SKRIPSI

ANALISIS KELANGSUNGAN HIDUP PASIEN KANKER PARU-PARU MENGGUNAKAN MODEL REGRESI *COX PROPORTIONAL HAZARD*

Santo Samuel Surja

NPM: 6161801049

PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022

FINAL PROJECT

SURVIVAL ANALYSIS OF LUNG CANCER PATIENTS USING COX PROPORTIONAL HAZARD REGRESSION MODEL

Santo Samuel Surja

NPM: 6161801049

DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022

LEMBAR PENGESAHAN

ANALISIS KELANGSUNGAN HIDUP PASIEN KANKER PARU-PARU MENGGUNAKAN MODEL REGRESI *COX PROPORTIONAL HAZARD*

Santo Samuel Surja

NPM: 6161801049

Bandung, 11 Januari 2022

Menyetujui,

Pembimbing 1	Pembimbing 2
Maria Anestasia, M.Si, MActSc	Rizky Reza Fauzi, D.Phil.Math

Ketua Tim Penguji	Anggota Tim Penguji
Agus Sukmana, MSc	Dr. Andreas Parama Wijaya

Mengetahui,

Ketua Program Studi

Dr. Livia Owen

PERNYATAAN

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

ANALISIS KELANGSUNGAN HIDUP PASIEN KANKER PARU-PARU MENGGUNAKAN MODEL REGRESI *COX PROPORTIONAL HAZARD*

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 11 Januari 2022

Santo Samuel Surja
NPM: 6161801049

ABSTRAK

Pada tahun 2020, kanker paru-paru memakan korban jiwa paling banyak dibandingkan jenis kanker lainnya. Oleh karena itu, akan dianalisis faktor-faktor yang berpengaruh terhadap kelangsungan hidup pasien kanker paru-paru menggunakan model regresi *Cox proportional hazard* (*Cox*). Skripsi ini memuat 3 model *Cox*, yaitu model *Cox* tanpa penalti, model *Cox* dengan penalti kuadratik (penalti $L2$), dan model *Cox* dengan penalti absolut (penalti $L1$). Model *Cox* tanpa penalti dan model *Cox* dengan penalti memiliki perbedaan dalam bias dan varians. Estimasi parameter dari model *Cox* dilakukan menggunakan *maximum partial likelihood estimator*. Ketiga model akan dimanfaatkan untuk menganalisis 2 *data set* yang berasal dari Amerika Serikat, yaitu *data set* pasien kanker paru-paru dengan ras Asia dan *data set* pasien kanker paru-paru keseluruhan ras. Berdasarkan nilai *concordance* dan log *partial likelihood*, diperoleh bahwa model *Cox* dengan penalti $L2$ merupakan model terbaik untuk memprediksi probabilitas hidup kedua *data set*. Berdasarkan hasil eksperimen, disimpulkan bahwa untuk suatu rentang waktu, pasien kanker paru-paru dengan ras Asia memiliki probabilitas hidup yang lebih tinggi dibandingkan pasien kanker paru-paru keseluruhan ras.

Kata-kata kunci: model *Cox*, *partial likelihood*, penalti, kanker paru-paru

ABSTRACT

In 2020, lung cancer killed the most people compared to other types of cancer. Therefore, the factors that influence the survival of lung cancer patients will be analyzed using Cox proportional hazard regression model. In this paper, there will be 3 Cox models, which are Cox model without penalty, Cox model with quadratic penalty ($L2$ penalty), and Cox model with absolute penalty ($L1$ penalty). The Cox model without penalty differ in terms of bias and variance to the Cox model with penalty. Parameter estimation of the Cox model is done by using maximum partial likelihood estimator. These 3 models will be utilized to analyze 2 data sets from United States, which are Asian lung cancer patient data set and lung cancer patient data set with all race included. Based on the concordance score and log partial likelihood, the $L2$ model is the best model to predict the survival probabilities of both data sets. Based on experimental results, for a certain time interval, the Asian lung cancer patients have a higher survival probability compared to the lung cancer patients with all race included.

Keywords: Cox model, partial likelihood, penalty, lung cancer

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yesus Kristus atas kasih karunia dan penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi ini. Skripsi yang berjudul **"Analisis Kelangsungan Hidup Pasien Kanker Paru-Paru dengan Model Regresi *Cox Proportional Hazard*"** disusun sebagai salah satu syarat wajib untuk menyelesaikan studi Strata-1 Program Studi Matematika, Fakultas Teknologi Informasi dan Sains (FTIS), Universitas Katolik Parahyangan, Bandung. Penulis berharap skripsi ini dapat menjadi karya yang bermanfaat bagi setiap orang yang membacanya.

