

**PREDIKSI KOLEKTIBILITAS TUNGGAKAN IURAN JAMINAN
KESEHATAN BAGI PEKERJA BUKAN PENERIMA UPAH (PBPU)
DENGAN PENERAPAN TREE-BASED ALGORITHM**

SKRIPSI



**DIVA ALVANIA ANDI ABDUH
H081201048**

**PROGRAM STUDI ILMU AKTUARIA
DEPARTEMEN MATEMATIKA
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITAS HASANUDDIN**

2024



**PREDIKSI KOLEKTIBILITAS TUNGGAKAN IURAN JAMINAN
KESEHATAN BAGI PEKERJA BUKAN PENERIMA UPAH (PBP)
DENGAN PENERAPAN TREE-BASED ALGORITHM**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Aktuaria pada Program Studi Ilmu Aktuaria Departemen Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin**

UNIVERSITAS HASANUDDIN

Diva Alvania Andi Abduh

H081201048

**PROGRAM STUDI ILMU AKTUARIA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

FEBRUARI 2024



LEMBAR PERNYATAAN KEOTENTIKAN

Yang bertanda tangan di bawah ini:

Nama : Diva Alvania Andi Abdul

Nim : H081201048

Program Studi : Ilmu Aktuaria

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya yang berjudul

PREDIKSI KOLEKTIBILITAS TUNGGAKAN IURAN JAMINAN KESEHATAN BAGI PEKERJA BUKAN PENERIMA UPAH (PBP) DENGAN PENERAPAN TREE-BASED ALGORITHM

adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa tulisan skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 26 Februari 2024



Divia Alvania Andi Abdul

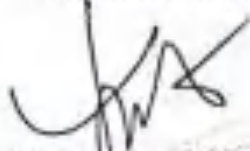
H081201048




PREDIKSI KOLEKTIBILITAS TUNGGAKAN IURAN JAMINAN KESEHATAN BAGI PEKERJA BUKAN PENERIMA UPAH (PBPU) DENGAN PENERAPAN TREE-BASED ALGORITHM

Disetujui oleh:

Pembimbing Utama


Dr. Amran, S.Si., M.Si.
NIP. 1978111011998021001

Pembimbing Pertama


Edy Saputra Rusdi, S.Si., M.Si.
NIP. 199104102020053001



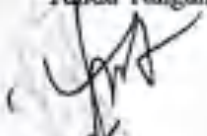

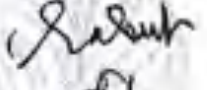

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh

Nama : Diva Alvania Andi Abdul
NIM : H081201048
Program Studi : Ilmu Aktuaria
Judul Skripsi : Prediksi Kolektibilitas Tunggakan Iuran Jaminan Kesehatan Bagi Pekerja Bukan Penerima Upah (PBPU) dengan Penerapan *Tree-Based Algorithm*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Aktuaria pada Program Studi Ilmu Aktuaria Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
Ketua	: Dr. Amran, S.Si., M.Si.	()
Sekretaris	: Edy Saputra Rusdi, S.Si., M.Si.	()
Anggota	: Prof. Dr. Hasmawati, M.Si.	()
Anggota	: Illuminata Wynnie, S.Si., M.Si.	()

Ditetapkan di : Makassar
Tanggal : 26 Februari 2024



HALAMAN PENGESAHAN

**PREDIKSI KOLEKTIBILITAS TUNGGAKAN IURAN JAMINAN
KESEHATAN BAGI PEKERJA BUKAN PENERIMA UPAH (PBPU)
DENGAN PENERAPAN TREE-BASED ALGORITHM**

Disusun dan diajukan oleh

Diva Alyania Andi Abdul
11081201048

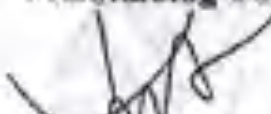
Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Ilmu Aktuaria Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Pada tanggal 26 Februari 2024


dan dinyatakan telah Memenuhi syarat kelulusan

Menyetujui,

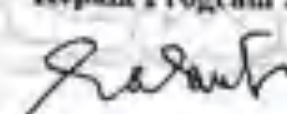
Pembimbing Utama


Dr. Amran, S.Si., M.Si.
NIP. 197011011998021001

Pembimbing Pertama


Edy Saputra Rusdi, S.Si., M.Si.
NIP. 199104102020053001

Kepala Program Studi


Prof. Dr. Hasmiwati, M.Si.
NIP. 196412311990032007



KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan kepada Allah SWT atas segala rahmat dan hidayah-nya dalam menyelesaikan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada junjungan Nabi Muhammad SAW, sebagai Nabi yang telah menjadi suri tauladan bagi seluruh umatnya sehingga penyusunan skripsi ini dapat terselesaikan dengan judul “**Prediksi Kolektibilitas Tunggakan Iuran Jaminan Kesehatan Bagi Pekerja Bukan Penerima Upah (PBPU) dengan Penerapan Tree-Based Algorithm**”.

Skripsi ini disusun dan diajukan sebagai tugas akhir untuk memenuhi syarat dalam menyelesaikan Pendidikan Strata Satu (S1) Sarjana Aktuaria pada Program Studi Ilmu Aktuaria Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Banyak pengalaman dan makna berharga yang dapat menjadi pelajaran bagi penulis dalam menyelesaikan skripsi ini. Oleh karena itu, dalam kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada ayahanda **alm. Ir. Andi Muhammad Abduh Rachman, M.Si.** dan ibunda **Rooschanty Robinson, S.E.**, serta adinda **Dashya Alaura Andi Abduh** selalu memberikan doa, dukungan, dan kasih sayang yang tak terhingga sehingga penulis dapat menyelesaikan skripsi ini. Selain itu, penulis menyadari bahwa penyusunan skripsi ini dapat selesai tidak lepas dari bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan hingga pada penyusunan skripsi ini. Oleh karena itu, pada kesempatan kali ini dengan segala ketulusan dan kerendahan hati penulis mengucapkan terimakasih kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin.
2. Bapak **Dr. Eng Amiruddin, M.Si.**, selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta jajarannya dan Staf Fakultas MIPA yang telah membantu dan memudahkan penulis dalam segala

administrasi.

Bapak **Dr. Firman, S.Si., M.Si.**, selaku Ketua Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Bapak dan Ibu Dosen Prodi



Ilmu Aktuaria yang dengan tulus telah memberikan banyak ilmu dan pelajaran berharga, serta Staf Departemen Matematika yang telah membantu proses administrasi dalam penyusunan skripsi ini.

4. Ibu **Prof. Hasmawati M.Si.**, selaku Ketua Program Studi Ilmu Aktuaria, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam.
5. Bapak **Dr. Amran S.Si., M.Si.** selaku pembimbing Utama penulis yang senantiasa membimbing dengan penuh kasih, serta ikhlas meluangkan banyak waktu di tengah kesibukan, sehingga penulisan skripsi ini dapat berjalan dengan lancar.
6. Bapak **Edy Saputra Rusdi, S.Si., M.Si.**, selaku Pembimbing Pertama penulis yang senantiasa sabar, tulus, membantu, dan memberikan arahan selama proses penulisan skripsi.
7. Ibu **Prof. Hasmawati M.Si.**, dan Ibu **Ibu Illuminata Wynnies, S.Si., M.Si.**, selaku Dosen Penguji yang senantiasa memberikan waktu dan kesediaannya untuk memberikan kritik dan masukan yang berharga dalam pelaksanaan seminar proposal hingga sidang skripsi.
8. Ibu **Ainun Mawaddah Abdal, S.Si., M.Si.**, selaku Penasehat Akademik penulis yang dengan tulus memberikan kesediaannya untuk membantu dan memberikan arahan dalam masa studi hingga proses penyusunan skripsi ini.
9. Bapak/Ibu **Dosen Program Studi Ilmu Aktuaria** yang telah mendidik dan memberikan ilmunya kepada penulis selama proses perkuliahan. Serta kepada **Staf dan Pegawai Departemen Matematika** yang telah membantu dalam proses administrasi.
10. **Muh Raihan Akham Arfan** adalah sosok yang tak pernah lelah mendengarkan cerita dan keluh kesah saya sebagai penulis. Memberi warna tersendiri dalam setiap langkah, memastikan saya tetap bersemangat dan fokus pada tujuan akhir. Tanpa kehadiran dan dukungannya, saya yakin skripsi ini tidak akan terselesaikan dengan baik dan membanggakan. Terima kasih, karena telah menjadi bagian tak terpisahkan dalam perjalanan penting ini.

...nda **Khristine Novi Palinggi** yang senantiasa membantu penulis dalam mencari topik skripsi dan kakak-kakak **Pegawai BPJS Kesehatan Kantor ... ang Makassar** yang telah memberikan kesempatan bagi penulis untuk



belajar selama proses magang dan mempermudah penulis untuk memperoleh data penelitian.

12. Terima kasih kepada keluarga **Bpk Taufik Hidayat** yang telah memberikan dukungan kepada penulis berupa motivasi dan materi selama proses penyelesaian skripsi ini.
13. Teman-teman penulis, “**IDEA**” dan “**JIMDEMA**” yang selalu menemani dan membantu penulis selama masa perkuliahan dan memberi motivasi untuk menyelesaikan proses penyusunan skripsi ini.
14. Terima kasih kepada **Muh. Raihan Arhab** yang telah membantu penulis sejak awal penulisan hingga penyelesaian skripsi ini.
15. Terima kasih kepada teman-teman “**Ilmu Aktuaria 2020**” yang telah hadir memberikan semangat, keceriaan, dan dukungan selama masa perkuliahan.
16. Terakhir, saya ingin berterima kasih kepada diri sendiri karena telah berjuang selama ini, telah mampu berproses dan tetap semangat dibalik masalah yang dihadapi, tetap tersenyum walau badai masalah menghampiri dan tidak pernah menyerah hingga akhir proses penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, mengingat adanya keterbatasan dalam pengetahuan dan kemampuan penulis. Oleh karena itu, penulis ingin mengucapkan permohonan maaf atas segala kesalahan dan kekurangan yang mungkin ada dalam penulisan skripsi ini, serta bersedia menerima kritik dan saran yang membangun.

