

**ANALISIS KLASIFIKASI *FINANCIAL DISTRESS* DENGAN
MEMBANDINGKAN METODE *EXTREME GRADIENT
BOOSTING* DAN *ARTIFICIAL NEURAL NETWORK*
MENGUNAKAN RASIO KEUANGAN PADA
PERUSAHAAN PERBANKAN TAHUN 2013 – 2022**



NUR FEBRIANTI BAKRI

H081201041

PROGRAM STUDI ILMU AKTUARIA

DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

2024



**ANALISIS KLASIFIKASI *FINANCIAL DISTRESS* DENGAN
MEMBANDINGKAN METODE *EXTREME GRADIENT
BOOSTING* DAN *ARTIFICIAL NEURAL NETWORK*
MENGUNAKAN RASIO KEUANGAN PADA
PERUSAHAAN PERBANKAN TAHUN 2013 – 2022**

SKRIPSI

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

NUR FEBRIANTI BAKRI

H081201041

PROGRAM STUDI ILMU AKTUARIA

DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2024



HALAMAN PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Analisis Klasifikasi *Financial Distress* dengan Membandingkan Metode *Extreme Gradient Boosting* dan *Artificial Neural Network* Menggunakan Rasio Keuangan pada Perusahaan Perbankan Tahun 2013 – 2022

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 28 Februari 2024



Nur Febrianti Bakri

H081201041



**ANALISIS KLASIFIKASI *FINANCIAL DISTRESS* DENGAN
MEMBANDINGKAN METODE *EXTREME GRADIENT BOOSTING*
DAN *ARTIFICIAL NEURAL NETWORK* MENGGUNAKAN RASIO
KEUANGAN PADA PERUSAHAAN PERBANKAN TAHUN
2013 – 2022**

Disetujui oleh:

UNIVERSITAS HASANUDDIN

Pembimbing Utama,



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

Pada 28 Februari 2024




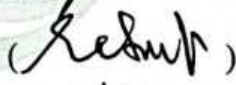

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh

Nama : Nur Febrianti Bakri
NIM : H081201041
Program Studi : Ilmu Aktuaria
Judul Skripsi : Analisis Klasifikasi *Financial Distress* dengan
Membandingkan Metode *Extreme Gradient Boosting*
dan *Artificial Neural Network* Menggunakan Rasio
Keuangan pada Perusahaan Perbankan Tahun 2013 –
2022

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

Ketua : Edy Saputra Rusdi, S.Si., M.Si. ()
Anggota : Prof. Dr. Hasmawati, M.Si. ()
Anggota : Ainun Mawaddah Abdal, S.Si., M.Si. ()



kan di : Makassar

l : 28 Februari 2024

HALAMAN PENGESAHAN

**ANALISIS KLASIFIKASI *FINANCIAL DISTRESS* DENGAN MEMBANDINGKAN
METODE *EXTREME GRADIENT BOOSTING* DAN *ARTIFICIAL NEURAL
NETWORK* MENGGUNAKAN RASIO KEUANGAN PADA PERUSAHAAN
PERBANKAN TAHUN 2013 – 2022**

Disusun dan diajukan oleh
NUR FEBRIANTI BAKRI
H081201041

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, 28 Februari 2024

Dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,



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

Ketua Program Studi,



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



KATA PENGANTAR

Bismillahirrahmanirrahim

Assalamu'alaikum Wr. Wb.

Puja dan puji syukur kehadiran Allah SWT atas anugerah, rahmat, kasih dan izin-Nya sehingga penulis dapat menyelesaikan penelitian dan penyusunan skripsi dengan judul “**Analisis Klasifikasi *Financial distress* dengan Membandingkan Metode *Extreme Gradient Boosting* dan *Artificial Neural Network* Menggunakan Rasio Keuangan pada Perusahaan Perbankan Tahun 2013 – 2022**”. Sebagai salah satu syarat untuk mendapatkan gelar Sarjana Aktuaria pada Program Studi Ilmu Aktuaria Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan yang disebabkan oleh keterbatasan penulis. Tanpa adanya bantuan, arahan dan bimbingan dari berbagai pihak, penyusunan skripsi ini tidak dapat terselesaikan dengan baik. Dengan penuh kerendahan hati dan keikhlasan, penulis haturkan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta yakni Bapak **Bakri Salim** dan Ibu **Nurwati** yang selalu memberikan doa, kasih sayang, semangat, perhatian, pengorbanan dan motivasi kepada penulis. Pada kesempatan ini pula, penulis hendak menyampaikan terima kasih kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya, serta Bapak **Dr. Eng. Amiruddin** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Bapak **Dr. Khaeruddin, M.Sc** selaku Wakil Bidang Akademik dan Kemahasiswaan serta para Staf Administrasi Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah membantu dan memudahkan penulis terhadap kendala dan hal administrasi yang banyak dialami oleh penulis.

Bapak **Dr. Firman, S.Si., M.Si.** selaku Ketua Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin terutama pada Bapak dan Ibu **Dosen Program Studi Ilmu Aktuaria** yang telah



memberikan banyak ilmu dan pengetahuan kepada penulis selama menjadi mahasiswa di Program Studi Ilmu Akuaria.

4. Ibu **Prof. Dr. Hasmawati, M.Si.**, selaku Ketua Program Studi Ilmu Aktuaria yang senantiasa membantu dalam hal memberikan arahan selama masa studi penulis hingga penyusunan skripsi maupun administrasi penulis.
5. Bapak **Edy Saputra Rusdi, S.Si., M.Si.** selaku Dosen Pembimbing Utama yang dengan sabar, tulus, dan ikhlas meluangkan banyak waktu di tengah kesibukan dan prioritasnya untuk membimbing dan memberi masukan serta motivasi dalam penulisan skripsi ini.
6. Ibu **Prof. Dr. Hasmawati, M.Si.** dan Ibu **Ainun Mawaddah Abdal, S.Si., M.Si.** selaku Tim Penguji yang telah meluangkan waktunya untuk memberikan masukan dan kritikan yang membangun terhadap penyempurnaan penulisan skripsi ini.
7. Saudara kandung penulis yakni kakak **Nur Ilham Bakri** dan Adik **Nur Febriana Bakri** dan **Nur Abdillah Bakri**, yang menjadi *special support system* penulis.
8. Sahabat penulis **Sahrul Yudha Fahrezi** yang telah banyak membantu dan memberikan motivasi kepada penulis untuk mendapatkan ide mulai dari proposal hingga penyelesaian skripsi ini.
9. Teman saya, **Wahyudiningrat, Nawa Ehza Fadhillah, Nur Indah Fauziah Tahir**, dan **Nunung Novitasari**, yang telah membantu penulis dalam memperbaiki tata kepenulisan dalam skripsi ini.
10. Sahabat **IPPMP-UH** penulis, Hasmawati, Riskiah Nurul Fajri, Hariati Ananda Bohari, Nasrullah, Saiful, Muh. Rezky Awaliyah, Fathul Kabir, Arif Al Imran, Ince Fachrul Islam S, yang telah kebersamai penulis mulai dari berorganisasi maupun kegiatan lainnya selama masa studi sarjana.
11. Sahabat **Shafiyah Akid** penulis, Dwi Sholihatin M, Illyin Fayza A.Z, Niar Nurwahyuni B, Putri Awaliah A, dan Zhuhrah Rizqah P, terima kasih atas kebersamaan dan dukungannya selama ini.



12. Organisasi **Ikatan Pemuda Pelajar Mahasiswa Pangkep Universitas Hasanuddin** yang telah menjadi rumah dan tempat berkembang selama penulis menempuh Pendidikan di Universitas Hasanuddin
13. Semua pihak yang tidak dapat di sebutkan satu per satu, yang telah memberikan do'a, dukungan, motivasi dan inspirasi bagi penulis dalam mengerjakan skripsi ini.

Akhir kata, penulis berharap semoga segala bentuk kebaikan yang telah diberikan bernilai ibadah dan mendapatkan balasan dari Allah SWT. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.



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

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

Nama : Nur Febrianti Bakri
Nim : H081201041
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:

Analisis Klasifikasi *Financial Distress* dengan Membandingkan Metode *Extreme Gradient Boosting* dan *Artificial Neural Network* Menggunakan Rasio Keuangan pada Perusahaan Perbankan Tahun 2013 – 2022

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 28 Februari 2024

Yang Menyatakan


Nur Febrianti Bakri



ABSTRAK

Financial distress mengacu pada tahap kemunduran kondisi keuangan perusahaan yang terjadi sebelum kebangkrutan. Klasifikasi *financial distress* merupakan hal yang menguntungkan bagi perusahaan, investor, dan kreditur. Fokus utama dalam penelitian ini adalah pada sektor perbankan di Indonesia. Variabel yang dipertimbangkan adalah rasio keuangan, antara lain CAR, ROA, ROE, NIM, LDR, BOPO, dan NPL. Penelitian ini menggunakan dua metode, yaitu *Extreme Gradient Boosting* dan *Artificial Neural Network*. Dengan membandingkan tingkat akurasi dari kedua metode tersebut, dapat dilihat metode mana yang memiliki kinerja yang lebih baik. Hasilnya menunjukkan bahwa metode *Extreme Gradient Boosting* mencapai tingkat akurasi 99.4%, sedangkan metode *Artificial Neural Network* mencapai tingkat akurasi 98,8%. Oleh karena itu, metode *Extreme Gradient Boosting* dikatakan lebih baik daripada metode *Artificial Neural Network* dalam mengklasifikasikan perusahaan sektor perbankan secara efektif sebagai perusahaan yang mengalami kesulitan keuangan atau tidak.

