

SKRIPSI

**APLIKASI *BAYESIAN ADDITIVE REGRESSION TREE*
(BART) UNTUK MEMPREDIKSI TINGKAT PEMULIHAN
KEHAMILAN ASURANSI DISABILITAS KELOMPOK
JANGKA PANJANG**



Stevanny Budiana

NPM: 6161801026

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

FINAL PROJECT

**APPLICATION OF BAYESIAN ADDITIVE REGRESSION
TREE (BART) FOR PREDICTING MATERNITY RECOVERY
RATE OF GROUP LONG TERM DISABILITY INSURANCE**



Stevanny Budiana

NPM: 6161801026

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

LEMBAR PENGESAHAN

APLIKASI *BAYESIAN ADDITIVE REGRESSION TREE* (BART) UNTUK MEMPREDIKSI TINGKAT PEMULIHAN KEHAMILAN ASURANSI DISABILITAS KELOMPOK JANGKA PANJANG

Stevanny Budiana

NPM: 6161801026

Bandung, 12 Juli 2022

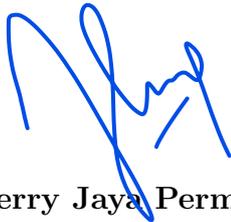
Menyetujui,

Pembimbing



Felivia Kurnadi, MActSc, ASAI

Ketua Tim Penguji



Dr. Ferry Jaya Permana, ASAI

Anggota Tim Penguji



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:

**APLIKASI *BAYESIAN ADDITIVE REGRESSION TREE* (BART) UNTUK
MEMPREDIKSI TINGKAT PEMULIHAN KEHAMILAN ASURANSI
DISABILITAS KELOMPOK JANGKA PANJANG**

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 12 Juli 2022



Stevanny Budiana
NPM: 6161801026

ABSTRAK

Asuransi disabilitas adalah asuransi yang memberikan manfaat apabila tertanggung mengalami cedera atau terkena suatu penyakit yang menyebabkan tertanggung tidak dapat bekerja dan memperoleh pendapatan. Setiap perusahaan yang menerbitkan produk asuransi disabilitas perlu mengetahui risiko yang ditanggung perusahaan dengan cara memprediksi tingkat pemulihan tertanggung yang akan digunakan untuk menentukan manfaat dan premi setiap tertanggung. Manfaat yang dibayarkan oleh perusahaan ditentukan berdasarkan tingkat pemulihan dari tertanggung menggunakan metode berbasis pohon keputusan. Data tingkat pemulihan yang digunakan adalah tingkat pemulihan untuk kasus disabilitas kehamilan yang diambil dari data asuransi disabilitas kelompok jangka panjang dari laman *Society of Actuaries* (SOA) dan metode utama yang akan digunakan untuk memprediksi tingkat pemulihan adalah metode *Bayesian Additive Regression Tree*. Prediksi tingkat pemulihan kehamilan akan diprediksi menggunakan beberapa metode berbasis pohon seperti *Bayesian Additive Regression Tree*, *Gradient Boosting Machine*, *Random Forest*, dan *Decision Tree* dan dibandingkan berdasarkan nilai *root mean squared error* (RMSE) dan waktu *runtime* dari masing-masing model yang terbentuk. Berdasarkan perbandingan hasil prediksi untuk setiap *dataset*, didapatkan bahwa metode *Bayesian Additive Regression Tree* memberikan hasil prediksi yang terbaik dengan nilai RMSE terkecil dan waktu *runtime* yang relatif singkat.

Kata-kata kunci: Asuransi Disabilitas, Tingkat Pemulihan Kehamilan, Metode Berbasis Pohon Keputusan, *Bayesian Additive Regression Tree*, *Root Mean Squared Error*, *Runtime*.

ABSTRACT

Disability insurance gives benefits to the insured if the insured has suffered from some injury or illness that unables the insured to work on their job and get some income. Every insurance company that releases disability insurance products needs to know the underlying risks by predicting the insured's recovery rate to determine the number of benefits and premiums for each insured. Benefits paid by a company are determined by the insured's predicted recovery rate using some decision tree-based method. The data used for simulation is the recovery rate for maternity disability cases from group long-term disability data taken from Society of Actuaries (SOA) and the main method used for predicting maternity recovery rate is Bayesian Additive Regression Tree method. The predicted maternity recovery rate will be predicted using some tree-based methods such as Bayesian Additive Regression Tree, Gradient Boosting Machine, Random Forest, and Decision Tree. Each model built using the said method will be compared to the other by comparing the mean squared error value and runtime from each model. According to the comparison between the model for every dataset built, Bayesian Additive Regression Tree gives the best prediction based on smaller Root Mean Squared Error values and relatively short runtime.