Akhir Kata, penulis berharap semoga segala bentuk kebaikan yang telah diberikan bernilai ibadah dan mendapatkan balasan dari Allah SWT. Semoga skripsi ini dapat memberikan manfaat dalam pengembangan ilmu

Makassar, 26 Februari 2024



Divia Alvania Andi Abdul



**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Diva Alvania Andi Abduh
NIM : H081201048
Program Studi : Ilmu Aktuaria
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non eksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul:

**Prediksi Kolektibilitas Tunggakan Iuran Jaminan Kesehatan Bagi
Pekerja Bukan Penerima Upah (PBPU) dengan Penerapan
*Tree-Based Algorithm***

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar Pada
Tanggal 26 Februari 2024
Yang Menyatakan



Divia Alvania Andi Abduh



ABSTRAK

Penelitian ini bertujuan untuk memprediksi kolektibilitas tunggakan iuran Jaminan Kesehatan Nasional (JKN) bagi Peserta Bukan Penerima Upah (PBPU) BPJS Kesehatan dengan menerapkan algoritma berbasis pohon keputusan. Penelitian ini dilakukan dalam rangka memelihara keberlanjutan finansial BPJS Kesehatan sebagai penyelenggara program JKN. Penelitian ini membandingkan performa dan evaluasi pemodelan data prediksi kolektibilitas tunggakan iuran peserta JKN bagi PBPU menggunakan algoritma berbasis pohon keputusan. Hasil penelitian menunjukkan bahwa aktivitas penagihan iuran kepada peserta PBPU, terutama melalui *telecollecting*, memiliki potensi meningkatkan keberhasilan program JKN. Analisis prediktif data mining pada data set *telecollecting* menunjukkan tingkat keberhasilan sebesar 89% dengan tingkat akurasi sebesar 80%. Lima faktor teridentifikasi sebagai berpengaruh terhadap kolektibilitas tunggakan iuran, yaitu lama waktu menunggak, usia peserta individu, waktu penagihan, jumlah keluarga yang bertanggung, dan wilayah peserta terdaftar fasilitas kesehatan. Dalam konteks pemodelan, algoritma *LightGBM* menunjukkan kinerja terbaik dengan komposisi data 90:10. Implikasi dari penelitian ini dapat menjadi landasan bagi BPJS Kesehatan untuk meningkatkan efektivitas kegiatan *telecollecting* dan strategi penagihan iuran, sehingga dapat meminimalisir piutang Dana Jaminan Sosial (DJS) dan menjaga keberlanjutan program JKN.

Kata Kunci: *Tree-Based Algorithm*, Tunggakan Iuran, *Telecollecting*

Judul : Prediksi Kolektibilitas Tunggakan Iuran Jaminan Kesehatan Bagi Pekerja Bukan Penerima Upah (PBPU) dengan Penerapan *Tree-Based Algorithm*
Nama : Diva Alvania Andi Abduh
NIM : H081201048
Program Studi : Ilmu Aktuaria



ABSTRACT

This study aims to predict the collectibility of arrears of the National Health Insurance (JKN) premiums for Non-Wage Recipients (PBPU) BPJS Health by applying a decision tree-based algorithm. This study was conducted in order to maintain the financial sustainability of BPJS Health as the organizer of the JKN program. This study compares the performance and evaluation of data modeling to predict the collectibility of arrears of JKN premiums for PBPU using a decision tree-based algorithm. The results of the study show that the activity of collecting premiums from PBPU participants, especially through telecollecting, has the potential to increase the success of the JKN program. Predictive data mining analysis on the telecollecting data set shows a success rate of 89% with an accuracy level of 80%. Five factors were identified as influencing the collectibility of arrears of premiums, namely the length of time overdue, the age of individual participants, the billing time, the number of family members covered, and the region of participants registered for health facilities. In the context of modeling, the LightGBM algorithm shows the best performance with a data composition of 90:10. The implications of this study can be a basis for BPJS Health to improve the effectiveness of telecollecting activities and premium billing strategies, so as to minimize the receivables of the Social Security Fund (DJS) and maintain the sustainability of the JKN program.

Keywords: *Tree-Based Algorithm, Premium Arrears, Telecollecting*

Title : Predicting the Collectibility of Arrears of Health Insurance Premiums for Non-Wage Workers (PBPU) with the Application of Tree-Based Algorithm

Name : Diva Alvania Andi Abduh

Student ID : H081201048

Study Program : Actuarial Science



DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN KEOTENTIKAN.....	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR	x
ABSTRAK	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xvi
DAFTAR TABEL.....	xviii
DAFTAR NOTASI.....	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penulisan	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Jaminan Sosial.....	6
2.2 Jaminan Kesehatan Nasional.....	6
2.2.1 Kepesertaan Jaminan Kesehatan Nasional.....	7
2.2.1.1 Turan Jaminan Kesehatan Nasional	9
2.2.1.2 Manfaat Jaminan Kesehatan Nasional	9
2.2.1.3 Badan Penyelenggara Jaminan Sosial Kesehatan	10



2.4	Kolektibilitas Iuran.....	11
2.5	<i>Tree-Based Algorithm</i>	13
2.5.1	<i>Random Forest</i>	14
2.5.2	<i>Light Gradient Boosting Machine (LightGBM)</i>	16
2.5.3	<i>Adaptive Boosting (AdaBoost)</i>	19
2.5.4	Evaluasi Model <i>Machine Learning</i>	21
BAB III METODE PENELITIAN.....		27
3.1	Pendekatan dan Jenis Penelitian.....	27
3.2	Waktu dan Tempat Penelitian	27
3.3	Objek Penelitian.....	27
3.4	Jenis dan Sumber Data.....	28
3.5	Metode Pengumpulan Data.....	28
3.6	Metode Analisis Data.....	28
3.7	Alur Kerja	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Pemahaman Program <i>Telecollecting</i>	31
4.2	Pemahaman Data.....	32
4.2.1	Statistik Deskriptif Data.....	35
4.2.2	<i>Fitting</i> Distribusi Variabel Numerik	36
4.2.3	Hubungan Karakteristik Variabel	39
4.2.4	Kelas <i>Imbalance data</i> , Anomali, dan <i>Outlier</i>	42
4.3	<i>Pre-Processing</i> Data	44
4.3.1	Pemodelan Awal <i>Imbalance data</i>	44
4.3.2	Menyeimbangkan Kelas Data Target.....	47
	Transformasi Data Kategorik.....	47
	Pemodelan Data.....	48



4.4.1	Hasil Pembelajaran 60% Data latih.....	48
4.4.2	Hasil Pembelajaran 70% Data latih.....	52
4.4.3	Hasil Pembelajaran 80% Data latih.....	56
4.4.4	Hasil Pembelajaran 90% Data latih.....	60
4.5	Evaluasi Model.....	64
4.5.1	Hasil Evaluasi <i>Precision, Recall, F1-Score</i> , dan <i>Accuracy</i>	64
4.5.2	Hasil Evaluasi <i>Area Under ROC Curve (AUC)</i>	67
4.5.3	Hasil Evaluasi <i>Cross Validation</i>	68
4.6	Hasil <i>Feature Importance</i>	69
BAB V KESIMPULAN DAN SARAN.....		71
5.1	Kesimpulan.....	71
5.2	Saran.....	71
5.2.1	Untuk Penyelenggara JKN (BPJS Kesehatan).....	71
5.2.2	Untuk Dinas Kesehatan dan Pembuat Kebijakan.....	72
5.2.3	Bagi Akademisi.....	72
DAFTAR PUSTAKA		73



DAFTAR GAMBAR

Gambar 2.1 Skema Prediksi <i>Random Forest</i>	14
Gambar 2.2 Leaf-Wise <i>Tree Growth</i>	18
Gambar 2.3 Ilustrasi <i>AdaBoost</i> Algorithm.....	21
Gambar 2.4 Ilustrasi ROC Curve dan AUC.....	25
Gambar 2.5 Skema <i>k-fold Cross Validation</i> dengan $k = 10$	26
Gambar 3.1 Diagram Alir Penelitian	30
Gambar 4.1 Komposisi <i>Dataset</i> Penelitian	31
Gambar 4.2 Visualisasi Histogram Variabel Numerik	36
Gambar 4.3 Distribusi Data Berdasarkan Usia Peserta.....	37
Gambar 4.4 Distribusi Data Berdasarkan Jumlah Bulan Menunggu	38
Gambar 4.5 Distribusi Data Berdasarkan Jumlah Anggota Keluarga.....	38
Gambar 4.6 Distribusi Data Total Iuran yang Ditagihkan	39
Gambar 4.7 Heatmap Plot Pearson Correlation	40
Gambar 4.8 Komposisi Kelas Target Berdasarkan Jenis Kelamin	41
Gambar 4.9 Komposisi Kelas Target Berdasarkan Wilayah	41
Gambar 4.10 Komposisi Kelas Target Berdasarkan Alasan Menunggu	42
Gambar 4.11 Box Plot Distribusi Total Tagihan	43
Gambar 4.12 <i>Random Forest</i> Klasifikasi <i>Imbalance Data</i>	44
Gambar 4.13 <i>LightGBM</i> Klasifikasi <i>Imbalance Data</i>	45
Gambar 4.14 <i>AdaBoost</i> Klasifikasi <i>Imbalance Data</i>	46
Gambar 4.15 Hasil Resampling dengan <i>RandomUnderSampler</i>	47
Gambar 4.16 Klasifikasi <i>Random Forest</i> 40% Data uji.....	49
Gambar 4.17 Klasifikasi <i>LightGBM</i> 40% Data uji	50
Gambar 4.18 Klasifikasi <i>AdaBoost</i> 40% Data uji.....	51
Gambar 4.19 Klasifikasi <i>Random Forest</i> 30% Data uji.....	53
Gambar 4.20 Klasifikasi <i>LightGBM</i> 30% Data uji	54
Gambar 4.21 Klasifikasi <i>AdaBoost</i> 30% Data uji.....	55
Gambar 4.22 Klasifikasi <i>Random Forest</i> 20% Data uji.....	57
Gambar 4.23 Klasifikasi <i>LightGBM</i> 20% Data uji	58
Gambar 4.24 Klasifikasi <i>AdaBoost</i> 20% Data uji.....	59
Gambar 4.25 Klasifikasi <i>Random Forest</i> 10% Data uji.....	61



