



TUGAS AKHIR - KA 184801

**KLASIFIKASI JENIS KECELAKAAN KERJA YANG
DITANGANI OLEH BPJS KETENAGAKERJAAN
MENGUNAKAN PENDEKATAN *MULTINOMIAL
LOGISTIC REGRESSION* DAN *RANDOM FOREST***

LAILIYA RISKY IMANIA
NRP 0631184000009

Dosen Pembimbing

Pratnya Paramitha Oktaviana, S.Si, M.Si.

NIP 1990201812015

**PROGRAM STUDI SARJANA SAINS AKTUARIA
DEPARTEMEN AKUARIA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2022**



TUGAS AKHIR - KA 184801

**KLASIFIKASI JENIS KECELAKAAN KERJA YANG
DITANGANI OLEH BPJS KETENAGAKERJAAN
MENGUNAKAN PENDEKATAN *MULTINOMIAL
LOGISTIC REGRESSION* DAN *RANDOM FOREST***

**LAILIYA RISKY IMANIA
NRP 0631184000009**

Dosen Pembimbing

Pratnya Paramitha Oktaviana, S.Si, M.Si.

NIP 1990201812015

**PROGRAM STUDI SARJANA SAINS AKTUARIA
DEPARTEMEN AKTUARIA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2022**



FINAL PROJECT - KA 184801

**THE CLASSIFICATION OF WORK ACCIDENTS
HANDLED BY BPJS KETENAGAKERJAAN USING
MULTINOMIAL LOGISTIC REGRESSION AND RANDOM
FOREST**

LAILIYA RISKY IMANIA
NRP 0631184000009

Advisor

Pratnya Paramitha Oktaviana, S.Si, M.Si.

NIP 1990201812015

**UNDERGRADUATE PROGRAMME ACTUARIAL SCIENCE
DEPARTEMENT OF ACTUARIAL SCIENCE
FACULTY OF SCIENCE AND DATA ANALYTICS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2022**

LEMBAR PENGESAHAN




KLASIFIKASI JENIS KECELAKAAN KERJA YANG DITANGANI OLEH BPJS KETENAGAKERJAAN MENGGUNAKAN PENDEKATAN *MULTINOMIAL LOGISTIC REGRESSION DAN RANDOM FOREST*

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Ilmu Aktuaria pada
Program Studi S-1 Sains Aktuaria
Departemen Aktuaria
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh : **LAILIYA RISKY IMANIA**
NRP. 063118 4000 0009

Disetujui oleh Tim Penguji Tugas Akhir :

- | | | |
|---|------------|---|
| 1. Pratnya Paramitha Oktaviana, S.Si, M.Si. | Pembimbing | () |
| 2. Imam Safawi Ahmad, S.Si, M.Si | Penguji | () |
| 3. Ulil Azmi, S.Si, M.Si. | Penguji | () |

SURABAYA

Juli, 2022

APPROVAL SHEET




THE CLASSIFICATION OF WORK ACCIDENTS HANDLED BY BPJS KETENAGAKERJAAN USING MULTINOMIAL LOGISTIC REGRESSION AND RADOM FOREST

FINAL PROJECT

Submitted to fulfill one of the requirements
for obtaining a degree Bachelor of Actuarial Science at
Undergraduate Study Program of Actuarial Science
Department of Actuarial Science
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember

By: **LAILIYA RISKY IMANIA**
NRP. 063118 4000 0009

Approved by Final Project Examiner Team:

- | | | |
|---|----------|---|
| 1. Pratnya Paramitha Oktaviana, S.Si, M.Si. | Advisor | () |
| 2. Imam Safawi Ahmad, S.Si, M.Si. | Examiner | () |
| 3. Ulil Azmi, S.Si, M.Si. | Examiner | () |

SURABAYA

July, 2022

PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NRP : Lailiya Risky Imania / 0631184000009

Program studi : S1-Sains Aktuaria

Dosen Pembimbing / NIP : Pratnya Paramitha Oktavianna, S.Si, M.Si.
/1990201812015

dengan ini menyatakan bahwa Tugas Akhir dengan judul “Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Menggunakan Pendekatan *Multinomial Logistic Regression* dan *Random Forest*” adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

Surabaya, Juli 2022

Mengetahui

Dosen Pembimbing

Mahasiswa



(Pratnya Paramitha Oktaviana, S.Si, M.Si.)

(Lailiya Risky Imania)

NIP. 1990201812015

NRP. 0631184000009

STATEMENT OF ORIGINALITY

The undersigned below:

Name of student / NRP : Lailiya Risky Imania / 0631184000009
Department : Actuarial
Advisor / NIP :Pratnya Paramitha Oktavianna, S.Si,
M.Si./1990201812015

Hereby declare that the Final Project with the title of “The Classification of Work Accidents Handled by BPJS Ketenagakerjaan using Multinomial Logistic Regression and Random Forest” is the result of my own work, is original, and is written by following the rules of scientific writing.

If in the future there is a discrepancy with statement then I am willing to accept sanctions in accordance with the provisions that apply at Institut Teknologi Sepuluh Nopember.

Surabaya, July 2022

Acknowledge

Advisor,



(Pratnya Paramitha Oktavianna, S.Si, M.Si.)

NIP. 1990201812015

Student,



(Lailiya Risky Imania)

NRP. 0631184000009

**KLASIFIKASI JENIS KECELAKAAN KERJA YANG DITANGANI OLEH BPJS
KETENAGAKERJAAN MENGGUNAKAN PENDEKATAN *MULTINOMIAL
LOGISTIC REGRESSION* DAN *RANDOM FOREST***

Nama Mahasiswa / NRP : Lailiya Risky Imania / 0631184000009
Departemen : Aktuaria
Dosen Pembimbing : Pratnya Paramitha Oktaviana, S.Si, M.Si.

Abstrak

Kecelakaan kerja merupakan masalah yang serius dimana suatu Negara harus menjamin adanya perlindungan tenaga kerja untuk memberikan perlindungan keselamatan bagi pekerja/buruh pada saat bekerja. Badan Penyelenggara Jaminan Sosial merupakan badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial sebagai salah satu bentuk perlindungan sosial untuk menjamin seluruh rakyat agar dapat memenuhi kebutuhan dasar hidupnya yang layak sehingga mendukung program jaminan sosial khususnya dalam pembangunan ketenagakerjaan. Banyaknya kasus kecelakaan kerja di Indonesia menjadi salah satu alasan untuk melakukan penelitian tentang klasifikasi kasus kecelakaan kerja berdasarkan jenis kecelakaan yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto. Variabel respon adalah jenis kecelakaan kerja yang merupakan variabel kategorik dengan kategori sebanyak 3, sedangkan variabel prediktor juga merupakan variabel kategorik. Sehingga analisis yang digunakan metode regresi logistik multinomial, selain itu dilakukan pula analisis menggunakan *random forest* yang nantinya digunakan untuk perbandingan. Hasil analisis didapatkan variabel jenis kelamin, jabatan, dan umur berpengaruh signifikan terhadap kecelakaan kerja. Ketepatan klasifikasi jenis kecelakaan kerja dengan regresi logistik multinomial data *training* sebesar 74,03% dan data *testing* sebesar 70,31%, sedangkan untuk *random forest* data *training* sebesar 87,63% dan data *testing* sebesar 78,72% sehingga dapat diambil kesimpulan bahwa metode yang terbaik dalam kasus ini adalah *random forest*.

Kata kunci: Kecelakaan kerja, Badan Penyelenggara Jaminan Sosial (BPJS), klasifikasi, regresi logistik multinomial (*multinomial logistic regression*), *random forest*.

**THE CLASSIFICATION OF WORK ACCIDENTS HANDLED BY BPJS
KETENAGAKERJAAN USING MULTINOMIAL LOGISTIC REGRESSION AND
RANDOM FOREST**

Student Name / NRP : Lailiya Risky Imania
Department : Aktuarial
Advisor : Pratnya Paramitha Oktaviana, S.Si, M.Si.

Abstract

Work accidents are a serious problem where a country must ensure the protection of workers to provide safety protection for workers/laborers at work. The Social Security Administering Body is a legal entity established to administer social security programs as a form of social protection to ensure that all people can meet their basic needs for a decent life so as to support social security programs, especially in employment development. The number of work accident cases in Indonesia is one reason to conduct research on the classification of work accident cases based on the type of accident handled by BPJS Ketenagakerjaan in the Mojokerto region. The response variable is the type of work accident which is a categorical variable with 3 categories, while the predictor variable is also a categorical variable. So that the analysis used the multinomial logistic regression method, besides that an analysis was also carried out using a random forest which would later be used for comparison. The results of the analysis showed that gender, position, and age had a significant effect on work accidents. The accuracy of the classification of work accident types with multinomial logistic regression of training data is 74.03% and testing data is 70.31%, while for random forest training data is 87.63% and testing data is 78.72%, so it can be concluded that the method the best in this case is random forest.

Keywords: Work accidents, Badan Penyelenggara Jaminan Sosial (BPJS), classification, multinomial logistic regression, random forest.

KATA PENGANTAR

Puji Syukur penulis panjatkan atas rahmat dan hidayah yang diberikan Allah SWT sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul “Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Menggunakan Pendekatan *Multinomial Logistic Regression* dan *Random Forest*”.

Tugas Akhir ini dapat terselesaikan tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, atas dukungan, saran, motivasi, semangat, serta bantuan yang telah diberikan kepada penulis, penulis menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberikan petunjuk, kekuatan, kesabaran serta keteguhan hati kepada penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan sangat baik tanpa melalaikan perintah-Nya.
2. Bapak Drs.Soehardjoepri, M.Si. selaku Kepala Departemen Sains Aktuaria sekaligus Dosen Wali yang selalu memberikan motivasi kepada anak didiknya.
3. Ibu Pratnya Paramitha Oktaviana, S.Si, M.Si. selaku dosen pembimbing tugas akhir yang telah sabar membimbing dan memberikan arahan dalam penyusunan laporan tugas akhir ini.
4. Bapak Imam Safawi Ahmad, S.Si, M.Si. dan Ibu Ulil Azmi, S.Si, M.Si. selaku dosen penguji yang telah memberikan kritik,saran, serta masukan dalam pengerjaan tugas akhir ini.
5. Seluruh dosen dan tendik Aktuaria yang mengajarkan serta membimbing anak didiknya.
6. Kedua orang tua dan sanak saudara yang selalu memberikan dukungan, baik moral maupun materi yang tidak pernah putus kepada penulis sehingga dapat berada di titik ini.
7. Devano Rizaldy selaku tunangan saya yang sudah menemani dan memberi semangat dalam pengerjaan tugas akhir ini.
8. Teman seperjuangan Hanifah Rizqy M dan I Gusti Ayu Mas Darsasasmitha Yani dalam menyelesaikan tugas akhir ini yang selalu memotivasi satu sama lain.
9. Segenap jajaran kepegawaian di BPJS Ketenagakerjaan Wilayah Mojokerto yang membantu penulis dalam proses pengumpulan data.

Penulis menyadari berbagai keterbatasan yang dimiliki sehingga laporan Tugas Akhir ini masih jauh dari kata sempurna. Untuk itu, penulis terbuka menerima kritik dan masukan yang bersifat membangun untuk perbaikan penyusunan laporan Tugas Akhir agar menjadi lebih baik. Semoga dukungan, bimbingan, dan kebaikan yang telah diberikan kepada penulis mendapatkan rida Allah SWT.

Surabaya, Juli 2022
Hormat kami,

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

LEMBAR PENGESAHAN	iii
APPROVAL SHEET	iv
PERNYATAAN ORISINALITAS	v
STATEMENT OF ORIGINALITY	vi
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
BAB 2 TINJAUAN PUSTAKA	5
2.1 Hasil Penelitian Terdahulu.....	5
2.2 Statistika Deskriptif.....	5
2.3 K-Fold Cross Validation.....	5
2.4 Regresi Logistik.....	6
2.5 Regresi Logistik Multinomial.....	6
2.6 Estimasi Parameter.....	7
2.7 Uji Signifikansi Parameter.....	8
2.7.1 Pengujian Serentak.....	8
2.7.2 Uji Individu.....	8
2.8 Interpretasi Model.....	9
2.9 Uji Kesesuaian Model.....	9
2.10 Random Forest.....	9
2.10.1 CART (Classification And Regression Trees).....	10
2.10.2 Algoritma Random Forest.....	10
2.10.3 Pohon Klasifikasi.....	11

2.10.4	Ilustrasi Random Forest	12
2.10.5	Ketepatan Klasifikasi	13
2.11	Penanganan Kecelakaan Kerja	13
BAB 3	METODOLOGI	15
3.1	Sumber Data.....	15
3.2	Variabel Penelitian	15
3.3	Metode Analisis	16
BAB 4	HASIL DAN PEMBAHASAN	19
4.1	Karakteristik Kecelakaan yang Ditangani oleh BPJS Ketenagakerjaan	19
4.2	Analisis Regresi Logistik Multinomial untuk Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Wilayah Mojokerto	21
4.2.1	Pemilihan Presentase Kombinasi Data <i>Training</i> dan Data <i>Testing</i> Terbaik dalam Analisis Regresi Logistik Multinomial.....	21
4.2.2	Pemilihan Subset Terbaik dari Kombinasi Data <i>Training</i> dan Data <i>Testing</i> Terpilih Menggunakan <i>K-Fold Cross Validation</i>	22
4.2.3	Uji Serentak Variabel-variabel yang Berpengaruh terhadap Jenis Kecelakaan Kerja	22
4.2.4	Uji Parsial Variabel-variabel yang Berpengaruh terhadap Jenis Kecelakaan Kerja.....	23
4.2.5	Model Regresi Logistik Multinomial.....	24
4.2.6	Uji Kesesuaian Model	25
4.2.7	Klasifikasi Jenis Kecelakaan Kerja	26
4.3	Analisis <i>Random Forest</i> untuk Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Wilayah Mojokerto	26
4.3.1	Analisis CART untuk Klasifikasi Jenis Kecelakaan Kerja	26
a.	Pembentukan Pohon Klasifikasi Maksimal.....	26
b.	Pemangkasan Pohon Klasifikasi Maksimal (<i>Prunning</i>).....	29
c.	Pemilihan Pohon Klasifikasi Optimal	29
4.3.2	Analisis <i>Random Forest</i> untuk Klasifikasi Jenis Kecelakaan Kerja.....	30
a.	Pemilihan Presentase Kombinasi Data <i>Training</i> dan Data <i>Testing</i> Terbaik dalam Analisis <i>Random Forest</i>	30

BAB 5 KESIMPULAN DAN SARAN	32
5.1 Kesimpulan	33
5.2 Saran.....	33
DAFTAR PUSTAKA	35
LAMPIRAN.....	37
BIODATA PENULIS	55

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Pembagian Data <i>Training-Testing</i> (Sumber : Astri Silvia, 2019)	6
Gambar 2.2 Algoritma <i>Random Forest</i> (Sumber : Abid, dkk, 2020).....	11
Gambar 2.3 Ilustrasi Pohon Klasifikasi (Sumber : Abid, dkk, 2020).....	11
Gambar 3.1 Diagram Alir Penelitian.....	17
Gambar 4.1 Jenis Kecelakaan Kerja.....	19
Gambar 4.2 Histogram Usia Korban Kecelakaan Kerja.....	20
Gambar 4.3 Pohon Klasifikasi Maksimal untuk Jenis Kecelakaan Kerja	28
Gambar 4.4 Konstruksi Pohon Klasifikasi Optimal Jenis Kecelakaan Kerja.....	29

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1 Data Sampel untuk Ilustrasi <i>Random Forest</i>	12
Tabel 2.2 Hasil Prediksi Kelas dengan <i>Random Forest</i>	12
Tabel 2.3 Klasifikasi Respon Multinomial Tiga Kategori.....	13
Tabel 3.1 Variabel Penelitian.....	15
Tabel 3.2 Struktur Data.....	16
Tabel 4.1 <i>Crosstabulation</i> Jenis Kelamin dan Jenis Kecelakaan Kerja.....	19
Tabel 4.2 <i>Crosstabulation</i> Jabatan dan Jenis Kecelakaan Kerja	19
Tabel 4.3 <i>Crosstabulation</i> Jam Kerja dan Jenis Kecelakaan Kerja	20
Tabel 4.4 <i>Crosstabulation</i> Waktu Kejadian Kecelakaan dan Jenis Kecelakaan Kerja.	20
Tabel 4.5 Perbandingan <i>Total Accuracy Rate</i> Kombinasi Data	21
Tabel 4.6 Pemilihan <i>Subset</i> Terbaik Regresi Logistik Multinomial Berdasarkan <i>Accuracy</i>	22
Tabel 4.7 Uji Serentak Regresi Logistik Multinomial.....	22
Tabel 4.8 Uji Parsial Regresi Logistik Multinomial.....	23
Tabel 4.9 Uji Serentak Regresi Logistik Multinomial Model Terbaik.....	23
Tabel 4.10 Uji Parsial Regresi Logistik Multinomial Model Terbaik	24
Tabel 4.11 Nilai <i>Odds Ratio</i> Variabel.....	25
Tabel 4.12 <i>Goodness of Fit</i> Jenis Kecelakaan Kerja	25
Tabel 4.13 Klasifikasi Jenis Kecelakaan Kerja.....	26
Tabel 4.14 Perhitungan Kemungkinan Jumlah Pemilah dari Setiap Variabel.....	27
Tabel 4.15 Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>).....	29
Tabel 4.16 Skor Variabel Terpenting dari Pohon Klasifikasi Optimal.....	29
Tabel 4.17 Perbandingan <i>Total Accuracy Rate</i> Kombinasi Data	30
Tabel 4.18 Ketepatan Klasifikasi <i>Random Forest</i> pada Kombinasi 500.....	30
Tabel 4.19 Perbandingan Hasil Klasifikasi Regresi Logistik Multinomial dan <i>Random Forest</i>	31

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1 <i>Library</i> yang digunakan	37
Lampiran 2 <i>Syntax</i> untuk <i>Preprocessing</i> Data, Statistika Deskriptif, <i>Crosstabulation</i> , dan Partisi Data	38
Lampiran 3 <i>Syntax</i> untuk Metode Regresi Logistik Multinomial	41
Lampiran 4 <i>Syntax</i> untuk Metode <i>Random Forest</i>	43
Lampiran 5 Data Penelitian	46
Lampiran 6 <i>Output</i> Statistika Deskriptif.....	49
Lampiran 7 <i>Output</i> Metode Regresi Logistik Multinomial.....	51
Lampiran 8 <i>Output</i> Metode <i>Random Forest</i>	53

(Halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kecelakaan kerja dapat diartikan sebagai suatu kejadian yang tak terduga dan tidak disengaja dalam suatu aktivitas yang dapat menimbulkan kerugian baik korban manusia dan/atau harta benda. Akibat yang ditimbulkan dari kecelakaan kerja antara lain seperti cedera atau kematian pada pekerja, harta benda atau properti, kerusakan lingkungan dan terganggunya proses produksi pada suatu perusahaan. Berdasarkan data dari International Labour Organization (ILO) pada tahun 2013 mencatat angka kematian yang dikarenakan kecelakaan dan penyakit akibat kerja (PAK sebanyak 2 juta kasus setiap tahun, dan menyatakan 1 (satu) orang pekerja di dunia meninggal setiap 15 detik karena kecelakaan kerja dan 160 pekerja mengalami sakit akibat kerja.