Kata Kunci : *Artificial Neural Network, Financial distress, Extreme Gradient Boosting*

Judul : Analisis Klasifikasi *Financial distress* dengan Membandingkan Metode *Extreme Gradient Boosting* dan *Artificial Neural Network* Menggunakan Rasio Keuangan pada Perusahaan Perbankan Tahun 2013 – 2022

Nama : Nur Febrianti Bakri

: H081201041

Studi : Ilmu Aktuaria



ABSTRACT

Financial distress refers to a stage of deterioration in a company's financial condition that occurs prior to bankruptcy. Predicting financial distress is advantageous for companies, investors, and creditors. In this study, the focus is on the banking sector in Indonesia. The variables considered are financial ratios, including CAR, ROA, ROE, NIM, LDR, BOPO, and NPL. The research employs two methods, namely Extreme Gradient Boosting and Artificial Neural Network, to analyze the data. By comparing the accuracy levels of both methods, it is possible to determine which one performs better. The results indicate that the Extreme Gradient Boosting method achieves an accuracy rate of 99.4%, while the Artificial Neural Network method achieves an accuracy rate of 98.8%. Therefore, the Extreme Gradient Boosting method outperforms the Artificial Neural Network method in effectively classifying banking sector companies as either facing financial distress or not.

Keywords : Artificial Neural Network, Financial distress, Extreme Gradient Boosting

Title : Analysis of Financial distress Classification by Comparing Extreme Gradient Boosting and Artificial Neural Network Methods Using Financial Ratios in Banking Companies in 2013 - 2022

Name : Nur Febrianti Bakri

ID : H081201041

Program : Actuarial Science



DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR.....	vii
PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	x
ABSTRAK	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL.....	xvii
DAFTAR NOTASI.....	xviii
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
Ruang Penelitian.....	6
Manfaat Penelitian.....	6



1.6 Sistematika Penulisan 7

BAB II. TINJAUAN PUSTAKA 8

2.1 Pengertian *Financial distress* 8

2.2 *Earning Per Share* 10

2.3 Rasio Keuangan..... 10

 2.3.1 Rasio Likuiditas 10

 2.3.2 Rasio Rentabilitas 11

 2.3.3 Rasio Aktiva Produktif..... 12

 2.3.4 Rasio Permodalan Bank..... 12

2.4 Statistika Deskriptif..... 14

2.5 Korelasi *Spearman* 14

2.6 Normalisasi Data 15

2.7 *Balancing Data* 15

2.8 Teknik Ensemble 17

2.9 *Boosting*..... 17

2.10 *Extreme Gradient Boosting (XGBoost)* 18

 2.10.1 Fungsi Objektif (*Objective Function*)..... 18

 2.10.2 *Decision tree Ensemble*..... 20

 2.10.3 Cara Kerja XGBoost..... 20



2.12 *Tunning Hyperparameter* 28

2.13 *Artificial Neural Network (ANN)*..... 30

 2.13.1 *Pengertian Artificial Neural Network* 30

 2.13.2 *Komponen Artificial Neural Network*..... 30

 2.13.3 *Jenis Fungsi Aktivasi* 31

 2.13.4 *Tahapan pada Metode Artificial Neural Network (ANN)*..... 33

2.14 *Metrik Klasifikasi pada Machine Learning*..... 35

BAB III. METODOLOGI PENELITIAN 38

 3.1 *Sumber Data* 38

 3.2 *Populasi dan Sampel Penelitian*..... 38

 3.3 *Variabel Penelitian*..... 39

 3.4 *Metode Analisis* 40

 3.5 *Tahapan Penelitian*..... 40

BAB IV. HASIL DAN PEMBAHASAN 43

 4.1 *Statistika Deskriptif*..... 43

 4.2 *Normalisasi Data* 48

 4.3 *Balancing Data* 48

 4.4 *Pembagian Data Testing dan Data Training* 49

 4.5 *Metode Extreme Gradient Boosting* 50



4.5.1 Model XGBoost.....	51
4.5.2 Prediksi XGBoost.....	52
4.6 Metode <i>Artificial Neural Network</i> (ANN)	53
4.6.1 Arsitektur <i>Artificial Neural Network</i> (ANN).....	54
4.6.2 Inisialisasi Bobot.....	55
4.6.3 Pelatihan Model	57
4.6.4 Pengujian Model.....	60
4.6.5 Perbandingan Klasifikasi Metode XGBoost dan ANN.....	62
BAB V. KESIMPULAN DAN SARAN.....	64
5. 1 Kesimpulan.....	64
5.2 Saran	65
DAFTAR PUSTAKA.....	66
LAMPIRAN.....	73



DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Proses *Ensemble* 17

Gambar 2. 2 Ilustrasi Proses *Boosting*..... 18

Gambar 2. 3 Ilustrasi Penggunaan *Regularization Term* 20

Gambar 2. 4 Contoh Membangun Model XGBoost..... 23

Gambar 2. 5 Contoh Perhitungan *Similarity* dan *Gain*..... 24

Gambar 2. 6 Contoh *Split* pada Pohon XGBoost 24

Gambar 2. 7 Contoh *Split* pada Turunan Percabangan 25

Gambar 2. 8 Contoh Perhitungan *Similarity* dan *Gain* pada *Split* Lanjutan..... 25

Gambar 2. 9 Proses *Prunning* pada Pohon 26

Gambar 2. 10 Perhitungan *Output Value* 26

Gambar 2. 11 Komponen ANN 30

Gambar 2. 12 Fungsi Aktivasi Sigmoid Biner..... 32

Gambar 2. 13 Fungsi Aktivasi ReLu 32

Gambar 2. 14 Contoh *Confusion matrix*..... 35

Gambar 3. 1 *Flowchart*..... 42

Gambar 4. 1 Perbandingan Data *Financial Distress* 43

Gambar 4. 2 Perbandingan Nilai CAR & NIM Tahun 2013 - 2022 44

Gambar 4. 3 Perbandingan Nilai ROA & ROE Tahun 2013 - 2022 45

Gambar 4. 4 Perbandingan Nilai BOPO dan LDR Tahun 2013 – 2022. 46

Gambar 4. 5 Perbandingan Nilai NPL Tahun 2013 - 2022..... 46

Gambar 4. 6 Hasil Korelasi Data Menggunakan Metode *Spearman*..... 47

Gambar 4. 7 Tampilan Dataset Setelah Dilakukan Normalisasi..... 48

Gambar 4. 8 Data Target *Before* dan *After* Dilakukan *Over Sampling*..... 49

Gambar 4. 9 *Confusion matrix* dari Hasil Pelatihan Model 53

Gambar 4. 10 Arsitektur ANN yang Akan Digunakan 54

Gambar 4. 11 Bobot dan Bias Awal dari *Input Layer* ke *Hidden layer* 56

Gambar 4. 12 Nilai Data *Input Training*..... 57

Gambar 4. 13 Tampilan Sinyal *Input Layer* ke *Hidden layer*..... 58



Gambar 4. 14 Tampilan Nilai *output Hidden layer* Pertama yang Telah diberikan Fungsi Aktivasi..... 59

Gambar 4. 15 Tampilan Nilai *Error* Setelah Melakukan Pelatihan Model Dengan 25 *Epoch*..... 60

Gambar 4. 16 *Confusion matrix* ANN 61

Gambar 4. 17 Perbandingan Akurasi Model XGBoost dan ANN 63



DAFTAR TABEL

Tabel 2. 1 Contoh Dataset Untuk Membangun Model XGBoost.....	21
Tabel 2. 2 Perhitungan Nilai <i>Error</i> atau <i>Residuals</i>	21
Tabel 2. 3 Hasil Perhitungan untuk Satu Pohon	27
Tabel 3. 1 Variabel Penelitian	39
Tabel 4. 1 Pembagian Data <i>Training</i> dan Data <i>Testing</i>	49
Tabel 4. 2 Rentang Nilai Parameter yang Digunakan	51
Tabel 4. 3 Hasil <i>Best Parameters</i> Setelah Melakukan <i>Tunning Parameters</i>	51
Tabel 4. 4 Hasil Evaluasi Setelah Model Dilatih.....	52
Tabel 4. 5 Hasil Evaluasi Setelah Model Dilatih.....	61
Tabel 4. 6 Perbandingan Akurasi Model	63



DAFTAR NOTASI

Notasi	Keterangan
d_i	: Perbedaan peringkat berpasangan
n	: Jumlah kasus.
i	: skor berpasangan.
X_i	: Nilai data asli.
X_{new}	: Nilai data hasil normalisasi.
X_{mean}	: Nilai rata-rata dari data asli.
T	: Total jumlah <i>decision tree</i>
w_j	: Berat pada <i>decision tree</i> ke- j
λ	: Parameter regularisasi untuk L_1 regularisasi
γ	: Parameter regularisasi untuk L_2 regularisasi
\hat{Y}	: <i>Residual</i> ke- i
λ	: Reg_lambda
$f_i(x)$: Probabilitas sebelumnya
$f_1(x)$: Probabilitas untuk data x pada iterasi ke – 1
$h_o(x)$: Prediksi awal yang dihasilkan
$h_1(x)$: Prediksi model tambahan yang ditambahkan pada iterasi pertama
	: Fungsi sigmoid
	: <i>Learning rate</i>



y : Label biner yang sebenarnya (0 atau 1).

p : Probabilitas prediksi dari model kelas positif (biasanya $0 \leq p \leq 1$)



BAB I

PENDAHULUAN

1.1 Latar Belakang

Stabilitas sistem keuangan merupakan hal yang sangat penting dalam suatu negara mengingat perannya dalam mengalokasikan dana dari pihak satu ke pihak lainnya. Stabilitas sistem keuangan terjadi ketika komponen-komponen ekonomi yang terkait dengan mekanisme penetapan harga, alokasi sumber daya, dan pengelolaan risiko berfungsi dengan baik dan mendukung pertumbuhan ekonomi (Otoritas Jasa Keuangan, 2017). Terdapat dua aspek utama yang berhubungan dengan stabilitas sistem keuangan, yaitu stabilitas harga dan stabilitas sektor keuangan. Stabilitas sektor keuangan mencakup lembaga keuangan dan pasar keuangan secara keseluruhan yang mendukung sistem keuangan. Sebagai sektor utama dalam industri keuangan, perbankan memegang peran penting dalam sistem keuangan di Indonesia terutama untuk menjaga stabilitas sektor keuangan. Sehingga untuk menjadi lembaga keuangan yang sehat, bank-bank harus memiliki modal yang memadai, likuiditas yang cukup, dan manajemen risiko yang baik.