Gambar 4.26 Klasifikasi *LightGBM* 10% Data uji 62

Gambar 4.27 Klasifikasi *AdaBoost* 10% Data uji..... 63

Gambar 4.28 Kurva ROC Hasil Klasifikasi 67

Gambar 4.29 *Feature Importance LightGBM* (90:10)..... 69

Gambar 4.30 *Feature Importance Random Forest* (90:10) 70

Gambar 4.31 *Feature Importance AdaBoost* (90:10) 70



DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	22
Tabel 4.1 Daftar Variabel yang Dianalisis	33
Tabel 4.2 Statistik Deskriptif Data Numerik.....	35
Tabel 4.3 Nilai Korelasi Antara Variabel Independen dan Dependen.....	40
Tabel 4.4 Evaluasi Model Komposisi Data 60:40	65
Tabel 4.5 Evaluasi Model Komposisi Data 70:30	65
Tabel 4.6 Evaluasi Model Komposisi Data 80:20	66
Tabel 4.7 Evaluasi Model Komposisi Data 90:10	66
Tabel 4.8 ROC-AUC <i>Score</i>	67
Tabel 4.9 Performa Akurasi Model dengan 10- <i>Fold Cross Validation</i>	68
Tabel 4.10 Performa Nilai AUC Model dengan 10- <i>Fold Cross Validation</i>	68



DAFTAR NOTASI

X_i	: Variabel independen atau fitur
Y_i	: Variabel dependen atau respon
x_n, y_n	: Sampel pelatihan ke-n
R	: Himpunan bilangan riil
ε	: Kesalahan-kesalahan klasifikasi minimum
\hat{Y}_I	: Output <i>Random Forest</i>
S_n	: <i>Training Set Random Forest</i>
$S_n^{\Theta_1}$: Sampel <i>bootstrap</i>
A	: Subset data gradien besar
B	: Subset data gradien kecil
A^c	: Komplemen subset data gradien besar
a	: Rasio sampling data gradien besar
b	: Rasio sampling data gradien kecil
$V_j(d)$: Estimasi <i>gain varians</i> setelah subset $A \cup B$
g_i	: Gradien setiap data
d	: Jumlah iterasi
α_t	: Bobot untuk pengklasifikasian pada pembelajar lemah ke-t
$w_n^{(t)}$: Bobot pada sampel pelatihan dengan iterasi ke-t
h_f	: Fungsi hipotesis <i>AdaBoost</i>
$\text{sgn}(x)$: Fungsi tanda atau fungsi signum



BAB I

PENDAHULUAN

1.1 Latar Belakang

Cita-cita bangsa Indonesia, sebagaimana tercantum dalam Pancasila dan Undang-undang Dasar Negara Republik Indonesia Tahun 1945, menyatakan bahwa kesehatan adalah hak yang melekat pada setiap manusia dan merupakan salah satu unsur penting dalam mencapai kesejahteraan. Setiap orang berhak mendapatkan perlindungan sosial, guna memenuhi kebutuhan dasar hidup yang layak. Bahwa untuk mewujudkan jaminan sosial yang menyeluruh, negara mengembangkan Sistem Jaminan Sosial Nasional (SJSN) bagi seluruh rakyat Indonesia, pemerintah telah menerapkan serangkaian kebijakan yang diatur dalam Undang-Undang Nomor 40 Tahun 2004. Penyelenggaraan SJSN dilakukan melalui Badan Penyelenggara Jaminan Sosial (BPJS). BPJS Kesehatan berfungsi menyelenggarakan program jaminan kesehatan.

Program Jaminan Kesehatan Nasional (JKN) melalui BPJS Kesehatan merupakan upaya dalam mendistribusikan layanan kesehatan dengan adil bagi semua lapisan masyarakat dengan iuran yang terjangkau dan cakupan layanan kesehatan yang luas, diselenggarakan dengan menggunakan prinsip asuransi kesehatan sosial, dimana setiap peserta JKN berkontribusi dalam bentuk iuran bulanan yang nilainya telah ditetapkan oleh pemerintah (Rizal, 2022). Keberadaan program JKN sesuai dengan salah satu komponen dari Program Indonesia Sehat, yaitu memastikan bahwa setiap individu dalam keluarga telah terdaftar sebagai peserta JKN (Kementerian Kesehatan Republik Indonesia, 2017).

Program JKN ditujukan untuk seluruh warga dengan partisipasi yang bersifat wajib dan didasarkan pada semangat gotong-royong yang dapat diimplementasikan melalui sistem saling bantu antar peserta yang mampu membantu peserta yang kurang mampu, peserta dengan risiko rendah membantu peserta dengan risiko tinggi, dan peserta sehat membantu peserta yang sakit. Dengan harapan dapat memperkuat keadilan sosial bagi seluruh rakyat Indonesia melalui prinsip

tersebut (Kementerian Kesehatan Republik Indonesia, 2013). Keberhasilan implementasi BPJS Kesehatan bergantung pada faktor kunci, yaitu iuran. Apabila jumlah iuran yang terkumpul sejalan dengan manfaat yang



disediakan untuk peserta BPJS Kesehatan, maka sistem jaminan sosial ini dapat beroperasi secara efektif tanpa menimbulkan kerugian bagi pihak yang terlibat di dalamnya. Kepesertaan seseorang tidak aktif jika orang tersebut tidak membayar iuran dan/atau memiliki tunggakan iuran.

Kolektibilitas tunggakan iuran peserta JKN merupakan salah satu penyebab piutang Dana Jaminan Sosial (DJS). Artinya, piutang Dana Jaminan Sosial (DJS) berbanding lurus dengan besarnya tunggakan iuran program JKN terkhusus bagi peserta Pekerja Bukan Penerima Upah (PBPU), hal ini mengacu pada kepatuhan peserta membayar iuran. Apabila tingkat kolektibilitas tunggakan iuran meningkat, maka diharapkan dapat meminimalisir piutang DJS. Sehingga, BPJS Kesehatan dapat menyelenggarakan program JKN yang bermutu dan berkesinambungan. Fokus utama BPJS Kesehatan dalam mengurangi piutang DJS, yaitu mengupayakan peningkatan kolektibilitas iuran JKN terutama bagi kelompok peserta mandiri atau peserta PBPU. Pengiriman informasi tagihan kepada peserta PBPU dilakukan melalui media komunikasi seperti SMS/WA/Email/Surat, *telecollecting*, agen institusi atau program kerja sama kemitraan, dan kader JKN.

Telecollecting merupakan upaya penagihan tunggakan iuran peserta PBPU yang paling efektif, yaitu melalui telepon dan/atau *WhatsApp* dengan hasil pendataan tercatat peserta PBPU tersebut tercatat detail data peserta, keluarga bertanggung, riwayat telepon, riwayat pelayanan, riwayat mutasi kepesertaan, dan kesediaan membayar tunggakan iurannya. *Telecollecting* menggunakan aplikasi berbasis situs web yang mengandalkan *database* dalam kegiatan penginputan, pemrosesan, pengeluaran, dan penilaiannya. Kegiatan *telecollecting*, selain mengingatkan mengenai kewajiban membayar iuran, petugas juga menginformasikan mengenai jumlah bulan tertunggak, besar tagihan iuran tertunggak pada anggota keluarga yang bertanggung, kemudahan akses pada berbagai kanal pembayaran, kemudahan mengakses *Mobile JKN*, serta *call center* BPJS Kesehatan, sehingga masyarakat juga lebih paham akan program JKN dan mempermudah informasi terkait pembayaran iurannya (Ossy *et al.*, 2022).

mulasi tunggakan iuran paling tinggi berasal dari kontribusi iuran peserta sehingga, indikasi bahwa besaran piutang Dana Jaminan Sosial (DJS) BPJS berbanding lurus dengan besarnya nilai agregat tunggakan iuran peserta



JKN terkhususnya peserta PBPU. Seluruh data penginputan, pemrosesan, pengeluaran, dan penilaian dari sistem *telecollecting* yang telah terekam dalam *database* dapat dianalisis menggunakan algoritma *machine learning* karena data yang diperoleh memiliki karakteristik volume yang besar, beragam, dan dinamis. Hasil dari analisis tersebut akan dijadikan acuan dalam menentukan tingkat kolektibilitas iuran, selanjutnya dapat diketahui mayoritas peserta yang diprediksikan membayar tunggakan iuran setelah kegiatan *telecollecting*. Sehingga, pada penelitian ini akan memprediksi kolektibilitas tunggakan iuran jaminan kesehatan peserta PBPU untuk memberikan gambaran terhadap piutang Dana Jaminan Sosial (DJS) BPJS Kesehatan dengan penerapan *tree-based algorithm*.

1.2 Rumusan Masalah

Rumusan masalah penelitian ini adalah bagaimana hasil prediksi kolektibilitas tunggakan iuran jaminan kesehatan peserta PBPU dapat ditingkatkan menggunakan penerapan *tree-based algorithm*. Penerapan *tree-based algorithm* dapat mempertimbangkan faktor-faktor tersebut untuk memprediksi kolektibilitas tunggakan iuran jaminan kesehatan peserta PBPU secara lebih akurat dan efektif.

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah:

1. Penelitian ini memprediksi kolektibilitas tunggakan iuran jaminan kesehatan bagi Pekerja Bukan Penerima Upah (PBPU) berdasarkan data peserta yang sudah ataupun belum membayar tunggakan iuran setelah dilakukan *telecollecting*.
2. *Dataset* untuk analisis data menggunakan data aktivitas dan hasil penginputan *telecollecting* yang berlangsung di BPJS Kesehatan Kantor Cabang Makassar dengan periode 1 Januari 2023 hingga 31 Oktober 2023.
3. Pengolahan data *telecollecting* hanya peserta PBPU dengan tunggakan iuran peserta jaminan kesehatan lebih dari 1 bulan hingga tunggakan 24 bulan yang telah dihubungi oleh pegawai di BPJS Kesehatan Kantor Cabang Makassar.



4. Prediksi dan analisis data dilakukan menggunakan implementasi *machine learning* dengan menggunakan *tree-based algorithm*, yaitu *Random Forest*, *Light Gradient Boosting Machine (LightGBM)*, dan *Adaptive Boosting (AdaBoost)*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan, tujuan penelitian yang dapat dirumuskan adalah:

1. Membandingkan performa dan evaluasi pemodelan data prediksi hasil kolektibilitas tunggakan iuran peserta Jaminan Kesehatan Nasional (JKN) bagi PBPU BPJS Kesehatan dengan menggunakan *tree-based algorithm*.
2. Mengetahui hubungan masing-masing karakteristik variabel terhadap kolektibilitas tunggakan iuran peserta JKN bagi PBPU.
3. Mengetahui variabel yang memiliki pengaruh signifikan terhadap kolektibilitas tunggakan iuran peserta JKN bagi PBPU berdasarkan hasil dari *feature importance*.

