index</i>	-	0,0084	H0 ditolak	<i>Karnofsky Index</i> signifikan
4	Ukuran Tumor	-	0,0183	H0 ditolak	Ukuran Tumor signifikan
5	Interaksi: <i>Gender</i> dengan Diagnosis	$sex_1 * diagnosis_1$	0,0559	H0 tidak ditolak	$sex_1 * diagnosis_1$ tidak signifikan
		$sex_1 * diagnosis_2$	0,1741	H0 tidak ditolak	$sex_1 * diagnosis_2$ tidak signifikan
		$sex_1 * diagnosis_3$	0,0232	H0 ditolak	$sex_1 * diagnosis_3$ signifikan

Setelah diketahui bahwa model dan kovariat-kovariat yang digunakan pada persamaan (4.1) signifikan, akan diuji asumsi PH untuk menentukan apakah model CPH pada persamaan (4.1) dapat diterapkan pada data pasien kanker otak. Seperti yang sudah dijelaskan pada subsubbab 3.3.2, diasumsikan hipotesis H_0 dan H_1 sebagai berikut:

H_0 : Kovariat bersifat tidak signifikan;

H_1 : Kovariat bersifat signifikan.

Apabila dalam model persamaan (4.1) terdapat satu atau lebih kovariat yang tidak memenuhi asumsi PH, maka model tersebut tidak dapat digunakan. Pengujian asumsi PH dilakukan dengan menggunakan perangkat lunak RStudio, *package* 'survival' dan perintah 'cox.zph'. Hasil dari uji asumsi PH dapat dilihat pada Tabel 4.9.

Tabel 4.9: Hasil uji asumsi PH model pada data pasien kanker otak

No	Kovariat	Variabel <i>Dummy</i>	Nilai <i>Chi-Square</i>	<i>p</i> -value	Keputusan	Interpretasi
1	<i>Gender</i>	sex_1	1,8469	0,17	H_0 tidak ditolak	Memenuhi asumsi PH
2	Diagnosis	$diagnosis$	2,2511	0,52	H_0 tidak ditolak	Memenuhi asumsi PH
3	<i>Karnofsky Index</i>	-	0,0648	0,80	H_0 tidak ditolak	Memenuhi asumsi PH
4	Ukuran Tumor	-	0,1051	0,75	H_0 tidak ditolak	Memenuhi asumsi PH
5	Interaksi: <i>Gender</i> dengan Diagnosis	$sex * diagnosis$	1,7830	0,62	H_0 tidak ditolak	Memenuhi asumsi PH

Karena uji asumsi PH terpenuhi, maka model CPH dapat digunakan untuk data pasien kanker otak. Setelah menganalisis tingkat *survival* pasien kanker otak, maka dapat diketahui bahwa kovariat Diagnosis, *Karnofsky Index*, Ukuran Tumor memengaruhi tingkat *survival* pasien secara signifikan dibandingkan kovariat *Gender*, Letak Tumor, dan Metode Terapi. Lebih lanjut, dapat dilihat pada Tabel 4.10 ringkasan dan interpretasi dari model CPH terbaik.

Tabel 4.10: Interpretasi hasil model CPH

No	Kovariat	Nilai β_a	Nilai $\exp(\beta_a)$	Interpretasi
1	Diagnosis: LG Glioma	2,43012	11,36026	Pasien dengan diagnosis LG glioma memiliki risiko 11,36026 kali lebih besar untuk mengalami kematian dibandingkan pasien dengan diagnosis Meningioma.
2	Diagnosis: HG Glioma	2,86159	17,48931	Pasien dengan diagnosis HG Glioma memiliki risiko 17,48931 kali lebih besar untuk mengalami kematian dibandingkan pasien dengan diagnosis Meningioma.
3	Diagnosis: Other	2,14071	8,50547	Pasien dengan diagnosis <i>Other</i> memiliki risiko 8,50547 kali lebih besar untuk mengalami kejadian kematian dibandingkan pasien dengan diagnosis Meningioma.
4	<i>Karnofsky Index</i>	-0,050027	0,95097	Setiap peningkatan unit pada <i>Karnofsky Index</i> menurunkan risiko kematian pasien sebesar 4,903%.
5	Ukuran Tumor	0,05120	1,05253	Setiap peningkatan unit pada GTV meningkatkan risiko kematian pasien sebesar 5,253%.
6	SexMale:DiagnosisOther	-2,68997	0,06788	Interaksi jenis kelamin laki-laki dan diagnosis lain mengurangi risiko kejadian kematian sebesar 93% dibandingkan efek utama.