Selama masa studi dan penyusunan skripsi, penulis telah mendapatkan banyak bantuan, ilmu, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin berterima kasih kepada:

1. Kedua orang tua, Sutiono Surja dan Lioe Lie Ing yang selalu mendoakan, mendukung, dan memberikan semangat kepada penulis. Kedua kakak, Sem Samuel Surja dan Simon Samuel Surja yang turut mendorong penulis dalam perjalanan hidup.
2. Ibu Maria Anestasia, M.Si, MActSc selaku dosen pembimbing 1 yang telah memberikan ilmu, arahan, dan saran dalam seluruh proses penyusunan skripsi.
3. Bapak Rizky Reza Fauzi, D.Phil.Math selaku dosen pembimbing 2 yang telah memberikan arahan, didikan, dan bantuan di setiap proses penyusunan skripsi ini.
4. Bapak Agus Sukmana, MSc selaku dosen penguji 1 dan Bapak Dr. Andreas Parama Wijaya selaku dosen penguji 2 yang telah memberikan saran, kritik, dan masukan sehingga skripsi ini dapat menjadi lebih baik.
5. Bapak Liem Chin, M.Si selaku koordinator skripsi yang telah memberikan ilmu, saran, bantuan, dan arahan selama perkuliahan dan penyusunan skripsi ini.
6. Bapak Iwan Sugiarto, MSi. selaku dosen wali penulis yang telah memberikan nasihat kepada penulis selama proses kuliah.
7. Seluruh dosen, staf tata usaha, dan karyawan FTIS yang memberikan ilmu, dukungan, dan bantuan selama masa perkuliahan.
8. Cynthia Maria yang selalu mendengarkan keluh kesah, menemani, dan memberikan semangat kepada penulis.
9. Simonyong, yaitu Kevin, Elbert, Patrick, Laurentco, dan Andrew yang menjadi tempat belajar dan bergaul sepanjang kuliah.
10. Travel buddies yang senantiasa berjalan bersama penulis selama kuliah berlangsung.
11. Angelica, Bryant, dan Tiffany yang telah melepaskan penat penulis sesaat sebelum sidang.
12. Teman-teman Matematika angkatan 2015, 2016, 2017, 2018, 2019, 2020, dan 2021 yang tidak dapat disebutkan satu per satu.
13. Semua pihak yang telah berjasa kepada penulis selama masa perkuliahan dan penyusunan skripsi.

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan dan jauh dari kesempurnaan. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun dari para pembaca agar skripsi ini dapat menjadi lebih baik. Akhir kata semoga skripsi ini dapat bermanfaat dan dapat dikembangkan menjadi karya yang lebih baik.

Bandung, Januari 2022

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Keprihatinan terhadap penyakit kanker secara umum muncul dari beban kanker yang terus tumbuh secara global, sehingga memberikan tekanan fisik, emosional, dan keuangan yang luar biasa pada individu, keluarga, komunitas, dan sistem kesehatan. Banyak sistem kesehatan di negara-negara berpenghasilan rendah dan menengah tidak siap untuk mengelola beban ini. Akibatnya sejumlah pasien kanker di dunia tidak memiliki akses untuk memperoleh pengobatan yang berkualitas. Di negara-negara dengan sistem kesehatan yang kuat, tingkat kematian pasien kanker lebih kecil berkat deteksi dini, pengobatan, dan perawatan yang berkualitas.

Menurut *Global Cancer Statistics* 2020 [1], terdapat kurang lebih 19,3 juta kasus kanker baru di dunia pada tahun 2020, sedangkan angka kematiannya berada pada nilai mendekati 10 juta. Fakta ini menunjukkan bahwa tingkat kematian seorang pasien kanker adalah sekitar 50%. Apabila diteliti lebih lanjut mengenai kanker paru-paru, tingkat kematian seorang pasien kanker paru-paru adalah 81%. Dibandingkan 35 jenis kanker lainnya, tingkat kematian pasien kanker paru-paru berada pada posisi kelima tertinggi. Kemudian, terdapat 2,2 juta kasus kanker paru-paru atau sekitar 11% dari total kasus kanker. Jumlah kasus ini merupakan kontribusi kedua terbesar terhadap kasus kanker baru di dunia. Dari 2,2 juta kasus kanker paru-paru, 1,8 juta diantaranya meninggal, ini membuat kanker paru-paru sebagai jenis kanker yang paling banyak memakan korban jiwa.