1.5 Manfaat Penulisan

Penelitian ini diharapkan mampu memberikan kontribusi dalam pemahaman yang lebih mendalam terhadap analisis data dan prediksi nilai menggunakan *machine learning* dan memberikan wawasan baru bagi para pembaca. Juga, diharapkan dapat menambah pengetahuan dan mengembangkan keterampilan dalam mengkaji implementasi pada bidang aktuarial terkhusus bidang ilmu asuransi kesehatan dan *data science*.

Selain itu, sebagai sumber referensi yang dapat digunakan untuk penelitian atau studi lanjutan terkait prediksi kolektibilitas tunggakan iuran. Diharapkan dapat membantu dalam mengkaji faktor-faktor yang mempengaruhi besaran kolektibilitas tunggakan iuran peserta JKN bagi PBPU yang menunggak, serta membantu dalam mengevaluasi dan merencanakan strategi yang tepat dalam dari kenaikan besar piutang Dana Jaminan Sosial (DJS).



1.6 Sistematika Penulisan

Adapun sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi uraian penjelasan tentang latar belakang masalah, rumusan masalah, tujuan, batasan masalah, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi uraian penjelasan tentang definisi dan konsep menurut para ahli yang menjadi dasar dari penelitian, meliputi paparan teori.

BAB III METODOLOGI PENELITIAN

Bab ini berisi uraian penjelasan tentang Pendekatan dan Jenis Penelitian, Waktu dan Tempat Penelitian, Objek Penelitian, Jenis dan Sumber Data, Metode Pengumpulan Data dan Alur Kerja.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi uraian penjelasan tentang hasil penelitian setelah proses penelitian dilakukan. Bagian ini mencakup gambaran umum tentang objek penelitian, hasil analisis data, hasil perhitungan statistik, dan proses pembahasannya. Penjelasan tersebut akan memberikan informasi mendalam tentang temuan yang diperoleh dari penelitian, serta interpretasi dan analisis yang terkait dengan hasil tersebut.

BAB V PENUTUP

Bab ini berisi uraian penjelasan tentang kesimpulan dari hasil penelitian dan pembahasan yang telah dilakukan. Kesimpulan tersebut akan disusun berdasarkan analisis data, tinjauan pustaka, dan pembahasan yang telah dijelaskan sebelumnya.



BAB II

TINJAUAN PUSTAKA

2.1 Jaminan Sosial

Berdasarkan Undang-Undang RI Nomor 40 Tahun 2004, Sistem Jaminan Sosial Nasional (SJSN) merupakan suatu sistem yang bertujuan untuk memberikan program jaminan sosial kepada seluruh rakyat Indonesia. Sistem ini dirancang untuk memastikan bahwa setiap warga negara Indonesia dapat memenuhi kebutuhan dasar hidup yang layak, termasuk jaminan terhadap risiko sosial ekonomi seperti kesehatan. Mekanisme SJSN diselenggarakan berdasarkan asas kemanusiaan, asas manfaat, dan asas keadilan sosial bagi seluruh rakyat Indonesia.

2.2 Jaminan Kesehatan Nasional

Jaminan kesehatan adalah perlindungan kesehatan yang diberikan kepada peserta agar memperoleh manfaat pemeliharaan kesehatan dan perlindungan dalam memenuhi kebutuhan dasar kesehatan. Jaminan ini diberikan kepada setiap peserta yang telah membayar iuran atau peserta yang iurannya dibayar oleh pemerintah. Pada tanggal 1 Januari 2014, program Jaminan Kesehatan Nasional (JKN) secara resmi diimplementasikan untuk memberikan jaminan kesehatan (Bappenas, 2015).

Jaminan Kesehatan Nasional (JKN) merupakan komponen dari Sistem Jaminan Sosial Nasional (SJSN), program ini bertujuan untuk memberikan akses pelayanan kesehatan yang merata dan terjangkau bagi seluruh penduduk Indonesia. Setiap peserta program JKN memenuhi kriteria tertentu dan mereka diwajibkan membayar iuran sesuai dengan kelas atau kemampuan ekonomi masing-masing. Untuk memastikan keberlanjutan program JKN, penting bagi semua peserta untuk membayar iuran secara teratur. Hal ini akan memastikan dana yang cukup untuk membiayai pelayanan kesehatan bagi semua peserta.

Berdasarkan Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional, JKN diselenggarakan secara nasional berdasarkan prinsip asuransi sosial dan prinsip ekuitas, yaitu memperoleh pelayanan secara merata sesuai kebutuhan medis tanpa terikat dengan besar iuran yang dibayarkan. Prinsip ekuitas diwujudkan dengan membayar iuran bagi mereka yang memperoleh upah, dan pemerintah membayarkan iuran bagi mereka yang kurang mampu.



2.2.1 Kepesertaan Jaminan Kesehatan Nasional

Berdasarkan Peraturan Presiden RI Nomor 82 Tahun 2018 tentang Jaminan Kesehatan, para peserta adalah setiap orang termasuk orang asing yang bekerja selama minimal 6 (enam) bulan di Indonesia, yang telah melunasi pembayaran Iuran Jaminan Kesehatan. Kepesertaan Jaminan Kesehatan dibagi menjadi 2 segmen berdasarkan sumber dana pembayaran iuran, yaitu peserta Penerima Bantuan Iuran (PBI) Jaminan Kesehatan dan peserta Bukan Penerima Bantuan Iuran (non-PBI) Jaminan Kesehatan.

- Peserta Penerima Bantuan Iuran (PBI) Jaminan Kesehatan meliputi peserta yang dianggap layak menerima bantuan iuran, yang mana tergolong fakir miskin dan orang tidak mampu sebagai peserta program Jaminan Kesehatan yang ditetapkan oleh Kementerian Sosial dan iurannya dibayarkan oleh pemerintah pusat. Sedangkan, terdapat pula kepesertaan yang didaftarkan oleh Pemerintah Daerah yang iurannya dibayarkan pula oleh Pemerintah Daerah.
- Peserta Bukan Penerima Bantuan Iuran (non-PBI) Jaminan Kesehatan meliputi peserta yang tidak menerima bantuan iuran dari pemerintah, antara lain seperti:
 - 1) Pekerja Penerima Upah (PPU) dan anggota keluarganya, terdiri dari:
 - a. Pejabat Negara;
 - b. Pimpinan dan anggota Dewan Perwakilan Rakyat Daerah (DPRD);
 - c. Pegawai Negeri Sipil (PNS);
 - d. Prajurit;
 - e. Anggota Polri;
 - f. Kepala desa dan perangkat desa;
 - g. Pegawai swasta;
 - h. Pekerja/pegawai yang tidak termasuk pada poin di atas yang menerima gaji/upah.

Pekerja Bukan Penerima Upah (PBPU) dan anggota keluarganya, terdiri dari:

- a. Pekerja di luar hubungan kerja atau pekerja mandiri; dan



- b. Pekerja yang tidak termasuk poin di atas yang bukan penerima gaji/upah.
- 3) Bukan Pekerja (BP) dan anggota keluarganya, terdiri dari:
- a. Investor
 - b. Pemberi kerja
 - c. Penerima pensiun, terdiri dari:
 - Pejabat negara yang berhenti dengan hak pensiun
 - PNS yang berhenti dengan hak pensiun
 - PNS yang berhenti dengan hak pensiun
 - Prajurit dan Anggota Polri yang berhenti dengan hak pensiun
 - Janda, duda, atau anak yatim dan/atau piatu dari penerima pensiun yang mendapat hak pensiun
 - Penerima pensiun selain poin di atas
 - Janda, duda, atau anak yatim dan/atau piatu dari penerima pensiun lain yang mendapat hak pensiun.
 - d. Veteran
 - e. Perintis Kemerdekaan
 - f. Janda, duda, atau anak yatim dan/atau piatu dari Veteran atau Perintis Kemerdekaan

Anggota keluarga yang ditanggung, yaitu:

- 1) Anggota keluarga inti dari peserta maksimal empat orang, diantaranya dari istri atau suami yang sah, anak kandung, anak tiri dari perkawinan yang sah, dan anak angkat yang sah.
- 2) Anak kandung, anak tiri dari perkawinan yang sah, dan anak angkat yang sah, dengan kriteria:
 - a. Belum menikah atau tidak mempunyai penghasilan sendiri.
 - b. Usia belum mencapai 21 tahun atau masih melanjutkan sekolah sebelum berusia 25 tahun.
- 3) Peserta dapat menambahkan anggota keluarga lainnya seperti anak ke-4 dan seterusnya, ayah, ibu, dan mertua.



2.2.2 Iuran Jaminan Kesehatan Nasional

Iuran Jaminan Kesehatan adalah sejumlah uang yang dibayarkan secara teratur oleh peserta, pemberi kerja, dan/atau Pemerintah Pusat atau Pemerintah Daerah untuk program Jaminan Kesehatan (Peraturan Presiden RI Nomor 82 Tahun 2018 tentang Jaminan Kesehatan). Beberapa perubahan terjadi terkait jumlah iuran yang wajib disetor oleh peserta Jaminan Kesehatan Nasional, hal ini diatur dalam Peraturan Presiden Republik Indonesia Nomor 64 Tahun 2020 mengenai Perubahan Kedua Atas Peraturan Presiden Republik Indonesia Nomor 82 Tahun 2018 tentang Jaminan Kesehatan.

Iuran JKN dihitung berdasarkan persentase upah/penghasilan bagi peserta penerima upah atau suatu jumlah nominal tertentu untuk peserta yang tidak menerima upah yang ditetapkan dalam peraturan terbaru yang berlaku. Iuran tambahan juga dikenakan kepada peserta yang mengikutsertakan anggota keluarga lebih dari lima orang.

2.2.3 Manfaat Jaminan Kesehatan Nasional

Berdasarkan Peraturan Presiden RI Nomor 82 Tahun 2018, hak setiap peserta yaitu memperoleh manfaat jaminan kesehatan. Manfaat tersebut juga berlaku untuk bayi baru lahir paling lama 28 hari setelah bayi tersebut dilahirkan.

