SKRIPSI

PREDIKSI TINGKAT PEMULIHAN KEHAMILAN PESERTA ASURANSI CACAT BERKELOMPOK JANGKA PANJANG MENGGUNAKAN METODE *RANDOM FORESTS*

Vivian

NPM: 2016710019

PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2020

FINAL PROJECT

PREDICTION OF MATERNITY RECOVERY RATES OF GROUP LONG TERM DISABILITY INSURANCE USING RANDOM FORESTS METHOD

Vivian

NPM: 2016710019

DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2020

LEMBAR PENGESAHAN

PREDIKSI TINGKAT PEMULIHAN KEHAMILAN PESERTA ASURANSI CACAT BERKELOMPOK JANGKA PANJANG MENGGUNAKAN METODE *RANDOM FORESTS*

Vivian

NPM: 2016710019

Bandung, 28 Juli 2020

Menyetujui,

Pembimbing 1	Pembimbing 2
Dr. Julius Dharma Lesmono	**Felivia Kusnadi, M.Act.Sc.**
Ketua Tim Penguji	**Anggota Tim Penguji**
Iwan Sugiarto, M.Si.	**Benny Yong, Ph.D.**

Mengetahui,

Ketua Program Studi

Dr. Erwinna Chendra

PERNYATAAN

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

PREDIKSI TINGKAT PEMULIHAN KEHAMILAN PESERTA ASURANSI CACAT BERKELOMPOK JANGKA PANJANG MENGGUNAKAN METODE *RANDOM FORESTS*

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 28 Juli 2020

Meterai
Rp. 6000

Vivian
NPM: 2016710019

ABSTRAK

Kehamilan adalah suatu masa yang dimulai dari konsepsi sampai lahirnya janin. Setiap kehamilan memiliki risiko yang bahkan dapat menyebabkan kematian. Angka Kematian Ibu (AKI) merupakan salah satu indikator yang menunjukkan kesejahteraan masyarakat di suatu negara. Oleh karena itu, perlu untuk memprediksi terjadinya tingkat pemulihan yang disebabkan oleh faktor-faktor pada kehamilan. Faktor pada kehamilan adalah variabel-variabel yang digunakan pada data *Group Long Term Disability* (GLTD) *Recovery Rates*. Variabel yang digunakan terdiri dari enam variabel prediktor, yaitu *Disability Category*, *Age Band*, *Duration*, *Own Occupation to Any Transition*, *Integration with Short Term Disability*, dan *Gross Indexed Benefit Amount* dan satu variabel respons, yaitu *Actual Recovery Rate*. Pada skripsi ini, akan dihitung tingkat pemulihan kehamilan dengan menggunakan metode *Random Forests*. Metode *Random Forests* merupakan metode pengembangan dari metode *Decision Trees*. Metode *Decision Trees* dapat diaplikasikan menjadi *Classification Trees* dan *Regression Trees*. Data yang digunakan pada skripsi ini memiliki variabel respons yang bersifat kuantitatif maka akan digunakan metode *Regression Trees*. Dalam metode *Random Forests* terdapat kumpulan metode *Regression Trees* yang akan dilakukan dengan cara *bootstrapping*. Metode *Random Forests* akan melakukan estimasi pada variabel yang digunakan dan memprediksi tingkat pemulihan kehamilan. Selanjutnya, akan dilakukan perbandingan pada metode *Regression Trees* dan metode *Random Forests*. Perbandingan kedua metode dilakukan dengan menghitung nilai *Mean Square Error* (MSE), yaitu MSE *Predicted*. MSE *Predicted* akan dilakukan dengan menggunakan *test* data setelah mencocokkan model pada *training* data. Berdasarkan analisis yang diperoleh dapat disimpulkan bahwa MSE *Predicted* pada metode *Random Forests* lebih baik digunakan untuk memprediksi data GLTD *Recovery Rates* pada data kehamilan karena memiliki nilai prediksi yang lebih kecil daripada metode *Regression Trees*.