Banyak resiko yang harus dihadapi oleh tenaga kerja dalam melakukan pekerjaannya. Oleh sebab itu, pentingnya suatu perlindungan bagi tenaga kerja yang mengatur keselamatan kerja dan tata cara penggantian ganti rugi dari kecelakaan kerja. Kecelakaan kerja merupakan kecelakaan yang terjadi saat jam kerja berlangsung, dimana kecelakaan ini terjadi di lokasi kerja maupun perjalanan dari rumah menuju tempat kerja atau sebaliknya. Untuk meminimalisir biaya yang dikeluarkan oleh tenaga kerja yang diakibatkan oleh adanya resiko-resiko tenaga kerja seperti sakit atau cacat karena kecelakaan kerja baik fisik maupun mental, maka diperlukan adanya jaminan kecelakaan kerja.

Berdasarkan Pasal 15 Ayat (1) Undang-Undang Nomor 24 Tahun 2011 tentang terbentuknya Badan Penyelenggara Jaminan Sosial dimana wajib bagi pemberi kerja untuk mendaftarkan dirinya dan tenaga kerjanya sebagai peserta pada badan penyelenggara jaminan sosial sesuai dengan program jaminan sosial yang diikuti. Badan Penyelenggara Jaminan Sosial dapat diartikan sebagai badan hukum yang dibentuk secara khusus dengan tujuan menyelenggarakan program jaminan sosial sebagai salah satu bentuk perlindungan sosial. Program ini menjamin seluruh rakyat untuk dapat memenuhi kebutuhan dasar hidupnya yang layak dalam pembangunan ketenagakerjaan.

Pembentukan BPJS Ketenagakerjaan yang merupakan bagian dari Badan Penyelenggara Jaminan Sosial mempunyai peran penting dalam kesejahteraan pekerja Indonesia. BPJS Ketenagakerjaan merupakan suatu perlindungan yang memiliki program-program perlindungan dasar, pemeliharaan dan peningkatan kesejahteraan bagi tenaga kerja dalam bentuk santunan berupa uang sebagai pengganti sebagian penghasilan yang hilang atau berkurang dan pelayanan sebagai perlindungan diri dari ketidakpastian seperti resiko sosial dan ekonomi yang bisa terjadi, misalnya resiko kecelakaan kerja, sakit, kematian, masa pensiun, dan lain sebagainya. Sehingga diharapkan program BPJS Ketenagakerjaan ini dapat membantu tenaga kerja dalam menanggung beban akan resiko yang terjadi.

Berdasarkan data Badan Penyelenggara Jaminan Sosial (BPJS) Ketenagakerjaan, sepanjang tahun 2018 di Indonesia jumlah peserta yang mengalami kecelakaan kerja sebanyak 147.000 kasus atau sebanyak 40.273 kasus setiap hari. Dari jumlah itu, terdapat 4.678 kasus (3,18%) yang berakibat kecacatan dan 2.575 kasus (1,75%) yang berakhir dengan kematian. Tingginya angka kasus klaim kecelakaan kerja pada BPJS Ketenagakerjaan menjadi dasar pemikiran tentang pentingnya dilakukan penelitian mengenai faktor-faktor yang memengaruhi terjadinya kecelakaan tersebut. Terdapat banyak penerapan metode statistika dalam melakukan klasifikasi suatu variabel respon terhadap faktor-faktor yang mempengaruhinya seperti *Classification and Regression Trees*

(CART), *Naïve Bayes*, regresi logistik dan *Random Forest*.

Metode *Classification and Regression Trees* (CART) merupakan metode yang digunakan untuk mengasumsikan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Kelebihan dari *Classification and Regression Trees* (CART) ialah mudah menginterpretasikan model klasifikasi, hasil yang akurat, dan perhitungan yang lebih cepat dibandingkan metode klasifikasi yang lain. Selain itu, metode ini juga dapat diterapkan dalam data yang mempunyai skala cukup besar dengan variabel yang banyak. Akan tetapi, metode *Classification and Regression Trees* (CART) juga memiliki beberapa kelemahan antara lain yaitu hasil yang kurang stabil karena metode ini sangat sensitive terhadap data baru, sangat berganung pada jumlah sampel yang diujikan. Apabila sampel data *training* dan *testing* berubah maka pohon keputusan yang dihasilkan juga ikut berubah (Pratiwi & Zain, 2014).

Naïve Bayes merupakan metode pengklasifikasi statistik yang dirancang untuk mengasumsikan bahwa antar satu kelas dengan kelas lain tidak saling tergantung (independen) dan dapat memprediksi probabilitas keanggotaan kelas suatu data tuple yang akan masuk ke dalam kelas tertentu. Kelebihan dari *Naïve Bayes* ialah menghasilkan nilai akurasi dan kecepatan yang baik apabila diterapkan pada database yang besar. Akan tetapi, metode *Naïve Bayes* memiliki kelemahan tidak berlaku jika probabilitas kondisionalnya bernilai nol dan hanya mengasumsikan variabel bebas saja.

Menurut Agresti (2007) menyatakan bahwa regresi logistik merupakan salah satu metode yang dapat digunakan untuk mencari hubungan variabel respon yang bersifat *dichotomous* (berskala nominal atau ordinal dengan dua kategori) atau *polychotomous* (mempunyai skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor. Untuk variabel respon pada regresi logistik bersifat kontinu atau kategorik.

Random Forest merupakan salah satu metode dalam *Decision Tree* yang digunakan pada klasifikasi data yang cukup besar. Klasifikasi *randm forest* dilakukan melalui penggabungan pohon (*tree*) dengan melakukan *training* pada sampel data yang dimiliki. Kelebihan dari metode ini mampu mengklasifikasikan data yang memiliki atribut yang tidak lengkap, menangani data sampel yang banyak. Akan tetapi kelemahannya ialah tidak terlalu bagus apabila digunakan untuk regresi.

Berdasarkan uraian yang telah dijelaskan sebelumnya, pada tugas akhir ini akan dilakukan penelitian mengenai klasifikasi jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan. Variabel respon yang digunakan yaitu jenis kecelakaan kerja memiliki tiga kategori. Dikarenakan kategori variabel respon lebih dari dua, maka digunakan regresi logistik multinomial (*multinomial logistic regression*). Variabel prediktor yang digunakan juga kategorik, antara lain jenis kelamin, jabatan, umur, jam kerja, jenis pengajuan. Dengan menggunakan lebih dari dua kategori, maka metode yang dapat digunakan adalah regresi logistik multinomial pada kasus pengajuan klaim kecelakaan kerja di BPJS Ketenagakerjaan. Selain itu juga dilakukan pula analisis data menggunakan *random forest* yang nantinya digunakan untuk perbandingan dengan regresi logistik multinomial.

1.2 Rumusan Masalah

Seperti yang telah dijelaskan sebelumnya pada latar belakang, kecelakaan kerja merupakan permasalahan yang serius. Oleh karena itu, perlu mencari tahu karakteristik data dari laporan klaim kasus kecelakaan kerja serta dilakukan klasifikasi faktor penyebab terjadinya kecelakaan kerja pada BPJS Ketenagakerjaanwilayah Mojokerto dengan metode regresi logistik multinomial dan *random forest*. Selain itu juga dilakukan perbandingan hasil ketepatan pemodelan regresi logistik multinomial dan *random forest* dalam klasifikasi

faktor yang mempengaruhi terjadinya kecelakaan kerja pada BPJS Ketenagakerjaan wilayah Mojokerto.

1.3 Batasan Masalah

Data yang digunakan dalam penelitian ini adalah data harian klaim kecelakaan kerja pada tahun Bulan September 2020 hingga Bulan Mei 2021 mengenai kasus pengajuan klaim kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto.

1.4 Tujuan

Tujuan yang akan dicapai dalam penelitian ini adalah sebagai berikut.

1. Untuk mengetahui karakteristik data harian klaim kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto.
2. Untuk mengetahui ketepatan klasifikasi faktor penyebab terjadinya kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto dengan metode regresi logistik multinomial dan *random forest*.
3. Untuk membandingkan hasil ketepatan pemodelan regresi logistik multinomial dan *random forest* dalam klasifikasi faktor yang mempengaruhi terjadinya kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto.

1.5 Manfaat

Manfaat yang diharapkan pada penelitian ini adalah sebagai berikut.

1. Menambah wawasan ilmu pengetahuan mengenai metode analisis data dengan Regresi Logistik Multinomial dan *Random Forest*.
2. Memberikan informasi mengenai faktor yang mempengaruhi terjadinya kecelakaan kerja khususnya untuk para pekerja yang memiliki asuransi Jaminan Kecelakaan Kerja (JKK) pada BPJS Ketenagakerjaan.

(Halaman ini sengaja dikosongkan)

BAB 2

TINJAUAN PUSTAKA

2.1 Hasil Penelitian Terdahulu

Penelitian mengenai kecelakaan kerja pernah dilakukan oleh Windy Pranita Sari, dkk (2015) yang membahas tentang faktor-faktor yang mempengaruhi potensi kecelakaan kerja pada pengemudi truk di PT. Berkas Nugraha Sinar Lestari menggunakan pendekatan Cross Sectional. Analisis data menggunakan uji *exact fisher* dan uji regresi logistik biner yang menunjukkan hasil penelitian dari uji *exact fisher* bahwa kondisi kendaraan, kondisi jalan, dan kondisi fisik tubuh memiliki hubungan akan potensi terjadinya kecelakaan kerja, selain itu hasil dari uji regresi logistik biner dengan metode *backward stepwise* menunjukkan bahwa variabel yang paling signifikan yang berpengaruh terhadap potensi kecelakaan kerja adalah kondisi fisik tubuh. Penelitian yang dilakukan oleh Yelvina Tanriono, dkk (2018) yang membahas hubungan kelelahan kerja, kualitas tidur, perilaku mengemudi dan status gizi dengan kecelakaan kerja pada pengemudi ojek di kota Bitung. Penelitian ini menggunakan kuesioner *Industrial Fatigue Research Committee (IFRC)*, kuesioner *Pittsburg Sleep Quality Index (PSQI)*, *Modified Motorcycle Rider Behavior Questionnaire (MRBQ)*, dan metode penghitungan Indeks Massa Tubuh (IMT). Analisis bivariat menggunakan uji *Chi Square* yang selanjutnya untuk menilai variabel yang paling dominan menggunakan uji regresi logistik berganda. Hasil dari penelitian ini ialah Faktor yang penting berpengaruh pada kecelakaan kerja pengemudi ojek di kota Bitung adalah kelelahan kerja dan perilaku pengemudi, dan tidak ditemukan hubungan bermakna antara status gizi dengan kecelakaan kerja. Dan yang terakhir, penelitian yang dilakukan oleh Ita Rakhmawati (2015) tentang klasifikasi faktor-faktor yang mempengaruhi korban kecelakaan lalu lintas di Surabaya menggunakan pendekatan regresi logistik multinomial dan *random forests*. Hasil dari penelitian ini didapatkan bahwa variabel jenis kecelakaan, peran korban, kendaraan lawan, waktu dan usia berpengaruh signifikan terhadap korban kecelakaan lalu lintas. Ketepatan klasifikasi korban kecelakaan dengan regresi logistik multinomial sebesar 63,33% sedangkan *random forests* sebesar 58,33% sehingga metode yang baik digunakan pada kasus korban kecelakaan lalu lintas adalah regresi logistik multinomial.

2.2 Statistika Deskriptif

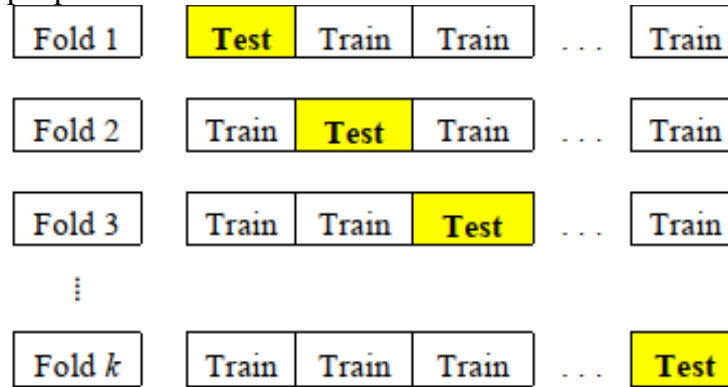
Menurut Sugiyono (2014), statistika deskriptif adalah statistik yang digunakan untuk menggambarkan atau menganalisis suatu statistik hasil penelitian, tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas. Menurut Walpole (2007), statistika deskriptif merupakan kumpulan jumlah tunggal statistik yang memberikan pengertian mengenai lokasi pusat data, keragaman data, dan sifat umum dari distribusi setiap pegamaan pada sampel.

2.3 K-Fold Cross Validation

Cross validation merupakan salah satu Teknik untuk menilai atau memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Model yang dibuat nantinya bertujuan untuk melakukan prediksi atau klasifikasi terhadap suatu data baru yang belum ada di dalam dataset. Data yang digunakan dalam proses pembangunan model disebut data latih (*training*), sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data tes (*testing*). *K-Fold Cross Validation* adalah salah satu metode dari *cross validation* yang digunakan untuk mempartisi data menjadi data *training* dan *testing*, dimana setiap data mendapat kesempatan menjadi data *testing*

(Refaeilzadeh, Tang, & Liu, 2009).

Dalam *K-Fold Cross Validation* data dibagi menjadi k buah segmen yang memiliki rasio yang sama atau hamper sama lalu dilakukan *training* dan validasi sebanyak k kali dengan tiap perulangannya mengambil satu segmen berbeda sebagai data *testing* atau validasi dan $k-1$ segmen sisanya sebagai data *training* kemudian diambil nilai rata-rata dari setiap hasil iterasi. Prosedur ini diulangi k kali sehingga setiap partisi digunakan untuk *testing* tepat satu kali. Nilai k yang sering digunakan adalah 10 karena dapat menghasilkan estimasi error yang paling baik dan membagi data menjadi proporsi seimbang (Witten,dkk, 2011). Ilustrasi pembagian data *training-testing* menggunakan *K-Fold Cross Validation* terdapat pada Gambar 2.1.



Gambar 2.1 Ilustrasi Pembagian Data *Training-Testing* (Sumber : Astri Silvia, 2019)

2.4 Regresi Logistik

Regresi merupakan metode yang digunakan untuk melengkapi analisis data yang menjelaskan hubungan antara variabel respon dengan satu atau lebih variabel prediktor atau penjelas. Sedangkan regresi logistik merupakan metode yang digunakan untuk mencari hubungan antara variabel respon yang bersifat dichotomus atau polychotomus dengan satu atau lebih variabel prediktor (Hosmer dkk, 2013).

2.5 Regresi Logistik Multinomial

Regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel dependen mempunyai skala yang bersifat polichotomus atau multinomial. Skala multinomial adalah suatu pengukuran yang dikategorikan menjadi lebih dari dua kategori.

Mengacu pada regresi logistik *trichotomous* untuk model regresi dengan variabel dependen berskala nominal tiga kategori digunakan kategori variabel hasil Y dikoding 0,1 dan 2. Variabel Y terparameterisasi menjadi dua fungsi logit. Sebelumnya perlu ditentukan kategori hasil mana yang digunakan untuk membandingkan. Pada umumnya digunakan $Y=0$ sebagai pembanding. Untuk membentuk fungsi logit, akan dibandingkan $Y=1$ dan $Y=2$ terhadap $Y=0$. Kedua fungsi logit dinotasikan sebagai berikut.

$$\begin{aligned}
 g_1 &= \ln \left[\frac{P(Y = 1|x)}{P(Y = 0|x)} \right] & (2.1) \\
 &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p \\
 &= x' \beta_1 \\
 g_2 &= \ln \left[\frac{P(Y = 2|x)}{P(Y = 0|x)} \right]
 \end{aligned}$$

$$\begin{aligned}
&= \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p \\
&= x'\beta_2
\end{aligned} \tag{2.2}$$

Dimana g_1 adalah fungsi logit untuk $Y=1$ dan g_2 adalah fungsi logit untuk $Y=2$. $\beta_{10}, \beta_{11}, \beta_{12}, \dots, \beta_{1p}$ adalah parameter (koefisien) yang akan ditaksir pada $g_1(x)$, sedangkan $\beta_{20}, \beta_{21}, \beta_{22}, \dots, \beta_{2p}$ adalah parameter (koefisien) yang akan ditaksir pada $g_2(x)$. x_1, x_2, \dots, x_p adalah variabel prediktor. Berdasarkan kedua fungsi logit tersebut maka didapatkan model regresi *logistic trichotomous* sebagai berikut.

$$\begin{aligned}
\pi_0(x) &= \frac{1}{1 + \exp g_1(x) + \exp g_2(x)} \\
\pi_1(x) &= \frac{\exp g_1(x)}{1 + \exp g_1(x) + \exp g_2(x)} \\
\pi_2(x) &= \frac{\exp g_2(x)}{1 + \exp g_1(x) + \exp g_2(x)}
\end{aligned} \tag{2.3}$$

Dimisalkan $P(Y=j | x) = \pi_j(x)$ untuk $j=0,1,2$. Setiap probabilitas adalah fungsi dari vektor dari $2(p+1)$ parameter $\beta' = (\beta'_1, \beta'_2)$. Persamaan umum untuk kasus 3 kategori adalah

$$\pi_j(x) = \Pr(Y = j|x) = \frac{\exp g_j(x)}{\sum_{k=0}^2 \exp g_k(x)} \tag{2.4}$$

Dimana vektor $\beta_0 = 0$ dan $g_0(x) = 0$ (Hosmer dkk, 2013).