Keywords: Disability Insurance, Maternity Recovery Rate, Decision Tree-Based Method, Bayesian Additive Regression Tree, Root Mean Squared Error, Runtime.

*Untuk semua pihak yang telah memberikan dukungan dan arahan
selama proses penulisan skripsi*

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya yang melimpah sehingga penulis dapat menyelesaikan proses penyusunan skripsi berjudul "Aplikasi *Bayesian Additive Regression Tree* (BART) Untuk Memprediksi Tingkat Pemulihan Kehamilan Asuransi Disabilitas Kelompok Jangka Panjang". Skripsi dengan judul tersebut disusun sebagai salah satu syarat wajib untuk menyelesaikan studi Strata-I Program Studi Matematika, Fakultas Teknologi Informasi dan Sains (FTIS), Universitas Katolik Parahyangan (UNPAR), Bandung. Penulis berharap skripsi ini dapat berguna bagi setiap orang yang membacanya.

Selama proses penyusunan skripsi, penulis mendapat banyak dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

- Kedua orang tua dan adik penulis yang selalu memberikan dukungan dan semangat dari awal perkuliahan hingga penyusunan skripsi berakhir.
- Keluarga besar yang selalu menyemangati penulis selama masa perkuliahan.
- Ibu Felivia Kusnadi, M.ActSc, ASAI selaku dosen pembimbing 1 yang sudah membimbing, memberikan ilmu, arahan, dan saran selama proses penyusunan skripsi.
- Bapak Dr. Ferry Jaya Permana, ASAI selaku dosen penguji 1 yang telah memberikan saran dan komentar sehingga skripsi ini selesai dengan baik.
- Bapak Dr. Andreas Parama Wijaya selaku dosen penguji 2 yang telah memberikan saran dan komentar sehingga skripsi ini selesai dengan baik.
- Bapak Liem Chin, M.Si. selaku dosen wali yang telah membimbing penulis selama perkuliahan.
- Bapak Dr. Daniel Salim, M.Si selaku koordinator skripsi, terima kasih atas segala saran, bantuan, dan informasi yang diberikan.
- Bapak Robyn Irawan Tjia, S.Si, M.Sc yang membantu dalam proses penulisan skripsi dan pemrograman.
- Seluruh dosen FTIS, terutama dosen Program Studi Matematika, terima kasih atas ilmu yang dibagikan selama masa perkuliahan.
- Seluruh staf Tata Usaha dan karyawan FTIS. Terima kasih atas segala bantuan selama masa perkuliahan.
- Teman-teman *Bunchies*: Celine, Devina, Evadne, Kimberley, dan Nadia yang selalu memberikan semangat dan berbagi cerita, informasi, dan waktu selama proses perkuliahan hingga masa penyusunan skripsi.
- Teman-teman KAZMAS: Angela, Kaschia, Melissa, dan Zefanya yang memberikan semangat dan mendengar keluh kesah penulis selama perkuliahan.
- Semua teman seangkatan 2018 atas pengalaman dan kebersamaan yang dilalui selama proses perkuliahan.
- Semua pihak yang telah membantu dalam masa perkuliahan dan penyusunan skripsi.

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

Bandung, Juli 2022

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 Asuransi Disabilitas	5
2.2 Pohon Keputusan	5
2.3 Teorema Bayes	7
2.3.1 Fungsi <i>Likelihood</i>	7
2.3.2 <i>Prior</i>	8
2.3.3 <i>Posterior</i>	9
2.4 <i>Markov Chain Monte Carlo</i>	9
2.4.1 Algoritma <i>Metropolis</i>	9
2.4.2 Algoritma <i>Metropolis-Hasting</i>	10
2.4.3 Algoritma MCMC	10
2.4.4 Contoh Penerapan Algoritma <i>Metropolis</i>	10
2.5 <i>Mean Squared Error</i> dan <i>Root Mean Squared Error</i>	11
2.6 Metode Pengambilan Sampel Ulang	11
2.6.1 Validasi Silang	11
2.6.2 <i>Bootstrap</i>	11
2.7 Metode Berbasis Pohon Keputusan	13
2.7.1 <i>Decision Tree</i>	13
2.7.2 <i>Random Forest</i>	13
2.7.3 <i>Gradient Boosting Machine</i>	13
3 <i>Bayesian Additive Regression Tree</i>	15
3.1 Bentuk Formula	15
3.2 <i>Prior</i> Untuk Metode <i>Bayesian Additive Regression Tree</i>	16
3.2.1 <i>Prior</i> T_j	16
3.2.2 <i>Prior</i> $\mu_{ij} T_j$	17
3.2.3 <i>Prior</i> σ	17
3.2.4 Banyaknya Pohon Yang Digunakan	18