Kata-kata kunci: Kehamilan, Tingkat Pemulihan, Metode *Regression Trees*, Metode *Random Forests*, *Mean Square Error* (MSE)

ABSTRACT

Maternity is a period that begins with conception until the birth of the fetus. Every maternity has a risk that can even result in death. Maternal Mortality Rate (MMR) is one of the indicators that shows the welfare of the people in a country. Therefore, it is important to predict the recovery rate that is caused by factors of maternity. The factors of maternity are the variables that are used in the data of Group Long Term Disability (GLTD) Recovery Rates. The variables that are used consists of six predictor variables, which is Disability Category, Age Band, Duration, Own Occupation to Any Transition, Integration with Short Term Disability, and Gross Indexed Benefit Amount and one response variable, which is Actual Recovery Rate. In this final project, will calculate the recovery rate of maternity using the Random Forests method. The Random Forests method is the method of developing the Decision Trees method. The Decision Trees method can be applied to Classification Trees and Regression Trees. Data that is used in this final project has a response variable that is quantitative then it will be used the Regression Trees method. In Random Forests method, there is a group of Regression Trees method that will be done through bootstrapping. The Random Forests method will estimate the variables that are used and predict recovery rates of maternity. Furthermore, a comparison will be done between the Regression Trees and the Random Forests methods. This comparison methods will be done by calculating the Mean Square Error (MSE) rate, which is MSE Predicted. MSE Predicted will be done using test data after fitting the model on the training data. According to the analysis that is obtained, it can be concluded that MSE Predicted in the Random Forests method is better used to predict GLTD Recovery Rates in maternity data because it has a smaller predictive value than the Regression Trees method.

Keywords: Maternity, Recovery Rates, Regression Trees Method, Random Forests Method, Mean Square Error (MSE)

God is able to do anything

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus karena kasih dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Prediksi Tingkat Pemulihan Kehamilan Peserta Asuransi Cacat Berkelompok Jangka Panjang Menggunakan Metode Random Forests"**. Skripsi ini disusun sebagai salah satu syarat unutuk menyelesaikan studi Strata-1 Program Studi Matematika, Fakultas Teknologi Informasi dan Sains (FTIS), Universitas Katolik Parahyangan, Bandung. Penulis berharap skripsi ini dapat berguna 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 menyampaikan ucapan terima kasih kepada :

1. Kedua orang tua (papi dan mami) dan Vilencia yang selalu mendoakan, mendukung, dan memberikan semangat sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

2. Bapak Dr. Julius Dharma Lesmono selaku dosen pembimbing-1 yang telah memberikan ilmu, arahan, dan saran dalam proses penyusunan skripsi.

3. Ibu Felivia Kusnadi, M.Act.Sc. selaku dosen pembimbing-2 yang telah memberikan bantuan, arahan, saran, dan nasihat sehingga skripsi ini dapat terselesaikan dengan baik.

4. Bapak Iwan Sugiarto, M.Si. selaku dosen penguji-1 dan Bapak Benny Yong, Ph.D. 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. Seluruh dosen, staf tata usaha, dan karyawan FTIS yang memberikan ilmu, dukungan, dan bantuan selama masa perkuliahan.

7. Ebenhaezer Hardani yang selalu mendengarkan keluh kesah dan menemani penulis, mendoakan, memberikan nasihat, dan semangat kepada penulis.

8. Angeline yang selalu mendengarkan keluh kesah dan menemani penulis, mendorong, memberikan semangat, dan dukungan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan tepat waktu.

9. Catherine Hauwanda, Theresa GJS, dan Felicia Kammal yang telah memberikan semangat kepada penulis.

10. Sahabat-sahabat "Cicans" : Fanny, Fenny, Jessica, dan Triny, yang telah menemani penulis, memberikan semangat, dan dukungan selama masa perkuliahan dan proses penulisan skripsi.

11. Keluarga komsel *Philadelphia Reborn* :

    - Jessica Elvina Tansil dan ko Yonathan Kristian Purnama yang membantu, memberikan semangat, dan dukungan kepada penulis.

- Natasia Pandora, Yovita Nathania, Michael Joshua, dan ci Windy Wilianti yang menyemangati penulis dalam penulisan skripsi.
- Dan saudara-saudara lain yang telah mendoakan, menyemangati, memberi dukungan serta kebersamaan.

12. Teman-teman Matematika 2016: Nadya, Melia, Leo, Laureen, JC, Ivan, Rudi, Fenny, Davyn, Claresta, Aretha, JT, Geraldy, Avel, Isa, Faza, Muti, Vheren, Niko, Julius, Chrestella, Yonathan, Alma, Al-vinda, Azka, Aldo, Felix, Salman, Asen, Widhiya, Evelyne, Edsel, Salomo, Triny, Nevan, Irsyad, Farand, Lucas, Febri, Deva, Adin, Khema, Bahri, Yohanes, Raisa, Wilbert, Daniel, Fransiskus, Janaka, Agnes, Suryani, Fanny, Nitya, Nur, dan Gresel. Terima kasih untuk setiap dukungan, kebersamaan, dan pengalaman yang telah kita lalui bersama.
13. Teman-teman Matematika angkatan 2014, 2015, 2017, 2018 yang tidak dapat disebutkan satu per satu.
14. Terakhir, kepada 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 dengan terbuka menerima 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, Juli 2020