Namun, sektor perbankan beroperasi dalam lingkungan yang penuh risiko dan kompleksitas. Hal tersebut terjadi diakibatkan oleh beberapa krisis keuangan baik dalam skala nasional, regional, maupun internasional. Ketika kita mengingat dua puluh lima tahun kebelakang, sektor perbankan dalam negeri mengalami penurunan yang sangat mengkhawatirkan selama krisis ekonomi dan moneter. Pada awal bulan Juli 1997, Indonesia menghadapi kesulitan likuiditas yang disebabkan oleh gejolak nilai tukar rupiah, hal ini menyebabkan bank kesulitan untuk memenuhi kewajibannya terhadap pihak luar negeri, serta menyulitkan nasabah dalam melunasi hutang mereka kepada bank. Setelah pencabutan izin usaha terhadap 16 bank pada bulan November kepercayaan terhadap perbankan semakin merosot. Akibatnya, terjadi penarikan ar-besaran. Keputusan untuk melikuidasi 16 bank tersebut dianggap sebagai



pemicu krisis kepercayaan yang kemudian mengakibatkan terpuruknya sektor perbankan (MFakhriansyah, 2023).

Pengalaman krisis ekonomi global menunjukkan bahwa semakin kuatnya keterkaitan antara aspek makroekonomi dan keuangan memiliki konsekuensi yang signifikan. Ketika sistem keuangan mengalami disfungsi, hal ini akan mengurangi efektivitas kebijakan moneter, mengganggu kelancaran aktivitas ekonomi, dan bahkan dapat menyebabkan perlambatan pertumbuhan ekonomi hingga kontraksi. Oleh karena itu, stabilitas sistem keuangan menjadi tanggungjawab bersama dari berbagai otoritas di sektor keuangan, termasuk Bank Indonesia. Kerentanan perbankan terhadap masalah baik dari segi eksternal maupun internal menjadi fokus perhatian bagi pihak yang terlibat, mengingat perbankan merupakan sektor vital bagi negara.

Komite Stabilitas Sistem Keuangan (KSSK), yang terdiri dari Otoritas Jasa Keuangan (OJK), Bank Indonesia (BI), Lembaga Penjamin Simpanan (LPS), dan Kementerian Keuangan Republik Indonesia, terus mengembangkan solusi untuk menghadapi berbagai permasalahan potensial di sektor perbankan. Salah satu solusi yang dikembangkan adalah membuat prediksi terhadap permasalahan yang mungkin terjadi di dunia perbankan. Prediksi ini bertujuan untuk menjadi sistem peringatan dini (*early warning system*) guna mencegah terjadinya masalah ekonomi yang lebih luas, termasuk potensi mengalami *financial distress* pada bank. *Financial distress* dapat didefinisikan sebagai kondisi penurunan keuangan yang terjadi sebelum terjadi kebangkrutan atau likuidasi (Platt & Platt, 2002). Penting untuk mengembangkan model yang dapat mengklasifikasikan *financial distress* suatu perusahaan agar dapat mencegah situasi tersebut, karena memiliki dampak negatif yang signifikan pada bisnis dan ekonomi secara keseluruhan. *Financial distress* tidak hanya mempengaruhi perusahaan itu sendiri tetapi juga berdampak pada pemangku kepentingan seperti bank, investor, dan pemasok (Sari & Hermi, 2023).

Penelitian mengenai *financial distress* dimulai oleh Beaver pada tahun 1966, kemudian dilanjutkan oleh Edward Altman pada tahun 1968, yang mengembangkan model untuk memprediksi *financial distress*. Altman menghasilkan



rumus yang dikenal sebagai *Z-Score*. Rumus ini adalah model rasio yang menggunakan analisis diskriminan berganda (MDA). Dalam metode MDA, lebih dari satu rasio keuangan yang terkait dengan kebangkrutan perusahaan digunakan. Altman menggunakan rasio-rasio seperti *working capital/total asset*, *retained earning/total asset*, *earning before interest* dan *tax/total asset*, *market value equity/total liabilities*, dan *sales/total asset*. Rumus *Z-Score* pertama yang dikembangkan oleh Altman berlaku untuk perusahaan manufaktur di Amerika Serikat yang memiliki saham yang diperdagangkan di bursa efek (Darmayanti *et al.*, 2023). Penelitian yang dilakukan oleh Zhang, Hu, Patuwo, dan Indro (1997) mengenai prediksi *financial distress* menggunakan metode *artificial neural network* (ANN) dengan *general framework* dan *cross-validation analyst* menunjukkan bahwa metode *Neural Network* lebih baik daripada metode regresi logistik. Hasil penelitian tersebut menunjukkan bahwa *overall classification rate* dari *Neural Network* berkisar antara 77,27% hingga 84,09%, sedangkan pada regresi logistik berkisar antara 75% hingga 81,82%.

Beberapa penelitian terkait pemodelan klasifikasi *financial distress*, teknik *data mining* banyak digunakan karena dengan teknik ini, pola dan tren tersembunyi dalam data perusahaan dapat diidentifikasi. Teknik *data mining* juga memungkinkan seleksi variabel yang relevan, pengolahan data yang kompleks, dan pengembangan model prediksi yang lebih akurat. Dalam penggunaannya, teknik *data mining* membantu dalam mengidentifikasi risiko potensial secara lebih awal, mengurangi kerugian, dan memperbaiki pengambilan keputusan dalam manajemen risiko dan keuangan perusahaan.

Penelitian ini memilih metode *Artificial Neural Network* (ANN) dan *Extreme Gradient Boosting* (XGBoost) dari berbagai teknik *data mining* yang ada. Metode ini digunakan untuk membandingkan dan mencari metode yang lebih efektif dalam membuat model klasifikasi *financial distress* dari sejumlah perusahaan. ANN dikenal salah satu model yang terkenal dalam *data mining* karena tingkat akurasi yang tinggi. Model ANN mensimulasikan interaksi saraf dalam memproses data dan pembelajaran dari pengalaman, mirip dengan fungsi otak manusia



(Shekhar & Xiong, 2008). ANN terdiri dari kumpulan unit pemrosesan sederhana yang saling terhubung, yang disebut neuron (Abiodun *et al.*, 2018). Dengan menggunakan algoritma ANN, kita dapat mengklasifikasikan perusahaan mana yang berpotensi mengalami kebangkrutan. Berdasarkan penelitian sebelumnya, ANN telah terbukti menjadi salah satu metode yang efektif dalam membuat model klasifikasi *financial distress* perusahaan (Sreedharan *et al.*, 2020)

Metode XGBoost, atau *Extreme Gradient Boosting*, adalah algoritma *Ensemble* yang juga digunakan untuk membuat model klasifikasi *financial distress* perusahaan. XGBoost merupakan metode yang sangat populer dalam komunitas *data science* karena kemampuannya dalam mengatasi masalah prediksi klasifikasi dan regresi dengan menggabungkan banyak model lemah menjadi model yang kuat. Dalam penelitian ini, XGBoost digunakan untuk membandingkan kinerja dan akurasi model klasifikasi dengan metode ANN. Terdapat banyak penelitian sebelumnya yang menggunakan XGBoost dalam berbagai bidang, termasuk dalam prediksi dan klasifikasi. Penelitian oleh Chen & Guestrin (2016) menggunakan XGBoost dalam prediksi penyakit jantung koroner. Mereka membandingkan performa XGBoost dengan beberapa metode klasifikasi lainnya dan menemukan bahwa XGBoost memberikan hasil yang lebih baik dalam mengklasifikasikan pasien dengan penyakit jantung koroner.