3.3	<i>Bayesian Backfitting Markov Chain Monte Carlo</i>	18
3.4	Inferensi <i>Posterior</i>	19
3.5	Ilustrasi Metode <i>Bayesian Additive Regression Tree</i>	20
3.6	Perhitungan Keakuratan Metode BART	27
3.7	Analisis Model	28
3.7.1	Uji Normalitas	28
3.7.2	Uji Heteroskedastisitas	29
3.7.3	Seleksi Variabel	29
4	STUDI KASUS DAN ANALISIS	33
4.1	Deskripsi Data	33
4.2	Pengolahan Data	34
4.3	Model Berbasis Pohon Keputusan Pada Perangkat Lunak <i>RStudio</i>	34
4.3.1	Pemrograman <i>Bayesian Additive Regression Tree</i>	35
4.3.2	Pemrograman <i>Bayesian Additive Regression Tree-Cross Validation</i>	35
4.3.3	Pemrograman <i>Gradient Boosting Machine</i>	35
4.3.4	Pemrograman <i>Random Forest</i>	35
4.3.5	Pemrograman <i>Decision Tree</i>	36
4.4	Hasil Studi Kasus	36
4.4.1	<i>Dataset Pertama</i>	36
4.4.2	<i>Dataset Kedua</i>	37
4.4.3	<i>Dataset Ketiga</i>	37
4.4.4	<i>Dataset Keempat</i>	37
4.5	Perbandingan Nilai RMSE dan <i>Runtime</i>	38
4.6	Model BART dan BART-CV	38
4.6.1	Model BART	39
4.6.2	Model BART-CV	42
5	KESIMPULAN DAN SARAN	47
5.1	Kesimpulan	47
5.2	Saran	47
	DAFTAR REFERENSI	49
	A UJI <i>Boxplot</i>	51

DAFTAR GAMBAR

2.1	Pohon Keputusan	6
2.2	Ilustrasi Penggunaan Validasi Silang	12
2.3	Ilustrasi Penggunaan Metode <i>Bootstrap</i>	12
3.1	Ilustrasi Struktur Pohon Saat Inisiasi	20
3.2	Ilustrasi Struktur Pohon Saat Iterasi Pertama	21
3.3	Ilustrasi Struktur Pohon Saat Iterasi Kedua	23
3.4	Ilustrasi Struktur Pohon Saat Iterasi Ketiga	24
3.5	Ilustrasi Struktur Pohon Saat Iterasi Keempat	25
3.6	Ilustrasi Struktur Pohon Saat Iterasi Kelima	26
3.7	<i>QQ Plot</i> Uji Normalitas	29
3.8	<i>Scatter Plot</i> Uji Heteroskedastisitas	30
4.1	Uji Residual Model BART	39
4.2	Struktur Pohon Model BART	40
4.3	Interval Kredibel Model BART	40
4.4	Interval Prediksi Model BART	41
4.5	<i>Variable Importance</i> Model BART	41
4.6	Seleksi Variabel Model BART	42
4.7	Uji Residual Model BART-CV	43
4.8	Struktur Pohon Model BART-CV	43
4.9	Interval Kredibel Model BART-CV	44
4.10	Interval Prediksi Model BART-CV	44
4.11	<i>Variable Importance</i> Model BART-CV	45
4.12	Seleksi Variabel Model BART-CV	45
A.1	Uji <i>Boxplot</i> Pada <i>Dataset 1</i>	51
A.2	Uji <i>Boxplot</i> Pada <i>Dataset 2</i>	52
A.3	Uji <i>Boxplot</i> Pada <i>Dataset 3</i>	53
A.4	Uji <i>Boxplot</i> Pada <i>Dataset 4</i>	54