2.6 Estimasi Parameter

Menurut Hosmer dkk (2013), untuk membuat *likelihood function* maka digunakan 3 variabel biner dikodekan 0 atau 1 untuk menandakan keanggotaan grup dari suatu pengamatan. Variabel dikodekan sebagai $Y=0$ maka $Y_0=1, Y_1=0, Y_2=0$; jika $Y=1$ maka $Y_0=0, Y_1=1, Y_2=0$; dan jika $Y=2$ maka $Y_0=0, Y_1=0, Y_2=1$. Menggunakan notasi ini maka *conditional likelihood function* untuk sampel n observasi independen adalah

$$l(\beta) = \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \tag{2.5}$$

Dimana $\pi_0(x_i)$ adalah $\Pr(Y = 0|x_i)$. $\pi_1(x_i)$ adalah $\Pr(Y = 1|x_i)$. $\pi_2(x_i)$ adalah $\Pr(Y = 2|x_i)$. y_{0i} adalah 1 apabila observasi ke- i berkategori $Y=0$ dan bernilai 0 jika observasi ke- i tidak berkategori $Y=0$. y_{1i} adalah 1 apabila observasi ke- i berkategori $Y=1$ dan bernilai 0 jika observasi ke- i tidak berkategori $Y=1$. y_{2i} adalah 1 apabila observasi ke- i berkategori $Y=2$ dan bernilai 0 jika observasi ke- i tidak berkategori $Y=2$. *Log-likelihood function*-nya adalah

$$L(\beta) = \sum_{i=1}^n \{Y_{1i}g_1(x_i) + Y_{2i}g_2(x_i)\} - \ln(1 + \exp(g_1(x_i)) + \exp(g_2(x_i))) \tag{2.6}$$

Likelihood equation dicari dengan menurunkan persamaan diatas dengan setiap $2(p+1)$ parameter yang tidak diketahui. Untuk mempermudah notasi maka $\pi_{ji} = \pi_j(x_i)$. Bentuk umum untuk persamaan ini adalah

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=j}^n x_{ki}(Y_{ij} - \pi_{ij}) \quad (2.7)$$

Untuk $j=1,2,\dots$ dan $k=1,2,\dots,p$, dengan $x_{0i} = 1$ untuk setiap subjek

Estimator *maximum likelihood*, $\hat{\beta}$, didapat dengan menyamakan persamaan 2.7 dengan 0 dan mencari hingga didapatkan $\hat{\beta}$. Solusi dari persamaan diatas memerlukan komputasi yang berulang untuk mendapatkan hasil biner. (Hosmer dkk, 2013).

2.7 Uji Signifikansi Parameter

Pengujian signifikansi parameter pada model digunakan untuk mengetahui apakah variabel-variabel yang digunakan mempunyai pengaruh secara signifikan terhadap model. Pengujian signifikansi parameter dilakukan secara serentak dan individu untuk mengetahui pengaruh variabel pada model secara serentak maupun individu.

2.7.1 Pengujian Serentak

Pengujian secara serentak dilakukan untuk mengetahui apakah semua variabel prediktor secara serentak mempengaruhi nilai signifikansi parameter pada model. Hipotesis yang digunakan pada pengujian serentak adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{Minimal terdapat satu } \beta_j \neq 0 \text{ dengan } j=1,2,\dots,p.$$

Statistik Uji yang digunakan adalah statistik uji G dengan rumus sebagai berikut.

$$G = -2 \ln \left[\frac{\left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_2}{n}\right)^{n_2}}{\prod_{i=1}^n \pi_0(x_i)^{n_{0i}} \pi_1(x_i)^{n_{1i}} \pi_2(x_i)^{n_{2i}}} \right] \quad (2.8)$$

dimana

n_0 = Banyak pengamatan dengan kategori $Y=0$

n_1 = Banyak pengamatan dengan kategori $Y=1$

n_2 = Banyak pengamatan dengan kategori $Y=2$

n = Banyak pengamatan

Statistik uji G merupakan *Likelihood Ratio Test* yang mengikuti distribusi *Chi-Square*. Pengambilan kesimpulan yang dilakukan adalah tolak H_0 jika nilai $G > \chi^2_{(a,v)}$ dengan v adalah derajat bebas yaitu banyaknya variabel prediktor dalam model.

2.7.2 Uji Individu

Pengujian individu dilakukan pada masing-masing variabel prediktor yang digunakan untuk mengetahui variabel prediktor mana yang berpengaruh signifikan terhadap model. Hipotesis yang digunakan dalam pengujian individu adalah sebagai berikut.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \text{ dengan } j=1,2,\dots,p.$$

Statistik uji yang digunakan adalah *uji Wald* dengan rumus sebagai berikut.

$$W^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \quad (2.9)$$

Dimana $\hat{\beta}_j$ adalah taksiran parameter ke- j dan $SE(\hat{\beta}_j)$ adalah taksiran Standard Error dari $(\hat{\beta}_j)$ yang didapatkan dari rumus $SE(\hat{\beta}_j) = [var(\hat{\beta}_j)]^{1/2}$ dengan $j = 0,1,2,\dots,p$. Nilai W^2 mengikuti distribusi *Chi-Square* dengan derajat bebas satu sehingga penarikan

kesimpulan dilakukan dengan tolak H_0 jika nilai $W^2 > \chi^2_{(\alpha,1)}$. (Hosmer dan Lemeshow, 2000).

2.8 Interpretasi Model

Interpretasi model dilakukan berdasarkan hasil pengujian pada parameter. Untuk mempermudah interpretasi, maka digunakan nilai *Odds Ratio* atau OR. *Odds Ratio* menjelaskan perbandingan antara kategori yang dihasilkan dengan kategori yang digunakan sebagai dasar (Hosmer, Lemeshow, & Sturdivant, 2013). *Odds Ratio* antara $Y = e$ atau kategori yang dibandingkan dengan $Y = d$ atau kategori yang dijadikan sebagai dasar untuk nilai $x = a$ dan $x = b$ adalah sebagai berikut.

$$OR_e(a, b) = \frac{\Pr(Y = e|x = a) / \Pr(Y = d|x = a)}{\Pr(Y = e|x = b) / \Pr(Y = d|x = b)} \quad (2.10)$$

Nilai OR yang dihasilkan selalu positif dan biasa digunakan untuk pendekatan nilai *Relative Risk*. Jika nilai $OR = 1$ maka risiko yang dihasilkan pada $x = a$ sama dengan $x = b$ untuk menghasilkan $Y = e$. Jika nilai yang dihasilkan $1 < OR < \infty$ maka risiko $x = a$ lebih besar dan $0 < OR < 1$ adalah sebaliknya.

2.9 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui apakah model dengan variabel dependen tersebut merupakan model yang sesuai. Uji kesesuaian model dapat menggunakan statistik uji *ChiSquare* dengan hipotesis sebagai berikut.

H_0 : model sesuai (tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

H_1 : model tidak sesuai (ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

Perhitungan statistik uji Chi-Square sebagai berikut.

$$\chi^2 = \sum_{k=1}^g \frac{(o_k - n'_k \pi_k)^2}{n'_k \pi_k (1 - \pi_k)} \quad (2.11)$$

dimana

$o_k = \sum_{j=1}^{n'_k} y_j$ (jumlah pengamatan pada grup ke- k)

$\bar{\pi}_k = \sum_{j=1}^{n'_k} \frac{m_j \pi_j}{n'_k}$ (rata-rata taksiran probabilitas)

g = jumlah grup (kombinasi kategori dalam model serentak)

m_j = banyaknya observasi yang memiliki nilai $\hat{\pi}_j$

n'_k = banyaknya observasi pada grup ke- k

Pengambilan keputusan didasarkan pada tolak H_0 jika $\chi^2_{hitung} \geq \chi^2_{(db,\alpha)}$ dengan $db = g - 2$.

2.10 Random Forest

Random forest merupakan metode klasifikasi yang *supervised*. Sesuai dengan namanya, metode ini menciptakan sebuah hutan (*forest*) dengan sejumlah pohon (*tree*). Secara umum, semakin banyak pohon (*tree*) pada sebuah hutan (*forest*) maka semakin kuat pula hutan tersebut terlihat. Pada kasus sama, semakin banyak *tree*, maka semakin besar pula akurasi yang dihasilkan (Polamuri, 2017).

Random forest dapat diartikan sebagai suatu metode klasifikasi yang terdiri dari gabungan pohon klasifikasi (CART) yang saling independen. Prediksi klasifikasi dari *random forest* melalui proses *voting* (jumlah terbanyak) berdasarkan pohon klasifikasi

yang terbentuk. *Random forests* merupakan pengembangan dari metode *ensemble* yang pertama kali dikembangkan oleh Leo Breiman (2001) dan digunakan untuk meningkatkan ketepatan klasifikasi. Bila dalam proses *bagging* digunakan *resampling bootstrap* untuk membangkitkan pohon klasifikasi dengan banyak versi yang kemudian mengkombinasikannya untuk memperoleh prediksi akhir, maka dalam *random forests* proses pengacakan untuk membentuk pohon klasifikasi tidak hanya dilakukan untuk data sampel saja melainkan juga pada pengambilan variabel prediktor. Sehingga, proses ini akan menghasilkan kumpulan pohon klasifikasi dengan ukuran dan bentuk yang berbeda-beda. Hasil yang diharapkan adalah suatu kumpulan pohon klasifikasi yang memiliki korelasi kecil antar pohon. Korelasi yang kecil akan menurunkan hasil kesalahan prediksi *Random Forests* (Breiman, 2001).

2.10.1 CART (Classification And Regression Trees)

CART merupakan pendekatan model nonparametric yang dapat menjelaskan variabel respon yang dipengaruhi oleh variabel predictor yang bersifat kontinu maupun kategorik. Data dependen tergantung dari partisi serangkaian *node* yang bercabang ke kanan dan ke kiri dapat disebut simpul anak (*child nodes*) yang berasal dari simpul utama (*parent nodes*). Setelah partisi berhenti, *child nodes* disebut *terminal nodes* (Zheng dkk, 2009).

2.10.2 Algoritma Random Forest

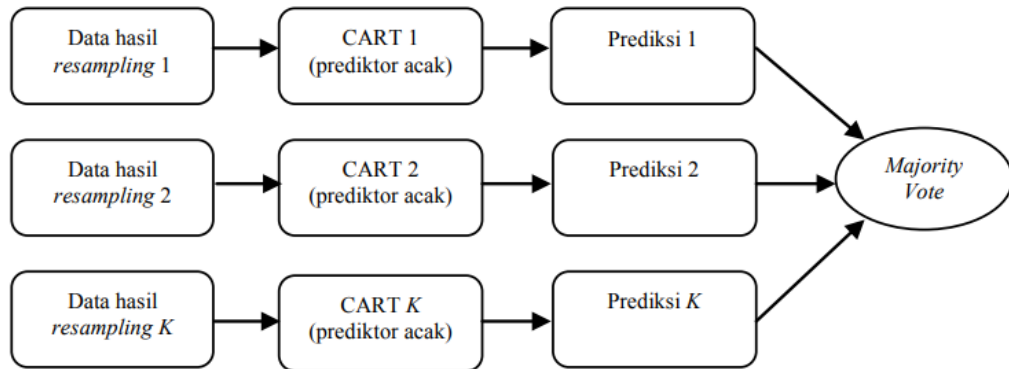
Secara umum, pengembangan *random forests* yang dilakukan dari proses *bagging* yaitu terletak pada proses pemilihan pemilah. Pada *random forest*, pemilihan pemilah hanya melibatkan beberapa variabel prediktor yang diambil secara acak. Algoritma *Random Forest* dijelaskan sebagai berikut.

- a. Mengambil n data sampel dari *dataset* awal dengan menggunakan teknik *resampling bootstrap* dengan pengembalian.
- b. Menyusun pohon klasifikasi dari setiap *dataset* hasil *resampling bootstrap*, dengan penentuan pemilah terbaik didasarkan pada variabel prediktor yang diambil secara acak. Jumlah variabel yang diambil secara acak dapat ditentukan melalui perhitungan $\log_2(M + 1)$ dimana M adalah banyak variabel prediktor (Breiman, 2001) atau \sqrt{p} dimana p adalah banyak variabel prediktor (Genuer, et al., 2009) atau \sqrt{d} dimana d adalah banyak variabel prediktor (Sartono, et al., 2010).
- c. Melakukan prediksi klasifikasi data sampel berdasarkan pohon klasifikasi yang terbentuk.
- d. Mengulangi langkah a-c hingga diperoleh sejumlah pohon klasifikasi yang diinginkan. Perulangan dilakukan sebanyak K kali.
- e. Melakukan prediksi klasifikasi data sampel akhir dengan mengkombinasikan hasil prediksi pohon klasifikasi yang diperoleh berdasarkan aturan *majority vote*.

Dalam analisis dengan menggunakan metode *random forest* dimulai dari pengambilan data dengan teknik *resampling bootstrap*. *Bootstrap* adalah suatu metode yang dapat bekerja tanpa membutuhkan asumsi distribusi karena sampel asli digunakan sebagai populasi. *Bootstrap* pertama kali diperkenalkan oleh Efron (1979) yang digunakan untuk mencari distribusi sampling dari suatu estimator dengan prosedur *resampling* dengan pengembalian dari data asli (Sungkono, 2013). Berikut adalah algoritma dari *resampling bootstrap*.

- a. Mengkonstruksi distribusi empiris dari suatu sampel dengan memberikan probabilitas $1/n$ pada setiap X_i dimana $i = 1, 2, \dots, n$.
- b. Mengambil sampel bootstrap berukuran n secara random dengan pengembalian dari

- distribusi empiris \hat{F}_n sebut sebagai sampel bootstrap pertama X^{*1} .
- Menghitung statistik sampel atau nilai estimasi $\hat{\theta}$ yang diinginkan dari sampel bootstrap X^{*1} disebut sebagai $\hat{\theta}_1^*$.
 - Mengulangi langkah b dan c hingga k kali diperoleh $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$
 - Mengkonstruksi suatu distribusi probabilitas dari $\hat{\theta}_B^*$ dengan memberikan probabilitas $1/B$ pada setiap $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$. Distribusi tersebut merupakan estimator bootstrap untuk distribusi sampling $\hat{\theta}$ dan dinotasikan dengan \hat{F}^* .
 - Pendekatan estimasi bootstrap adalah $\hat{\theta}^* = \sum_{b=1}^B \hat{\theta}_b^* \frac{1}{B}$

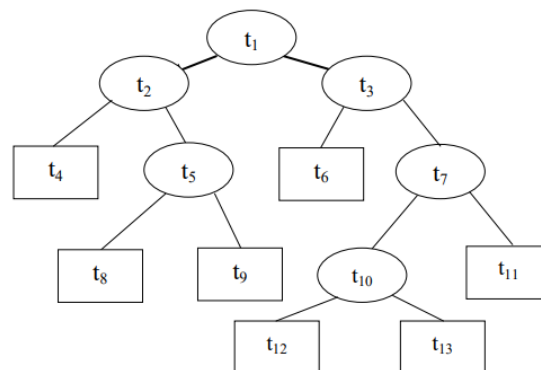


Gambar 2.2 Algoritma Random Forest (Sumber : Abid, dkk, 2020)

2.10.3 Pohon Klasifikasi

Sistem pohon keputusan (*decision trees*) berbeda proses yang terjadi pada metode regresi logistik, pohon keputusan (*decision trees*) bukanlah model parametrik sehingga tidak memiliki parameter. Pohon disajikan berupa rangkaian yang terdiri dari pilihan dan keputusan. Keputusan digambarkan seperti pohon asli namun ditampilkan secara terbalik. Ada dua macam pohon keputusan yakni pohon klasifikasi dan pohon regresi. Pada penelitian ini akan digunakan pohon klasifikasi.

Pohon klasifikasi seperti yang diilustrasikan pada Gambar 2.2 dibuat dari kumpulan simpul yang saling terhubung. Masing-masing simpul mewakili sebuah keputusan, dimana keputusan didasarkan pada salah satu variabel prediktor. Untuk setiap keputusan, proses akan turun satu level dari pohon ke simpul lainnya hingga mencapai simpul terminal yang dinamakan daun (*leaf*). Daun berisi level dari respon dan diakhiri dengan menentukan kelas.



Gambar 2.3 Ilustrasi Pohon Klasifikasi (Sumber : Abid, dkk, 2020)

2.10.4 Ilustrasi Random Forest

Ilustrasi *random forests* (Sumber : Rahmawati, 2015) berikut menggunakan data sebanyak 10. Ilustrasi yang terdiri dari variabel respon (Y) dengan tiga kategori dan sebanyak tujuh variabel prediktor.

Tabel 2.1 Data Sampel untuk Ilustrasi *Random Forest*

Data Ilustrasi Ke-	Data Awal				Data Resampling			
	Y	X_1	...	X_7	Y	X_1	X_5	X_7
1	2	0		18	0	0	3	34
2	1	1		13	0	0	1	26
3	1	0		13	1	3	0	18
4	0	4		34	2	3	2	21
5	1	1		26	2	2	2	36
6	2	3		18	2	4	0	18
7	0	0		36	1	1	3	13
8	2	4		21	0	1	3	13
9	2	2		18	0	2	1	18
10	1	2		36	1	3	0	21

Tahapan *random forests* diawali dengan melakukan *resampling* dengan pengembalian pada data ilustrasi awal. Data hasil *resampling* tersebut kemudian digunakan untuk membentuk pohon klasifikasi (CART). Namun, pemilah terbaik dipilih berdasarkan variabel prediktor yang diambil secara acak (misal $p=3$). Selain itu, pemilihan dilakukan hingga diperoleh pohon maksimal tanpa dilakukan pemangkasan pada pohon. Dari pohon yang terbentuk tersebut kemudian dilakukan prediksi klasifikasi dengan memberikan label kelas pada setiap terminal node yang dihasilkan. Proses ini diulang hingga sejumlah K replikasi sesuai keinginan peneliti. Bila hasil prediksi dari semua replikasi telah diperoleh, maka selanjutnya dilakukan *voting* mayoritas (*majority vote*) untuk menentukan prediksi klasifikasi akhir yang kemudian akan digunakan untuk menghitung kesalahan klasifikasinya. Berikut diberikan contoh hasil prediksi dengan menggunakan sebanyak empat kali replikasi pengambilan data dengan pengembalian ($K=4$). Hasil prediksi ditunjukkan pada Tabel 2.2 berikut.