Manfaat jaminan kesehatan yang dimaksud meliputi manfaat medis dan manfaat non medis. Manfaat medis diberikan sesuai dengan indikasi medis dan standar pelayanan serta tidak dibedakan berdasarkan besaran iuran Peserta. Sedangkan, manfaat non medis diberikan berdasarkan besaran iuran Peserta. Pelayanan kesehatan yang dijamin dalam manfaat jaminan kesehatan terdiri dari:

- a. Pelayanan kesehatan tingkat pertama, meliputi pelayanan kesehatan nonspesialistik. Misal: administrasi pelayanan; pelayanan promotif dan preventif; pemeriksaan, pengobatan, dan konsultasi medis; tindakan medis nonspesialistik, baik operatif maupun nonoperatif; pelayanan obat, alat kesehatan, dan bahan medis habis pakai; pemeriksaan penunjang diagnostik laboratorium tingkat pertama; dan rawat inap tingkat pertama sesuai dengan indikasi medis. Selanjutnya, manfaat pelayanan promotif dan preventif meliputi pemberian pelayanan:



- 1) Penyuluhan kesehatan perorangan
 - 2) Imunisasi rutin
 - 3) Keluarga Berencana (KB)
 - 4) Skrining riwayat kesehatan dan pelayanan penapisan (skrining tertentu)
 - 5) Peningkatan kesehatan bagi peserta dengan penyakit kronis
- b. Pelayanan kesehatan rujukan tingkat lanjutan, meliputi: administrasi pelayanan; pemeriksaan, pengobatan, dan konsultasi medis dasar; pemeriksaan, pengobatan, dan konsultasi spesialisik; tindakan medis spesialisik, baik bedah maupun nonbedah sesuai dengan indikasi medis; pelayanan obat, alat kesehatan, dan bahan medis habis pakai; pelayanan penunjang diagnostik lanjutan sesuai dengan indikasi medis; rehabilitasi medis; pelayanan darah; pemulasaran jenazah peserta yang meninggal di fasilitas kesehatan; pelayanan keluarga berencana; perawatan inap nonintensif; dan perawatan inap di ruang intensif.
- c. Pelayanan ambulans darat atau air.

2.3 Badan Penyelenggara Jaminan Sosial Kesehatan

Menurut Undang-Undang RI Nomor 24 Tahun 2011, Badan Penyelenggara Jaminan Sosial (BPJS) adalah badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial. BPJS bertujuan untuk mewujudkan terselenggaranya pemberian jaminan sosial dan terpenuhinya kebutuhan dasar hidup yang layak bagi setiap peserta dan anggota keluarganya, Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan adalah lembaga yang bertanggung jawab atas penyelenggaraan program JKN. Pada tanggal 1 Januari 2014 BPJS Kesehatan mulai beroperasi dalam melakukan program jaminan kesehatan.

Dalam pemisahan aset menurut Undang-Undang RI Nomor 24 Tahun 2011, BPJS mengelola dua aset, yaitu aset BPJS dan aset Dana Jaminan Sosial. Setiap bulan, BPJS Kesehatan mengumpulkan iuran jaminan sosial termasuk bantuan

na ini selanjutnya dikelola oleh BPJS Kesehatan kemudian dialokasikan Dana Jaminan Sosial (DJS) kepada masing-masing fasilitas kesehatan yang bekerja sama dengan BPJS Kesehatan. Nominal yang masuk ke



dalam rekening Dana Jaminan Sosial (DJS) sepenuhnya digunakan untuk tujuan pengembangan jaminan sosial dan untuk sebesar-besarnya kepentingan peserta, bukan untuk keperluan operasional BPJS Kesehatan itu sendiri. Tujuan jaminan sosial yang dimaksud yaitu untuk pembayaran manfaat atau pembiayaan operasional penyelenggara program jaminan sosial. Artinya, aset Dana Jaminan Sosial (DJS) bukan merupakan aset BPJS.

BPJS Kesehatan mengelola dana yang dikumpulkan dari iuran peserta dan mengkoordinasikan pemberian pelayanan kesehatan disebut Dana Jaminan Sosial (DJS). Menurut Direktur Utama BPJS Kesehatan, Fachmi Idris, peserta PBPJ adalah kelompok peserta yang seringkali memiliki tunggakan iuran dan memiliki tingkat pembayaran yang rendah. Tunggakan iuran tersebut sejalan dengan besaran nilai piutang Dana Jaminan Sosial (DJS) BPJS Kesehatan yang akan berperan besar dalam determinan nilai defisit atau nilai surplus Dana Jaminan Sosial (DJS), hal itu menjadi salah satu tantangan untuk meningkatkan kolektibilitas iuran peserta.

2.4 Kolektibilitas Iuran

Tagihan iuran PBPJ dan BP bersifat kolektif untuk seluruh anggota keluarga sebagaimana yang terdaftar pada Kartu Keluarga dan/atau yang sudah didaftarkan sebagai anggota keluarga. Metode penggabungan tersebut adalah dengan menggabungkan masing-masing total tagihan peserta. Saldo iuran yang terdapat pada salah satu anggota keluarga tidak dapat dibagikan kepada anggota keluarga lainnya. Berdasarkan Peraturan Presiden RI Nomor 82 Tahun 2018 tentang Jaminan Kesehatan, BPJS Kesehatan wajib mencatat dan menagih tunggakan iuran sebagai piutang BPJS Kesehatan paling banyak untuk 24 (dua puluh empat) bulan.

Secara umum, kolektibilitas tunggakan iuran merujuk pada sejauh mana suatu organisasi atau lembaga berhasil mengumpulkan iuran atau kontribusi yang seharusnya telah dibayar oleh peserta atau anggotanya. Ini adalah indikator penting dalam mengukur keberlanjutan dan keuangan suatu program atau layanan, khususnya dalam konteks jaminan sosial seperti BPJS Kesehatan.

collecting merupakan upaya penagihan iuran JKN peserta menunggakannya PBPJ. Kegiatan ini dilaksanakan melalui panggilan telepon atau WhatsApp dengan tujuan memperoleh konfirmasi dari peserta yang tunggakan apakah mereka akan melakukan pembayaran dan memastikan



mereka melakukan pembayaran iuran yang masih menunggak. Tunggakan iuran tersebut merupakan kewajiban keuangan atau piutang Dana Jaminan Sosial (DJS) BPJS Kesehatan.

Telecollecting adalah salah satu metode implementasi dari konsep pengumpulan iuran (*revenue collection*) dalam industri jaminan sosial kesehatan. Program ini juga menjadi sarana bagi BPJS Kesehatan untuk memberikan pemahaman kepada peserta terkait pentingnya kepatuhan membayar iuran secara rutin sehingga meningkatkan rasa tanggung jawab, meningkatkan jiwa gotong royong, dan meningkatkan rasa kepemilikan terhadap program JKN. *Telecollecting* bukan hanya bertujuan untuk menginformasikan terkait tunggakan iuran. Namun, *telecollecting* juga memberikan informasi lainnya, seperti:

- Memberikan informasi terkait program pembayaran terbaru BPJS Kesehatan atau pemerintah terkait kemudahan dalam pembayaran;
- Memberikan informasi singkat tentang tata cara melakukan pembayaran sesuai dengan metode pembayaran yang ingin dilakukan peserta;
- Memastikan peserta memiliki kesanggupan dalam melakukan pembayaran sesuai dengan prosedur agar tunggakan tidak berkepanjangan;
- Memberikan fasilitas dan solusi atas keluhan dan masalah terkait kesulitan memiliki biaya untuk pembayaran piutang iuran.

Beberapa studi dan penelitian telah dilakukan untuk mengevaluasi efektivitas *telecollecting* dalam meningkatkan kolektibilitas iuran JKN. Hasil penelitian menunjukkan bahwa *telecollecting* dapat meningkatkan tingkat pembayaran iuran dan mengurangi jumlah tunggakan, hal ini sejalan dengan upaya pengurang piutang Dana Jaminan Sosial (DJS) BPJS Kesehatan. Berikut ini tahapan dalam melakukan *telecollecting*:

1. Setiap bulannya, petugas *telecollecting* mengambil data secara acak berdasarkan wilayah dari daftar peserta menunggak untuk dilakukan penagihan melalui telepon atau pesan *WhatsApp*.
2. Melakukan penagihan sesuai dengan standar operasional prosedur dan berdasarkan strategi komunikasi yang telah ditetapkan.

Sistem akan melakukan pencatatan seluruh aktivitas kegiatan *telecollecting* yang telah dilakukan.



4. Selanjutnya, sistem akan mengevaluasi *output* dan status pembayaran iuran peserta yang telah menjalani proses penagihan.

Meskipun *telecollecting* dapat membantu meningkatkan kolektibilitas iuran BPJS Kesehatan, namun berdasarkan fakta yang terjadi di lapangan, masih terdapat beberapa tantangan dalam implementasi *telecollecting*, seperti tingkat respons dari peserta, kemampuan teknis petugas *telecollecting*, dan pengelolaan data yang tepat. Oleh karena itu, evaluasi dan perbaikan terus-menerus dalam pelaksanaan *telecollecting* sangat penting untuk memaksimalkan efektivitasnya dalam meningkatkan kolektibilitas iuran BPJS Kesehatan.

2.5 *Tree-Based Algorithm*

Pohon dan struktur pohon adalah konsep penting dalam metode *tree-based algorithm*. Pohon adalah struktur data non linier yang merepresentasikan hubungan hirarkis antara elemen-elemen. Struktur pohon terdiri dari satu simpul khusus yang disebut akar (*root*) dan beberapa subtree yang terhubung ke akar melalui cabang (*link*). Setiap simpul dalam pohon dapat memiliki informasi dan derajat (jumlah cabang yang keluar dari simpul tersebut). *Tree-based algorithms* adalah algoritma yang menggunakan struktur pohon untuk melakukan operasi pengolahan dan penyimpanan data.

Tree-based algorithms merupakan kelompok algoritma *machine learning* yang menggunakan struktur pohon untuk membuat keputusan atau prediksi. Dalam konteks ini, pohon keputusan (*decision tree*) adalah model yang mengambil serangkaian keputusan untuk mencapai hasil akhir. Struktur pohon ini memungkinkan algoritma untuk memecah ruang fitur input menjadi segmen-segmen yang lebih kecil dan lebih mudah diinterpretasi.