Penelitian ini terinspirasi dari penelitian Shely Ila Amaliah (2020) dengan judul “Analisis Perbandingan Klasifikasi *Financial distress* Perusahaan Menggunakan *Support Vector Machine* dan *Artificial Neural Network* Pada Perusahaan Pertambangan 2017-2018”, dimana pada penelitiannya menggunakan data perusahaan pertambangan, dan variabel penelitian yang digunakan berdasarkan penyesuaian Altman yaitu *working capital/total asset*, *retained earning/total asset*, *earning before interest* dan *tax/total asset*, dan *market value equity/total liabilities*. Sebagai perbandingan,

yang saya lakukan menggunakan data perbankan dari tahun 2013 hingga variabel yang saya gunakan meliputi CAR, ROA, ROE, NIM, LDR, BOPO, dan mana saya menggunakan variabel tersebut karena terdapat beberapa penelitian



yang juga menggunakan variabel tersebut seperti penelitian yang dilakukan oleh Hayuningtyas Pramesti Dewi (2022) dengan judul “*Financial distress* pada Sektor Bank Campuran di Indonesia dengan Rasio CAMEL Sebagai Alat Ukur Periode Studi 2014 – 2019”. Dari hasil penelitiannya beberapa dari rasio keuangan yang dijadikan variabel mempunyai pengaruh yang signifikan dalam melakukan klasifikasi *financial distress*. Dengan mengkombinasikan beberapa penelitian sebelumnya, penelitian ini bertujuan untuk menemukan metode yang lebih efektif dalam membuat model klasifikasi *financial distress* perusahaan. Hal ini diharapkan memberikan wawasan yang berharga dalam mengidentifikasi perusahaan yang berpotensi mengalami kesulitan keuangan dan memungkinkan pengambilan tindakan yang tepat secara dini.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini fokus pada pembentukan model klasifikasi *financial distress* yang akurat untuk perusahaan perbankan di Indonesia. Permasalahan yang ingin diselesaikan secara khusus dapat dirumuskan sebagai berikut:

1. Apa karakteristik dari rasio keuangan yang membedakan bank yang sehat dan bank yang mengalami masalah keuangan di Indonesia?
2. Variabel mana yang memiliki pengaruh signifikan terhadap kondisi keuangan bank?
3. Manakah algoritma paling akurat dalam melakukan klasifikasi *financial distress* menggunakan rasio keuangan pada perusahaan sektor perbankan 2013 – 2022?

1.3 Batasan Masalah

Untuk membatasi ruang lingkup penelitian ini, berikut adalah batasan permasalahan yang diberikan:

Data yang digunakan terbatas pada rasio keuangan perusahaan sektor perbankan dalam rentang 2013 – 2022.



2. Variabel yang akan digunakan dalam penelitian ini terdiri dari rasio keuangan CAR, ROA, ROE, NIM, LDR BOPO, NPL.
3. Parameter batasan rasio keuangan merupakan parameter ambang batas yang telah ditetapkan oleh Bank Indonesia.
4. Metode yang akan digunakan dalam penelitian ini adalah *Extreme Gradient Boosting* (XGBoost) dan *Artificial Neural Network* (ANN).

1.4 Tujuan Penelitian

Berdasarkan perumusan permasalahan yang telah disusun, tujuan secara keseluruhan dari penelitian ini adalah untuk mendapatkan model klasifikasi *financial distress* yang akurat untuk perusahaan perbankan di Indonesia. Secara spesifik, tujuan penelitian ini adalah sebagai berikut:

1. Menggambarkan karakteristik rasio keuangan yang membedakan bank yang sehat dan bank yang mengalami masalah keuangan di Indonesia.
2. Mengidentifikasi variabel yang memiliki pengaruh signifikan terhadap kondisi keuangan bank.
3. Membandingkan model klasifikasi *financial distress* menggunakan rasio keuangan untuk perusahaan perbankan di Indonesia menggunakan algoritma *Extreme Gradient Boosting* dan *Artificial Neural Network*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak, termasuk:

1. Menyediakan sistem peringatan dini (*early warning system*) bagi pihak-pihak terkait seperti pemerintah, Bank Indonesia, OJK (Otoritas Jasa Keuangan), LPS (Lembaga Penjamin Simpanan), dan pihak-pihak lain yang mungkin terdampak oleh *financial distress* perusahaan. Hal ini akan membantu mereka untuk mengambil tindakan preventif dan mengurangi dampak negatif yang mungkin timbul.



2. Memberikan pemahaman tentang tingkat kredit (*credit rate*) sebagai bagian dari manajemen risiko investasi dalam sektor perbankan. Penelitian ini akan memberikan gambaran yang jelas tentang faktor-faktor yang mempengaruhi kondisi keuangan bank, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan terkait investasi dan manajemen risiko.
3. Penelitian ini diharapkan dapat meningkatkan pemahaman penulis dalam menerapkan metode *machine learning* menggunakan algoritma *Extreme Gradient Boosting* dan *Artificial Neural Network*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan penelitian ini adalah sebagai berikut:

1. **BAB I PENDAHULUAN**

Bab ini menguraikan latar belakang masalah, rumusan masalah, tujuan, batasan masalah, manfaat penelitian dan sistematika penulisan.

2. **BAB 2 TINJAUAN PUSTAKA**

Bab ini membahas tentang paparan mengenai definisi dan konsep menurut para ahli yang menjadi dasar dari penelitian.

3. **BAB 3 METODOLOGI PENELITIAN**

Bab ini menjelaskan tentang pendekatan dan jenis penelitian, waktu dan tempat penelitian, objek penelitian, jenis dan sumber data, metode pengumpulan data dan alur kerja.

4. **BAB 4 HASIL DAN PEMBAHASAN**

Pada bab ini dijelaskan pembahasan hasil penelitian berisi uraian analitis dan terpadu mengenai hasil penelitian yang disajikan secara jujur.

5. **BAB V KESIMPULAN DAN SARAN**

Bab ini menjelaskan tentang kesimpulan dari hasil dan pembahasan penelitian serta saran yang bisa bermamfaat bagi berbagai pihak.



BAB II

TINJAUAN PUSTAKA

2.1 Pengertian *Financial distress*

Menurut Curry dan Banjarnahor (2018), *financial distress* merujuk pada kondisi ketidaksehatan keuangan perusahaan atau penurunan kondisi keuangan sebelum terjadinya kebangkrutan atau likuidasi. *Financial distress* dapat terjadi pada berbagai perusahaan dan dapat menjadi indikasi atau sinyal bahwa perusahaan tersebut berpotensi mengalami kebangkrutan. Ketika perusahaan berada dalam kondisi *financial distress*, manajemen harus berhati-hati karena perusahaan dapat berada pada tahap yang mengarah kebangkrutan. Namun, sering terjadi kesalahan umum dalam menganggap *financial distress* sama dengan kebangkrutan. Padahal, ini tidak benar. *Financial distress* merupakan faktor penyebab kebangkrutan suatu perusahaan, namun bukan berarti semua perusahaan yang mengalami *financial distress* akan mengalami kebangkrutan.

Menurut Brigham dan Daves (2013) *financial distress* terjadi karena serangkaian kesalahan, pengambilan keputusan yang tidak tepat, dan kelemahan yang saling berhubungan dapat berakibat secara langsung maupun tidak langsung terhadap manajemen. *Financial distress* dapat dilihat dari indikator kinerja bank yaitu kekurangan dana untuk melunasi kewajiban perusahaan jangka pendek (likuiditas) sampai pada kekurangan dana untuk melunasi seluruh kewajiban perusahaan (solvabilitas).

Salah satu kasus kurangnya manajemen risiko pada bank adalah terjadinya krisis moneter Indonesia tahun 1998. Menurut Williams (2013) krisis moneter yang disebabkan oleh salah satunya tidak likuidnya aliran uang kepada masyarakat, merupakan salah satu dampak dari kurangnya manajemen risiko dalam pengelolaan risiko dari tidak likuidnya aliran uang kepada masyarakat dalam skala besar ujung pada risiko sistemik. Risiko sistemik merupakan kondisi ketika bank



kesulitan dalam melunasi kewajibannya yang menimbulkan kepanikan dalam sistem keuangan dan mempengaruhi lembaga lainnya.

Indikator *financial distress* dapat dilihat dari analisis aliran kas, analisis strategi perusahaan, dan laporan keuangan perusahaan (Rizki, 2019). Tanda-tanda perusahaan mengalami *financial distress* dapat dilihat dari laporan keuangan perusahaan. Beberapa tanda-tanda *financial distress* yang dapat dilihat dari laporan keuangan perusahaan antara lain:

1. Laba yang menurun secara signifikan: perusahaan mengalami penurunan laba secara signifikan selama beberapa periode, yang menunjukkan bahwa perusahaan mengalami kesulitan dalam menghasilkan pendapatan.
2. Arus kas operasi yang negatif: perusahaan mengalami arus kas operasi yang negatif selama beberapa periode, yang menunjukkan bahwa perusahaan tidak menghasilkan arus kas yang cukup dari operasinya.
3. Rasio keuangan yang buruk: perusahaan memiliki rasio keuangan yang buruk, seperti rasio utang terhadap ekuitas yang tinggi atau rasio lancar yang rendah.
4. Kenaikan utang jangka pendek: perusahaan mengalami kenaikan utang jangka pendek selama beberapa periode, yang menunjukkan bahwa perusahaan mengalami kesulitan dalam membayar utangnya.
5. Penjualan yang menurun: perusahaan mengalami penurunan penjualan selama beberapa periode, yang menunjukkan bahwa perusahaan mengalami kesulitan dalam menghasilkan pendapatan.

Jika perusahaan mengalami tanda-tanda tersebut, maka perusahaan harus segera mengambil tindakan untuk mengatasi masalah keuangan tersebut. Beberapa tindakan yang dapat dilakukan antara lain melakukan restrukturisasi utang, memotong biaya operasional, mencari pendanaan tambahan, atau menjalin kemitraan strategis.

Tujuannya adalah untuk mengembalikan kesehatan keuangan perusahaan dan menjaga kinerja bisnis (Meiza, 2022).



2.2 Earning Per Share

EPS (*Earning per Share*) atau laba per lembar saham adalah tingkat keuntungan bersih untuk tiap lembar saham yang mampu diraih perusahaan pada saat menjalankan operasinya. *Earning Per Share* memberikan informasi kepada para pihak luar seberapa jauh kemampuan perusahaan menghasilkan laba untuk tiap lembar saham yang beredar di pasaran. Laba per lembar saham atau EPS diperoleh dari laba yang tersedia bagi pemegang saham biasa dibagi dengan jumlah rata-rata saham biasa yang beredar.