Tabel 2.2 Hasil Prediksi Kelas dengan *Random Forest*

Data Ilustrasi ke-	Y	X_1	...	X_7	Prediksi CART 1	Prediksi CART 2	Prediksi CART 3	Prediksi CART 4	Prediksi Kelas *
1	2	0		18	1	0	0	2	2
2	1	1		13	2	1	1	1	1
3	1	0		13	1	2	1	1	1
4	0	4		34	0	1	2	1	1
5	1	1		26	0	0	1	0	0
6	2	3		18	0	1	1	1	1
7	0	0		36	1	1	2	2	2
8	2	4		21	2	1	1	1	1
9	2	2		18	1	1	1	1	1
10	1	2		36	1	2	2	1	1

*Hasil *majority vote* dari prediksi CART 1, CART 2, CART 3 dan CART 4

Apabila dibandingkan antara hasil pengamatan dengan prediksi, ada 6 pengamatan yang salah diklasifikasikan yaitu pengamatan ke-4, 5, 6, 7, 8, 9 sedangkan 4 pengamatan yang lainnya tepat diklasifikasikan.

2.10.5 Ketepatan Klasifikasi

Menurut Johnson (2007) untuk menghitung ketepatan klasifikasi pada hasil pengelompokkan digunakan *apparent error rate* (APER). Nilai APER menyatakan representasi proporsi sampel yang salah diklasifikasikan. Dalam penelitian kali ini digunakan respon multinomial tiga kategori sehingga penentuan kesalahan klasifikasi dapat dihitung dari tabel klasifikasi berikut.

Tabel 2.3 Klasifikasi Respon Multinomial Tiga Kategori

Aktual	Prediksi			Total
	1	2	3	
1	n_{11}	n_{12}	n_{13}	$n_{.1}$
2	n_{21}	n_{22}	n_{23}	$n_{.2}$
3	n_{31}	n_{32}	n_{33}	$n_{.3}$
Total	N_1	N_2	N_3	N

Nilai APER dihitung sebagai berikut.

$$APER(\%) = \frac{n_{12} + n_{13} + n_{23} + n_{21} + n_{31} + n_{32}}{N} \times 100\% \quad (2.12)$$

$$Ketepatan\ klasifikasi = 1 - APER \quad (2.13)$$

keterangan :

- n_{11} : jumlah observasi dari kelas 1 yang tepat diprediksi sebagai kelas 1
- n_{21} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 1
- n_{31} : jumlah observasi dari kelas 3 yang salah diprediksi sebagai kelas 1
- n_{12} : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 2
- n_{22} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 2
- n_{32} : jumlah observasi dari kelas 3 yang salah diprediksi sebagai kelas 2
- n_{13} : jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 3
- n_{23} : jumlah observasi dari kelas 2 yang salah diprediksi sebagai kelas 3
- n_{33} : jumlah observasi dari kelas 3 yang salah diprediksi sebagai kelas 3
- N_1 : jumlah observasi dari kelas 1
- N_2 : jumlah observasi dari kelas 2
- N_3 : jumlah observasi dari kelas 3
- N : jumlah observasi

2.11 Penanganan Kecelakaan Kerja

Negara dalam hal ini pemerintah telah berupaya untuk memberikan suatu jaminan khususnya dalam pembangunan ketenagakerjaan melalui program Jaminan Sosial yang diselenggarakan oleh BPJS Ketenagakerjaan, secara khusus mengatur jaminan sosial bagi tenaga kerja meliputi jaminan kecelakaan kerja, jaminan kematian, jaminan hari tua, dan jaminan pensiun. Dengan demikian upaya pemenuhan jaminan sosial yang adil dan merata untuk seluruh rakyat Indonesia dapat dilaksanakan sejalan dengan program pembangunan nasional Indonesia yang sesuai dengan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945.

BPJS Ketenagakerjaan, sebagai lembaga yang diberi amanah untuk melaksanakan jaminan sosial bagi tenaga kerja di Indonesia, memberikan perlindungan kepada pekerja formal (Penerima Upah/PU) maupun informal (Bukan Penerima Upah/BPU).

Pembentukan BPJS melalui Undang-Undang Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial merupakan pelaksanaan program jaminan social nasional yang bertujuan untuk memberikan jaminan kesejahteraan bagi seluruh rakyat Indonesia. Badan Penyelenggaraan Jaminan Sosial terdiri dari BPJS Kesehatan dan BPJS Ketenagakerjaan sebagaimana diatur dalam Pasal 5 ayat (2) Undang-Undang Nomor 24 Tahun 2011 tentang BPJS. Jaminan Sosial Tenaga Kerja (JAMSOSTEK) selanjutnya berubah menjadi BPJS Ketenagakerjaan mulai beroperasi pada tanggal 1 Januari 2014, tidak lagi menyelenggarakan program jaminan pemeliharaan kesehatan melainkan hanya terfokus pada empat program yaitu Jaminan Kecelakaan Kerja, Jaminan Kematian, Jaminan Hari Tua, dan Jaminan Pensiun.

BAB 3 METODOLOGI

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder laporan klaim kecelakaan kerja pada BPJS Ketenagakerjaan wilayah Mojokerto. Periode data laporan klaim kecelakaan kerja yang akan diteliti adalah dari Bulan September 2020 hingga Bulan Mei 2021, diambil sebanyak 320 data harian klaim kecelakaan kerja

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah jenis kecelakaan kerja sebagai variabel respon dengan 3 kategorik dan 4 variabel prediktor berskala nominal dan 1 variabel prediktor yang berskala rasio. Variabel prediktor yang berskala nominal terdiri dari jenis kelamin, jabatan, jam kerja, jenis kendaraan sedangkan variabel prediktor yang berskala rasio yaitu umur.

Definisi dari variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

- a. Jenis kecelakaan kerja
 - Di dalam lokasi kerja yaitu dimana kejadiannya terjadi di tempat lokasi kerja seperti di pabrik atau dikantor pada saat jam kerja.
 - Di luar lokasi kerja yaitu dimana kejadiannya terjadi di luar tempat lokasi kerja dan diluar jam kerja.
 - Di lalu lintas yaitu kejadiannya terjadi di jalan raya dan kecelakaan terjadi karena peristiwa tertabrak atau menabrak bisa dikatakan (ada lawan) pada saat jam kerja.
- b. Jenis kelamin digunakan untuk melihat dari faktor demografi.
- c. Jabatan
 - Bagian produksi yaitu profesi yang memungkinkan resiko paling tinggi terhadap kecelakaan kerja dimana langsung berhubungan dengan mesin pabrik.
 - Bagian office yaitu profesi dibagian perkantoran (staff)
- d. Umur digunakan untuk melihat dari faktor demografi.
- e. Jam kerja
Merupakan pelaksanaan kerja yang terbagi menjadi >8 jam atau <8 jam.
- f. Waktu Kejadian Kecelakaan
Merupakan jam atau waktu terjadinya kecelakaan dimana terbagi waktu terang (06.00-18.00 WIB) dan waktu gelap (19.00-05.00 WIB).

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel	Skala	Kategori
Y	Jenis Kecelakaan Kerja	Nominal	0 = Lalu Lintas 1 = Didalam lokasi kerja 2 = Diluar lokasi kerja
X ₁	Jenis Kelamin	Nominal	0 = Laki-laki 1 = Perempuan
X ₂	Jabatan	Nominal	0 = Produksi 1 = Office
X ₃	Umur	Rasio	
X ₄	Jam Kerja	Nominal	0 = >8 jam 1 = <8 jam
X ₅	Waktu Kejadian Kecelakaan	Nominal	0 = Waktu terang (06.00-18.00 WIB) 1 = Waktu Gelap (19.00-05.00 WIB)

Struktur data dalam penelitian ini dapat dilihat pada Tabel 3.2 dengan variabel respon sebanyak tiga kategori dan lima kategori variabel prediktor.

Tabel 3.2 Struktur Data

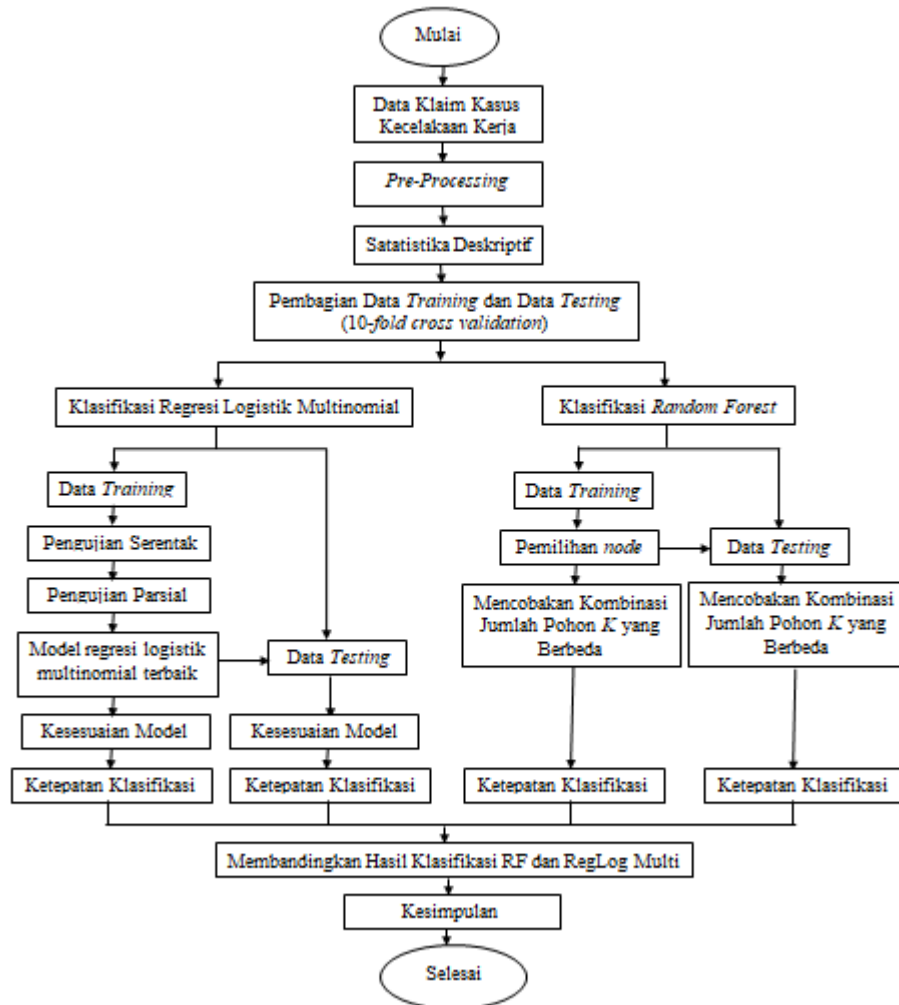
Y_i	X_1	X_2	X_3	X_4	X_5
Y_1	$X_{(1)1}$	$X_{(1)2}$	$X_{(1)3}$	$X_{(1)4}$	$X_{(1)5}$
Y_2	$X_{(2)1}$	$X_{(2)2}$	$X_{(2)3}$	$X_{(2)4}$	$X_{(2)5}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Y_n	$X_{(n)1}$	$X_{(n)2}$	$X_{(n)3}$	$X_{(n)4}$	$X_{(n)5}$

3.3 Metode Analisis

Analisis yang digunakan dalam penelitian ini adalah analisis regresi logistik multinomial dan *random forests* yang digunakan untuk pengklasifikasian jenis kecelakaan kerja pada BPJS Ketenagakerjaan. Berikut tahapan analisis dalam penelitian ini.

1. Pengumpulan data pengajuan klaim kasus kecelakaan kerja oleh peserta Jaminan Kecelakaan Kerja (JKK) BPJS Ketenagakerjaan wilayah Mojokerto.
2. Melakukan *pre-processing* data penelitian.
Pre-processing data dilakukan meliputi pengkodean data sesuai dengan kategori yang telah ditentukan.
3. Melakukan statistika deskriptif pada data pengajuan klaim kasus kecelakaan kerja oleh peserta Jaminan Kecelakaan Kerja (JKK) BPJS Ketenagakerjaan wilayah Mojokerto.
4. Membagi data menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* dan data *testing* dibagi menggunakan *10-fold cross validation*.
5. Melakukan analisis regresi logistik multinomial melalui tahapan berikut.
 - a. Melakukan uji serentak terhadap variabel-variabel yang berpengaruh terhadap kasus kecelakaan kerja pada BPJS Ketenagakerjaan.
 - b. Melakukan uji parsial terhadap variabel-variabel yang berpengaruh terhadap kasus kecelakaan kerja pada BPJS Ketenagakerjaan.
 - c. Membuat persamaan model regresi logistik multinomial dan menginterpretasi model.
 - d. Uji kesesuaian model.
 - e. Menghitung ketepatan klasifikasi hasil bentukan dengan menggunakan data *training* dan mengevaluasinya dengan cara menjalankan data *testing* pada persamaan regresi logistik yang terbentuk.
 - f. Membandingkan hasil ketepatan klasifikasi dari setiap kombinasi data *training* dan data *testing*.
6. Melakukan analisis *Random Forest* melalui tahapan berikut.
 - a. Mengambil n sampel bootstrap dengan pengembalian dari data *training*.
 - b. Menentukan jumlah variabel prediktor yang akan dilakukan pengambilan secara acak dalam proses penentuan pemilah saat pembentukan pohon klasifikasi
$$\log_2(M + 1) \text{ atau } \sqrt{p}$$
 - c. Membentuk pohon klasifikasi dimana pemilihan *node* terbaik dilakukan berdasarkan variabel-variabel prediktor yang diambil secara acak.
 - d. Melakukan prediksi klasifikasi untuk data *training*.
 - e. Mengulangi tahap b sampai tahap d hingga K kali replikasi.
 - f. Mengulangi tahap b sampai tahap d dengan mencobakan kombinasi jumlah pohon (K) yang berbeda, yaitu sebesar 100, 500 dan 1000.
 - g. Menghitung ketepatan klasifikasi data *training*.

- h. Menghitung ketepatan klasifikasi data *testing*.
 - i. Memilih kombinasi jumlah pohon yang memiliki keakurasian tertinggi.
7. Memilih model terbaik dengan cara membandingkan tingkat akurasi antara regresi logistik multinomial dengan *Random Forests*. Klasifikasi yang mempunyai 1-APER terbesar dipilih sebagai tingkat klasifikasi yang lebih baik dalam melakukan klasifikasi.
- Langkah-langkah analisis dalam penelitian ini diilustrasikan dengan diagram alir pada Gambar 3.1



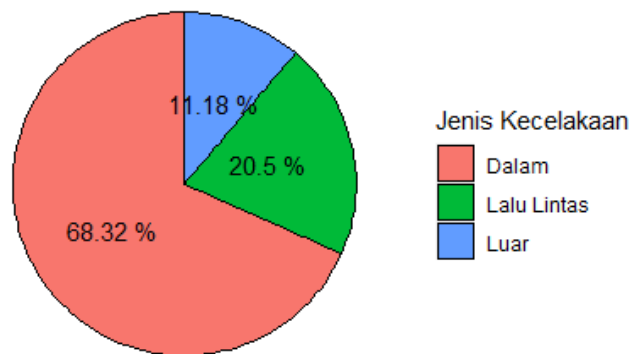
Gambar 3.1 Diagram Alir Penelitian

(Halaman ini sengaja dikosongkan)

BAB 4 HASIL DAN PEMBAHASAN

4.1 Karakteristik Kecelakaan yang Ditangani oleh BPJS Ketenagakerjaan

Analisis dan pembahasan mengenai statistika deskriptif terhadap jenis kecelakaan, jenis kelamin, jabatan, usia, jam kerja, dan waktu kejadian kecelakaan pada Bulan September 2020 hingga Bulan Mei 2021 diharapkan dapat berguna bagi masyarakat khususnya untuk para pekerja yang memiliki asuransi Jaminan Kecelakaan Kerja (JKK) pada BPJS Ketenagakerjaan wilayah Mojokerto sebagai referensi antisipasi jatuhnya korban kecelakaan kerja.



Gambar 4.1 Jenis Kecelakaan Kerja

Gambar 4.1 menunjukkan jenis kecelakaan kerja terbagi menjadi tiga kategori yaitu dalam lokasi kerja, lalu lintas, luar lokasi kerja. Jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto yang jumlah terbanyak adalah di dalam lokasi kerja yaitu sebesar 68,32% sedangkan jumlah terbanyak kedua adalah di lalu lintas yaitu sebesar 20,5% dan yang terakhir adalah di luar lokasi kerja sebanyak 11,18%. Pada Tabel 4.1 menunjukkan bahwa korban kecelakaan di dalam lokasi kerja terbanyak adalah berjenis kelamin laki-laki dengan jumlah 206 orang, begitu pula di lalu lintas dengan jumlah 43 orang. Sedangkan untuk korban kecelakaan di luar lokasi kerja jumlah korban laki-laki dan perempuan sama yaitu 18 orang.

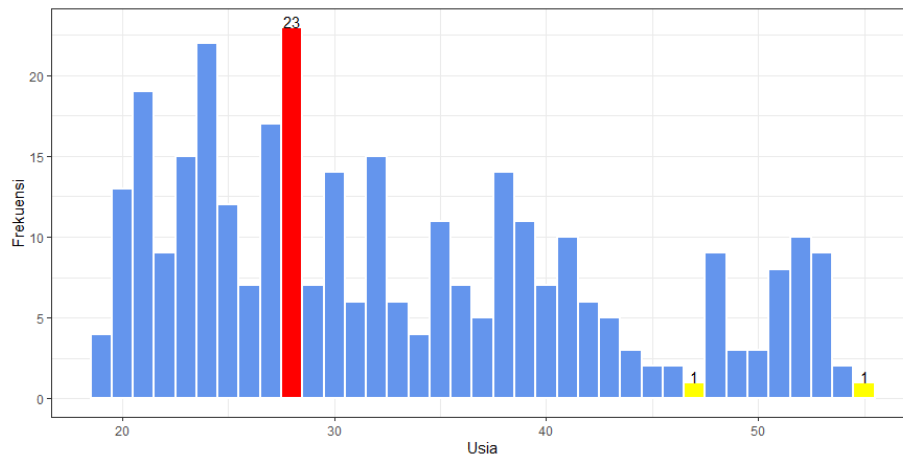
Tabel 4.1 Crosstabulation Jenis Kelamin dan Jenis Kecelakaan Kerja

Jenis Kelamin	Jenis Kecelakaan Kerja		
	Dalam Lokasi Kerja	Lalu Lintas	Luar Lokasi Kerja
Laki-laki	206	43	18
Perempuan	14	23	18

Tabel 4.2 Crosstabulation Jabatan dan Jenis Kecelakaan Kerja

Jabatan	Jenis Kecelakaan Kerja		
	Dalam Lokasi Kerja	Lalu Lintas	Luar Lokasi Kerja
Produksi	174	19	17
Office	46	47	19

Dapat dilihat pula pada Tabel 4.2 korban kecelakaan kerja dari jabatan yang diduduki pekerja, dimana dalam hal ini korban kecelakaan di dalam lokasi kerja yang terbanyak adalah yang bekerja pada bidang produksi yaitu sebanyak 174 orang, sedangkan korban kecelakaan kerja di lalu lintas dan di luar lokasi kerja yang terbanyak pada bidang *office* dengan jumlah 47 orang dan 19 orang.