4.5 Analisis Hasil

Dari hasil metode estimator Kaplan-Meier, diperoleh informasi bahwa tiga dari empat variabel yang dipertimbangkan, yaitu Diagnosis, Lokasi Tumor, dan Metode Terapi pasien memiliki pengaruh yang berbeda pada peluang *survival* pasien kanker otak dari awal hingga akhir waktu pengamatan. Secara spesifik, pada variabel Diagnosis, kelompok yang memiliki peluang *survival* paling tinggi berturut-turut adalah Meningioma, LG glioma, *Other*, dan HG Glioma. Untuk variabel Lokasi Tumor, lokasi tumor *supratentorial* memiliki risiko kematian lebih tinggi dibandingkan dengan lokasi tumor *infratentorial*. Kemudian, berdasarkan variabel Metode Terapi, pasien dengan metode terapi SRS lebih berisiko untuk mengalami kematian dibandingkan dengan pasien dengan metode terapi SRT.

Dari model *survival tree*, diperoleh hasil yang mencerminkan pola hierarki dalam menentukan peluang *survival* pasien kanker otak. Model *survival tree* yang dihasilkan menunjukkan bahwa kovariat Diagnosis adalah kovariat utama yang perlu diperhatikan dalam menentukan peluang *survival* pasien kanker otak. Di antara kategori yang terdapat pada kovariat Diagnosis, HG glioma memiliki peluang *survival* paling buruk dibandingkan dengan kategori diagnosis lainnya terlepas dari pengaruh kovariat lain. Pada pasien dengan kategori diagnosis Meningioma, LG glioma, dan *other*, Ukuran Tumor menjadi kovariat yang dipertimbangkan dalam mengestimasi peluang *survival* pasien kanker otak. Pasien dengan ukuran tumor kecil ($< 10,585$) memiliki peluang *survival* yang lebih tinggi dibandingkan dengan pasien dengan ukuran tumor yang lebih besar ($\geq 10,585$). Dalam model *survival tree*, kombinasi antara Diagnosis dan Ukuran Tumor menjadi indikator utama dalam mengestimasi peluang *survival* pasien kanker otak. Hasil akhir dari model *survival tree* menunjukkan bahwa peluang *survival* pasien kanker otak terbaik ditemukan pada pasien dengan kategori diagnosis meningioma dengan ukuran tumor kecil, sedangkan peluang *survival* terburuk diperoleh pada pasien dengan diagnosis HG glioma tanpa mempertimbangkan kovariat lain.

Lebih lanjut, diperoleh informasi berdasarkan kurva Kaplan-Meier dari model *survival tree*, pasien dengan diagnosis HG glioma memiliki peluang *survival* yang sangat rendah terlepas dari hubungannya dengan kovariat lain. Terlihat bahwa dalam 20 bulan awal pengamatan, peluang *survival* pasien kanker otak dengan diagnosis HG glioma sudah di bawah 30%. Sementara pasien dengan diagnosis Meningioma, LG glioma, dan *Other* yang memiliki ukuran tumor lebih besar

dari 10,585 menunjukkan peluang *survival* yang sedikit lebih baik, meskipun tidak berbeda secara signifikan dibandingkan dengan pasien dengan diagnosis HG glioma. Terlihat pada kurva *survival*nya bahwa pada 20 bulan awal pengamatan, peluang *survival* pasien masih di atas 50% walaupun dalam jangka panjang, kemungkinan pasien untuk *survive* tidak tinggi. Selanjutnya, pasien dengan diagnosis LG glioma yang memiliki ukuran tumor lebih kecil dari 10,585 menunjukkan peluang *survival* yang lebih baik dalam jangka pendek, tetapi cenderung memburuk dalam jangka panjang. Peluang *survival* terbaik ditemukan pada kelompok pasien dengan diagnosis Meningioma dan *Other* yang memiliki ukuran tumor lebih kecil dari 10,585, di mana dalam seluruh waktu pengamatan, peluang *survival* pasien masih berada di atas 80%.