Classification and Regression Tree (CART) adalah salah satu jenis pohon keputusan yang dapat digunakan untuk tugas klasifikasi (mengelompokkan data ke dalam kategori) dan regresi (memprediksi nilai kontinu). Pohon keputusan adalah struktur berhierarki yang memecah *dataset* menjadi subset yang semakin kecil

dan serangkaian keputusan biner yang diambil pada setiap simpulnya. Pada setiap simpul dalam pohon *CART*, suatu keputusan dibuat berdasarkan nilai dari



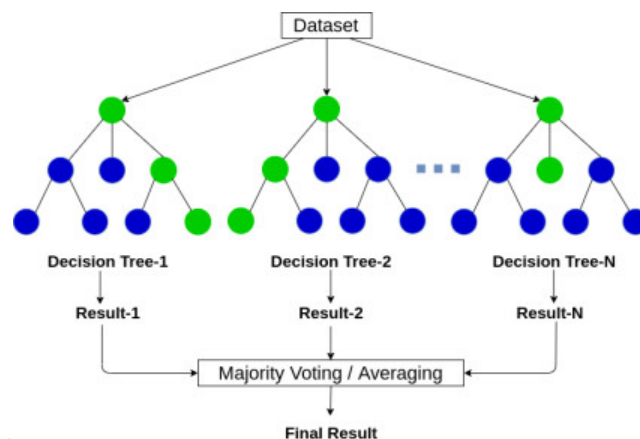
suatu fitur, dan pohon terus bercabang hingga mencapai simpul daun yang memberikan hasil prediksi.

Ensemble learning adalah paradigma di mana beberapa model digabungkan untuk meningkatkan kinerja dan kestabilan prediksi dibandingkan dengan model tunggal. Dalam konteks *tree-based algorithms*, dua pendekatan ensemble learning yang umum digunakan adalah *Random Forest* dan *Gradient Boosting*. Beberapa algoritma berbasis pohon keputusan *CART* dengan model *ensemble learning* yang sering digunakan adalah *Random Forest*, *Light Gradient Boosting Machine (LightGBM)*, dan *Adaptive Boosting (AdaBoost)*.

2.5.1 *Random Forest*

Random Forest merupakan pengembangan dari metode *Classification and Regression Tree (CART)* dan metode *ensemble learning* yang digunakan untuk klasifikasi dan dapat digunakan untuk regresi dalam *machine learning*. Model dasar tersebut dikombinasikan dengan *bagging (bootstrap aggregation)*.

Kumpulan data sampel akan dibagi menjadi beberapa sampel *bootstrap* menghasilkan data latih, kemudian *random forest* menguraikan data latih ke dalam beberapa pohon keputusan secara acak, dan seluruh pohon tersebut digabungkan. Proses penentuan pohon terbaik melibatkan mekanisme *majority voting* untuk klasifikasi atau *average* untuk regresi dari prediksi yang dihasilkan oleh setiap pohon. Hasilnya, prediksi dari masing-masing pohon tersebut dikombinasikan ke dalam satu model yang lebih kokoh Breiman, L. (2001).



Gambar 2.1 Skema Prediksi *Random Forest*



Seperti ilustrasi pada Gambar 2.1 Skema Prediksi Random Forest masing-masing simpul akan mewakili sebuah keputusan, Keputusan didasarkan pada salah satu variabel prediksi saja, dimana untuk setiap keputusan proses akan turun satu tingkat dari pohon ke simpul lainnya hingga mencapai simpul pangkal atau biasanya disebut dengan *leaf* (daun). Daun berisi tingkatan dari respon dan diakhiri dengan menentukan hasil yang kemudian dapat dilakukan pemilihan berdasarkan hasil mayoritas (untuk klasifikasi) dan rata-rata dari seluruh kelas (untuk regresi).

Input sampel dapat dinyatakan sebagai X yang berisi m fitur ($X = x_1, x_2, \dots, x_m$) sedangkan output dinyatakan sebagai Y sehingga *training set* dapat dinyatakan sebagai S_n yang terdiri dari n observasi, secara matematis dapat dinyatakan dengan persamaan (Li *et al.*, 2018):

$$S_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}, X \in \mathbb{R}^m, Y \in \mathbb{R} \quad (2.1)$$

Pada tahap pelatihan setiap pohon keputusan akan melakukan pemecahan data untuk diteruskan ke *decision node* sebagai input baru (X) yang dilakukan berulang-ulang sampai ke tahap *leaf nodes*. Proses *training* akan berhenti jika jumlah maksimum yang ditentukan telah mencapai atau jika *node* telah berisi jumlah sampel yang telah ditentukan. Nilai *output* dari *random forest* untuk prediksi dinyatakan sebagai \hat{Y} yang diperoleh dari nilai rata-rata pohon keputusan, secara matematis dinyatakan sebagai berikut (Li *et al.*, 2018):

$$\hat{Y} = \frac{1}{q} \sum_{l=1}^q \hat{Y}_l = \frac{1}{q} \sum_{l=1}^q \hat{h}(X, S_n^{\theta_l}) \quad (2.2)$$

dimana \hat{Y}_l merupakan *output* ke- l dari pohon ($l = 1, 2, \dots, q$), q merupakan banyaknya pohon, dan $\hat{h}(X, S_n^{\theta_l})$ merupakan fungsi prediksi dari variabel respon pada setiap pohon.

Random forest dapat digunakan untuk menyelesaikan tugas klasifikasi maupun regresi. Beberapa kelebihan *random forest* adalah sebagai berikut (G. Ke *et al.*, 2017):

- Kemampuan dalam mengatasi data pencilan (*outliers*) ataupun *missing data*

na memiliki metode untuk membangun pohon dari data dengan nilai
ut yang hilang.



- Kemampuan untuk memproses data besar secara efisien dengan berbagai karakteristik dan kategori.
- Tidak peka terhadap penskalaan atau transformasi data nilai dalam bentuk apapun.
- Karena *random forest* dapat digunakan untuk klasifikasi dan regresi, maka dapat menangani data diskrit dan kontinu dengan baik.
- Memiliki metode untuk mengevaluasi signifikansi variabel independen dalam model.

Selain kelebihan yang dimiliki, algoritma *random forest* juga memiliki kekurangan yaitu *random forest* membangun sejumlah besar pohon keputusan yang dapat membuat algoritma berjalan lambat dan tidak efisien untuk memprediksi data.

2.5.2 *Light Gradient Boosting Machine (LightGBM)*

Light Gradient Boosting Machine adalah suatu metode *gradient boosting* yang memiliki kecepatan tinggi, terdistribusi, dan kinerja yang unggul, berbasis pada struktur pohon keputusan. *LightGBM* merupakan salah satu metode *ensemble* yang melakukan agregasi terhadap prediksi dari metode *CART*, dengan menambahkan kontribusi dari setiap pohon. *LightGBM* adalah implementasi dari *Gradient Boosting Decision Tree (GBDT)* dengan dua teknik untuk mengurangi kompleksitas dengan menurunkan sampling data dan fitur menggunakan dua teknik yaitu, yaitu *Exclusive Feature Bundling (EFB)* dan *Gradient-Based One-Side Sampling (GOSS)*.

Exclusive Feature Bundling (EFB) adalah metode *LightGBM* yang dapat menggabungkan beberapa fitur (*feature bundling*) yang *mutually exclusive*, yaitu fitur yang tidak saling memengaruhi menjadi satu fitur. Metode ini dapat mengurangi kompleksitas atau *overhead* model dan meningkatkan efisiensi pelatihan terutama untuk *dataset* dengan banyak fitur.

Sedangkan, *LightGBM* menggunakan teknik *Gradient-based One-Side Sampling (GOSS)* adalah suatu teknik yang digunakan dalam *LightGBM* untuk

malkan proses pelatihan. Teknik *GOSS* melakukan penarikan sampel tak hanya dari satu sisi distribusi gradien untuk mengurangi waktu. Teknik ini membantu *LightGBM* untuk mengurangi jumlah data yang



dijadikan sampel selama pelatihan, yang juga membantu meningkatkan efisiensi sehingga berfokus pada bagian yang paling sulit dari ruang fitur yang sedang dipelajari. Gradien pada teknik GOSS nilai yang menunjukkan laju perubahan fungsi loss terhadap model saat ini. Secara matematis, Microsoft memperkenalkan formula estimasi *gain variance* pada teknik GOSS dirumuskan sesuai persamaan (G. Ke *et al.*, 2017):

- 1) Setelah nilai gradien dihitung, metode GOSS memberi data latih peringkat sesuai dengan nilai absolut gradiennya dan diurutkan menurun.
- 2) Data teratas $a \times 100\%$ dengan gradien besar yang lebih besar disimpan, kemudian diperoleh data subset A .
- 3) Untuk himpunan A^c yang tersisa yang terdiri dari $(1 - a) \times 100\%$ data dengan gradien yang lebih kecil, dipilih secara acak membentuk subset B berukuran $b \times |A^c|$.
- 4) Data akan dibagi berdasarkan nilai perkiraan varian pada vektor $V_j(d)$ yang didefinisikan sebagai berikut:

$$V_j(d) = \frac{1}{n} \left(\frac{\left(\sum_{x_i \in A_l} r_i + \frac{1-a}{b} \sum_{x_i \in B_l} r_i \right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} r_i + \frac{1-a}{b} \sum_{x_i \in B_r} r_i \right)^2}{n_r^j(d)} \right) \quad (2.3)$$

dimana:

$V_j(d)$ = Estimasi *gain variance* setelah subset $A \cup B$,

$A_l = \{x_i \in A: x_{ij} \leq d\}, A_r = \{x_i \in A: x_{ij} > d\}$,

$B_l = \{x_i \in B: x_{ij} \leq d\}, B_r = \{x_i \in B: x_{ij} > d\}$,

$n_l^j(d) = \sum l(x_i \in (A_l \cup B_l)), n_r^j(d) = \sum l(x_i \in (A_r \cup B_r))$,

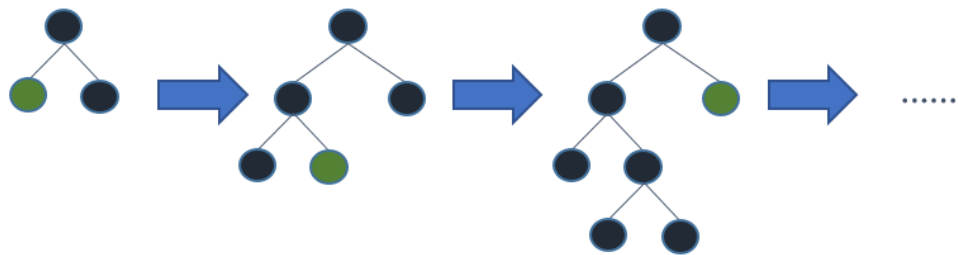
$\frac{1-a}{b}$ = Koefisien yang digunakan untuk menormalkan jumlah gradien pada B ke ukuran A^c , dengan A adalah himpunan data dengan gradien besar dan B adalah himpunan data dengan gradien kecil. g_i merupakan gradien dari setiap data, a adalah rasio sampling dari data gradien besar, b adalah rasio sampling dari data gradien kecil, dan d adalah jumlah iterasi.