Gambar 4.2 Histogram Usia Korban Kecelakaan Kerja

Usia korban kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto yang terbanyak adalah korban di usia kisaran 21 tahun sampai 28 tahun, yang dapat diambil titik tengahnya yaitu 24 tahun. Jadi, korban kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto terbanyak kurang lebih berusia 24 tahun yang terdapat pada histogram usia korban kecelakaan kerja pada Gambar 4.2. Korban kecelakaan kerja tertinggi sebanyak 23 kasus pada usia 28 tahun, dan korban kecelakaan kerja terendah sebanyak 1 kasus pada usia 47 tahun dan 55 tahun. Rata-rata (*mean*) usia korban kecelakaan kerja adalah 33 tahun artinya jumlah korban kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto 50 persen dibawah 33 tahun dan 50 persen berusia diatas 33 tahun.

Kecelakaan di dalam lokasi kerja terbanyak terjadi pada korban yang jam kerjanya lebih dari 8 jam dengan jumlah 153 orang, begitu pula dengan lalu lintas dan luar lokasi kerja dengan jumlah 47 orang dan 19 orang yang dapat dilihat pada Tabel 4.3.

Tabel 4.3 Crosstabulation Jam Kerja dan Jenis Kecelakaan Kerja

Jam Kerja	Jenis Kecelakaan Kerja		
	Dalam Lokasi Kerja	Lalu Lintas	Luar Lokasi Kerja
>8jam	153	47	19
<=8jam	67	19	17

Tabel 4.4 Crosstabulation Waktu Kejadian Kecelakaan dan Jenis Kecelakaan Kerja

Waktu Kejadian Kecelakaan	Jenis Kecelakaan Kerja		
	Dalam Lokasi Kerja	Lalu Lintas	Luar Lokasi Kerja
Terang (06.00-18.00 WIB)	106	22	15
Gelap (19.00-05.00 WIB)	114	44	21

Waktu kejadian kecelakaan di dalam lokasi kerja tertinggi terjadi pada waktu malam hari (gelap) pada pukul 19.00 WIB sampai 05.00 WIB dengan jumlah 114 orang, begitu pula pada lalu lintas dan di luar lokasi kerja dengan jumlah 44 orang dan 21 orang.

4.2 Analisis Regresi Logistik Multinomial untuk Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Wilayah Mojokerto

Data pengamatan yang digunakan untuk mengklasifikasi jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto adalah sebesar 322 pengamatan korban kecelakaan kerja, data tersebut dapat dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Pembagian data menjadi dua bagian bertujuan untuk mendapatkan suatu model yang optimal dengan melakukan percobaan klasifikasi menggunakan *10-Fold Cross Validation* dengan kombinasi data *training* dan data *testing* yang dicobakan dalam penelitian ini adalah 75%:25%, 80%:20%, 85%:15%, 90%:10%.

Hasil klasifikasi yang diperoleh dari tiap-tiap presentase kombinasi selanjutnya akan dilakukan perbandingan terhadap ketepatan klasifikasi data *training* dan data *testing* dengan 10 macam kombinasi (*subset fold*). Kombinasi proporsi data *training* dan data *testing* terbaik dipilih berdasarkan nilai ketepatan klasifikasi yang paling besar. Data *training* digunakan untuk mengestimasi parameter model dan data *testing* digunakan untuk mencerminkan kebaikan model dalam mengklasifikasikan data baru.

4.2.1 Pemilihan Presentase Kombinasi Data *Training* dan Data *Testing* Terbaik dalam Analisis Regresi Logistik Multinomial

Pemilihan presentase kombinasi data *training* dan data *testing* terbaik didasarkan pada nilai rata-rata *total accuracy rate*. Rata-rata diperoleh dari nilai (1-APER) masing-masing presentase kombinasi yang ditunjukkan pada Tabel 4.5.

Tabel 4.5 Perbandingan *Total Accuracy Rate* Kombinasi Data

Kombinasi Data <i>Training</i> dan Data <i>Testing</i>	<i>Total Accuracy Rate</i> (1-APER)	
	<i>Data Training</i> (%)	<i>Data Testing</i> (%)
75%:25%	71,07	71,25
80%:20%	72,09	60,93
85%:15%	74,90	65,95
90%:10%	69,07	70,96

Tabel 4.5 menunjukkan bahwa kombinasi data *training* dan data *testing* yang mampu memberikan ketepatan klasifikasi yang tinggi adalah pada kombinasi data *training* sebesar 75% dan data *testing* sebesar 25%. Ketepatan klasifikasi yang dipilih dari beberapa kombinasi data *training* dan data *testing* adalah ketepatan klasifikasi pada data *training* dan data *testing* yang paling tinggi. Apabila pada data *training* memberikan nilai ketepatan klasifikasi yang paling tinggi akan tetapi pada data *testing* tidak memberikan nilai ketepatan klasifikasi yang paling tinggi begitu juga sebaliknya sehingga yang pilih adalah kombinasi data *training* dan data *testing* yang memberikan nilai ketepatan klasifikasi tertinggi pada ketepatan akurasi data *testing*.

Ketepatan klasifikasi untuk data *testing* yang diperoleh adalah sebesar 71,25% yang merupakan nilai tertinggi diantara semua kombinasi data. Untuk ketepatan klasifikasi data *training* diperoleh ketepatan klasifikasi sebesar 71,07%. Oleh karena itu, hasil analisis yang digunakan untuk menjelaskan jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto adalah hasil klasifikasi yang menggunakan kombinasi

data *training* dan data *testing* terbaik, yaitu data *training* sebesar 75% dan data *testing* sebesar 25%.

4.2.2 Pemilihan Subset Terbaik dari Kombinasi Data *Training* dan Data *Testing* Terpilih Menggunakan *K-Fold Cross Validation*

Berdasarkan nilai tertinggi dari ketepatan klasifikasi, kombinasi data *training* dan data *testing* yang terpilih yaitu sebesar 75% data *training* dan 25% data *testing*. Selanjutnya dilakukan metode *cross validation* dengan 10 macam kombinasi (*subset fold*) yang nantinya dilakukan pemilihan *subset* terbaik dengan metode regresi logistik multinomial menggunakan data awal pada *training* berdasarkan nilai *accuracy* tertinggi yang ditunjukkan pada Tabel 4.6.

Berdasarkan Tabel 4.6 pada *subset* ke-6 mempunyai performa klasifikasi yang terbaik dengan nilai *accuracy* sebesar 0,82, sehingga dengan kombinasi *training* dan *testing* data pada *subset* ini akan dilakukan pemodelan.

Tabel 4.6 Pemilihan *Subset* Terbaik Regresi Logistik Multinomial Berdasarkan *Accuracy*

<i>Subset 10-Fold</i>	<i>Accuracy</i>
<i>Subset</i> ke-1	0,66
<i>Subset</i> ke-2	0,73
<i>Subset</i> ke-3	0,70
<i>Subset</i> ke-4	0,75
<i>Subset</i> ke-5	0,68
<i>Subset</i> ke-6	0,82
<i>Subset</i> ke-7	0,60
<i>Subset</i> ke-8	0,66
<i>Subset</i> ke-9	0,72
<i>Subset</i> ke-10	0,60

4.2.3 Uji Serentak Variabel-variabel yang Berpengaruh terhadap Jenis Kecelakaan Kerja

Langkah awal dalam analisis regresi logistik multinomial adalah uji serentak terhadap variabel-variabel yang berpengaruh terhadap jenis kecelakaan kerja. Pada analisis ini digunakan data *training* sebesar 291 data dan data *testing* sebesar 32 data.

Tabel 4.7 Uji Serentak Regresi Logistik Multinomial

Model	<i>Likelihood Ratio Test</i>			
	<i>Log Likelihood</i>	<i>df</i>	<i>Chi-square</i>	<i>P-value</i>
<i>Intercept Only</i>	-211,03			
<i>Final</i>	-267,28	10	112,51	$2,2 \times 10^{-16}$

Berdasarkan Tabel 4.7 nilai statistik uji serentak diperoleh sebesar 112,51. Nilai tersebut akan dibandingkan dengan *Chi-Square* tabel ($\chi^2_{(0,05,2)}$) sebesar 0,103 dimana nilai tersebut lebih kecil dari statistik uji yang didapat sehingga diputuskan Tolak H_0 . Hasil uji serentak juga bisa diperoleh dengan membandingkan *p-value* sebesar $2,2 \times 10^{-16}$ dimana nilai tersebut kurang dari 5% atau 0,05. Hal ini berarti paling sedikit terdapat satu variabel dependen yang secara statistik signifikan mempengaruhi variabel independen. Dikarenakan uji serentak signifikan, maka dapat dilanjutkan ke uji parsial.

4.2.4 Uji Parsial Variabel-variabel yang Berpengaruh terhadap Jenis Kecelakaan Kerja

Langkah selanjutnya setelah uji serentak, dilakukan uji parsial terhadap variabel-variabel yang berpengaruh terhadap jenis kecelakaan kerja. Berdasarkan dari hasil analisis uji serentak didapatkan bahwa variabel yang berpengaruh adalah jenis kelamin dan jabatan. Akan tetapi, tidak diketahui kategori dari dua variabel yang berpengaruh terhadap jenis kecelakaan kerja sehingga dilakukan uji parsial untuk mengetahui variabel prediktor yang mempengaruhi jenis kecelakaan kerja dengan berbagai macam kategorinya. Berikut uji parsial untuk mengetahui variabel yang berpengaruh terhadap kecelakaan lalu lintas ditampilkan pada Tabel 4.8.

Tabel 4.8 Uji Parsial Regresi Logistik Multinomial

Variabel Respon	Variabel Prediktor	P-value
Lalu Lintas	Jenis Kelamin (1)	$4,072 \times 10^{-3}$
	Jabatan (1)	$3,087 \times 10^{-9}$
	Umur	0,613
	Jam Kerja (1)	0,973
	Waktu Kejadian Kecelakaan (1)	0,061
Luar Lokasi Kerja	Jenis Kelamin (1)	$1,175 \times 10^{-6}$
	Jabatan (1)	$3,520 \times 10^{-3}$
	Umur	0,243
	Jam Kerja (1)	0,548
	Waktu Kejadian Kecelakaan (1)	0,816

Berdasarkan hasil Tabel 4.8 menunjukkan bahwa masih banyak variabel yang tidak signifikan dikarenakan *p-value*nya lebih dari 0,05. Oleh karena itu, digunakan fungsi *STEPAIC* pada *software* R untuk mendapatkan model terbaik. Pengujian serentak dilakukan terlebih dahulu untuk memeriksa koefisien β secara serentak pada model yang terbaik. Nilai uji serentak pada model terbaik dapat dilihat pada Tabel 4.9.

Tabel 4.9 Uji Serentak Regresi Logistik Multinomial Model Terbaik

Model	Likelihood Ratio Test			
	Log Likelihood	df	Chi-square	P-value
<i>Intercept Only</i>	-213,03			
<i>Final</i>	-267,28	2	108,5	$2,2 \times 10^{-16}$

Berdasarkan Tabel 4.9 nilai statistik uji serentak diperoleh sebesar 108,5. Nilai tersebut akan dibandingkan dengan *Chi-Square* tabel ($\chi^2_{(0,05,2)}$) sebesar 0,103 dimana nilai tersebut lebih kecil dari statistik uji yang didapat sehingga diputuskan Tolak H_0 . Hasil uji serentak juga bisa diperoleh dengan membandingkan *p-value* sebesar $2,2 \times 10^{-16}$ dimana nilai tersebut kurang dari 5% atau 0,05. Hal ini berarti paling sedikit terdapat satu variabel dependen yang secara statistik signifikan mempengaruhi variabel independen. Dikarenakan uji serentak signifikan, maka dapat dilanjutkan ke uji parsial.

Tabel 4.10 Uji Parsial Regresi Logistik Multinomial Model Terbaik

Variabel Respon	Variabel Prediktor	Std.Error	Estimasi	P-value
	<i>Intercept</i>	0,670	0,127	0,849
Lalu Lintas	Jenis Kelamin (1)	0,459	1,935	$2,512 \times 10^{-5}$
	Jabatan (1)	0,375	-2,211	$3,971 \times 10^{-9}$
	Umur	0,018	-0,015	0,415
	<i>Intercept</i>	0,872	-2,772	0,001
Luar Lokasi Kerja	Jenis Kelamin (1)	0,501	2,814	$1,979 \times 10^{-8}$
	Jabatan (1)	0,463	-1,341	$3,801 \times 10^{-3}$
	Umur	0,021	0,031	0,136

4.2.5 Model Regresi Logistik Multinomial

Setelah dilakukan uji serentak dan uji parsial model terbaik maka selanjutnya membuat persamaan model regresi logistik multinomial yang digunakan untuk pengklasifikasian baik data *training* maupun data *testing* yang selanjutnya dilakukan interpretasi terhadap model tersebut.

Berdasarkan Tabel 4.10 variabel yang signifikan terhadap terjadinya kecelakaan kerja adalah jenis kelamin (X_1), jabatan (X_2), dan umur (X_3). Oleh karena itu, persamaan *logit* untuk jenis kecelakaan lalu lintas $g_1(x)$ dan jenis kecelakaan luar lokasi kerja $g_2(x)$ adalah :

$$g_1(x) = 0,127 + 1,935 \text{ Jenis Kelamin}(1) - 2,211 \text{ Jabatan}(1) - 0,015 \text{ Umur}$$

$$g_2(x) = -2,772 + 2,814 \text{ Jenis Kelamin}(1) - 1,341 \text{ Jabatan}(1) + 0,031 \text{ Umur}$$

Sebagai contoh estimasi Y pada data pertama dimana terdapat peserta berjenis kelamin laki-laki ($X_1=0$), mempunyai jabatan pada produksi ($X_2=0$), dan umur ($X_3=22$) sehingga dapat dituliskan persamaan logit sebagai berikut :

$$g_1(x) = 0,127 + 1,935(0) - 2,211(0) - 0,015(22) = -0,203$$

$$g_2(x) = -2,772 + 2,814(0) - 1,341(0) + 0,031(22) = -2,09$$

dengan demikian, nilai $\pi_0(x)$, $\pi_1(x)$ dan $\pi_2(x)$ adalah :

$$\pi_0(x) = \frac{1}{1 + \exp(-0,203) + \exp(-2,09)} \approx 0,5154$$

$$\pi_1(x) = \frac{\exp(-0,203)}{1 + \exp(-0,203) + \exp(-2,09)} \approx 0,4208$$

$$\pi_2(x) = \frac{\exp(-2,09)}{1 + \exp(-0,203) + \exp(-2,09)} \approx 0,0638$$

Karena nilai $\pi_0(x)$ yang diperoleh merupakan nilai terbesar, maka Y pada data pertama adalah faktor 0 (Kecelakaan Dalam Lokasi Kerja). Perlu dicatat bahwa jumlah antara $\pi_0(x)$, $\pi_1(x)$, dan $\pi_2(x)$ adalah 1.

Selanjutnya dari variabel tersebut akan dilakukan interpretasi model menggunakan *odds ratio* yang terdapat pada Tabel 4.11.

Tabel 4.11 Nilai *Odds Ratio* Variabel

Variabel Respon	Variabel Prediktor	<i>Odds Ratio</i> Exp(B)
Lalu Lintas (1)	Jenis Kelamin (1)	6,929
	Jabatan (1)	0,109
	Umur	0,985
Luar Lokasi Kerja (2)	Jenis Kelamin (1)	16,680
	Jabatan (1)	0,261
	Umur	1,032

Nilai *odds ratio* pada jenis kecelakaan lalu lintas pada variabel jenis kelamin perempuan sebesar 6,929 yang artinya jenis kelamin perempuan memiliki peluang sebesar 6,929 kali lebih tinggi dari jenis kelamin laki-laki terhadap terjadinya kecelakaan kerja. Nilai *odds ratio* pada variabel jabatan *office* sebesar 0,109 yang artinya jabatan *office* memiliki peluang sebesar 0,109 kali lebih tinggi dari jabatan produksi terhadap terjadinya kecelakaan kerja. Dan nilai *odds ratio* pada variabel umur sebesar 0,985 yang artinya umur memiliki peluang sebesar 0,985 kali lebih tinggi terhadap terjadinya kecelakaan kerja. Pada jenis kecelakaan kerja diluar lokasi kerja nilai *odds ratio* jenis kelamin perempuan sebesar 16,680 yang artinya jenis kelamin perempuan memiliki peluang sebesar 16,680 kali lebih tinggi dari jenis kelamin laki-laki terhadap terjadinya kecelakaan kerja. Nilai *odds ratio* pada variabel jabatan *office* sebesar 0,261 yang artinya jabatan *office* memiliki peluang sebesar 0,261 kali lebih tinggi dari jabatan produksi terhadap terjadinya kecelakaan kerja. Dan nilai *odds ratio* pada variabel umur sebesar 1,032 yang artinya umur memiliki peluang sebesar 1,032 kali lebih tinggi terhadap terjadinya kecelakaan kerja. Berdasarkan faktor-faktor yang mempengaruhi korban kecelakaan kerja yang telah dijelaskan diatas maka dapat dijadikan sebagai masukan atau peringatan khususnya bagi para pekerja yang memiliki asuransi Jaminan Kecelakaan Kerja (JKK) pada BPJS Ketenagakerjaan wilayah Mojokerto.

4.2.6 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui model regresi logistik multinomial yang terbentuk pada jenis kecelakaan kerja sudah sesuai atau belum yang akan digunakan dalam analisis pengklasifikasian jenis kecelakaan kerja. Uji kesesuaian model ditampilkan pada Tabel 4.12.