Dari hasil model CPH, dapat disimpulkan bahwa model valid digunakan untuk analisis kovariat yang memengaruhi pasien kanker otak karena asumsi PH dipenuhi. Dari hasil uji serentak menunjukkan bahwa model bersifat signifikan dengan $p\text{-value} < 0,05$. Secara parsial, kovariat yang berpengaruh dalam tingkat *survival* pasien adalah Diagnosis, *Karnofsky Index*, dan Ukuran Tumor. Dapat dilihat juga bahwa Diagnosis merupakan kovariat yang paling signifikan dalam menentukan tingkat *survival* pasien kanker otak. Pasien dengan diagnosis HG glioma memiliki risiko kematian tertinggi, yaitu 17,48931 kali lebih tinggi dibandingkan pasien dengan diagnosis meningioma. Diagnosis lain juga memiliki pengaruh pada tingkat *survival* pasien, di mana pasien dengan LG glioma dan *Other* memiliki risiko 11,36026 dan 8,50547 kali lebih besar dibandingkan dengan pasien dengan diagnosis Meningioma.

Selain Diagnosis, *Karnofsky Index* juga memberikan pengaruh yang signifikan, di mana setiap peningkatan satu unit pada *Karnofsky Index* memberikan pengaruh positif pada tingkat *survival* pasien sebesar 4,9%, yang mengindikasikan bahwa kondisi fisik pasien yang lebih baik berkaitan dengan peluang *survival* yang lebih tinggi. Sebaliknya, untuk kovariat Ukuran Tumor menunjukkan pengaruh yang berlawanan di mana setiap peningkatan satu unit pada Ukuran Tumor meningkatkan risiko kematian pasien sebesar 5,3%. Hal ini mengindikasikan bahwa ukuran tumor yang lebih besar meningkatkan risiko kematian pada pasien. Interaksi antar kovariat Gender dan Diagnosis menunjukkan bahwa pasien laki-laki dengan diagnosis *Other* memiliki risiko kematian yang lebih rendah dibandingkan dengan diagnosis lain. Hal ini terlihat dari nilai koefisien sebesar 0,06788 yang mengindikasikan bahwa kombinasi laki-laki dengan diagnosis *Other* mengurangi risiko kematian hingga 93%.

Berdasarkan hasil dari ketiga model, dapat disimpulkan bahwa Diagnosis merupakan kovariat paling signifikan dalam menentukan tingkat *survival* pasien kanker otak, di mana pasien dengan HG glioma secara konsisten memiliki tingkat *survival* paling buruk di semua metode dan model. Metode estimator Kaplan-Meier memberikan gambaran awal tentang hubungan antar kovariat dan tingkat *survival*, sementara model *survival tree* memperjelas pola yang menunjukkan pengaruh kombinasi antar kovariat, terutama Diagnosis dan Ukuran Tumor. Model CPH di sisi lain menegaskan hubungan antar kovariat terhadap tingkat *survival*. Aplikasi dari hasil ini dapat digunakan dalam bidang medis. Sebagai contoh, hasil ini dapat membantu memprioritaskan pengobatan pada pasien dengan HG glioma serta mempertimbangkan ukuran tumor dan kondisi fisik pasien (*Karnofsky Index*). Lebih lanjut, model *survival tree* dapat digunakan untuk mengidentifikasi risiko klaim berdasarkan kombinasi faktor Diagnosis dan Ukuran Tumor. Sebagai contoh, pasien kanker otak dengan diagnosis HG glioma yang memiliki tingkat *survival* paling rendah dapat diklasifikasikan

sebagai pemegang polis risiko tinggi, sehingga premi yang dikenakan lebih besar atau cadangan klaim ditingkatkan.



BAB 5

KESIMPULAN DAN SARAN

Pada bab ini dibahas kesimpulan yang diperoleh dari Bab 4, dan saran yang dapat dilakukan untuk penelitian berikutnya.

5.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, berikut merupakan beberapa kesimpulan yang diperoleh.