Menurut Darmadji dan Fakhrudin (2016) semakin tinggi nilai EPS semakin besar laba yang disediakan untuk pemegang saham dan kemungkinan peningkatan jumlah dividen yang diterima pemegang saham juga akan meningkat. Pengukuran *Earning Per Share* pada penelitian ini diukur dengan persamaan berikut:

$$Earning\ Per\ Share = \frac{\text{Laba Bersih Setelah Pajak}}{\text{Jumlah Saham Beredar}}$$

2.3 Rasio Keuangan

Rasio keuangan melibatkan perbandingan angka-angka yang terdapat dalam laporan keuangan guna memahami posisi keuangan suatu perusahaan dan mengevaluasi kinerja manajemen dalam periode tertentu. Rasio keuangan adalah angka yang diperoleh melalui perbandingan antara satu pos dalam laporan keuangan dengan pos lain yang memiliki keterkaitan yang relevan dan signifikan (Evi *et al.*, 2019). Perbandingan tersebut dapat dilakukan antara komponen dalam satu laporan keuangan maupun antara komponen yang ada di berbagai laporan keuangan. Selanjutnya, angka yang dibandingkan dapat berupa angka dalam satu periode atau beberapa periode. Terdapat beberapa jenis rasio keuangan, termasuk:

2.3.1 Rasio Likuiditas

Rasio ini memberikan informasi tentang kemampuan bank untuk memenuhi kewajiban yang harus segera dibayar. Salah satu rasio likuiditas bank yang digunakan untuk mengukur rasio kredit terhadap dana pihak ketiga atau *loan to deposit ratio* (LDR). Berikut persamaan yang digunakan untuk menghitung nilai rasio LDR:



$$\text{LDR} = \frac{\text{Kredit}}{\text{Dana Pihak Ketiga}}$$

2.3.2 Rasio Rentabilitas

Rasio ini memberikan informasi tentang kemampuan bank dalam menghasilkan keuntungan melalui operasi bisnisnya. Salah satu rasio rentabilitas yang digunakan adalah *return on assets* (ROA). Berikut adalah persamaan yang digunakan untuk menghitung rasio ROA

$$\text{ROA} = \frac{\text{Laba Sebelum Pajak}}{\text{Rata – Rata Total Aset}}$$

Rasio ROA menunjukkan berapa banyak keuntungan yang dihasilkan perusahaan dari setiap unit aset yang digunakan. Semakin tinggi rasio ROA, semakin efisien perusahaan dalam memanfaatkan asetnya.

Rasio rentabilitas juga mencakup rasio *Return on Equity* (ROE): Rasio ROE mengukur tingkat pengembalian investasi bagi pemegang saham. Rumus umum untuk ROE adalah sebagai berikut:

$$\text{ROE} = \frac{\text{Laba Setelah Pajak}}{\text{Rata – Rata Equity}}$$

Rasio ROE menggambarkan seberapa besar keuntungan yang diperoleh perusahaan bagi setiap unit ekuitas yang dimiliki pemegang saham. Semakin tinggi rasio ROE, semakin menguntungkan investasi bagi pemegang saham.

Net interest margin (NIM) atau marjin bunga bersih adalah salah satu rasio keuangan yang digunakan dalam analisis rentabilitas bank. Rasio ini mengukur perbedaan antara pendapatan bunga yang diperoleh dari aktiva bunga dengan biaya bunga yang dibayar atas kewajiban bunga. Persamaan untuk menghitung nilai NIM adalah sebagai berikut:

$$\text{NIM} = \frac{\text{Pendapatan Bunga Bersih}}{\text{Rata – Rata Aktiva Produktif}}$$



Rasio terakhir yang termasuk dalam kategori rasio rentabilitas bank adalah rasio beban operasional terhadap pendapatan operasional (BOPO). Rasio BOPO mengukur proporsi beban operasional yang dikeluarkan oleh bank dibandingkan dengan pendapatan operasional yang diperoleh. Berikut adalah persamaan yang digunakan untuk menghitung nilai rasio BOPO:

$$\text{BOPO} = \frac{\text{Total Beban Operasional}}{\text{Total Pendapatan Operasional}}$$

2.3.3 Rasio Aktiva Produktif

Rasio ini digunakan untuk mengevaluasi tingkat risiko yang dihadapi oleh bank dalam menjalankan operasinya. Rasio-rasio yang termasuk dalam kategori rasio risiko usaha bank meliputi rasio aktiva produktif bermasalah dan rasio kredit bermasalah terhadap total kredit, juga dikenal sebagai non-performing loan (NPL).

$$\text{NPL} = \frac{\text{Kredit Bermasalah}}{\text{Total Kredit}}$$

2.3.4 Rasio Permodalan Bank

Rasio ini juga dikenal sebagai rasio solvabilitas, memberikan informasi tentang kemampuan modal sebuah bank untuk mendukung kegiatan operasionalnya dan menyerap risiko yang tak terhindarkan. Salah satu rasio yang termasuk dalam kategori ini adalah rasio kewajiban penyediaan modal minimum (KPMM), yang juga dikenal sebagai capital adequacy ratio (CAR).

$$\text{CAR} = \frac{\text{Modal}}{\text{Aktiva Tertimbang Menurut Resiko}} \quad (2.8)$$

Pada penelitian ini lebih memfokuskan tanda *financial distress* pada rasio keuangan perusahaan. Rasio keuangan yang digunakan terdiri dari CAR, ROA, ROE, LDR, NIM, BOPO, NPL. Berikut adalah penjelasan mengenai beberapa rasio dan ciri-cirinya ketika perusahaan mengalami *financial distress*:



1. Capital Adequacy Ratio (CAR): Menurut Peraturan Bank Indonesia Nomor 3/21/PBI/2001 besarnya CAR perbankan untuk saat ini minimal 8%, sehingga di bawah dari itu Kesehatan keuangan bank dikatakan buruk.
2. Return on Assets (ROA): Ketika perusahaan mengalami *financial distress*, ROA cenderung menurun. Penurunan ROA dapat disebabkan oleh penurunan pendapatan atau peningkatan beban yang mengurangi profitabilitas perusahaan. Angka ROA berdasarkan ketentuan BI minimal 0.5% agar bank tersebut dapat dikatakan kondisi sehat.
3. Return on Equity (ROE): Ketika perusahaan mengalami *financial distress*, ROE cenderung menurun secara signifikan. Penurunan ROE dapat disebabkan oleh penurunan laba atau peningkatan utang yang mengurangi ekuitas perusahaan. Dalam Surat Edaran Bank No 6/23/DPNP tanggal 31 Mei 2004 Return on Equity dihitung dengan membagi laba setelah pajak dengan rata-rata modal inti. Rasio yang baik berkisar dari 5% - 12,5%.
4. Loan to Deposit Ratio (LDR): Ketika perusahaan mengalami *financial distress*, LDR cenderung meningkat. Hal ini dapat terjadi karena perusahaan mengandalkan pinjaman untuk memenuhi kebutuhan likuiditas dan menghadapi kesulitan dalam memperoleh dana dari simpanan. Berdasarkan (Bank Indonesia, 2013) batas bawah aman LDR yaitu sebesar 78% dan batas aman sebesar 92%.
5. Net Interest Margin (NIM): Ketika perusahaan mengalami *financial distress*, NIM cenderung menurun. Penurunan NIM dapat disebabkan oleh peningkatan biaya bunga atau penurunan pendapatan bunga yang mengurangi profitabilitas perusahaan. Adapun Standar yang ditetapkan Bank Indonesia untuk ratio Net Interest Margin (NIM) adalah 6% keatas (Pandia, 2012).
6. Beban Operasi terhadap Pendapatan Operasional (BOPO): Ketika perusahaan mengalami *financial distress*, BOPO cenderung meningkat. Peningkatan BOPO dapat disebabkan oleh peningkatan biaya operasional yang tidak banding dengan pendapatan, sehingga mengurangi profitabilitas perusahaan. Menurut ketentuan Bank Indonesia efisiensi operasi diukur dengan BOPO



(Biaya Operasional Pendapatan Operasional) dengan batas maksimum BOPO (Biaya Operasional Pendapatan Operasional) yaitu 94%.

7. Non-Performing Loan (NPL): Ketika perusahaan mengalami *financial distress*, NPL cenderung meningkat. Peningkatan NPL menunjukkan adanya risiko kredit yang tinggi dan kemungkinan adanya kerugian yang signifikan bagi perusahaan. Menurut peraturan Bank Indonesia Nomor 15/2/PBI/2013 tentang penetapan status dan tindak lanjut pengawanan bank umum konvensional, batas maksimal rasio ini sebesar 5% jika melebihi maka akan mempengaruhi tingkat kesehatan bank yang bersangkutan.

2.4 Statistika Deskriptif

Statistika deskriptif adalah metode yang berkaitan dengan pengumpulan, pengklasifikasian, penyajian, dan analisis data untuk memberikan informasi yang berguna. Statistik deskriptif digunakan untuk menganalisis data dengan tujuan memberikan gambaran yang ringkas dan informatif tentang data tersebut. Metode-metode dalam statistik deskriptif membantu kita memahami karakteristik dasar dari data yang diamati, seperti pemusatan (misalnya, rata-rata, median), sebaran (misalnya, rentang, simpangan baku), bentuk distribusi (misalnya, *skewness*, kurtosis), serta hubungan antara variabel (misalnya, korelasi). Dengan menganalisis data menggunakan statistik deskriptif, kita dapat mengidentifikasi pola, tren, anomali, atau perbedaan dalam data (Walpole, 1995). Statistik deskriptif berfungsi untuk mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data sampel atau populasi (Sugiyono, 2007).

