SKRIPSI

PEMODELAN TINGKAT PEMULIHAN ASURANSI GRUP CACAT JANGKA PANJANG MENGGUNAKAN *GRADIENT BOOSTING MACHINE* (GBM)

Agnes G. Mercyana S.

NPM: 2016710053

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

FINAL PROJECT

GROUP LONG TERM DISABILITY RECOVERY RATE MODELING USING GRADIENT BOOSTING MACHINE (GBM)

Agnes G. Mercyana S.

NPM: 2016710053

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

LEMBAR PENGESAHAN

PEMODELAN TINGKAT PEMULIHAN ASURANSI GRUP CACAT JANGKA PANJANG MENGGUNAKAN *GRADIENT BOOSTING MACHINE* (GBM)

Agnes G. Mercyana S.

NPM: 2016710053

Bandung, 27 Juli 2020

Menyetujui,

Pembimbing 1	Pembimbing 2
Dr. Julius Dharma Lesmono	**Felivia Kusnadi, M.Act.Sc.**
Ketua Tim Penguji	**Anggota Tim Penguji**
Farah Kristiani, Ph.D.	**Dr. Andreas Parama Wijaya**

Mengetahui,

Ketua Program Studi

Dr. Erwinna Chendra

PERNYATAAN

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

PEMODELAN TINGKAT PEMULIHAN ASURANSI GRUP CACAT JANGKA PANJANG MENGGUNAKAN *GRADIENT BOOSTING MACHINE* (GBM)

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 27 Juli 2020

Meterai
Rp. 6000

Agnes G. Mercyana S.
NPM: 2016710053

ABSTRAK

Suatu perusahaan asuransi perlu mengetahui bagaimana tingkat pemulihan cacat setiap kliennya maka dari itu perlu dibuat suatu model yang dapat memprediksi tingkat pemulihan cacat. *Society of Actuaries* (SOA) membuat model prediksi tingkat pemulihan cacat tersebut dengan menggunakan pendekatan Pohon Keputusan. Pohon Keputusan adalah metode pembelajaran mesin untuk membangun model prediksi dari data. Model diperoleh dengan membuat partisi terhadap ruang prediktor (ruang yang dibentuk oleh variabel independen) menjadi sejumlah bagian sederhana. Partisi tersebut dapat direpresentasikan secara grafis sebagai pohon keputusan (*decision tree*). Guna menghasilkan model prediksi yang lebih akurat dari metode Pohon Keputusan, dikembangkan metode *ensemble*. Metode tersebut membangun beberapa model prediktif kemudian mengintegrasikan model-model tersebut agar diperoleh model dengan performansi prediksi yang lebih baik. Salah satu metode *ensemble* yang populer adalah *Boosting*. *Boosting* meliputi banyak algoritma, dua di antaranya adalah *Gradient Boosting Machine* (GBM) dan *AdaBoost*. Pada skripsi ini akan dibahas mengenai pemodelan tingkat pemulihan cacat jangka panjang dengan menggunakan metode GBM. GBM mampu meningkatkan performansi prediksi yang dihasilkan dari Pohon Keputusan yang telah dibahas dalam jurnal *Predicting Group Long Term Disability Recovery and Mortality Rate Using Tree Models*. Akan digunakan *Mean Square Error* (MSE) untuk memvalidasi model prediksi yang diperoleh menggunakan GBM. Kemudian akan dibandingkan nilai MSE dari model GBM dan nilai MSE dari model Pohon Keputusan. Setelah dilakukan simulasi ternyata nilai MSE dari model GBM lebih kecil dibanding MSE dari model Pohon Keputusan. Maka dapat disimpulkan bahwa GBM mampu meningkatkan performansi model dalam memprediksi tingkat pemulihan cacat. GBM juga dapat disesuaikan untuk kebutuhan aplikasi tertentu dengan menyesuaikan *loss-function* yang digunakan.

Kata-kata kunci: Tingkat pemulihan, *Decision Tree*, Metode *Ensemble*, *Boosting*, *Gradient Boosting Machine* (GBM), *Mean Square Error* (MSE), *Loss-function*.