1. Metode Kaplan-Meier digunakan untuk mengestimasi dan membandingkan peluang *survival* pasien kanker otak berdasarkan kategori dari suatu variabel selama waktu pengamatan, sementara model *survival tree* digunakan untuk mengestimasi peluang *survival* pasien kanker otak secara bertahap berdasarkan nilai *deviance* yang diperoleh dari kovariat yang dipertimbangkan. Diperoleh estimasi peluang *survival* berdasarkan metode Kaplan-Meier dan model *survival tree* sebagai berikut:
 - (a) Dari metode Kaplan-Meier diperoleh kesimpulan bahwa variabel Diagnosis dengan kategori HG glioma, variabel Lokasi Tumor dengan kategori Supratentorial, dan variabel Metode terapi dengan kategori SRT memiliki peluang *survival* yang paling rendah. Sementara tidak dapat dilihat dengan jelas kategori yang memiliki peluang *survival* paling rendah pada variabel *Gender* dikarenakan peluang *survival* untuk kedua kategori dalam *Gender* ada di nilai yang sama menuju akhir waktu pengamatan.
 - (b) Dari model *survival tree*, diperoleh kesimpulan bahwa pasien kanker otak dengan diagnosis Meningioma dan Other dengan ukuran tumor yang kecil memiliki peluang *survival* yang tinggi, sementara pasien dengan diagnosis HG glioma memiliki peluang *survival* paling rendah tanpa perlu mempertimbangkan kovariat lainnya.
2. Berdasarkan hasil dari model regresi Cox *Proportional Hazard*, diperoleh kesimpulan bahwa kovariat-kovariat yang secara signifikan memengaruhi tingkat *survival* pasien kanker otak, yaitu Diagnosis, *Karnofsky Index*, Ukuran Tumor. Selain itu dapat disimpulkan juga bahwa interaksi antara kovariat *Gender* kategori Laki-laki dengan Diagnosis kategori Other memiliki pengaruh yang signifikan dalam menentukan tingkat *survival* pasien kanker otak.

5.2 Saran

Berikut merupakan saran untuk pengembangan penelitian selanjutnya:

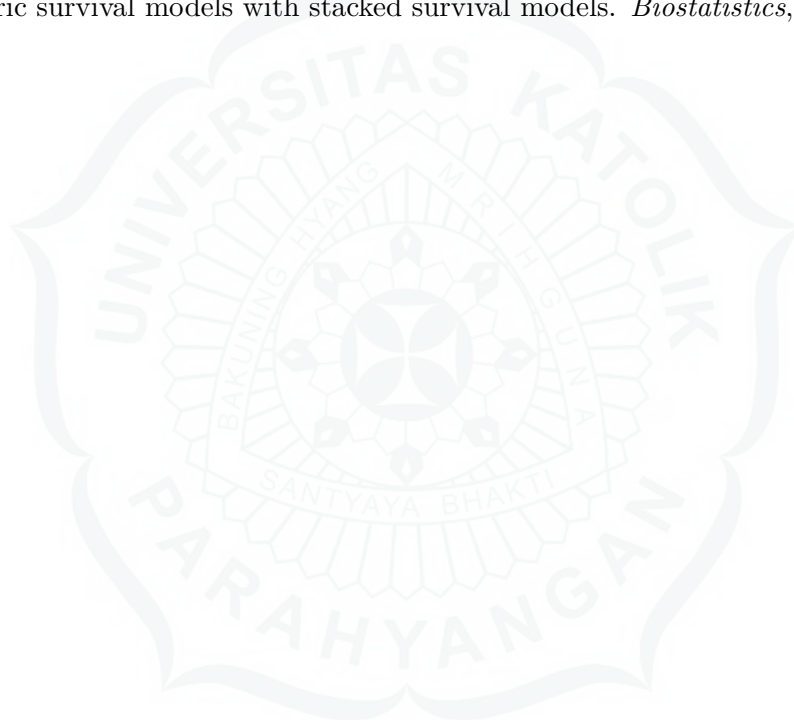
1. Menambahkan kovariat yang berpotensi menjadi pengaruh tingkat *survival* pasien kanker otak seperti Usia, Berat badan, dan stadium tumor;
2. Mengembangkan model *survival tree* yang lebih kompleks seperti *Random Survival Forest* [21];
3. Menggunakan model *survival* parametrik untuk interpretasi dan analisis yang lebih lengkap [22].