Konsep *leaf-wise growth* merupakan salah satu teknik dalam *LightGBM*

merupakan modifikasi dari algoritma *CART* dimana semua daun yang diuji sebelum melakukan pembelahan, dan pembelahan hanya dilakukan pada daun yang memiliki delta kesalahan terbesar. Delta kesalahan terbesar



adalah nilai yang menunjukkan perbedaan antara kesalahan fungsi loss sebelum dan sesudah melakukan pembelahan pada suatu daun. Hal ini cenderung menghasilkan pohon asimetris yang sangat dalam yang dapat dengan mudah menyesuaikan data. Pendekatan ini tidak hanya membantu menghindari penggunaan daya komputasi yang berlebihan tetapi juga mengurangi risiko *overfitting* pada model, sehingga mampu mengontrol kedalaman model. Konsep dari pertumbuhan *leaf-wise* akan diilustrasikan dalam Gambar 2.2 Leaf-Wise Tree Growth



Gambar 2.2 Leaf-Wise Tree Growth

LightGBM memiliki kecepatan yang tinggi dalam memproses data dengan tingkat akurasi yang lebih baik dibandingkan dengan model pohon lainnya. Beberapa kelebihan penggunaan *LightGBM* dalam *machine learning* yaitu:

- *LightGBM* menggunakan algoritma berbasis histogram untuk memasukkan nilai fitur berkelanjutan ke dalam bin diskrit, yang menghasilkan proses pelatihan yang lebih cepat dan efisien.
- *LightGBM* mengganti nilai persisten ke lokasi terpisah, hal ini dapat mengurangi penggunaan memori.
- *LightGBM* menghasilkan pohon yang jauh lebih kompleks dengan mengikuti pemisahan daun daripada metode level, yang merupakan faktor kunci dalam mencapai akurasi yang lebih tinggi dibandingkan metode *boosting* lainnya.
- Memungkinkan untuk mengelola *dataset* berukuran besar dengan waktu pelatihan yang berkurang secara signifikan.

Namun, model *LightGBM* juga memiliki kelemahan yaitu cukup sensitif *imbalance data* yang dapat mengakibatkan *overfitting* yang membuat pembelajaran akan lebih mengarah pada kelas data yang lebih dominan.



2.5.3 Adaptive Boosting (AdaBoost)

Algoritma Adaptive Boosting (AdaBoost) adalah sebuah teknik pembelajaran mesin yang digunakan untuk meningkatkan akurasi dan kinerja model prediksi dengan menggabungkan beberapa model sederhana (*weak learners*) menjadi satu model kuat (*strong learner*). Algoritma ini termasuk dalam metode ensemble, yaitu metode yang menggabungkan banyak elemen menjadi satu kesatuan. AdaBoost merupakan salah satu variasi dari berbagai algoritma boosting, yaitu algoritma yang bekerja dengan cara memberikan bobot lebih pada data yang salah diklasifikasikan oleh model sebelumnya. Berbeda dengan LightGBM, algoritma yang menggunakan pohon-pohon yang kompleks dan berdasarkan gradien, algoritma AdaBoost berfokus pada pengumpulan pohon-pohon yang lemah dalam memprediksi. Pohon-pohon yang lemah adalah pohon keputusan yang memiliki akurasi prediksi yang rendah, hampir sama dengan tebakan acak. Selanjutnya, pohon-pohon klasifikasi tersebut diberi bobot yang sama untuk setiap pengamatan. Pengamatan adalah data yang digunakan untuk melatih model. Setelah mengevaluasi pohon pertama, bobot ditambahkan pada pengamatan yang sulit diklasifikasikan, dan diturunkan pada pengamatan yang mudah diklasifikasikan. Proses ini diulang hingga menghasilkan pohon terbaik untuk jumlah iterasi yang ditentukan. Iterasi adalah jumlah kali model dilatih dan diuji. Algoritma *Adaboost* dapat dijelaskan melalui tahapan-tahapan berikut (X. Liu *et al.*, 2017):

Input:

Diberikan kumpulan data latih dengan label $S = \{(x_n, y_n), n = 1, 2, \dots, N\}$, di mana $y_n \in \{-1, +1\}$

Inisialisasi:

Nilai bobot sampel latih pada iterasi pertama $w_n^{(1)} = \frac{1}{N}$, untuk $n = 1, 2, \dots, N$; $t = 1$ dengan jumlah maksimum pembelajar lemah T .

Ketika $t \leq T$, maka:

1) *Training Resampling* merupakan proses memilih ulang sebagian data latih dengan menggunakan bobot sampel sebagai probabilitas pemilihan dinyatakan

dan $TR_t = \{(x_n, y_n), n = 1, 2, \dots, N\}$;

melakukan pembelajar lemah ke- t dengan melatih TR_t menggunakan algoritma



3) Hitung bobot error (ε_t)

$$\varepsilon_t = \sum_{n=1}^N w_n^{(t)} \cdot I(h_t(x_n) \neq y_n); \quad (2.4)$$

di mana bobot error adalah proporsi sampel yang salah diklasifikasikan oleh pembelajar lemah ke- t .

4) Jika $\varepsilon_t > 0,5$ atau $\varepsilon_t = 0$ maka buat bobot konsisten

$$w_n^{(1)} = \frac{1}{N}, n = 1, 2, \dots, N \quad (2.5)$$

5) Hitung bobot voting (α_t) pada pembelajar lemah ke- t dengan menggunakan rumus berikut:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right); \quad (2.6)$$

di mana bobot voting adalah nilai yang menunjukkan seberapa besar kontribusi pembelajar lemah pada model akhir. Bobot voting digunakan untuk menggabungkan prediksi dari semua pembelajar lemah dengan cara pemungutan suara tertimbang.

6) Perbarui bobot sampel latih pada iterasi berikutnya

$$w_n^{(t+1)} = w_n^{(t)} \exp(-\alpha_t h_t(x_n) y_n); \quad (2.7)$$

7) Normalisasi

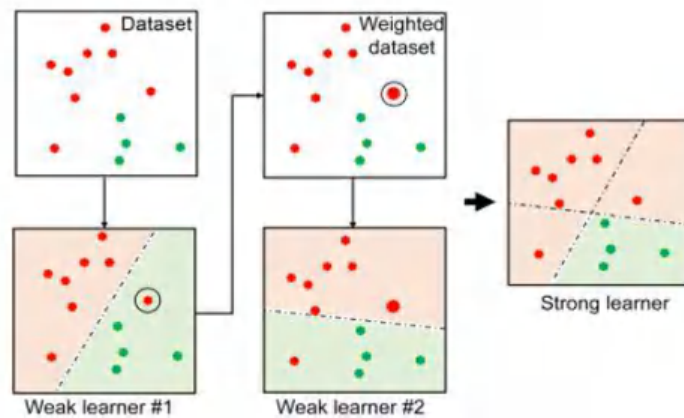
$$w_n^{(t+1)} = \frac{w_n^{(t+1)}}{\sum_{n=1}^N w_n^{(t+1)}}; t = (t + 1) \quad (2.8)$$

Output:

Setelah mencapai maksimum iterasi, maka output *hypothesis final* (h_f) diperoleh dengan pemungutan suara tertimbang untuk contoh pengujian x_n , di mana T adalah jumlah maksimum pembelajar lemah dan $h_t(x_n)$ adalah prediksi pembelajar lemah ke- t pada sampel ke- n :

$$h_f(x_n^*) = \text{sgn} \left(\sum_{t=1}^T \alpha_t h_t(x_n) \right) \quad (2.9)$$





Gambar 2.3 Ilustrasi AdaBoost Algorithm

Keuntungan *AdaBoost* adalah mudah digunakan tanpa harus mengubah banyak parameter, *AdaBoost* dapat mengurangi risiko *overfitting*, dan mampu mengatasi *imbalance data*. Kekurangannya adalah sensitif terhadap *noise* atau *outlier* pada data, memerlukan lebih banyak waktu dalam prosesnya karena bersifat iterative, dan *Adaboost* membutuhkan sampel pelatihan yang cukup panjang.

Random Forest, *Light Gradient Boosting Machine (LightGBM)*, dan *Adaptive Boosting (AdaBoost)* merupakan algoritma berbasis pohon atau (*tree-based algorithms*) yang sering digunakan dalam metode *ensemble* ini. Algoritma berbasis pohon ini digunakan secara luas dalam berbagai tugas seperti klasifikasi, regresi, dan penggalian data. Keuntungan utama mereka meliputi kemampuan interpretasi, ketahanan terhadap data yang tidak teratur, dan kemampuan untuk menangani fitur kategori. Namun, perlu diperhatikan bahwa mereka dapat menjadi rentan terhadap *overfitting*, dan beberapa teknik seperti *pruning* atau pengaturan parameter yang baik diperlukan untuk mengoptimalkan kinerjanya.