DAFTAR TABEL

2.1	Pemilihan Distribusi <i>Prior</i>	8
3.1	Tabel Data Ilustrasi	20
3.2	Residual Parsial Untuk Iterasi Pertama	23
3.3	<i>Bayesian Backfitting</i> MCMC Pada Iterasi Pertama	23
3.4	Residual Parsial Untuk Iterasi Kedua	23
3.5	<i>Bayesian Backfitting</i> MCMC Pada Iterasi Kedua	24
3.6	Residual Parsial Untuk Iterasi Ketiga	24
3.7	<i>Bayesian Backfitting</i> MCMC Pada Iterasi Ketiga	24
3.8	Residual Parsial Untuk Iterasi Keempat	25
3.9	<i>Bayesian Backfitting</i> MCMC Pada Iterasi Keempat	25
3.10	Residual Parsial Untuk Iterasi Kelima	26
3.11	<i>Bayesian Backfitting</i> MCMC Pada Iterasi Kelima	26
3.12	Data <i>Automobile</i> (1)	27
3.13	Data <i>Automobile</i> (2)	27
3.14	Data <i>Automobile</i> (3)	27
3.15	Data <i>Automobile</i> (4)	27
3.16	Subset Data <i>Automobile</i> Untuk Menghitung Nilai MSE	28
4.1	Deskripsi Variabel	33
4.2	Perincian Data Pada Setiap <i>Dataset</i>	34
4.3	RMSE Serta <i>Runtime</i> Model Pada <i>Dataset</i> 1	36
4.4	RMSE Serta <i>Runtime</i> Model Pada <i>Dataset</i> 2	37
4.5	RMSE Serta <i>Runtime</i> Model Pada <i>Dataset</i> 3	37
4.6	RMSE Serta <i>Runtime</i> Model Pada <i>Dataset</i> 4	37
4.7	Perbandingan Nilai RMSE Pada Berbagai <i>Dataset</i>	38
4.8	Perbandingan <i>Runtime</i> Pada Berbagai <i>Dataset</i> Dalam Menit	38
4.9	Variabel Signifikan Model BART	42
4.10	Variabel Signifikan Model BART-CV	43

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada tahun 2017, tercatat terjadi kenaikan sebesar 52% kasus disabilitas dibandingkan dengan tahun 1990 yang mayoritas disebabkan oleh penyakit tidak menular seperti penyakit pernapasan, kanker, penyakit pencernaan, diabetes, dan gangguan muskuloskeletal. *Institute for Health Metrics and Evaluation* (IHME) dan *The World Health Organization* (WHO) juga menyatakan bahwa usia penderita disabilitas tertinggi terdapat pada usia kerja dari 20 tahun sampai 70 tahun. Berdasarkan jenis kelamin, perempuan lebih rentan mengalami disabilitas dibandingkan laki-laki. Hal tersebut dapat terjadi karena perempuan mengalami proses kehamilan yang merupakan salah satu penyebab disabilitas yang tidak dialami oleh laki-laki. Selain itu, disabilitas untuk kaum perempuan juga cenderung disebabkan karena tekanan darah tinggi, kadar gula darah tinggi, dan indeks massa tubuh (BMI) yang tinggi [1, hlm. 12–17].

Asuransi disabilitas berfokus untuk melindungi tertanggung yang tidak bisa bekerja karena suatu kondisi tertentu seperti cedera atau terkena penyakit sehingga tidak bisa mendapatkan pendapatan. Perlindungan asuransi disabilitas dibagi berdasarkan banyaknya tertanggung yaitu asuransi disabilitas individu yang menanggung hanya satu orang dan asuransi disabilitas kelompok yang menanggung beberapa individu pada suatu golongan. Perlindungan asuransi disabilitas dibagi berdasarkan durasi pertanggungannya yakni asuransi disabilitas jangka pendek yang memberikan perlindungan selama beberapa bulan hingga 2 tahun dan asuransi disabilitas jangka panjang yang menanggung kerugian hingga usia pensiun.

Contoh asuransi disabilitas adalah *Social Security Disability Insurance* (SSDI) yang merupakan program asuransi jangka panjang dari negara untuk warga negara Amerika. Asuransi tersebut memberikan perlindungan untuk masyarakat yang tidak dapat bekerja selama satu tahun atau lebih dikarenakan oleh kecacatan seperti penyakit menular, tidak menular, cedera fisik, penyakit mental, dan penyakit karena kehamilan. Pemilik asuransi SSDI perlu memiliki beberapa kualifikasi sebelum dirinya mendapatkan manfaat yang dibayarkan setiap bulan ketika tidak bekerja. Kualifikasi tersebut di antaranya bekerja di bidang tertentu selama beberapa tahun sebelum mengalami disabilitas, tidak dapat bekerja pada bidang pekerjaan apapun setelah mengalami disabilitas, dan jenis disabilitas yang diderita diperkirakan berlangsung selama minimal 12 bulan atau dapat berakibat kematian¹.