DAFTAR REFERENSI

- [1] Badža, M. M. dan Barjaktarović, M. Č. (2020) Classification of brain tumors from MRI images using a convolutional neural network. *Applied Sciences*, **10**, 19–99.
- [2] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., dan Jemal, A. (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, **74**, 229–263.
- [3] Abdel-Sater, R. (2022) Marriage dissolution in the United States: A survival analysis approach. *Journal of Divorce & Remarriage*, **63**, 235–261.
- [4] Obed, S. A., Mohammed, P. A., dan Kadir, D. H. (2021) The estimation of (Covid-19) cases in Kurdistan region using Nelson-Aalen estimator. *Cihan University-Erbil Scientific Journal*, **5**, 24–31.
- [5] Pahlevi, M. R., Mustafid, M., dan Wuryandari, T. (2016) Model Regresi Cox Stratified pada Data Ketahanan. *Jurnal Gaussian*, **5**, 455–464.
- [6] Wang, P., Li, Y., dan Reddy, C. K. (2019) Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, **51**, 1–36.
- [7] Dickson, D. C., Hardy, M. R., dan Waters, H. R. (2009) *Actuarial Mathematics for Life Contingent Risks*, 1st edition. Cambridge University Press, New York.
- [8] London, D. (1997) *Survival Models and Their Estimation*, 3rd edition. ACTEX Publications, Connecticut.
- [9] McFaline-Figueroa, J. R. dan Lee, E. Q. (2018) Brain tumors. *The American journal of medicine*, **131**, 874–882.
- [10] Leavitt, D. (1998) Beam shaping for SRT/SRS. *Medical Dosimetry*, **23**, 229–236.
- [11] Frees, E. W. (2010) *Regression Modelling With Actuarial and Financial Applications*, 1st edition. Cambridge University Press, New York.
- [12] Ross, S. M. (2010) *Introduction to Probability Models 10th edition*, 1st edition. Elsevier, San Diego.
- [13] Atkinson, E. J. dan Therneau, T. M. (2000) An introduction to recursive partitioning using the RPART routines. *Rochester: Mayo Foundation*, **3**, 1–57.
- [14] Davis, R. B. dan Anderson, J. R. (1989) Exponential survival trees. *Statistics in medicine*, **8**, 947–961.
- [15] Bremner, A. P. dan Taplin, R. H. (2002) Theory & methods: modified classification and regression tree splitting criteria for data with interactions. *Australian & New Zealand Journal of Statistics*, **44**, 169–176.

-
- [16] Prodromidis, A. L. dan Stolfo, S. J. (2001) Cost complexity-based pruning of ensemble classifiers. *Knowledge and Information Systems*, **3**, 449–469.
- [17] Bou-Hamad, I., Larocque, D., dan Ben-Ameur, H. (2011) A review of survival trees. *Statistics Surveys*, **5**, 44–71.
- [18] Kleinbaum, D. G. dan Klein, M. (2012) *Survival Analysis A Self-Learning Text, 3rd Edition*. Springer, Berlin.
- [19] Collett, D. (2023) *Modelling survival data in medical research*, 1st edition. Chapman and Hall/CRC, Boca Raton.
- [20] Muhammad, M., Yuwaningsih, D., dkk. (2019) Estimating Survival Time of Dengue Haemorrhagic Fever Using Extended Cox Model. *Journal of Physics*, **118**, 12–17.
- [21] Ishwaran, H. dan Kogalur, U. B. (2007) Random survival forests for R. *R news*, **7**, 25–31.
- [22] Wey, A., Connett, J., dan Rudser, K. (2015) Combining parametric, semi-parametric, and non-parametric survival models with stacked survival models. *Biostatistics*, **16**, 537–549.