ABSTRACT

An insurance company needs to know the recovery rate of each client's disability therefore, it is necessary to build a model that can predict the rate of recovery of disability. The Society of Actuaries (SOA) builds a predictive model for the recovery rate of disability using Decision Tree. Decision Tree is a machine learning method to build predictive models from data. The model is obtained by partitioning the predictor space (the space formed by independent variables) into a number of simple regions. The partition can be represented graphically as a decision tree. In order to obtain a more accurate prediction model than the Decision Tree method, an ensemble method was developed. It builds a number of predictive models and then integrates these models to get a model with better prediction performance. One popular ensemble method is Boosting. The two popular Boosting algorithms are Gradient Boosting Machine (GBM) and AdaBoost. This final project will discuss the modelling of long-term disability recovery rates using GBM method. GBM is able to improve the performance of predictions resulting from Decision Tree discussed in journal Predicting Group Long Term Disability Recovery and Mortality Rates Using Tree Models. Mean Square Error (MSE) will be used to validate the prediction models obtained using GBM. Then the MSE from the GBM and Decision Tree model will be compared. Based on the simulation results it turns out that the MSE value of GBM model is smaller than the MSE of Decision Tree model. It can be concluded that GBM can improve the performance of the model in predicting recovery rate of disability. GBM also can be used for specific application needs by adjusting the loss-function.

Keywords: Recovery rate, Decision Tree, Ensemble Method, Boosting, Gradient Boosting Machine (GBM), Mean Square Error (MSE), Loss-function.

Untuk aku dan orang-orang terkasih...

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan Yesus Kristus atas kasih dan karunia-Nya yang melimpah dalam setiap langkah hidup penulis terutama saat penyusunan skripsi hingga penulis mampu menyelesaikan skripsi tepat pada waktunya. Skripsi yang berjudul "Pemodelan Tingkat Pemulihan Asuransi Grup Cacat Jangka Panjang Menggunakan *Gradient Boosting Machine* (GBM)" disusun sebagai salah satu syarat wajib untuk menyelesaikan studi Strata-I Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung. Penulis berharap skripsi ini dapat berguna bagi setiap orang yang membacanya.

Selama masa perkuliahan dan penyusunan skripsi, penulis telah mendapatkan banyak ilmu, bantuan, dukungan serta doa dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada:

- Ir. Heren Helena (Mama) yang tak pernah lelah mendoakan, mencintai, membimbing, menghibur, dan mendukung penulis sehingga dapat menyelesaikan masa kuliah dan skripsi ini dengan baik. Dan kepada Alm. Katamso Benhur (Papa) di surga yang selalu mendoakan penulis.
- Kakak dan adik tercinta, Harenka Paulina Agatha, Rodo Gabriel Oloanki, dan Fx. Reynaldo Bagas yang selalu mendoakan, memberi semangat, membantu dan selalu sabar mendengarkan keluh kesah penulis.
- Bapak Dr. Julius Dharma Lesmono dan Ibu Felivia Kusnadi, M.Act.Sc. selaku Dosen Pembimbing 1 dan 2 yang telah membimbing, memberikan ilmu, saran, dan nasihat kepada penulis sehingga skripsi ini dapat selesai dengan baik.
- Ibu Farah Kristiani, Ph.D. selaku dosen penguji-1 dan Bapak Dr. Andreas Parama Wijaya selaku dosen penguji-2 yang telah memberikan ilmu, saran dan kritik sehingga skripsi ini menjadi lebih baik.
- Bapak Dr. Ferry Jaya Permana, ASAI selaku Dosen Wali yang telah membimbing dan memberikan ilmu kepada penulis selama masa perkuliahan.
- Bapak Liem Chin, M.Si selaku koordinator skripsi, terima kasih atas segala saran, bantuan, dan informasi yang diberikan.
- Seluruh dosen, staf Tata Usaha Fakultas Teknologi Informasi dan Sains (FTIS), terutama dosen Program Studi Matematika, terima kasih atas segala ilmu yang telah diberikan selama masa perkuliahan.
- Christopher Aryo Pambudi yang selalu mendoakan, memberi semangat dan saran, juga selalu sabar mendengarkan keluh kesah penulis. Terima kasih karena telah hadir dan membawa suka cita ke dalam hidup penulis.
- Aretha Belicia, Suryani, dan Nitya Salsabila sebagai sahabat yang selalu menemani, memberikan semangat, membantu, mendengarkan segala cerita senang maupun sedih selama masa perkuliahan. Terima kasih karena telah hadir dan membawa suka cita ke dalam hidup penulis.