Tabel 4.12 *Goodness of Fit* Jenis Kecelakaan Kerja

	<i>Chi-Square</i>	<i>df</i>	<i>P-value</i>
<i>Pearson</i>	46,017	4	$3,001 \times 10^{-15}$

Model regresi logistik multinomial yang telah terbentuk tidak sesuai yang artinya terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model. Hal ini ditunjukkan dengan *p-value* sebesar $3,001 \times 10^{-15}$ yang nilainya lebih kecil dari 0,05. Artinya, bahwa model regresi logistik dari data *training* tidak dapat digunakan dalam analisis pengklasifikasian jenis kecelakaan kerja.

4.2.7 Klasifikasi Jenis Kecelakaan Kerja

Klasifikasi jenis kecelakaan kerja digunakan untuk mengetahui ketepatan model yang telah terbentuk dalam mengestimasi parameter model menggunakan data *training* dan untuk mengetahui ketepatan dalam mengklasifikasikan data baru menggunakan data *testing*. Berikut klasifikasi jenis kecelakaan kerja.

Tabel 4.13 Klasifikasi Jenis Kecelakaan Kerja

Observasi	Prediksi			Total	1-APER (%)	
	0	1	2			
Data <i>Training</i>	0	171	40	16	227	74,03
	1	1	13	6	20	
	2	4	0	7	11	
Data <i>Testing</i>	0	42	11	6	59	70,31
	1	2	2	0	4	
	2	0	0	1	1	

Ketepatan klasifikasi untuk data *training* sebesar 74,03% sedangkan ketepatan klasifikasi untuk data *testing* sebesar 70,31% dapat dilihat pada Tabel 4.14. Model regresi logistik multinomial untuk jenis kecelakaan kerja mempunyai ketepatan 74,03% dalam mengestimasi parameter model sedangkan model regresi logistik multinomial untuk jenis kecelakaan kerja mempunyai ketepatan 70,31% dalam mengklasifikasikan data baru.

4.3 Analisis *Random Forest* untuk Klasifikasi Jenis Kecelakaan Kerja yang Ditangani oleh BPJS Ketenagakerjaan Wilayah Mojokerto

Random forest merupakan suatu metode *ensemble* yang dapat meningkatkan akurasi klasifikasi dari sebuah pemilah tunggal yang tidak stabil dengan cara mengkombinasikan banyak pemilah dari suatu metode yang sama (CART) melalui proses *voting*. Jadi, hasil klasifikasi *random forest* berasal dari pengulangan metode CART, sehingga perlu dilakukan analisis menggunakan CART terlebih dahulu.

4.3.1 Analisis CART untuk Klasifikasi Jenis Kecelakaan Kerja

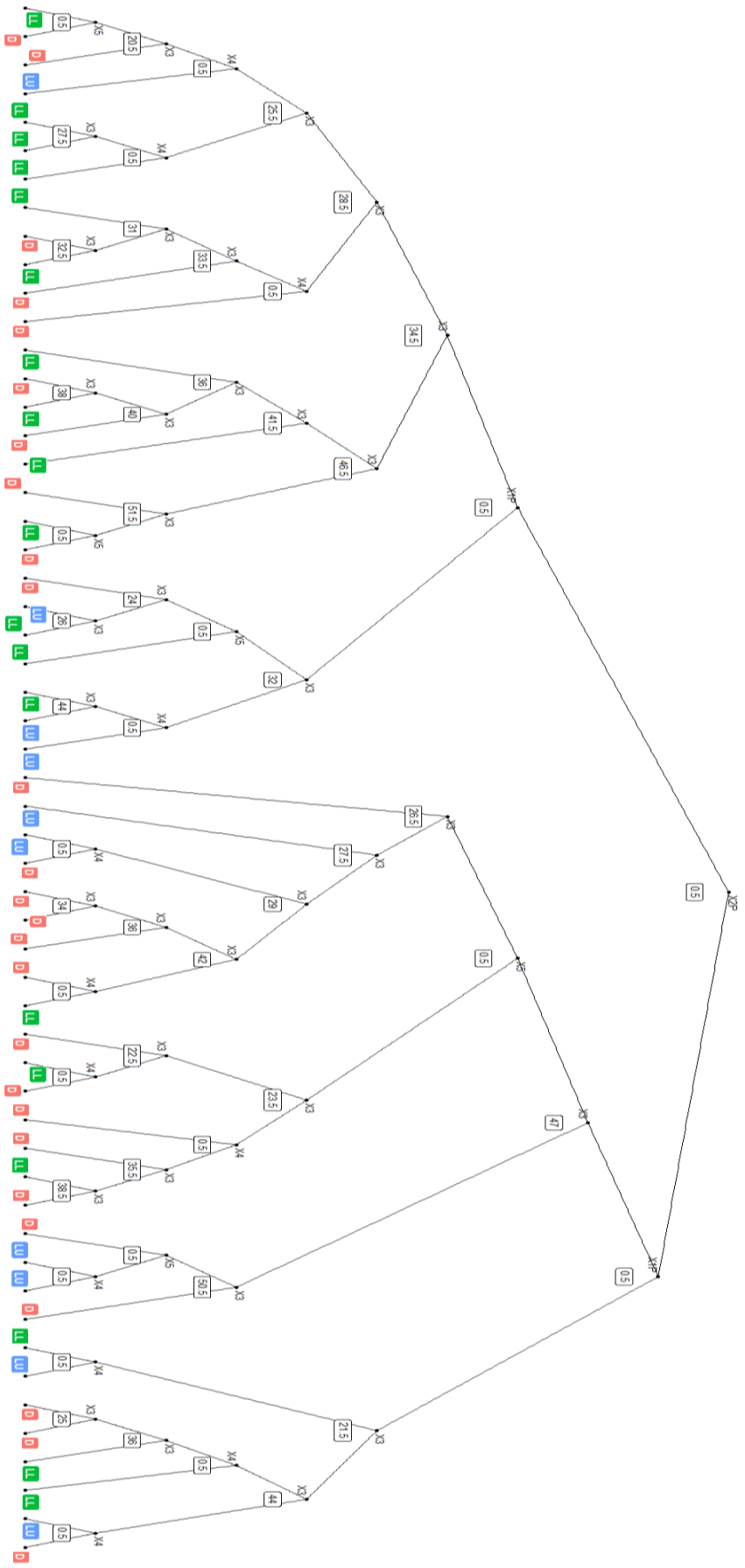
Pada analisis CART data sebanyak 320 korban kecelakaan kerja menggunakan 10-*Fold Cross Validation* dengan kombinasi data *training* dan data *testing* yang dicobakan dalam penelitian ini adalah 75%:25%, 80%:20%, 85%:15%, 90%:10%. Analisis ini diawali dengan langkah pembentukan pohon klasifikasi yang maksimal. Berikut serangkaian tahapan analisis CART.

a. Pembentukan Pohon Klasifikasi Maksimal

Tahapan awal yang dilakukan untuk membentuk pohon klasifikasi adalah dengan menentukan variabel pemilah dan nilai variabel (*threshold*). Variabel pemilah dan *threshold* dipilih dari beberapa kemungkinan pemilah dari masing-masing variabel. Kemungkinan banyaknya pemilah untuk variabel prediktor kategorik menggunakan banyak kategori dari variabel tersebut. Perhitungan banyaknya kemungkinan pemilah ditampilkan pada Tabel 4.14.

Tabel 4.14 Perhitungan Kemungkinan Jumlah Pemilah dari Setiap Variabel

Variabel	Nama Variabel	Skala Data	Banyaknya Kategori	Kemungkinan Pemilah
X_1	Jenis Kelamin	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_2	Jabatan	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_3	Umur	Rasio	37	$37 - 1 = 36$ pemilah
X_4	Jam Kerja	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_5	Waktu Kejadian Kecelakaan	Nominal	2	$2^{2-1} - 1 = 1$ pemilah



Gambar 4.3 Pohon Klasifikasi maksimal untuk Jenis Kecelakaan Kerja

b. Pemangkasan Pohon Klasifikasi Maksimal (*Prunning*)

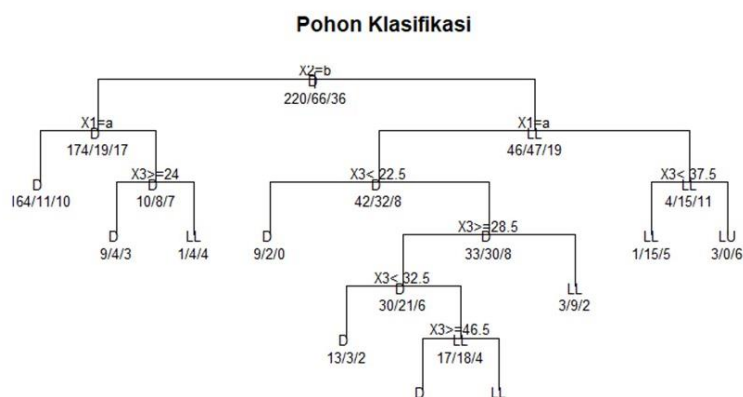
Untuk mempermudah proses analisis, pohon klasifikasi maksimal yang dihasilkan kemudian dilakukan pemangkasan secara iteratif. Setiap hasil pemangkasan memiliki nilai *relative error* tertentu, sehingga kemudian dipilih hasil pemangkasan dengan nilai *relative error* yang minimum yang dapat dilihat pada Tabel 4.15.

Tabel 4.15 Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Terminal Nodes</i>	<i>Relative Error</i>	<i>xerror</i>	<i>xstd</i>
0	1,000	1,000	0,081
3	0,833	1,009	0,082
7	0,715	0,941	0,080
9	0,686	0,921	0,079

c. Pemilihan Pohon Klasifikasi Optimal

Hasil pemangkasan yang diperoleh dari Tabel 4.15 selanjutnya digunakan untuk memilih pohon klasifikasi yang optimal dengan jumlah *terminal nodes* 9 dan nilai *relative error* sebesar 0,686. Setiap simpul pada pohon klasifikasi yang terbentuk telah dilakukan pelabelan kelas menjadi tiga yaitu “D” untuk kecelakaan yang terjadi di dalam lokasi kerja, “LL” untuk kecelakaan yang terjadi di lalu lintas, dan “LU” untuk kecelakaan yang terjadi diluar lokasi pabrik yang ditunjukkan pada Gambar 4.4.



Gambar 4.4 Konstruksi Pohon Klasifikasi Optimal Jenis Kecelakaan Kerja

Pohon klasifikasi optimal dipengaruhi oleh semua variabel prediktor. Besarnya kontribusi variabel sebagai pemilah baik pemilah utama maupun pengganti pada pohon klasifikasi optimal yang terbentuk ditunjukkan melalui suatu angka skor yang ditampilkan pada Tabel 4.16.

Tabel 4.16 Skor Variabel Terpenting dari Pohon Klasifikasi Optimal

Variabel	Skor Variabel
X_2	52,03
X_1	51,18
X_3	38,90
X_4	19,47
X_5	3,51

Tabel 4.16 menunjukkan bahwa semua variabel menjadi pembangun pohon klasifikasi. Akan tetapi, berdasarkan skor yang dihasilkan diketahui bahwa variabel yang terpenting

dalam mengklasifikasikan jenis kecelakaan kerja adalah jabatan (X_2) karena memiliki skor paling tinggi yaitu sebesar 52,03. Selain itu, terdapat beberapa variabel yang juga berpengaruh dalam melakukan pemilihan yaitu jenis kelamin sebesar 51,18 begitu juga seterusnya sampai waktu kejadian kecelakaan (X_5).

4.3.2 Analisis *Random Forest* untuk Klasifikasi Jenis Kecelakaan Kerja

CART merupakan suatu metode pengklasifikasian yang mampu bekerja untuk data yang berdimensi tinggi. Metode ini masih memiliki kelemahan dimana apabila terjadi perubahan pada data *training* maka pohon yang dihasilkan akan memiliki perbedaan yang cukup besar dan pohon tidak stabil sehingga perlu dicobakan suatu solusi yang mampu mengatasi keterbatasan tersebut. Salah satunya dapat menggunakan metode *ensemble* yaitu *random forest*. *Random forest* dapat meningkatkan akurasi prediksi dari pohon klasifikasi yang dihasilkan melalui proses pembentukan pohon dan prediksi perhitungan komputasi yang cepat. Algoritma *random forest* menggunakan *10-Fold Cross Validation* dengan kombinasi data *training* dan data *testing* yang dicobakan dalam penelitian ini adalah 75%:25%, 80%:20%, 85%:15%, 90%:10%. Dimana *subset* yang terpilih yaitu *subset* ke-10 yang dapat dilihat pada Lampiran 7. Kombinasi proporsi data *training* dan data *testing* terbaik dipilih berdasarkan nilai ketepatan klasifikasi yang paling besar.

Dalam penelitian ini parameter kontrol yang ditentukan adalah variabel prediktor yang diambil secara acak dari lima variabel amatan yang digunakan sebagai pembentuk pohon dalam setiap pemilihan adalah tiga variabel yang diperoleh dari hasil akar banyaknya variabel prediktor yaitu 2,58 variabel yang dapat dibulatkan menjadi tiga variabel atau $\log_2 8 = 3$ dengan jumlah pohon yang dibentuk akan dicobakan pada kombinasi yaitu 100,500, dan 1000 pohon (Rahmawati, 2015).

a. Pemilihan Presentase Kombinasi Data *Training* dan Data *Testing* Terbaik dalam Analisis *Random Forest*

Pemilihan presentase kombinasi data *training* dan data *testing* terbaik didasarkan pada nilai rata-rata *total accuracy rate*. Rata-rata diperoleh dari nilai (1-APER) masing-masing presentase kombinasi yang ditunjukkan pada Tabel 4.17.

Tabel 4.17 Perbandingan *Total Accuracy Rate* Kombinasi Data

Kombinasi Data <i>Training</i> dan Data <i>Testing</i>	<i>Total Accuracy Rate (1-APER)</i>					
	K=100		K=500		K=1000	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
75%:25%	91,75	77,41	89,69	74,19	89,34	70,96
80%:20%	88,37	76,56	87,98	78,12	83,33	76,56
85%:15%	88,36	76,59	87,63	78,72	91,63	68,08
90%:10%	89,66	76,25	89,6	73,75	92,56	73,75

Tabel 4.18 Ketepatan Klasifikasi *Random Forest* pada Kombinasi 500

Observasi	Prediksi			Total	1-APER (%)
	0	1	2		
Data <i>Training</i>	0	174	13	9	196
	1	2	39	6	47
	2	0	1	14	15
Data <i>Testing</i>	0	41	6	2	49
	1	3	7	3	13
	2	0	0	2	2

Tabel 4.17 menunjukkan bahwa kombinasi data *training* dan data *testing* yang mampu memberikan ketepatan klasifikasi yang tinggi adalah pada kombinasi data *training* sebesar 85% dan data *testing* sebesar 15%. Ketepatan klasifikasi yang dipilih dari beberapa kombinasi data *training* dan data *testing* adalah ketepatan klasifikasi pada data *training* dan data *testing* yang paling tinggi. Apabila pada data *training* memberikan nilai ketepatan klasifikasi yang paling tinggi akan tetapi pada data *testing* tidak memberikan nilai ketepatan klasifikasi yang paling tinggi begitu juga sebaliknya sehingga yang pilih adalah kombinasi data *training* dan data *testing* yang memberikan nilai ketepatan klasifikasi tertinggi pada ketepatan akurasi data *testing*.

Ketepatan klasifikasi untuk data *testing* yang diperoleh adalah sebesar 78,72% yang merupakan nilai tertinggi diantara semua kombinasi data. Untuk ketepatan klasifikasi data *training* diperoleh ketepatan klasifikasi sebesar 87,63%. Oleh karena itu, hasil analisis yang digunakan untuk menjelaskan jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto adalah hasil klasifikasi yang menggunakan kombinasi data *training* dan data *testing* terbaik, yaitu data *training* sebesar 85% dan data *testing* sebesar 15%.

4.4. Perbandingan Hasil Klasifikasi Regresi Logistik Multinomial dan Random Forest

Metode yang digunakan dalam analisis jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto adalah regresi logistik multinomial dan *random forest*. Kriteria yang digunakan untuk membandingkan kedua metode tersebut adalah *total accurate rate* (1-APER) dari masing-masing data *training* dan *testing*. Metode klasifikasi yang terbaik dipilih dari nilai *total accurate rate* (1-APER) terbesar.

Tabel 4.19 Perbandingan Hasil Klasifikasi Regresi Logistik Multinomial dan *Random Forest*

Metode Klasifikasi	Data	1-APER (%)
Regresi Logistik Multinomial	<i>Training</i>	74,03
	<i>Testing</i>	70,31
<i>Random Forest</i>	<i>Training</i>	87,63
	<i>Testing</i>	78,72

Tabel 4.19 menunjukkan bahwa metode klasifikasi yang mempunyai nilai 1-APER terbesar untuk data *training* dan data *testing* adalah metode *random forest*. Hal ini berarti dalam penelitian kali ini, metode yang terbaik adalah *random forest* dengan presentase nilai ketepatan klasifikasi data *training* sebesar 87,63% dan data *testing* sebesar 78,72%.

(Halaman ini sengaja dikosongkan)

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Berdasarkan hasil deskripsi mengenai jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto diketahui bahwa kecelakaan yang paling banyak terjadi adalah di dalam lokasi pabrik dengan presentase sebesar 68,32% yang berjenis kelamin laki-laki dengan kisaran usia kurang lebih 24 tahun.
2. Klasifikasi jenis kecelakaan kerja menggunakan regresi logistik multinomial dengan menggunakan *10-Fold Cross Validation* kombinasi data *training* dan data *testing* sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10% dimana ketepatan klasifikasi yang terpilih berdasarkan nilai 1-APER terbesar adalah kombinasi 75%:25% dengan ketepatan untuk data *training* sebesar 74,03% dan data *testing* sebesar 70,31%. Variabel yang berpengaruh secara signifikan terhadap jenis kecelakaan kerja adalah jenis kelamin dan jabatan. Pada analisis klasifikasi jenis kecelakaan kerja menggunakan *random forest* dengan kombinasi data *training* dan data *testing* sebesar 75%:25%, 80%:20%, 85%:15%, 90%:10% dimana ketepatan klasifikasi yang terpilih berdasarkan nilai 1-APER terbesar adalah kombinasi 85%:15% dengan ketepatan untuk data *training* pada kombinasi pohon 500 sebesar 87,63% dan data *testing* pada kombinasi pohon 500 sebesar 78,72%.
3. Perbandingan kedua metode regresi logistik multinomial dan *random forest* dalam pengklasifikasian jenis kecelakaan kerja yang ditangani oleh BPJS Ketenagakerjaan wilayah Mojokerto memberikan hasil ketepatan klasifikasi yang berbeda baik untuk data *training* maupun data *testing*. Ketepatan klasifikasi untuk data *training* maupun data *testing* terbesar menggunakan metode *random forest* yaitu sebesar 87,63% untuk data *training* dan 78,72% untuk data *testing*.