LAMPIRAN A

DATA

Tabel A.1: Data pasien kanker otak

No	sex	diagnosis	loc	ki	gtv	stereo	status	time (bulan)	j	r_j	q_j
1	Male	HG glioma	Supratentorial	70	33,69	SRT	1	0,07	1	87	1
2	Male	LG glioma	Infratentorial	90	30,41	SRT	0	1,18			
3	Male	Meningioma	Supratentorial	70	0,97	SRT	1	1,41	2	86	1
4	Female	Meningioma	Infratentorial	70	2,94	SRS	0	1,54			
5	Female	Meningioma	Supratentorial	80	1,57	SRT	0	2,03	3	86	0
6	Female	HG glioma	Supratentorial	40	22,87	SRT	1	3,38	4	85	1
7	Male	Meningioma	Supratentorial	80	12,63	SRT	1	4,16	5	83	2
8	Female	HG glioma	Supratentorial	80	7,59	SRT	1	4,56			
9	Male	HG glioma	Supratentorial	80	29,27	SRT	0	5,15	6	83	0
10	Female	Meningioma	Supratentorial	90	7,3	SRT	0	5,51			
11	Female	LG glioma	Supratentorial	80	0,85	SRS	1	6,1			
12	Male	HG glioma	Supratentorial	90	9,95	SRT	1	6,23			
13	Male	HG glioma	Supratentorial	90	5,06	SRT	1	6,3	7	77	6
14	Male	HG glioma	Supratentorial	80	19,81	SRT	1	6,39			
15	Female	Other	Infratentorial	60	24	SRT	1	6,82			
16	Male	Meningioma	Supratentorial	80	13,49	SRS	1	6,92			
17	Female	HG glioma	Supratentorial	70	11,38	SRS	1	7,05			
18	Female	Other	Supratentorial	70	14,26	SRT	0	7,18	8	75	2
19	Female	HG glioma	Supratentorial	80	12,08	SRT	1	7,25			
20	Male	HG glioma	Supratentorial	90	19,35	SRT	1	8,98	9	74	1
21	Male	HG glioma	Supratentorial	90	2,5	SRT	1	9,77	10	73	1
22	Female	Meningioma	Supratentorial	80	6,93	SRS	0	10,1			
23	Female	Other	Infratentorial	80	2,11	SRS	0	10,49	11	72	1
24	Female	Other	Infratentorial	70	13,45	SRT	1	10,82			
25	Female	HG glioma	Supratentorial	80	0,19	SRT	1	11,02			
26	Male	HG glioma	Supratentorial	100	2,53	SRT	0	11,48	12	70	2
27	Female	LG glioma	Supratentorial	80	0,19	SRS	0	11,51			
28	Male	Other	Supratentorial	60	34,64	SRT	1	11,57			
29	Male	Other	Infratentorial	80	0,11	SRT	0	13,9	14	69	1
30	Male	HG glioma	Supratentorial	70	14,44	SRT	1	14			
31	Female	Meningioma	Supratentorial	70	6,7	SRT	0	14,56			
32	Male	Meningioma	Supratentorial	80	11,51	SRT	1	14,62	15	68	1
33	Female	Meningioma	Supratentorial	80	6,6	SRT	0	14,75			
34	Male	HG glioma	Supratentorial	90	0,28	SRT	1	16,43	17	66	2
35	Female	HG glioma	Supratentorial	80	0,63	SRT	1	16,92			
36	Female	Meningioma	Supratentorial	90	2,47	SRT	0	17,57	18	66	0
37	Male	Other	Infratentorial	90	3,12	SRT	0	18,95	19	66	0
38	Male	Meningioma	Supratentorial	70	8,45	SRT	0	19,41			
39	Male	Other	Supratentorial	100	24,91	SRT	0	19,74	20	65	1
40	Male	HG glioma	Supratentorial	80	3,75	SRT	1	19,9			