- Febrizio Willem Ong yang selalu membantu dan bersedia berbagi ilmu kepada penulis selama masa perkuliahan.
- Teman-teman angkatan 2016: Nadya, Melia, Leo, Laureen, Jessica C, Ivan, Rudi, Fenny, Davyn, Claresta, Aretha, Jesicca T, Gerald, Avel, Isa, Faza, Muti, Vheren, Vivian, 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, Suryani, Fanny, Nitya, dan Nur. Terima kasih untuk kebersamaan yang pernah dilalui bersama selama masa perkuliahan.
- Teman-teman Matematika angkatan 2014, 2015, 2017, yang tak dapat disebutkan satu persatu.
- 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 mengharapkan masukkan saran dan kritik 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 lagi. Terima kasih.

Bandung, Juli 2020

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada dasarnya manusia tidak dapat memprediksi kapan suatu musibah akan terjadi. Musibah tersebut bisa menimbulkan risiko yang dapat mengancam jiwa atau harta benda. Salah satu cara untuk mengatasi risiko tersebut adalah dengan mengikuti program asuransi. Asuransi terbagi menjadi dua jenis berdasarkan banyaknya jumlah tertanggung yaitu asuransi individu dan asuransi kelompok. Asuransi individu adalah polis asuransi yang hanya melindungi satu orang, sedangkan asuransi kelompok adalah polis asuransi yang melindungi sekelompok orang seperti anggota atau karyawan dari suatu perusahaan. Perusahaan yang baik biasanya sudah memiliki sistem proteksi untuk melindungi karyawannya dari risiko kerugian akibat kehilangan pendapatan karena tidak mampu bekerja akibat suatu musibah seperti sakit, cedera, atau kecelakaan. Sistem proteksi yang dimaksud ialah asuransi kelompok cacat. Asuransi kelompok cacat terbagi menjadi dua jenis berdasarkan jangka waktu pemberian manfaat yaitu asuransi kelompok cacat jangka pendek dan asuransi kelompok jangka panjang. Asuransi kelompok cacat jangka pendek membayarkan manfaat untuk periode waktu yang singkat biasanya tiga bulan, enam bulan, atau satu tahun, setelah 1-14 hari periode eliminasi, sedangkan pada asuransi kelompok cacat jangka panjang manfaat dibayarkan untuk jangka waktu yang lebih lama, yakni dua tahun, lima tahun, 10 tahun, hingga usia 65 tahun, atau seumur hidup, tergantung pada kebijakan. Manfaat ini dibayarkan setelah 10-53 minggu periode eliminasi. Semakin lama masa pertanggungan, semakin tinggi premi. Skripsi ini akan berfokus pada asuransi kelompok cacat jangka panjang.

Suatu perusahaan asuransi yang menawarkan program asuransi kelompok cacat jangka panjang perlu mengetahui bagaimana tingkat pemulihan cacat setiap kliennya untuk beberapa kepentingan. Oleh karena itu, perlu dibuat suatu model yang dapat memprediksi tingkat pemulihan cacat. Pohon Keputusan merupakan metode pembelajaran mesin untuk membangun model prediksi dari data dalam bentuk struktur pohon. Terdapat banyak algoritma untuk membangun Pohon Keputusan salah satunya adalah *Iterative Dichotomiser* atau disebut juga $ID3$. Algoritma $ID3$ digunakan untuk memodelkan kasus klasifikasi maupun kasus regresi. Model diperoleh dengan membagi ruang prediktor (ruang yang dibentuk oleh variabel independen) menjadi sejumlah bagian sederhana. Akibatnya, partisi dapat direpresentasikan secara grafis sebagai pohon keputusan.