5.2 Saran

Saran yang dapat diberikan dari hasil penelitian ini adalah sebagai berikut.

1. Untuk dijadikan sebagai rekomendasi kepada peneliti selanjutnya yaitu penambahan faktor-faktor yang mempengaruhi jenis kecelakaan kerja serta penambahan data untuk mendapatkan hasil klasifikasi yang lebih informatif.
2. Untuk mengantisipasi terjadinya kecelakaan kerja perlu dilakukan sosialisasi dan pengarahan keselamatan kerja melalui *safety meeting* dari pihak perusahaan kepada para pekerja serta melakukan pembinaan K3 untuk memperhatikan tentang pentingnya keselamatan dan kesehatan kerja.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- Astri, Silvia Rahmaningrum. (2019). Klasifikasi Sentimen Terhadap Review Layanan Hotel Bintang Tiga di Surabaya pada Situs Traveloka Menggunakan *Naïve Bayes Classifier* (NBC) dan Regresi Logistik Biner. Eprints.umm.ac.id. 2021. [online] Available at: <<https://eprints.umm.ac.id/44035/3/BAB%20II.pdf>> [Accessed 12 January 2022].
- Fre-web-08.novadine.com. 2022. *Loading*. [online] Available at: https://fre-web-08.novadine.com/w/content/Z3Q2C5/an-introduction-to-categorical-data-analysis-alan-agresti-solution-manual_pdf > [Accessed 12 January 2022].
- Id.wikipedia.org. 2021. *BPJS Ketenagakerjaan - Wikipedia bahasa Indonesia, ensiklopedia bebas*. [online] Available at: <https://id.wikipedia.org/wiki/BPJS_Ketenagakerjaan> [Accessed 5 January 2022].
- Medium. 2021. *Analisis Regresi Logistik Multinomial dengan SPSS*. [online] Available at: <<https://medium.com/@wahyudhizainal/analisis-regresi-logistik-multinomial-dengan-spss-cb65fb246b69>> [Accessed 5 January 2022].
- Pradipta, Aldo. (2021). Pendekatan *Multinomial Logistic Regression-Bsyesian Network Hybrid* untuk Analisis Produk Asuransi pada Perusahaan PT AJ Central Asia Raya.
- Pratiwi, F. E & Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Trees) di Provinsi Sulawesi Utara, Jurnal Sains dan Seni POMITS, Vol.3, No.1, 2337-3520.
- Repository.its.ac.id.2021. [online] Available at:<<https://repository.its.ac.id/62783/1/1312100134-Undergradute%20Thesis.pdf>> [Accessed 12 January 2022].
- Repository.its.ac.id. 2022. [online] Available at: <<https://repository.its.ac.id/555/1/1313030042-Presentation.pdf>> [Accessed 12 January 2022].
- Research-indonesia.blogspot.com. 2021. *REGRESI MULTINOMIAL DENGAN SPSS*. [online] Available at: <<https://research-indonesia.blogspot.com/2016/06/regresi-multinomial-dengan-spss.html>> [Accessed 10 January 2022].
- Repository.its.ac.id. 2022. *Klasifikasi Faktor-Faktor Yang Mempengaruhi Korban Kecelakaan Lalu Lintas Di Surabaya Dengan Pendekatan Regresi Logistik Multinomial Dan Random Forests - ITS Repository*. [online] Available at: <<https://repository.its.ac.id/51951/>> [Accessed 12 January 2022].
- Repository.its.ac.id. 2022. *Perbandingan Metode Random Forest Classification Dan Support Vector Machine Untuk Deteksi Epilepsi Menggunakan Data Rekaman Electroen Cephalograph (EGG) - ITS Repository*. [online] Available at: <<https://repository.its.ac.id/72224/>> [Accessed 12 January 2022].
- Statslab-rshiny.fmipa.unej.ac.id. 2021. [online] Available at: <<http://statslab-rshiny.fmipa.unej.ac.id/JORS/logmult/>> [Accessed 24 January 2022].
2022. [online] Available at: <<https://www.neliti.com/publications/14564/faktor-faktor-yang-mempengaruhi-potensi-kecelakaan-kerja-pada-pengemudi-truk-di-p>> [Accessed 24 January 2022].

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1 *Library* yang digunakan

```
library(quantmod)
library(xts)
library(portfolio.optimization)
library(PerformanceAnalytics)
library(fitdistrplus)
library(MASS)
library(GLDEX)
library(forecast)
library(ggplot2)
library(ggpubr)
library(ggtext)
library(caret)
library(xlsx)
library(crosstable)
library(rpart)
```

Lampiran 2 Syntax untuk *Preprocessing* Data, Statistika Deskriptif, *Crosstabulation*, dan Partisi Data

```
#Preprocessing Data
df <- read.xlsx("D:\\LAPORAN BPJS.xlsx",sheetIndex = 2)
df <- df[,-1]
colnames(df) <- c("Y","X1","X2","X3","X4","X5")
head(df)

df$Y <- as.factor(df$Y)
df$X1 <- as.factor(df$X1)
df$X2 <- as.factor(df$X2)

levels(df$Y);levels(df$X1);levels(df$X2)

#Statistika Deskriptif
x <- c("D","LL","LU")
y <- c(sum(df$Y=="D"),sum(df$Y=="LL"),sum(df$Y=="LU"))
daf <- data.frame(y,x)

ggplot(daf, aes(x = "", y = y, fill = x)) +
  geom_col(color = "black") +
  geom_text(aes(label = paste(round(y/sum(y)*100,2,"%")),
    position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_discrete(name="Jenis Kecelakaan",labels = c("Dalam", "Lalu Lintas",
"Luar"))+
  theme(axis.text = element_blank(),axis.ticks = element_blank(),
    panel.grid = element_blank())

x <- c("L","P")
y <- c(sum(df$X1=="L"),sum(df$X1=="P"))
daf <- data.frame(y,x);daf

ggplot(daf, aes(x = "", y = y, fill = x)) +
  geom_col(color = "black") +
  geom_text(aes(label = paste(round(y/sum(y)*100,2,"%")),
    position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_discrete(name="Jenis Kelamin",labels = c("Laki-laki", "Perempuan"))+
  theme(axis.text = element_blank(),axis.ticks = element_blank(),
    panel.grid = element_blank())

x <- c("P","O")
y <- c(sum(df$X2=="P"),sum(df$X2=="O"))
daf <- data.frame(y,x)

ggplot(daf, aes(x = "", y = y, fill = x)) +
  geom_col(color = "black") +
```

Lampiran 2 (Lanjutan 1)

```
geom_text(aes(label = paste(round(y/sum(y)*100,2),"%")),
          position = position_stack(vjust = 0.5)) +
coord_polar(theta = "y") +
scale_fill_discrete(name="Jabatan",labels = c("Office", "Produksi"))+
theme(axis.text = element_blank(),axis.ticks = element_blank(),
       panel.grid = element_blank())

summary(df$X3)
daf <- data.frame(df$X3);daf
coll <- c(rep("cornflowerblue",9),"red",rep("cornflowerblue",18),"yellow",rep("cornflowerblue",7),
          "yellow")
ggplot(daf, aes(x = df.X3,label=..count..)) +
  theme_bw() +
  geom_histogram(binwidth=1,colour = "white", fill = coll, size = 1) +
  scale_x_continuous(name = "Usia")+
  scale_y_continuous(name = "Frekuensi")+
  stat_bin(binwidth=1,
           aes(label=ifelse(..count..==23,"23",ifelse(..count..==1,"1","")), vjust=-0)
  geom="text",

x <- c("P","O")
y <- c(sum(df$X4==0),sum(df$X4==1))
daf <- data.frame(y,x);daf

ggplot(daf, aes(x = "", y = y, fill = x)) +
  geom_col(color = "black") +
  geom_text(aes(label = paste(round(y/sum(y)*100,2),"%")),
           position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_discrete(name="Jam Kerja",labels = c("<= 8 jam", "> 8 Jam"))+
  theme(axis.text = element_blank(),axis.ticks = element_blank(),
       panel.grid = element_blank())

x <- c("P","O")
y <- c(sum(df$X5==0),sum(df$X5==1))
daf <- data.frame(y,x);daf
ggplot(daf, aes(x = "", y = y, fill = x)) +
  geom_col(color = "black") +
  geom_text(aes(label = paste(round(y/sum(y)*100,2),"%")),
           position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_discrete(name="Waktu Kejadian Kecelakaan",labels = c("Gelap", "Terang"))+
  theme(axis.text = element_blank(),axis.ticks = element_blank(),

       panel.grid = element_blank())
```

Lampiran 2 (Lanjutan 2)

```
#Crosstabulation
crosstable(df,cols = "Y",by = "X1")
crosstable(df,cols = "Y",by = "X2")
crosstable(df,cols = "Y",by = "X4")
crosstable(df,cols = "Y",by = "X5")

#Partisi Data
set.seed(99)
idx90 = createDataPartition(df$Y, p = .90,list = FALSE)
trn90 = df[idx90,]
tst90 = df[-idx90, ]
idx85 = createDataPartition(df$Y, p = .85,list = FALSE)
trn85 = df[idx85,]
tst85 = df[-idx85, ]
idx80 = createDataPartition(df$Y, p = .80,list = FALSE)
trn80 = df[idx80,]
tst80 = df[-idx80, ]
idx75 = createDataPartition(df$Y, p = .75,list = FALSE)
trn75 = df[idx75,]
tst75 = df[-idx75, ]
ctrl <- trainControl(method = "cv", number = 10)
```

Lampiran 3 *Syntax* untuk Metode Regresi Logistik Multinomial

```
#Pemodelan Multinomial
#training 90:10
sim_multinom_modeltr90 <- train(Y~., data = trn90, method = "multinom", trControl =
ctrl)
multinom_model90 <- sim_multinom_modeltr90$finalModel
acc90 <- confusionMatrix(predict(sim_multinom_modeltr90,trn90),trn90$Y)
acc90ts <- confusionMatrix(predict(sim_multinom_modeltr90,tst90),tst90$Y)
#training 85:15
sim_multinom_modeltr85 <- train(Y~., data = trn85, method = "multinom", trControl =
ctrl)
multinom_model85 <- sim_multinom_modeltr85$finalModel
acc85 <- confusionMatrix(predict(sim_multinom_modeltr85,trn85),trn85$Y)
acc85ts <- confusionMatrix(predict(sim_multinom_modeltr85,tst85),tst85$Y)
#training 80:20
sim_multinom_modeltr80 <- train(Y~., data = trn80, method = "multinom", trControl =
ctrl)
multinom_model80 <- sim_multinom_modeltr80$finalModel
acc80 <- confusionMatrix(predict(sim_multinom_modeltr80,trn80),trn80$Y)
acc80ts <- confusionMatrix(predict(sim_multinom_modeltr80,tst80),tst80$Y)
#training 75:25
sim_multinom_modeltr75 <- train(Y~., data = trn75, method = "multinom", trControl =
ctrl)
multinom_model75 <- sim_multinom_modeltr75$finalModel
acc75 <- confusionMatrix(predict(sim_multinom_modeltr75,trn75),trn75$Y)
acc75ts <- confusionMatrix(predict(sim_multinom_modeltr75,tst75),tst75$Y)

#Memilih Model Multinomial Terbaik
name <- c("75:25","80:20","85:15","90:10")
acc.tr <- c(acc75$overall[1],acc80$overall[1],acc85$overall[1],acc90$overall[1])
acc.ts <-
c(acc75ts$overall[1],acc80ts$overall[1],acc85ts$overall[1],acc90ts$overall[1])
acc <- data.frame(name,acc.tr,acc.ts)
acc #kombinasi 75:25 yang dipilih

acc.resampl <- sim_multinom_modeltr75$resample;acc.resampl

#Pengujian Asumsi Model Multinomial
sim_multinom_model <- train(Y~X1+X2+X3, data = trn75, method = "multinom",
trControl = ctrl)
multinom_model <- sim_multinom_model$finalModel
summary(multinom_model)

#Odds Ratio
odds <- summary(multinom_model)
odds <- odds$wts[8:length(odds$wts)];odds <- exp(odds[c(-5,-4)]);odds
```

Lampiran 3 (Lanjutan 1)

```
dff <- df[,1:6]
mm <- multinom(Y~X1+X2+X3, data = dff)
lrtest(mm,c("X1","X2","X3"))

#Uji serentak
Anova(multinom.model)

#Uji Parsial
z
summary(multinom.model)$coefficients/summary(multinom.model)$standard.errors <-
pval <- (1 - pnorm(abs(z), 0, 1)) * 2
pval

#Uji kesesuaian model
chisq.test(predict(sim_multinom_modeltr75,trn75),trn75$Y)

#Klasifikasi
#train
cof.mat.tr
confusionMatrix(predict(sim_multinom_modeltr75,trn75),trn75$Y);cof.mat.tr <-
cof.mat.tr$table

#test
cof.mat.ts
confusionMatrix(predict(sim_multinom_modeltr75,tst75),tst75$Y);cof.mat.ts <-
cof.mat.ts$table
```


Lampiran 4 Syntax untuk Metode *Random Forest*

```
#Membuat Pohon Klasifikasi
#Banyaknya kemungkinan usia
length(levels(as.factor(df$X3)))

# grow tree
fit <- rpart(Y ~ .,method="class", data=df)
fit$stable

# plot tree
plot(fit, uniform=TRUE,
     main="Pohon Klasifikasi")
text(fit, use.n=TRUE, all=TRUE, cex=0.8)

#Pemodelan Random Forest
sim_randomForest_model100 <- train(Y~., data = df, method = "rf", trControl = ctrl,
                                   mytry=3, ntree=100,set.seed(99),maxnodes=9)

acc.rf.tr <- matrix(rep(0,12),ncol = 3)
acc.rf.ts <- matrix(rep(0,12),ncol = 3)
in.p <- c(0.1,0.15,0.2,0.25)
in.t <- c(100,500,1000)
max.acc <- 0
for (i in 1:4) {
  idx = createDataPartition(df$Y, p = 1-in.p[i],list = FALSE)
  trn = df[idx,]
  tst = df[-idx,]
  for (j in 1:3) {
    sim_randomForest_modeltr <- train(Y~., data = trn, method = "rf",
                                       trControl = ctrl,
                                       mytry=3, ntree=in.t[j],set.seed(1000))
    acc.rf1 <- confusionMatrix(predict(sim_randomForest_modeltr,trn),trn$Y)
    acc.rf2 <- confusionMatrix(predict(sim_randomForest_modeltr,tst),tst$Y)
    acc.rf.tr[i,j] <- acc.rf1$overall[1]
    acc.rf.ts[i,j] <- acc.rf2$overall[1]
    if (acc.rf.ts[i,j]>max.acc) {
      tabb1 <- acc.rf1$table
      tabb2 <- acc.rf2$table
    }
    max.acc <- acc.rf.ts[i,j]
  }
}
acc.rf.tr
acc.rf.ts
```

Lampiran 4 (Lanjutan 1)

```
sim_randomForest_modeltr <- train(Y~., data = trn75, method = "rf",
                                trControl = ctrl,
                                mytry=3, ntree=100,set.seed(1000),importance=TRUE)
sim_randomForest_modeltr$resample
importance(sim_randomForest_modeltr$finalModel,type = 1)
confusionMatrix(predict(sim_randomForest_modeltr,trn75),trn75$Y)
#Viasualisasi Hasil Klasifikasi Random Forest
tree_func <- function(final_model,
                      tree_num) {

# get tree by index
tree <- randomForest::getTree(final_model,
                              k = tree_num,
                              labelVar = TRUE) %>%
  tibble::rownames_to_column() %>%
  # make leaf split points to NA, so the 0s won't get plotted
  mutate(`split point` = ifelse(is.na(prediction), `split point`, NA))

# prepare data frame for graph
graph_frame <- data.frame(from = rep(tree$rowname, 2),
                          to = c(tree$`left daughter`, tree$`right daughter`))

# convert to graph and delete the last node that we don't want to plot
graph <- graph_from_data_frame(graph_frame) %>%
  delete_vertices("0")

# set node labels
V(graph)$node_label <- gsub("_", " ", as.character(tree$`split var`))
V(graph)$leaf_label <- as.character(tree$prediction)
V(graph)$split <- as.character(round(tree$`split point`, digits = 5))

# plot
plot <- ggraph(graph, 'dendrogram') +
  theme_bw() +
  geom_edge_link() +
  geom_node_point() +
  geom_node_text(aes(label = node_label), na.rm = TRUE, repel = TRUE) +
  geom_node_label(aes(label = split), vjust = 2.5, na.rm = TRUE, fill = "white") +
  geom_node_label(aes(label = leaf_label, fill = leaf_label), na.rm = TRUE,
repel = TRUE, colour = "white", fontface = "bold", show.legend = FALSE) +
```

Lampiran 4 (Lanjutan 2)

```
theme(panel.grid.minor = element_blank(),
      panel.grid.major = element_blank(),
      panel.background = element_blank(),
      plot.background = element_rect(fill = "white"),
      panel.border = element_blank(),
      axis.line = element_blank(),
      axis.text.x = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks = element_blank(),
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      plot.title = element_text(size = 18))
print(plot)
}
tree_func(sim_randomForest_modeltr$finalModel,100)
```