No	sex	diagnosis	loc	ki	gtv	stereo	status	time (bulan)	j	r_j	q_j
41	Male	LG glioma	Supratentorial	90	2,64	SRT	0	20,13			
42	Male	HG glioma	Supratentorial	80	0,16	SRT	1	20,69	21	64	1
43	Female	Other	Supratentorial	80	11,83	SRT	1	22,03	22	63	1
44	Male	Other	Infratentorial	70	0,01	SRS	0	23,67	23	63	0
45	Female	Meningioma	Supratentorial	80	2,82	SRS	0	24,39	24	63	0
46	Male	LG glioma	Supratentorial	90	4,23	SRT	1	25,02	25	62	1
47	Female	Meningioma	Infratentorial	70	7,95	SRS	0	26,46			
48	Female	Meningioma	Infratentorial	90	9,24	SRT	0	26,85	26	62	0
49	Male	Other	Infratentorial	80	12,51	SRT	1	29,7	29	61	1
50	Female	Meningioma	Supratentorial	90	2,5	SRT	0	30,16	30	61	0
51	Male	Meningioma	Supratentorial	80	20,93	SRT	1	31,15			
52	Male	Meningioma	Supratentorial	60	7,09	SRS	1	31,25	31	59	2
53	Female	HG glioma	Supratentorial	90	0,04	SRT	0	31,67			
54	Female	Meningioma	Supratentorial	70	3,63	SRT	0	32,82			
55	Male	Meningioma	Supratentorial	90	2,5	SRT	0	32,82	32	59	0
56	Female	Meningioma	Infratentorial	90	0,63	SRS	0	32,98			
57	Female	Meningioma	Supratentorial	70	6,48	SRT	1	33,41			
58	Male	HG glioma	Infratentorial	80	0,22	SRS	0	33,67	33	58	1
59	Female	Meningioma	Infratentorial	90	6,54	SRS	0	34,26			
60	Female	Meningioma	Supratentorial	70	12,16	SRT	0	34,66	34	58	0
61	Male	LG glioma	Supratentorial	80	0,14	SRT	1	35,93	35	57	1
62	Male	Meningioma	Supratentorial	60	3,81	SRT	0	36,1	36	57	0
63	Female	Meningioma	Supratentorial	100	9,66	SRT	0	39,25			
64	Female	Meningioma	Supratentorial	80	26,31	SRT	1	39,54	39	56	1
65	Male	Meningioma	Infratentorial	80	24,41	SRT	0	39,54			
66	Female	Meningioma	Supratentorial	60	9,18	SRT	0	41,44	41	56	0
67	Male	Meningioma	Supratentorial	90	2,02	SRS	0	42,07			
68	Male	LG glioma	Supratentorial	80	12,37	SRT	0	42,1	42	56	0
69	Male	Meningioma	Supratentorial	90	2,56	SRS	0	44,39	44	56	0
70	Female	Meningioma	Supratentorial	90	2,54	SRT	0	45,74	45	56	0
71	Female	HG glioma	Supratentorial	80	15,45	SRT	1	46,16	46	55	1
72	Female	Other	Supratentorial	90	1,82	SRT	0	47,11			
73	Female	LG glioma	Supratentorial	80	7,61	SRT	1	47,8	47	54	1
74	Female	Meningioma	Supratentorial	90	7,26	SRT	0	49,05	49	53	0
75	Male		Supratentorial	90	6,38	SRT	0	50,85	50	53	0
76	Female	Meningioma	Infratentorial	100	2,13	SRS	1	51,02	51	53	1
77	Female	Meningioma	Supratentorial	100	4	SRT	0	52,23	52	53	0
78	Female	Meningioma	Supratentorial	80	4,82	SRS	0	52,75			
79	Male	Other	Infratentorial	90	0,48	SRS	0	54,43	54	53	0
80	Male	Meningioma	Supratentorial	80	3,19	SRT	0	55,8	55	53	0

No	sex	diagnosis	loc	ki	gtv	stereo	status	time (bulan)	j	r_j	q_j
81	Male	Other	Infratentorial	80	6,81	SRS	0	57,11			
82	Female	Meningioma	Supratentorial	80	31,74	SRT	0	57,25	57	53	0
83	Female	Meningioma	Infratentorial	90	6,11	SRS	0	57,64		53	
84	Female	Meningioma	Supratentorial	90	4,72	SRS	0	65,02	65	53	0
85	Male	Meningioma	Supratentorial	80	4,77	SRT	0	67,38	67	53	0
86	Female	Meningioma	Supratentorial	80	2,39	SRS	0	73,74	73	53	0
87	Male	LG glioma	Supratentorial	80	9,58	SRT	0	78,75	78	53	0
88	Male	HG glioma	Supratentorial	90	10,8	SRT	0	82,56	83	53	0