Contoh asuransi pemerintah di Indonesia yang memberikan santunan karena disabilitas adalah BPJS ketenagakerjaan dengan 50.696.599 peserta² atau 18,6% dari total populasi penduduk Indonesia sebanyak 272.229.372 jiwa³. Sejauh ini, program asuransi pemerintah seperti BPJS Kesehatan dan BPJS Ketenagakerjaan serta perusahaan asuransi swasta belum dapat memberikan manfaat untuk pekerja yang mengalami disabilitas karena kehamilan. Namun, BPJS Kesehatan memberikan manfaat untuk ibu hamil dengan menanggung biaya pelayanan pemeriksaan selama kehamilan dan

¹<https://www.investopedia.com/social-security-disability-insurance-ssdi-definition-5223624toc-understanding-social-security-disability-insurance-ssdi> [Diakses pada 26 April 2022]

²https://www.bpjsketenagakerjaan.go.id/assets/uploads/laporan_tahunan/BPJS_IR2020_LO16_Lowres.pdf [Diakses pada 19 Desember 2021]

³<https://dukcapil.kemendagri.go.id/berita/baca/809/distribusi-penduduk-indonesia-per-juni-2021-jabar-terbanyak-kaltara-paling-sedikit> [Diakses pada 12 Desember 2021]

pasca kehamilan, layanan ultrasonografi atau USG, biaya persalinan atau operasi *caesar*⁴.

Data yang akan digunakan untuk studi kasus adalah data disabilitas kelompok jangka panjang (GLTD) tahun 2008 yang didapatkan dari laman *Society of Actuaries* (SOA). Data tersebut terdiri dari beberapa penyebab disabilitas seperti kanker, diabetes, penyakit saraf, kehamilan, penyakit pencernaan, dan lainnya. Dari beberapa penyebab disabilitas tersebut, akan digunakan data disabilitas karena kehamilan. Disabilitas karena kehamilan dapat terjadi karena penyakit atau cedera yang disebabkan oleh kehamilan yang muncul saat proses kehamilan hingga masa setelah melahirkan dan dapat berlangsung selama beberapa bulan hingga seumur hidup. Beberapa contoh kasus yang sering terjadi adalah pendarahan berlebih yang menyebabkan anemia, infeksi pada organ reproduksi, kerusakan atau pergeseran organ dalam tubuh, tekanan darah tinggi, dan masalah mental seperti depresi⁵.

Setiap perusahaan asuransi perlu mengetahui besarnya risiko yang ditanggung oleh perusahaan apabila terjadi klaim. Pada kasus asuransi disabilitas karena kehamilan, perusahaan asuransi perlu mengetahui besarnya manfaat yang diberikan oleh asuransi dan premi yang dibayarkan oleh tertanggung ketika tertanggung mengalami disabilitas karena kehamilan dan menyebabkan tertanggung tidak dapat bekerja dan memperoleh penghasilan. Besaran biaya tersebut dapat diketahui dengan cara memprediksi tingkat pemulihan tertanggung ketika mengalami cedera atau penyakit setelah melahirkan.

Untuk memprediksi tingkat pemulihan kehamilan, akan digunakan metode *Bayesian Additive Regression Tree* (BART). Metode BART dipilih karena dianggap memiliki kelebihan dibandingkan dengan beberapa model lainnya [2, hlm. 15–25]. Kelebihan tersebut di antaranya menggunakan *prior* sehingga waktu pembentukan model menjadi lebih singkat, memberikan nilai akurasi yang lebih baik dibanding metode berbasis pohon keputusan lainnya, dan hasil prediksi yang didapatkan mendekati nilai sesungguhnya.