Lampiran 5 Data Penelitian

Y	X1	X2	X3	X4	X5
0	0	0	22	0	0
1	0	0	31	0	0
1	0	0	41	0	0
1	0	0	27	0	0
2	1	1	25	0	0
1	0	0	54	0	1
1	0	0	27	0	1
1	0	1	28	0	1
1	0	1	32	1	0
1	0	1	49	1	0
1	0	0	35	0	0
1	0	0	32	0	1
2	0	1	52	0	0
1	0	0	40	1	0
1	0	0	53	0	1
1	0	0	24	0	0
2	1	0	50	0	0
1	0	0	39	0	0
2	0	0	48	0	1
1	0	0	24	1	0
1	1	0	51	1	0
0	0	1	42	1	1
0	1	0	38	0	1
1	0	1	32	0	0
1	0	0	39	0	0
1	0	0	30	0	1
0	0	0	23	0	1
1	0	0	28	1	1
1	0	0	39	1	1
1	0	0	30	1	1
1	0	0	32	0	1
2	0	1	24	1	1
1	0	0	26	0	1
2	0	0	27	1	0
0	0	1	35	0	0
0	0	1	45	0	1
0	1	1	28	0	1
0	0	1	44	0	1
2	0	0	28	0	0
0	0	0	38	0	0

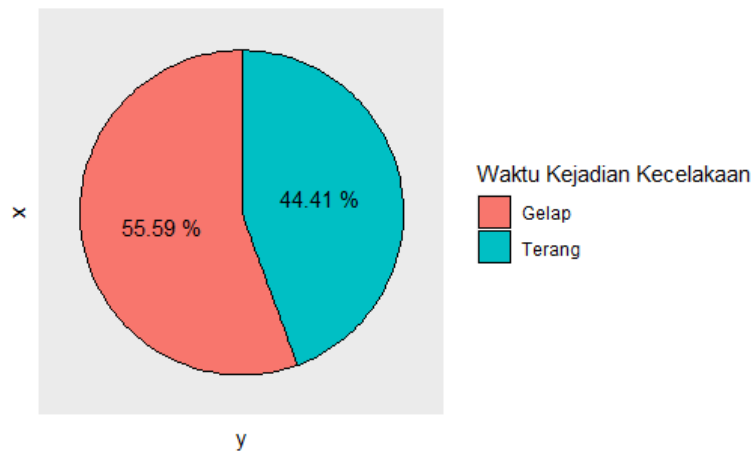
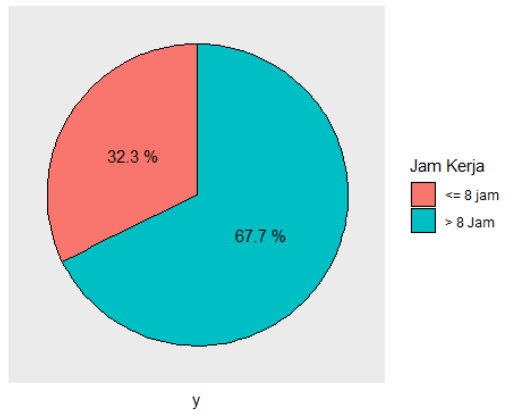
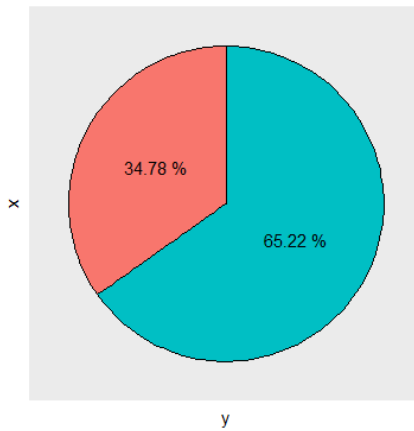
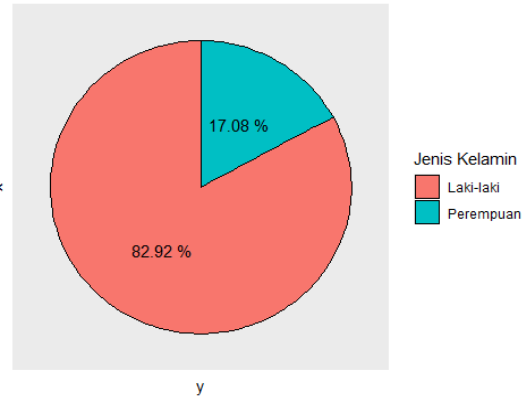
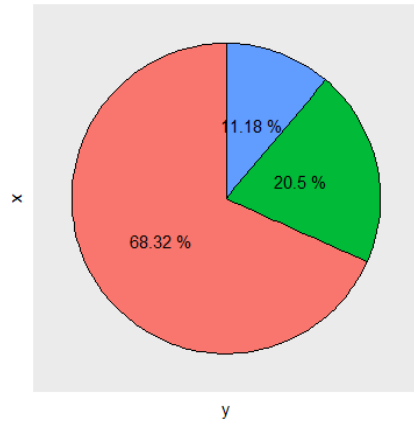
Lampiran 5 (Lanjutan 1)

Y	X1	X2	X3	X4	X5
1	0	0	24	0	1
1	0	0	30	0	0
2	1	1	52	0	0
1	0	0	33	0	0
1	0	0	38	0	0
1	0	0	53	1	0
1	0	0	22	0	0
1	0	0	47	0	0
0	0	0	43	1	0
1	0	0	26	0	1
1	1	0	23	0	1
1	0	0	19	0	1
1	0	0	24	0	0
2	1	0	20	1	1
2	1	0	20	1	0
1	0	0	21	1	1
1	0	0	48	1	1
1	0	0	20	0	1
1	0	0	23	0	1
1	1	0	25	0	0
1	1	0	34	0	0
1	0	1	34	0	0
1	0	0	22	0	1
1	0	0	48	0	0
1	0	0	21	0	1
1	0	0	51	0	1
1	0	0	25	0	1
1	0	0	26	0	0
0	1	0	25	1	0
0	0	0	35	0	0
0	1	0	20	0	1
1	0	0	19	0	1
0	1	1	25	1	0
1	0	0	29	1	0
0	0	1	20	1	0
1	0	1	37	0	1
1	0	0	46	0	1
1	0	0	24	0	1
1	0	0	23	0	1

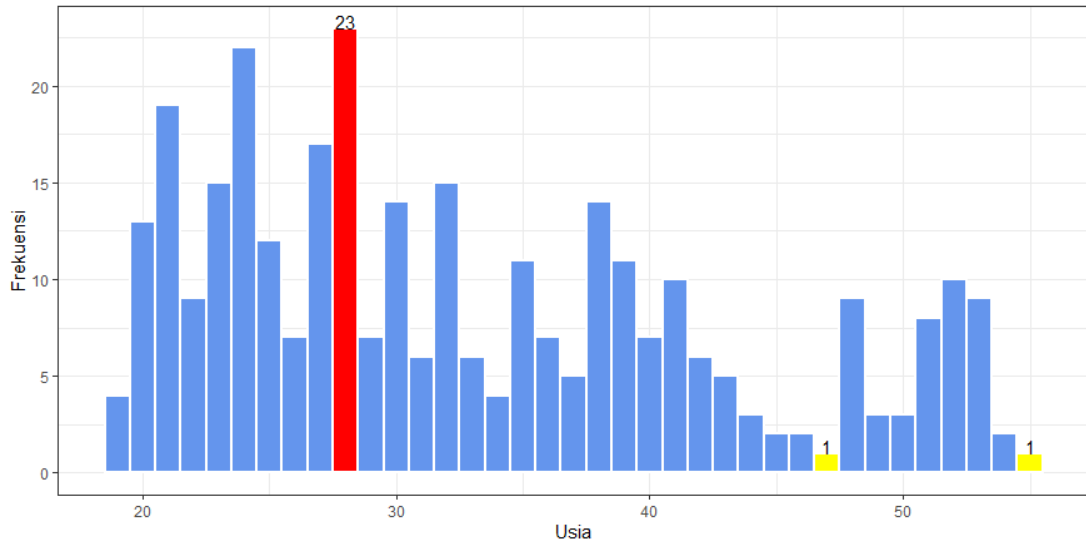
Lampiran 5 (Lanjutan 2)

Y	X1	X2	X3	X4	X5
0	0	1	28	1	0
1	0	0	38	0	0
1	0	0	36	0	0
0	0	1	33	0	0
1	0	1	20	0	1
1	0	1	21	0	1
1	0	0	22	1	1
1	0	0	27	0	1
0	1	1	21	0	1
1	0	0	38	0	1
1	0	1	30	1	1
1	0	1	21	0	1
0	0	1	24	1	1
0	0	1	39	1	0
1	0	1	48	0	1
1	0	0	21	0	0
1	0	0	24	1	0
0	0	0	41	1	1
1	0	0	53	0	1
0	0	1	28	0	1
1	0	0	32	0	0
1	0	0	25	0	0
0	0	1	30	0	0
0	0	1	27	0	1
2	0	1	35	0	0
2	0	0	27	0	0
0	0	1	52	0	0
1	0	0	30	1	0
1	0	0	24	1	1
0	1	1	28	0	1
0	0	1	51	0	1
1	0	0	32	1	1
1	0	0	40	0	1
1	0	0	38	1	1
0	1	1	36	0	0
1	0	1	51	0	0
1	0	0	41	1	1
1	0	0	21	1	0
1	0	0	43	0	0
1	0	0	22	0	1
⋮	⋮	⋮	⋮	⋮	⋮

Lampiran 6 Output Statistika Deskriptif



Lampiran 6 (Lanjutan 1)



Jenis Kelamin variable			Lak	Perem
<chr>	<chr>	<chr>	<chr>	<chr>
1	Y	D	206 (93.64%)	14 (6.36%)
2	Y	LL	43 (65.15%)	23 (34.85%)
3	Y	LU	18 (50.00%)	18 (50.00%)

Jabatan variable			Off	Prod
<chr>	<chr>	<chr>	<chr>	<chr>
1	Y	D	46 (20.91%)	174 (79.09%)
2	Y	LL	47 (71.21%)	19 (28.79%)
3	Y	LU	19 (52.78%)	17 (47.22%)

Jam Kerja variable			>8	<=8
<chr>	<chr>	<chr>	<chr>	<chr>
1	Y	D	153 (69.55%)	67 (30.45%)
2	Y	LL	45 (68.18%)	21 (31.82%)
3	Y	LU	20 (55.56%)	16 (44.44%)

Wakto variable			Terang	Gelap
<chr>	<chr>	<chr>	<chr>	<chr>
1	Y	D	106 (48.18%)	114 (51.82%)
2	Y	LL	22 (33.33%)	44 (66.67%)
3	Y	LU	15 (41.67%)	21 (58.33%)

Lampiran 7 Output Metode Regresi Logistik Multinomial

Pemilihan kombinasi Training Testing (1-APER)

> acc #kombinasi 75:25 yang dipilih

name	acc.tr	acc.ts
1 75:25	0.7107438	0.7125000
2 80:20	0.7209302	0.6093750
3 85:15	0.7490909	0.6595745
4 90:10	0.6907216	0.7096774

Pemilihan subset

>acc.resampl<sim_multinom_modeltr75\$resample;acc.resampl #Mencari fold terbaik

	Accuracy	Kappa	Resample
1	0.6666667	0.18644068	Fold01
2	0.7391304	0.31000000	Fold02
3	0.7083333	0.32530120	Fold05
4	0.7500000	0.41700405	Fold04
5	0.6800000	0.20000000	Fold03
6	0.8260870	0.61825726	Fold06
7	0.6000000	0.13793103	Fold09
8	0.6666667	0.07246377	Fold08
9	0.7200000	0.41666667	Fold07
10	0.6000000	0.18831169	Fold10

UJI SIMULTAN

Loglikelihood Ratio Test

#LogLik	Df	Chisq	Pr(>Chisq)
1	-211.03		
2	-267.28	10	112.51 < 2.2e-16 ***

UJI PARSIAL

	(Intercept)	X1P	X2P	X3	X4	X5
LL	0.70713649	4.072247e-03	3.087951e-09	0.6130709	0.9734683	0.06109275
LU	0.00680754	1.175599e-06	3.520513e-03	0.2432180	0.5483309	0.81611049

LR test model setelah stepAIC

#LogLik Df Chisq Pr(>Chisq)

1	-213.03			
2	-267.28	2	108.5	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Lampiran 7 (Lanjutan 1)

Model Terbaik

```
> summary(multinom_model)
```

Call:

```
nnet::multinom(formula = .outcome ~ ., data = dat, decay = param$decay)
```

Coefficients:

```
(Intercept)  X1P  X2P  X3  
LL  0.1272003 1.935763 -2.211964 -0.01500443  
LU -2.7722201 2.814246 -1.341180  0.03150400
```

Std. Errors:

```
(Intercept)  X1P  X2P  X3  
LL  0.6701234 0.4594051 0.3758402 0.01842390  
LU  0.8728875 0.5013096 0.4634028 0.02113455
```

Uji Parsial

```
(Intercept)  X1P  X2P  X3  
LL 0.849453103 2.512997e-05 3.971275e-09 0.4154154  
LU 0.001493625 1.979448e-08 3.801275e-03 0.1360561
```

Odds Ratio (X1, X2, dan X3 untuk masing-masing LL dan LU)

```
LL : 6.9293264 0.1094854 0.9851076  
LU : 16.6805861 0.2615369 1.0320055
```

Uji Kesesuaian Model

Pearson's Chi-squared test

```
data: predict(sim_multinom_modeltr75, trn75) and trn75$Y  
X-squared = 46.017, df = 4, p-value = 3.001e-15
```

Hasil Klasifikasi Training

```
Reference  
Prediction D LL LU  
D 171 40 16  
LL 1 13 6  
LU 4 0 7  
(1-APER)= 74,03%
```

Hasil Klasifikasi Testing

```
Reference  
Prediction D LL LU  
D 42 11 6  
LL 2 2 0  
LU 0 0 1  
(1-APER)=70,31%
```

Lampiran 8 Output Metode Random Forest

Tabel kemungkinan

X1 = 2
 X2 = 2
 X3 = (37-1) = 36
 X4 = 2
 X5 = 2

Perbandingan Ketepatan Klasifikasi (1-APER)

	[k=100]	[k=500]	[k=1000]
[90:10]	0.9175258	0.7741935	0.8969072
[85:15]	0.8836364	0.7659574	0.8763636 0.7872340
[80:20]	0.8837209	0.7656250	0.8798450
[75:25]	0.8966942	0.7625000	0.8966942

Pemilihan Subset

	Accuracy	Kappa	Resample
1	0.8214286	0.6164384	Fold01
2	0.7500000	0.4060606	Fold02
3	0.8148148	0.6052632	Fold05
4	0.8148148	0.5312500	Fold04
5	0.7307692	0.4770115	Fold03
6	0.7142857	0.4120735	Fold06
7	0.6428571	0.2200557	Fold09
8	0.8214286	0.6419437	Fold08
9	0.6551724	0.1238671	Fold07

Skor Variabel Terpenting

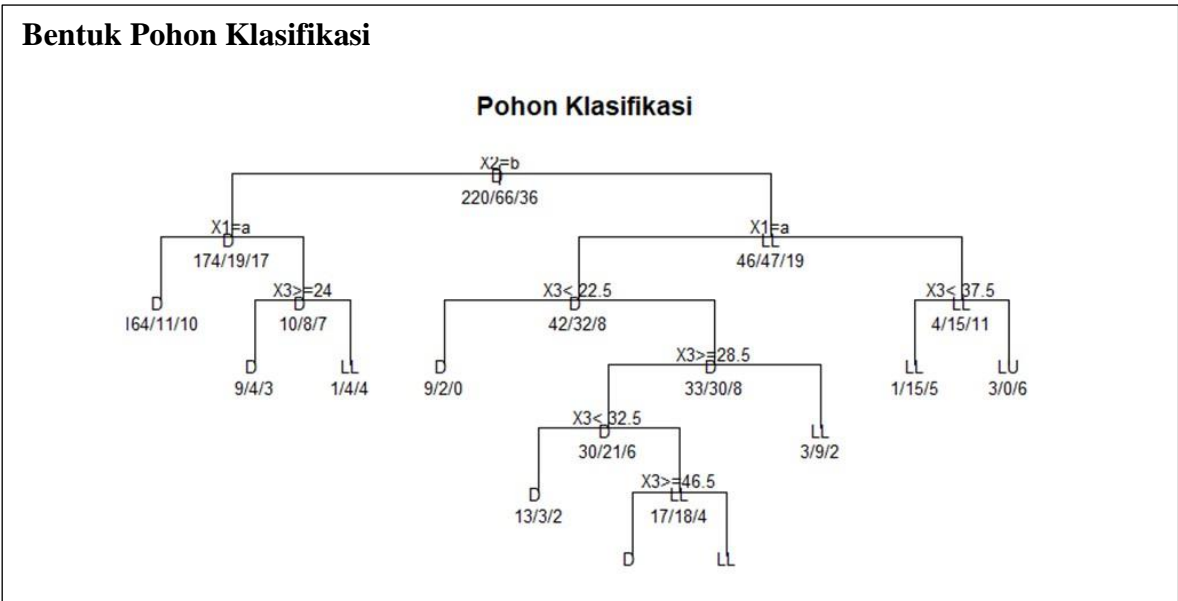
X1	51.187031
X2	52.031905
X3	38.909765
X4	19.475199
X5	3.519233

Pruning

Relatif Error			
	CP	nsplit	rel error
1	0.05392157	0	1.0000000
2	0.02941176	3	0.8333333
3	0.01470588	7	0.7156863
4	0.01000000	9	0.6862745

Lampiran 8 (Lanjutan 1)

Bentuk Pohon Klasifikasi



Training

Prediksi	Observasi		
	D	LL	LU
D	174	13	9
LL	2	39	6
LU	0	1	14

Testing

Prediksi	Observasi		
	D	LL	LU
D	41	6	2
LL	3	7	3
LU	0	0	2

BIODATA PENULIS



Penulis bernama Lailiya Risky Imania dilahirkan di Mojokerto, 30 Juni 2000 dan merupakan anak kedua dari 3 bersaudara. Penulis telah menempuh pendidikan formal yaitu di TK Dharmawanita Watukenongo, SDN 2 Watukenongo, SMPN 1 Mojosari dan SMAN 1 Puri Mojokerto. Setelah lulus dari SMAN tahun 2018, Penulis mengikuti SNMPTN dan diterima di Departemen Sains Aktuaria FSAD - ITS pada tahun 2018 dan terdaftar dengan NRP 06311840000009. Di Departemen Sains Aktuaria, penulis sempat aktif di beberapa kegiatan Seminar yang diselenggarakan oleh Departemen, Himpunan Mahasiswa Sains Aktuaria (HIMASAKTA) dan aktif pada UKM Voli ITS. Apabila pembaca ingin memberi kritik dan saran mengenai laporan tugas akhir ini, dapat menghubungi penulis melalui laiiyarimania@gmail.com