2.5 Korelasi Spearman

Korelasi *Spearman* adalah suatu ukuran statistik yang digunakan untuk mengevaluasi hubungan antara dua variabel. Dalam bahasa yang lebih sederhana, korelasi *Spearman* menggambarkan sejauh mana dua variabel bergerak bersama-sama ke arah yang serupa.

Ada dua metode untuk menghitung korelasi *Spearman* tergantung pada apakah: (1) data Anda tidak memiliki peringkat terikat atau (2) data Anda memiliki peringkat terikat.



- Rumus Untuk data tidak memiliki peringkat terikat

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

- Rumus yang digunakan ketika ada peringkat terikat adalah:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (2.1)$$

Dalam konteks korelasi *Spearman*, jika data memiliki peringkat terikat, yang berarti beberapa data memiliki nilai yang sama dan diberi peringkat yang sama.

2.6 Normalisasi Data

Normalisasi data adalah teknik yang digunakan untuk mengorganisir data dalam sebuah basis data relasional dengan tujuan untuk mengurangi atau mencegah terjadinya redudansi data dan memperbaiki susunan rancangan tabel. Normalisasi biasanya diperlukan ketika atribut-atribut dalam dataset memiliki skala atau rentang yang berbeda (Trivusi, 2022).

Metode normalisasi data yang akan digunakan yaitu *Standard Scaler*. *Standard Scaler* adalah salah satu teknik normalisasi yang umum digunakan dalam analisis data dan pemrosesan data. Metode ini mengubah data dengan mengurangi *mean* dan membaginya dengan standar deviasi, sehingga data akan memiliki *mean* nol dan standar deviasi satu. Dengan menggunakan *Standard Scaler*, data akan diubah menjadi distribusi yang lebih simetris dan memiliki skala yang serupa.

$$X_{new} = \frac{X_i - X_{mean}}{Standar\ Deviasi} \quad (2.2)$$

2.7 Balancing Data

Dalam membuat pemodelan klasifikasi sering kali menghadapi tantangan yang tidak seimbang, di mana kelas mayoritas secara signifikan lebih besar dari kelas minoritas. Ketidakseimbangan ini menghalangi kemampuan model



untuk belajar secara efektif dari kelas minoritas. Masalah ini menjadi sangat penting ketika informasi yang terkandung dalam kelas minoritas memiliki signifikansi yang lebih besar, seperti dalam dataset deteksi penyakit, dataset *churn*, dan dataset deteksi penipuan.

Teknik *Synthetic Minority Oversampling Technique* (SMOTE) mengatasi masalah ini dengan membuat sampel sintetis baru dari kelas minoritas. Metode ini melibatkan sintesis contoh-contoh baru berdasarkan data kelas minoritas yang ada. SMOTE memiliki berbagai macam adaptasi, salah satunya *SMOTE-Tomek*. Metode *SMOTE-Tomek* pertama kali diperkenalkan oleh Batista et al. (2003). Metode ini menggabungkan kemampuan SMOTE untuk menghasilkan data sintetis untuk kelas minoritas dan kemampuan *Tomek Links* untuk menghapus data yang diidentifikasi sebagai tautan *Tomek* dari kelas mayoritas, yaitu sampel data dari kelas mayoritas yang paling dekat dengan data kelas minoritas. Proses *SMOTE-Tomek Links* terdiri dari langkah-langkah berikut:

1. Memulai dengan teknik SMOTE, lalu memiliki secara acak data dari kelas minoritas.
2. Menghitung jarak antara data yang dipilih secara acak dan k tetangga terdekatnya. Mengalikan selisih jarak dengan bilangan acak antara 0 dan 1, lalu menambahkan hasilnya ke kelas minoritas sebagai sampel sintetis.
3. Mengulang langkah 2-3 hingga proporsi kelas minoritas tercapai. Ini menandai akhir dari proses SMOTE.
4. Memulai dari *Tomek Links*, pilih secara acak data dari kelas mayoritas.
5. Jika tetangga terdekat dari data yang dipilih secara acak adalah data dari kelas minoritas, yang berarti terbentuknya tautan *Tomek*, maka hapus tautan *Tomek* tersebut.

Dengan menggabungkan langkah-langkah SMOTE dan *Tomek Links*, metode *Tomek* bertujuan untuk mencapai keseimbangan antara kelas mayoritas dan minoritas dengan menghasilkan data sintetis untuk kelas minoritas dan menghapus data yang berdekatan antara kedua kelas. Hal ini dapat membantu



meningkatkan kemampuan model untuk belajar dan memprediksi data dari kelas minoritas dengan lebih baik, serta mengurangi bias yang mungkin timbul akibat ketidakseimbangan data.

2.8 Teknik *Ensemble*

Teknik *Ensemble* adalah sebuah pendekatan dalam pembelajaran mesin yang melibatkan beberapa model pengklasifikasi atau prediktor. Model-model ini digunakan untuk mengklasifikasikan data baru berdasarkan bobot prediksi yang dihasilkan sebelumnya. Proses *Ensemble* umumnya melibatkan penggabungan prediksi dari model-model tersebut dengan menggunakan metode seperti bobot dan *voting*. Metode *Ensemble* telah lama dikembangkan dan terus ditingkatkan untuk menciptakan metode *Ensemble* yang lebih baik. Hal ini disebabkan karena metode *Ensemble* dianggap mampu menghasilkan akurasi yang lebih baik daripada menggunakan hanya satu pengklasifikasi.



Gambar 2. 1 Ilustrasi Proses *Ensemble*

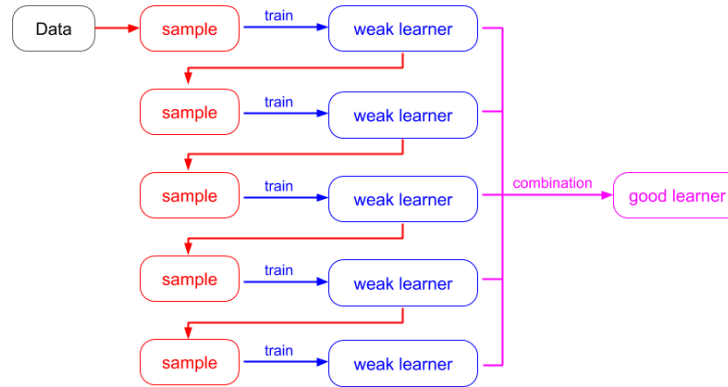
Untuk menciptakan *Ensemble* terbaik dengan menggunakan metode *Ensemble*, kita dapat menggabungkan beberapa model pengklasifikasi yang berbeda jenis. Pilihan model pengklasifikasi dapat ditentukan berdasarkan kebutuhan khusus dan jenis data yang sedang diproses. Dua teknik *Ensemble* yang umum digunakan adalah *Boosting* dan *bagging*.

ting

boosting adalah sebuah teknik *Ensemble* yang powerful. Teknik ini melibatkan dengan *weak learners*, yang juga sering disebut sebagai *base learners*, untuk



membentuk *strong learners* sehingga menghasilkan model yang lebih baik. Berbeda dengan teknik *bagging* yang melakukan proses secara paralel, teknik *Boosting* melatih *weak learners* secara sekuensial dan kemudian menggabungkannya. Contoh algoritma yang menggunakan teknik *Boosting* termasuk *adaboost*, *gradient Boosting*, dan XGBoost.



Gambar 2. 2 Ilustrasi Proses Boosting

Sumber : Jcop (2021)

2.10 Extreme Gradient Boosting (XGBoost)

XGBoost, yang dikembangkan oleh Chen dan Guestrin pada tahun 2016, adalah metode *Extreme Gradient Boosting*. XGBoost adalah salah satu teknik *Boosting* di mana setiap pohon yang dibangun bergantung pada pohon sebelumnya. Pada awalnya, pohon pertama dalam XGBoost memiliki performa klasifikasi yang lemah dan probabilitas inisialisasi yang ditentukan oleh penulis. Namun, setiap pohon yang dibangun kemudian melakukan *update* bobotnya, sehingga menghasilkan kumpulan pohon klasifikasi yang kuat.

2.10.1 Fungsi Objektif (*Objective Function*)

Dalam rangka mengevaluasi sejauh mana model cocok dengan data pelatihan, digunakan penggunaan fungsi objektif. Fungsi objektif ini memiliki dua komponen yaitu kerugian pelatihan (*training loss*) dan istilah regularisasi, yang terlihat dalam rumus di bawah ini:



$$obj(\theta) = L(\theta) + \Omega(\theta)$$

Dimana L adalah fungsi *training loss*, Ω adalah fungsi *Regularization Term*, dan θ adalah parameter berupa model terkait (Gouveia & Correia, 2020). *Training loss* mengukur seberapa prediktif model tersebut sehubungan dengan data latih. Fungsi *training loss* secara umum dapat ditulis seperti pada persamaan 2.12 berikut ini:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i)$$

Dimana y_i adalah nilai aktual yang dianggap benar dan \hat{y}_i adalah hasil prediksi model terkait sedangkan n adalah jumlah iterasi nilai *input* untuk model terkait. Dalam keseluruhan, rumus tersebut menggambarkan fungsi objektif yang mengukur total kerugian atau kesalahan model dalam memprediksi nilai target aktual pada setiap contoh data dalam dataset pelatihan. Tujuan utama adalah untuk menemukan nilai parameter θ yang menghasilkan nilai fungsi objektif (kerugian) yang lebih rendah, sehingga model menjadi lebih baik dalam memprediksi nilai target aktual. Formula umum yang biasa digunakan dari pengukuran *training loss* adalah *cross entropy loss* yang dapat dilihat pada persamaan 2.13 berikut ini:

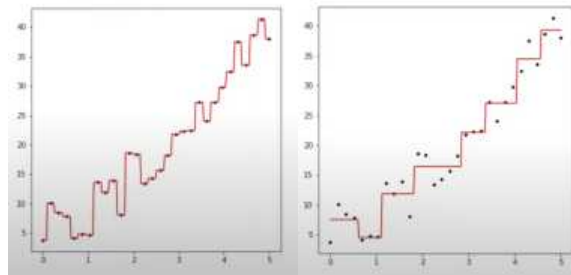
$$L(\theta) = [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Regularization Term berfungsi untuk mengendalikan kompleksitas model dan membantu mencegah *overfitting*.

$$\Omega(\theta) = \frac{1}{2} \lambda \sum_{j=1}^T w^2 + \gamma T$$

Ilustrasi tentang penggunaan *Regularization Term* dalam konteks keterkaitan pengguna terhadap topik seiring berjalannya waktu dapat dilihat pada Gambar 2.3.