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada suatu perusahaan asuransi terdapat beberapa jenis asuransi, salah satunya adalah asuransi kesehatan. Asuransi kesehatan merupakan produk asuransi yang menangani masalah kesehatan tertanggung karena suatu penyakit serta menanggung biaya proses perawatan. Pada dasarnya, biaya penyebab sakit tertanggung yang dapat ditanggung oleh perusahaan asuransi adalah cedera yang parah, cacat, sakit hingga kematian karena kecelakaan [1]. Kecacatan seseorang dapat disebabkan oleh berbagai faktor, salah satunya cacat pada kehamilan sehingga terbentuk asuransi disabilitas (cacat) pada kehamilan.

Asuransi cacat adalah jenis asuransi yang akan memberikan penghasilan jika pihak tertanggung (peserta asuransi) tidak dapat melakukan pekerjaan mereka karena cacat. Asuransi cacat terdapat 2 jenis, yaitu asuransi cacat jangka pendek dan asuransi cacat jangka panjang. Asuransi cacat jangka pendek adalah polis asuransi yang melindungi pihak tertanggung dari kehilangan pendapatan jika mereka sementara tidak dapat bekerja karena sakit, cedera, atau kecelakaan, sedangkan asuransi cacat jangka panjang jika pihak tertanggung tidak dapat bekerja untuk jangka waktu yang lama. Setiap polis asuransi cacat memiliki kondisi yang berbeda untuk pembayaran, penyakit atau kondisi yang sudah terjadi sebelumnya yang dapat dikecualikan dan berbagai kondisi lain yang membuat polis tersebut lebih atau kurang bermanfaat bagi pihak tertanggung [2].

Pada skripsi ini, kehamilan tergolong dalam asuransi cacat jangka panjang (kebijakan kelompok di Amerika Serikat) karena pada umumnya mereka yang memberikan perlindungan setelah tunjangan pada asuransi cacat jangka pendek berhenti membayar. Dengan demikian, tidak terdapat perbedaan mendasar antara seseorang menjadi cacat karena komplikasi kehamilan daripada penyakit lainnya. Dalam hal ini, kehamilan akan pulih lebih cepat dibandingkan penyakit yang lain [3].

Menurut *World Health Organization* (WHO), Angka Kematian Ibu (AKI) merupakan salah satu indikator kesehatan suatu bangsa. Kematian ibu merupakan kematian wanita yang dapat disebabkan pada saat kondisi hamil atau menjelang 42 hari setelah persalinan. Hal ini dapat terjadi akibat suatu kondisi yang diperberat oleh kehamilannya maupun dalam masa pemulihan, tetapi bukan termasuk kematian ibu hamil yang diakibatkan karena kecelakaan. AKI diakibatkan karena risiko (faktor) yang dihadapi oleh seorang ibu selama masa kehamilan hingga persalinan.

Kesehatan calon ibu adalah salah satu aspek yang penting untuk diperhatikan dalam siklus kehidupan seorang perempuan karena sepanjang masa kehamilannya dapat terjadi komplikasi yang tidak diharapkan. Setiap calon ibu akan menghadapi risiko yang bisa mengancam jiwanya. Oleh karena itu, setiap calon ibu akan memprediksi bahwa mereka akan tetap pulih terhadap risiko yang terjadi.

Pada tahun 2011, komite *Society of Actuaries* (SOA) menerbitkan Tabel Laporan *Group Long Term Disability* (GLTD) *Experience* tahun 2008 yang merinci pada model matematika untuk menentukan tingkat pemulihan. Dalam skripsi ini, akan menggunakan metode *Random Forests* untuk menghitung prediksi tingkat pemulihan kehamilan dari data GLTD *Recovery Rates* pada tahun 2008 [3]. Pada dasarnya, digunakan metode *Random Forests* adalah untuk menjaga akurasi tetap baik dari jumlah data yang besar dengan variabel yang banyak dan meminimalkan kesalahan

(*error*) untuk mengurangi *overfitting*.