Pada skripsi, akan diperlihatkan perbandingan performa antara model *bayesian additive regression tree* dan *bayesian additive regression tree-cross validation* dengan model *gradient boosting machine*, *random forest*, dan *decision tree* pada beberapa *dataset*. Model *decision tree* merupakan dasar dari model berbasis pohon keputusan, model yang dibentuk berkemungkinan memiliki variansi yang besar. Untuk mengatasi variansi yang terlalu besar, dibentuk sebuah model yaitu model *random forest* [3] yang sudah dilakukan pada skripsi Lucas dan Vivian. Performa model tersebut masih belum bisa dikatakan baik karena nilai MSE yang lebih besar dari model *decision tree* pada skripsi Lucas. Alasan model *random forest* memiliki performa yang lebih buruk daripada *decision tree* adalah menggunakan semua penyebab disabilitas sehingga banyaknya data menjadi terlalu besar. Sedangkan pada skripsi Vivian, performa model *random forest* lebih baik dibanding *decision tree* karena penyebab disabilitas yang digunakan hanyalah kehamilan. Selanjutnya dibentuk juga model *gradient boosting machine* [4] yang sudah dilakukan pada skripsi Agnes. Performa dari model tersebut lebih baik dibandingkan model *decision tree*.

Untuk semua *dataset* yang digunakan, akan diaplikasikan model-model berbasis pohon keputusan dan akan dipilih satu buah model dengan performa terbaik. Model terbaik dipilih dengan cara membandingkan nilai MSE dan lama waktu *runtime* yang digunakan oleh setiap model. Kemudian akan dibahas analisis untuk model *bayesian additive regression tree* dan *bayesian additive regression tree-cross validation*.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana cara memprediksi tingkat pemulihan disabilitas kelompok jangka panjang?
2. Bagaimana cara mendapatkan metode berbasis pohon keputusan terbaik?

⁴<https://hellosehat.com/sehat/asuransi/panduan-bpjs-untuk-ibu-hamil/> [Diakses pada 26 April 2022]

⁵<https://www.prb.org/resources/hidden-suffering-disabilities-from-pregnancy-and-childbirth-in-less-developed-countries/>

3. Bagaimana cara mencari variabel yang memengaruhi tingkat pemulihan kehamilan?

1.3 Tujuan

Tujuan yang akan dicapai adalah sebagai berikut:

1. Mengetahui cara memprediksi tingkat pemulihan menggunakan metode *Bayesian Additive Regression Tree* (BART).
2. Mengetahui metode berbasis pohon keputusan terbaik dengan membandingkan *root mean squared error* dan *runtime* dari metode *Bayesian Additive Regression Tree*, *Bayesian Additive Regression Tree-Cross Validation*, *Gradient Boosting Machine*, *Random Forest*, dan *Decision Tree*.
3. Mengetahui variabel yang memengaruhi tingkat pemulihan kehamilan dari nilai proporsi seleksi variabel.

1.4 Batasan Masalah

Batasan masalah yang akan digunakan adalah sebagai berikut:

1. Data yang digunakan adalah data asuransi disabilitas kelompok jangka panjang.
2. Data disabilitas yang digunakan adalah data disabilitas karena kehamilan.
3. Data kehamilan yang digunakan hanya untuk jenis kelamin perempuan.
4. Data yang mengandung nilai *unknown* tidak akan digunakan.
5. Diasumsikan pekerja yang cacat akan sembuh kembali, tidak meninggal.

1.5 Sistematika Pembahasan

Bab 1 : Pendahuluan

Bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, serta sistematika pembahasan.

Bab 2 : Landasan Teori

Bab ini membahas tentang teori pendukung seperti definisi asuransi disabilitas, pohon keputusan, teorema Bayes, *Markov Chain Monte Carlo* (MCMC), *mean squared error* dan *root mean squared error*, metode pengambilan sampel ulang, dan metode berbasis pohon keputusan.

Bab 3 : *Bayesian Additive Regression Tree* (BART)

Bab ini membahas bentuk formula *Bayesian Additive Regression Tree*, *prior* untuk metode *Bayesian Additive Regression Tree*, *Bayesian Backfitting Markov Chain Monte Carlo*, inferensi posterior, ilustrasi metode *Bayesian Additive Regression Tree*, perhitungan keakuratan, dan analisis model.

Bab 4 : Studi Kasus dan Analisis

Bab ini membahas tentang deskripsi data, pengolahan data, model berbasis pohon keputusan pada perangkat lunak *RStudio*, hasil studi kasus, perbandingan nilai RMSE dan *runtime*, dan analisis model BART dan BART-CV.

Bab 5 : Kesimpulan dan Saran

Bab ini berisi kesimpulan dari bab-bab sebelumnya dan juga saran untuk pengembangan penelitian selanjutnya.