Kanker merupakan kelompok penyakit yang diakibatkan oleh pertumbuhan sel secara abnormal pada suatu bagian tubuh dan menyebar ke bagian tubuh lain (*metastasis*). Kanker paru-paru merupakan jenis kanker yang selnya mulai tumbuh secara abnormal pada bagian paru-paru. Kematian akibat kanker paru-paru didorong oleh banyak faktor seperti usia, ukuran tumor, dan *metastasis*. Besar pengaruh faktor-faktor yang mendorong kematian akibat kanker paru-paru akan diselidiki sehingga diperoleh probabilitas hidup pasien kanker paru-paru. Probabilitas hidup pasien kanker paru-paru dapat diperoleh menggunakan model *Kaplan-Meier* [2] dan model regresi *Cox proportional hazard* (*Cox*) [3]. Model yang digunakan dalam skripsi ini adalah model *Cox* karena keunggulannya dari model *Kaplan-Meier*, yaitu model *Cox* dapat memprediksi probabilitas hidup yang memperhitungkan faktor-faktor yang mendorong kematian akibat kanker paru-paru [4].

Data set yang digunakan untuk membangun model *Cox* adalah *data set* pasien kanker paru-paru dari Institut Kanker Nasional di Amerika Serikat[1]. *Data set* ini mencatat data dari 564.611 pasien kanker paru-paru dari tahun 2004-2015. Di dalam *data set* ini, terdapat informasi mengenai ras dari pasien kanker, ras yang tercatat di dalamnya adalah ras kulit putih, kulit hitam, Asia, dan Amerika India/Alaska. Dari *data set* ini, akan dibentuk model *Cox* untuk memprediksi probabilitas hidup pasien kanker keseluruhan ras. Lalu dari *data set* pasien kanker keseluruhan ras, akan diambil pasien kanker dengan ras Asia agar probabilitas hidupnya juga dapat diprediksi. Pasien ras Asia menjadi ketertarikan di dalam skripsi ini karena penelitian ini berlangsung di negara Indonesia

[1] *Surveillance, Epidemiology, and End Results* (SEER) *Program* (www.seer.cancer.gov) SEER*Stat *Database: Incidence* - SEER *Research Data*, 18 *Registries, Nov* 2020 *Sub* (2000-2018) - *Linked To County Attributes - Time Dependent* (1990-2018) *Income/Rurality*, 1969-2019 *Counties, National Cancer Institute*, DCCPS, *Surveillance Research Program, released April* 2021, *based on the November* 2020 *submission.*

yang mayoritas penduduknya merupakan ras Asia. Alasan diprediksinya probabilitas hidup pasien kanker paru-paru keseluruhan ras dan ras Asia adalah ingin diselidiki perbandingan probabilitas hidup dari pasien ras Asia dengan pasien keseluruhan ras.

1.2 Rumusan Masalah

Rumusan masalah dalam skripsi ini adalah sebagai berikut.

1. Bagaimana cara membangun model *Cox* untuk memprediksi probabilitas hidup pasien kanker paru-paru?
2. Bagaimana perbandingan kelangsungan hidup antara pasien kanker paru-paru dengan ras Asia dan pasien kanker keseluruhan ras?

1.3 Tujuan

Tujuan yang ingin dicapai dalam skripsi ini diberikan sebagai berikut.

1. Membangun model *Cox* untuk memprediksi probabilitas hidup pasien kanker paru-paru.
2. Menyelidiki perbandingan antara probabilitas hidup pasien kanker paru-paru ras Asia dengan pasien kanker paru-paru keseluruhan ras.

1.4 Batasan Masalah

Batasan masalah dari skripsi ini diberikan sebagai berikut.

1. Variabel terikat dan variabel bebas pada model *Cox* diasumsikan memiliki hubungan eksponensial.
2. *Data set* yang digunakan adalah data pasien kanker paru-paru di Amerika Serikat dari Institut Kanker Nasional.

1.5 Sistematika Pembahasan

Skripsi ini terdiri atas 5 bab yaitu:

1. **Bab 1 : Pendahuluan**
   Pada bab ini akan dibahas mengenai latar belakang masalah, rumusan masalah, tujuan penulisan, batasan masalah, dan sistematika pembahasan.
2. **Bab 2 : Landasan Teori**
   Bagian ini berisi dasar teori yang digunakan dalam skripsi ini, yaitu model *survival*, model regresi linear, dan distribusi campuran.
3. **Bab 3 : Model Regresi *Cox Proportional Hazard***
   Pada bab ini akan dibahas mengenai pembentukan model regresi *Cox proportional hazard* dan pemilihan model terbaik.
4. **Bab 4 : Hasil Eksperimen dan Analisis**
   Pada bab ini akan dibahas mengenai tahap-tahap yang dilakukan sebelum eksperimen dan hasil eksperimen menggunakan model *Cox* yang disertai analisisnya.
5. **Bab 5 : Kesimpulan dan Saran**
   Bab ini berisi kesimpulan dari skripsi ini dan saran untuk pengembangan yang dapat dilakukan pada penelitian selanjutnya.