Metode *Random Forests* merupakan pengembangan dari metode CART (*Classification and Regression Trees*). Data GLTD *Recovery Rates* menggunakan variabel respons yang bersifat numerik maka akan digunakan metode *Regression Trees*. Metode *Regression Trees* merupakan salah satu metode prediksi yang menggunakan representasi struktur pohon (*tree*) dan yang paling populer digunakan. Selain itu, pembangunannya yang relatif cepat, hasil dari model yang dibangun mudah untuk diinterpretasikan.

Di dalam metode *Random Forests* terdapat kumpulan metode *Regression Trees* yang dibangun dari sampel acak sehingga membentuk hutan (*forests*). Metode *Random Forests* menghasilkan hasil akhir berupa simulasi numerik yaitu sebuah nilai galat kuadrat rata-rata (MSE). Dengan demikian, menggunakan metode *Random Forests* diharapkan dapat menghasilkan model yang baik untuk memprediksi tingkat pemulihan kehamilan. Model tersebut memiliki banyak pohon dengan setiap simpul merupakan atribut terpilih karena kemampuannya yang baik untuk memprediksi. Atribut yang dijadikan simpul pada pohon dapat dipertimbangkan sebagai faktor-faktor (variabel) yang penting dalam tingkat pemulihan.

1.2 Rumusan Masalah

Masalah yang akan dibahas pada skripsi ini adalah:

1. Bagaimana mengetahui faktor-faktor yang mempengaruhi tingkat pemulihan kehamilan dengan menggunakan metode *Regression Trees* dan *Random Forests* ?
2. Bagaimana memprediksi tingkat pemulihan kehamilan dengan menggunakan metode *Random Forests* ?
3. Bagaimana perbandingan hasil yang diperoleh dari metode *Regression Trees* dan metode *Random Forests* ?

1.3 Tujuan

Tujuan dari penulisan skripsi ini adalah:

1. Mengetahui faktor-faktor yang mempengaruhi tingkat pemulihan kehamilan dengan menggunakan metode *Regression Trees* dan *Random Forests*.
2. Mengetahui prediksi tingkat pemulihan kehamilan dengan menggunakan metode *Random Forests*.
3. Membandingkan hasil yang diperoleh dari metode *Regression Trees* dan metode *Random Forests*.

1.4 Batasan Masalah

Batasan masalah yang digunakan dalam skripsi ini adalah:

1. Data yang diolah adalah data peserta asuransi cacat berkelompok jangka panjang.
2. Orang yang hamil akan sembuh kembali (tidak meninggal).
3. Rentang usia ibu hamil antara 20 hingga 70 tahun.

1.5 Metodologi

Pada skripsi ini, akan mempelajari metode *Regression Trees* dan *Random Forests*. Selanjutnya, menerapkan kedua metode tersebut melalui simulasi numerik dan membandingkan hasil yang diperoleh dari kedua metode, serta melakukan analisis terhadap pengaruh faktor yang menyebabkan perubahan pada tingkat pemulihan kehamilan. Data yang digunakan adalah *Group Long Term Disability Recovery Rates* (2008) [3]. Untuk memudahkan perhitungan, perangkat lunak yang digunakan yaitu *R* dan Microsoft Excel.

1.6 Sistematika Pembahasan

Sistematika pembahasan pada skripsi ini terdiri dari 5 bab, yakni:
BAB 1 : Pendahuluan
Pada bab ini akan dibahas latar belakang, rumusan masalah, tujuan penulisan, batasan masalah, metodologi, dan sistematika pembahasan.
BAB 2 : Landasan Teori
Pada bab ini akan membahas teori pendukung seperti *Machine Learning*, *Decision Trees*, *Mean Square Error* (MSE), dan contoh sederhana dari penggunaan metode *Regression Trees*.
BAB 3 : *Random Forests*
Pada bab ini akan membahas mengenai teori metode *Random Forests*, *Bootstrap*, *Out-of-Bag Error Estimation*, *Variable Importance*, dan contoh sederhana dari penggunaan metode *Random Forests*.
BAB 4 : Hasil Perhitungan dan Analisis Hasil Prediksi
Pada bab ini memuat hasil perhitungan dengan menggunakan kedua metode dan menganalisis hasil prediksi dari pengaruh faktor terhadap tingkat pemulihan kehamilan yang telah dibahas pada Bab 2 dan Bab 3.
BAB 5 : Kesimpulan dan Saran
Pada bab ini berisi kesimpulan yang diperoleh dari hasil perhitungan dan analisis pada Bab 4 dan saran untuk pengembangan yang dapat dilakukan pada penelitian selanjutnya.