Gambar 2. 3 Ilustrasi Penggunaan Regularization Term

Sumber : Jcop (2021)

2.10.2 Decision tree Ensemble

Decision tree adalah metode pembelajaran yang menggunakan struktur grafik berbentuk pohon untuk memodelkan data yang memiliki jenis variabel kontinu atau kategorikal. *Decision tree* terdiri dari serangkaian pertanyaan biner yang, ketika dijawab, menghasilkan prediksi tentang data yang sedang diproses. Model *tree Ensemble* terdiri dari himpunan *decision tree* (*classification and regression trees*, CART) yang digunakan secara bersama-sama. XGBoost dikembangkan dari *Classification and Regression Trees* (CART) atau juga dengan nama lain *Decision trees*. Dimana metode ini merupakan gabungan dari dua jenis pohon, yaitu *classification tree* dan juga *regression tree*. Jika variabel dependen yang dimiliki bertipe kategorik maka CART menghasilkan pohon klasifikasi (*classification tree*) sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka CART menghasilkan pohon regresi (*regression tree*).

2.11 Cara Kerja XGBoost

Algoritma *Gradient Boosting* bekerja dengan menggabungkan beberapa model yang lemah menjadi model yang lebih kuat. Algoritma ini menggunakan pendekatan iteratif, di mana setiap iterasi bertujuan untuk meningkatkan model sebelumnya dengan menambahkan model baru. Proses ini terus diulang hingga model yang dihasilkan memiliki nilai *loss function* yang cukup kecil. Proses iteratif dalam algoritma *Gradient Boosting* terdiri dari beberapa tahap (Gouveia *et al.*, 2020)



Diberikan contoh dataset dengan dua variabel yaitu variabel dosis obat, dan efektifitas.

Tabel 2. 1 Contoh Dataset Untuk Membangun Model XGBoost

X	Y
2	0
8	1
12	1
18	0

Dengan X menggambarkan dosis obat dan Y menggambarkan efektifitas (nilai 0 = tidak efektif & nilai 1 = efektif)

Parameter-parameter yang digunakan yaitu `n_estimators` atau model yang dibangun sebanyak 2, `max_depth = 2`, `learning_rate = 1`, `gamma = 2`, `reg_lambda = 0`.

1. Prediksi awal

Mesin prediksi awal secara random atau `base_score` sebesar 0.5 untuk seluruh titik data dalam dataset dimana,

$$f_0(x) = h_0(x) = 0.5$$

2. Perhitungan *errors* atau *residuals*

Hitunglah *residuals* \hat{Y} untuk semua titik data dari prediksi sebelumnya dimana

$$\hat{Y} = y - f_0(x)$$

Tabel 2. 2 Perhitungan Nilai Error atau Residuals

X	Y	$f_0(x)$	$\hat{Y} = y - f_0(x)$



2	0	0.5	-0.5
8	1	0.5	0.5
12	1	0.5	0.5
18	0	0.5	-0.5

3. Training Model

Melatih model pertama (pelatihan model merujuk pada pembangunan pohon) yaitu M1 dengan menggunakan $[X, \hat{Y}]$ data dan modelnya adalah pohon XGBoost khusus yang dibangun berbeda jika dibandingkan dengan *decision tree*. Terdapat beberapa istilah sebelum membangun pohon XGBoost dengan menurunkan rumus dalam menyelesaikan masalah optimasi pada XGBoost sebagai berikut:

$$Gain = (Left_{similarity} + Right_{similarity}) - Root_{similarity}$$

$$Similarity\ Score = \frac{(\sum \hat{Y}_i)^2}{\sum [Previous\ f_i(x) \cdot (1 - Previous\ f_i(x))] + \lambda}$$

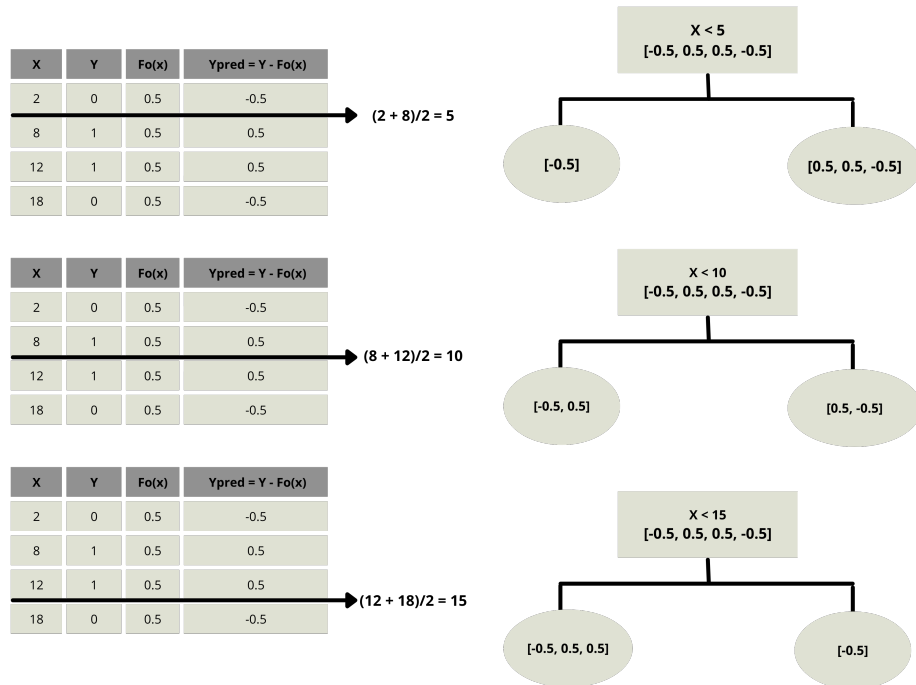
$$Output\ Score = \frac{(\sum \hat{Y}_i)}{\sum [Previous\ f_i(x) \cdot (1 - Previous\ f_i(x))] + \lambda}$$

- Perhitungan nilai *gain* hanya untuk menghitung akar pohon
- Perhitungan *similarity* untuk semua *node*
- Perhitungan *output value* hanya untuk *leaf node*

Lambda merupakan parameter, apabila nilai lambda meningkat akan menghasilkan pruning lebih banyak node pada pohon yang dibangun.



- a. Pohon dibangun dengan membagi data menjadi dua partisi dari berbagai kemungkinan pemisahan atau *split*.

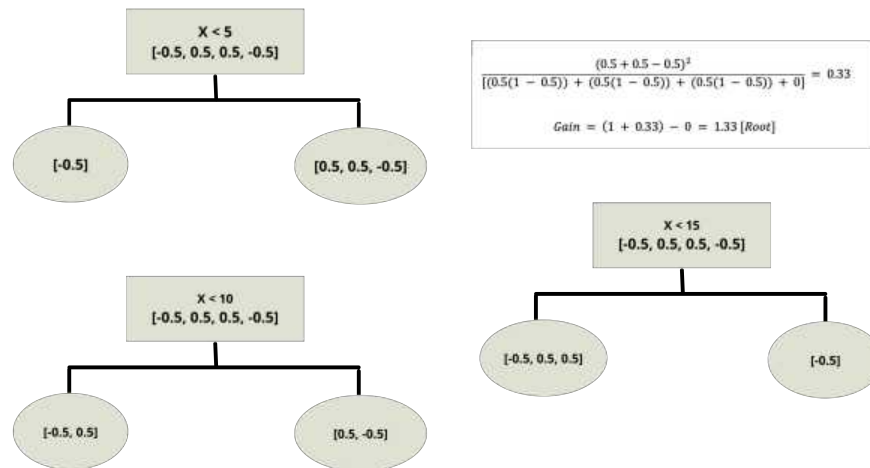


Gambar 2. 4 Contoh Membangun Model XGBoost

Menentukan batasan untuk *root* dihitung dengan mengambil nilai rata-rata antara dua titik percabangan atau *splits* dan sisanya menuju pada masing-masing *leaf node*. Jika dataset berisi n data, maka sejumlah $n - 1$ pohon dapat dibangun.

- b. Menghitung nilai *similarity* dan *gain* pada seluruh pohon yang dibangun untuk menemukan pohon dengan *split* yang optimal.