2.5.4 Evaluasi Model *Machine Learning*

Untuk memahami sejauh mana model tersebut dapat memberikan prediksi yang akurat dan berguna, maka diterapkan evaluasi model machine learning yang merupakan tahap penting dalam pengembangan model. Pada fase ini, dilakukan evaluasi terhadap hasil pengujian tiap model, dan selanjutnya dipilih model yang

sesuai untuk diterapkan pada kasus dan jenis data yang ada. Oleh karena itu, yang telah diidentifikasi sebagai paling akurat untuk diimplementasikan. Ketidaktepatan dalam kegiatan *telecollecting* harus dapat diinterpretasi, diukur,



dan mampu memenuhi tujuan yang telah ditetapkan dalam konteks penelitian ini. Sehingga, dapat disimpulkan dan disarankan langkah-langkah yang tepat terkait penggunaan model dalam kegiatan *telecollecting*.

1) *Confusion Matrix*

Confusion matrix adalah tabel yang umumnya digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model terhadap nilai sebenarnya dari data uji. Terdapat empat sel dalam *confusion matrix*, masing-masing mewakili kategori hasil prediksi (*predicted class*) dan kategori sebenarnya (*actual class*). Keempat kategori tersebut adalah:

- *True Positive (TP)*: Nilai yang muncul ketika model prediksi menunjukkan hasil positif dan nilai aktualnya benar, maka *correctly detected*.
- *True Negative (TN)*: Nilai yang muncul ketika model prediksi menunjukkan hasil negatif dan nilai aktualnya salah, maka *correctly rejected*.
- *False Positive (FP)*: Nilai yang muncul ketika model prediksi menunjukkan hasil positif dan nilai aktualnya salah (Kesalahan Tipe I), maka *incorrectly detected*.
- *False Negative (FN)*: Nilai yang muncul ketika model prediksi menunjukkan hasil negatif dan nilai aktualnya benar (Kesalahan Tipe II), maka *incorrectly rejected*.

Tabel 2.1 *Confusion Matrix*

		Predicted Class	
		0	1
Actual Class	0	True Negative (TN)	<i>False Positive (FP) Type I Error</i>
	1	<i>False Negative (FN) Type II Error</i>	True Positive (TP)

Confusion matrix membantu memberikan gambaran yang lebih rinci tentang model, khususnya dalam konteks klasifikasi biner. Dengan menggunakan metrik yang dihasilkan dari *confusion matrix*, seperti *accuracy*, *precision*,



recall, dan *F1-score* dapat mendapatkan pemahaman yang lebih komprehensif tentang kemampuan model. Berikut ini penjelasan mengenai ukuran evaluasi yang merupakan turunan dari *confusion matrix*:

a) *Accuracy*

Akurasi (*accuracy*) adalah suatu metrik evaluasi yang mengukur sejauh mana model klasifikasi memberikan prediksi yang benar secara keseluruhan. Metrik ini dihitung dengan membagi jumlah prediksi yang benar (*True Positive* dan *True Negative*) dengan jumlah *total instance* dalam *dataset*. Secara matematis, akurasi dapat dijelaskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.10)$$

Penting untuk dicatat bahwa, meskipun akurasi memberikan gambaran umum tentang kinerja model, itu mungkin tidak mencerminkan secara akurat kinerja model pada kelas minoritas atau dalam situasi ketidakseimbangan kelas.

b) *Precision*

Precision adalah salah satu metrik evaluasi yang digunakan dalam konteks klasifikasi pada *machine learning*. Metrik ini mengukur sejauh mana prediksi positif model benar-benar relevan atau akurat. *Precision* dihitung dengan menggunakan rumus:

$$Precision = \frac{TP}{TP + FP} \quad (2.11)$$

Precision memberikan gambaran tentang seberapa akurat model dalam mengidentifikasi kelas positif. Jika *precision* tinggi, berarti sebagian besar dari yang diprediksi sebagai positif oleh model memang benar-benar positif.

c) *Recall/Sensitivity*

Recall, juga dikenal sebagai *sensitivitas* atau *true positive rate*, adalah metrik evaluasi yang mengukur kemampuan model untuk mengidentifikasi semua *instance* positif yang seharusnya. Secara matematis, *recall* dihitung sebagai rasio antara *True Positive* (TP) dan jumlah total *instance* positif, yang dapat dijelaskan dengan rumus:



$$Recall = \frac{TP}{TP + FN} \quad (2.12)$$

Recall memberikan gambaran tentang sejauh mana model dapat menangkap atau "mengingat" *instance* positif yang seharusnya. Sebuah nilai *recall* yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik untuk mengidentifikasi *instance* positif, tetapi perlu diperhatikan bahwa *trade-off* dengan metrik lain seperti presisi perlu dipertimbangkan tergantung pada konteks aplikasi.

d) *F1-score*

F1-score adalah suatu metrik evaluasi yang mengukur keseimbangan antara presisi (*precision*) dan *recall* pada suatu model klasifikasi. *F1-score* adalah *harmonic mean* dari *precision* dan *recall*, memberikan gambaran komprehensif tentang performa model, terutama ketika terdapat ketidakseimbangan antara kelas positif dan negatif.

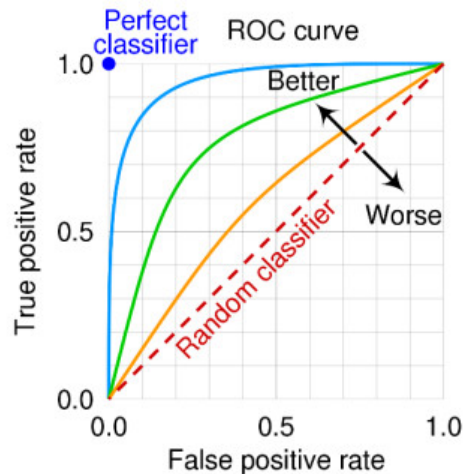
$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (2.13)$$

F1-score memberikan skor yang tinggi jika dan hanya jika baik presisi maupun *recall* memiliki nilai yang tinggi. Oleh karena itu, *F1-score* cocok digunakan ketika kita tertarik untuk memiliki keseimbangan yang baik antara kedua metrik tersebut.

2) *Area Under the Curve (AUC)*

Area Under the Curve (AUC) adalah metrik evaluasi yang umum digunakan untuk mengukur kinerja model klasifikasi, terutama pada kurva *Receiver Operating Characteristic (ROC)*. *ROC curve* membandingkan *trade-off* antara *sensitivity (true positive rate)* dan *specificity (true negative rate)* pada berbagai nilai ambang batas prediksi.



Gambar 2.4 Ilustrasi *ROC Curve* dan *AUC*

AUC mencerminkan seberapa baik model dapat membedakan antara kelas positif dan negatif. Nilai AUC berkisar antara 0 hingga 1, semakin besar nilai AUC, maka semakin akurat yang memberikan gambaran umum tentang seberapa baik model dapat memprediksi kelas secara keseluruhan. Dapat diinterpretasikan dalam formula berikut:

$$AUC = \int_0^1 ROC(t) dt \quad (2.14)$$

3) *k-Fold Cross Validation*

k-Fold Cross Validation adalah teknik validasi model yang umum digunakan dalam *machine learning*. Teknik ini membagi data menjadi k bagian yang seimbang atau *fold*. Pada setiap iterasi, satu kelompok yang berbeda akan diambil sebagai data uji, dan $k-1$ kelompok sisanya digunakan untuk melatih model. *k-Fold Cross Validation* akan mengevaluasi algoritma untuk setiap sampel dan memberikan akurasi rata-rata untuk *dataset* (Karas, 2020). Manfaat dari *k-Fold Cross Validation* adalah memberikan gambaran yang lebih baik tentang seberapa baik model akan berperilaku pada data yang belum pernah dilihat sebelumnya dan mengurangi risiko *overfitting*.

Cara kerja *k-Fold Cross Validation* adalah sebagai berikut:

1. *Dataset* akan dibagi ke dalam n *fold*.

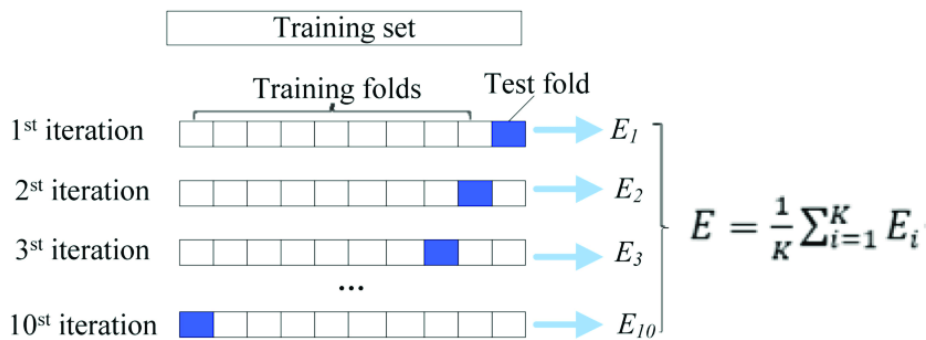
fold ke-1 adalah ketika *fold* pertama menjadi data uji dan lipatan kemudian diuji kepada data latih. Pada data latih, dilakukan pemodelan, kemudian



diuji kepada data latih. Selanjutnya, dilakukan perhitungan tingkat akurasi dari model yang digunakan terhadap data.

$$Accuracy = \frac{\sum \text{data uji yang diprediksi benar}}{\sum \text{data uji}} \quad (2.15)$$

3. *Fold* ke-2 adalah ketika lipatan ke-2 menjadi data uji dan *fold* lainnya menjadi data latih. Selanjutnya dilakukan perhitungan akurasi.
4. Pemilihan data uji dan data latih dilakukan seterusnya sampai dengan jumlah *fold* ke- k .
5. Selanjutnya, dilakukan perhitungan rata-rata dan standar deviasi dari k akurasi yang telah didapatkan untuk dijadikan akurasi final.



Gambar 2.5 Skema k -fold Cross Validation dengan $k = 10$

Skema *10-fold cross validation* adalah salah satu variasi dari skema *k-fold cross validation* yang menggunakan nilai $k = 10$. Artinya, data dibagi menjadi 10 bagian yang seimbang, dan setiap bagian mendapat giliran untuk menjadi data uji, sementara sembilan bagian lainnya menjadi data latih. Proses ini diulang sebanyak 10 kali, dan hasil evaluasi model diambil rata-rata dari semua iterasi. Skema *10-fold cross validation* memberikan estimasi kinerja model yang lebih stabil dan kurang bias dibandingkan dengan skema validasi lainnya. Skema ini juga dapat mengurangi waktu komputasi karena tidak perlu menggunakan semua data untuk melatih model.