Berdasarkan [1] *The Society of Actuaries (SOA) Committee* menerbitkan model matematika untuk menentukan tingkat pemulihan dan kematian dengan menggunakan pendekatan Pohon Keputusan. Namun model Pohon Keputusan memiliki kekurangan yaitu tidak fleksibel terhadap data baru sehingga hal ini menyebabkan performansi model dalam memprediksi data baru kurang baik.

Guna meningkatkan performansi prediksi dari metode Pohon Keputusan, dikembangkan metode *ensemble* yang membangun sebuah model prediktif dengan mengintegrasikan beberapa model sederhana menjadi suatu model tunggal yang mampu menghasilkan prediksi yang lebih akurat. Salah satu metode *ensemble* yang populer adalah *Boosting*. Ide utama dari *Boosting* adalah membentuk model aditif dari suatu model sederhana (*base-learner*) secara berurutan yang bertujuan untuk memperbaiki kesalahan prediksi dari model sebelumnya. *Boosting* meliputi banyak algoritma

dua di antaranya adalah *Gradient Boosting Machine* (GBM) dan *AdaBoost*.

Pada skripsi ini akan dibahas mengenai pemodelan tingkat pemulihan cacat jangka panjang berdasarkan data pada [1] dengan menggunakan metode GBM. Metode GBM dipilih karena mampu menangani kelemahan dari model Pohon Keputusan yang telah disebutkan sebelumnya [2]. GBM mampu meningkatkan performansi dari model Pohon Keputusan dengan cara menggabungkan beberapa model sederhana secara berurutan yang dimana model sederhana tersebut dibangun berdasarkan *negative gradient* dari suatu *loss-function*. Pada skripsi ini juga akan dibandingkan nilai *mean square error* dari kedua model tersebut.

1.2 Rumusan Masalah

Masalah yang dibahas pada skripsi ini adalah:

1. Bagaimana model untuk memprediksi tingkat pemulihan menggunakan *Gradient Boosting Machine* (GBM)?
2. Bagaimana perbandingan MSE dari model Pohon Keputusan berdasarkan [1] dengan MSE dari model *Gradient Boosting Machine* (GBM)?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah dijelaskan, tujuan dari penulisan skripsi ini adalah:

1. Membuat model untuk memprediksi tingkat pemulihan dengan menggunakan *Gradient Boosting Machine* (GBM).
2. Membandingkan nilai MSE yang diperoleh dari model Pohon Keputusan berdasarkan [1] dengan *Gradient Boosting Machine* (GBM).

1.4 Batasan Masalah

Batasan masalah pada skripsi ini adalah :

1. Jenis asuransi grup yang dibahas adalah asuransi kelompok cacat jangka panjang saja.
2. Diasumsikan pekerja yang cacat akan sembuh kembali, tidak meninggal.
3. Metode yang dipakai adalah *Gradient Boosting Machine* (GBM).
4. *Base-learner* yang dipakai adalah Pohon Keputusan.
5. *Loss-function* yang dipakai adalah *Least-squared Loss-Function*.
6. Data yang digunakan untuk simulasi hanya data untuk kasus regresi.

1.5 Sistematika Pembahasan

Sistematika pembahasan pada skripsi ini terdiri dari lima bab, yaitu :

Bab 1: Pendahuluan
Bab ini terdiri dari latar belakang, rumusan masalah, tujuan, batasan masalah dan sistematika pembahasan.

Bab 2: Landasan Teori Bab ini membahas teori-teori yang mendukung dalam penulisan makalah ini, yaitu *regression tree, classification tree*, *tree pruning* dan *least-squared loss-function.*

Bab 3: *Gradient Boosting Machine* (GBM)
Bab ini membahas *Boosting*, *Steepest Descent* dan *Gradient Boosting Machine* (GBM).

Bab 4: Simulasi Data
Bab ini membahas simulasi data yang didapat dari *Society of Actuaries* (SOA) [1] dan mengevaluasi performansi model prediksi yang diperoleh dengan menggunakan *Gradient Boosting Machine* (GBM). Kemudian membandingkan MSE dari model GBM dan model Pohon Keputusan.

Bab 5: Kesimpulan dan Saran
Bab ini berisi kesimpulan dari skripsi dan saran untuk pengembangan lebih lanjut.