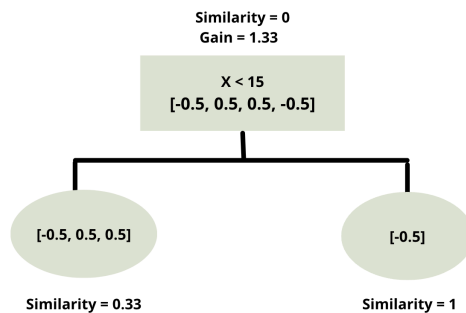




Gambar 2. 5 Contoh Perhitungan Similarity dan Gain

Memilih nilai dari suatu pohon yang memiliki nilai *gain* optimal, pada gambar pohon dengan $x < 15$ memiliki nilai *gain* maksimal sebesar 1.33

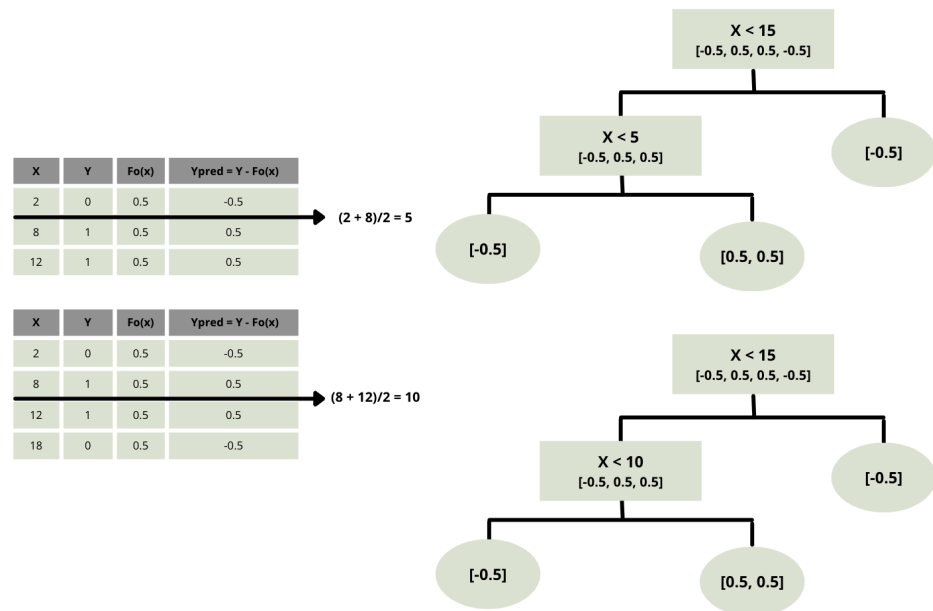
- c. Pemisahan kembali untuk pohon yang memiliki nilai *gain* maksimal hingga *max_depth* untuk konstruksi pohon lengkap.



Gambar 2. 6 Contoh Split pada Pohon XGBoost

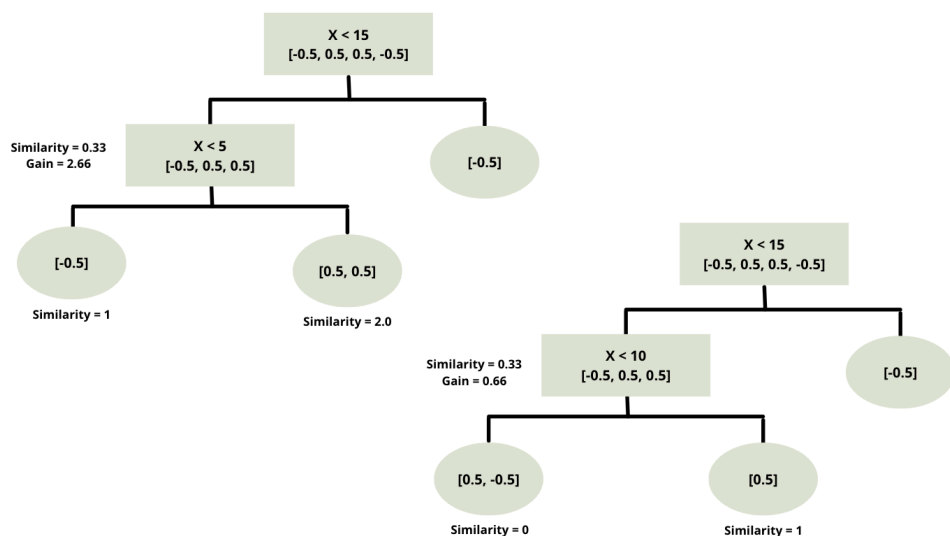
Karena nilai *max_depth* yang digunakan sebanyak dua, selanjutnya akan dibangun pohon kembali dengan memisahkan data yang ada pada *leaf* sebelah kiri.





Gambar 2. 7 Contoh Split pada Turunan Percabangan

Menghitung kembali nilai *similarity* dan *gain* untuk *split* baru pada percabangan lanjutan untuk memilih *internal root* yang memiliki nilai *gain* maksimal.

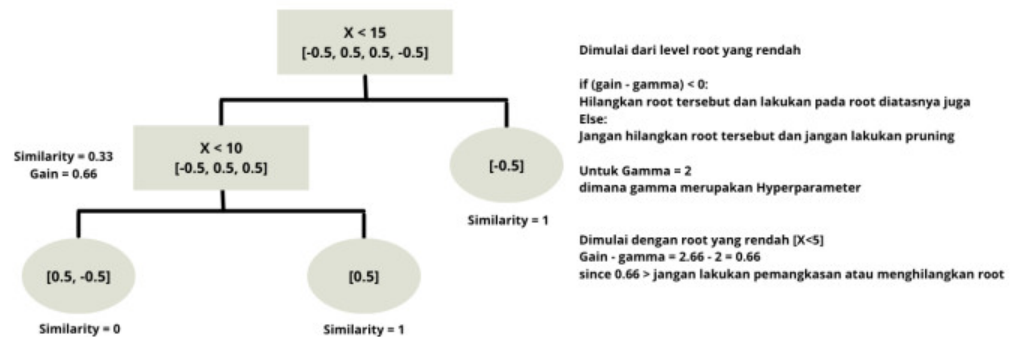


Gambar 2. 8 Contoh Perhitungan Similarity dan Gain pada Split Lanjutan



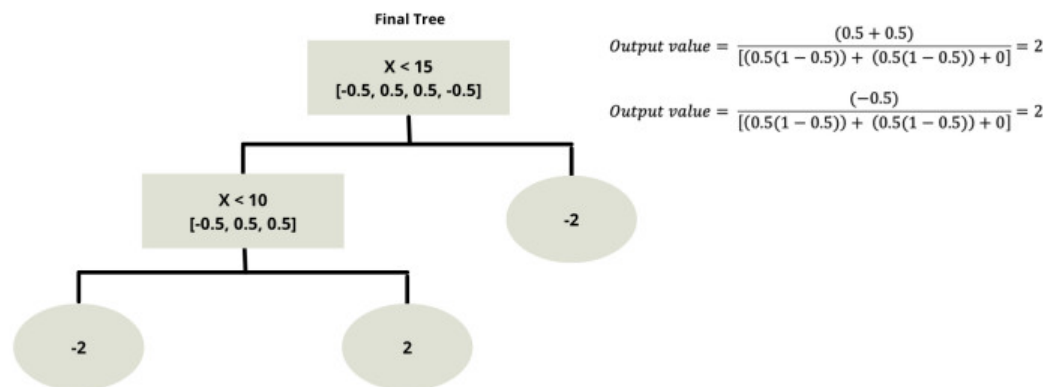
Internal *root* $x < 5$ terpilih untuk konstruksi pohon XGBoost karena menghasilkan nilai *gain* maksimal sebesar 2.66.

- d. Setelah membangun pohon, langkah selanjutnya adalah melakukan *pruning* atau pemangkasan yang bertujuan untuk memperkecil ukuran pohon keputusan dengan menghilangkan bagian pohon yang memiliki kekuatan yang kurang untuk mengklasifikasikan kejadian.



Gambar 2. 9 Proses Pruning pada Pohon

Menghitung nilai *output* untuk semua *leaf* untuk mendapatkan pohon terakhir pada model 1, karena beberapa *leaf* memiliki lebih dari satu *residual*.



Gambar 2. 10 Perhitungan Output Value



Menghitung prediksi dari model 1, seluruh titik data melewati pohon terakhir untuk hitung prediksi $f_1(x)$ dan nilai *residual*. Diberikan nilai $learning_rate = 1.0$

Maka,

$$f_1(x) = \sigma \left[\log \left(\frac{h_o(x)}{1 - h_o(x)} \right) + (\eta x h_1(x)) \right]$$

Memecahkan $f_1(x)$ pada klasifikasi didapatkan:

$$f_1(x) = \sigma (0 + 1 x (h_1(x)))$$

Fungsi sigmoid:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

Tabel 2. 3 Hasil Perhitungan untuk Satu Pohon

X	Y	$h_o(x)$	$f_1(x) = \sigma(0 + 1 x (h_1(x)))$	$\hat{Y} = y - f_1(x)$
2	0	0.5	0.35	-0.35
8	1	0.5	0.65	0.35
12	1	0.5	0.65	0.35
18	0	0.5	0.35	-0.35

4. Ulangi Langkah 3 untuk membangun pohon yang lain.

Salah satu faktor yang paling penting dibalik kesuksesan XGBoost adalah lebih efisien dan scalable pada berbagai skenario, karena dapat menyelesaikan berbagai fungsi seperti, regresi, klasifikasi, maupun ranking. Scalable atau alability sendiri disebabkan karena optimasi pada algoritma sebelumnya. Algoritma XGBoost melakukan optimasi 10 kali lebih cepat dibandingkan



dengan implementasi GBM lainnya. Kesuksesan tersebut dibuktikan dengan metode XGBoost menjadi metode yang banyak digunakan pada kompetisi *machine learning* (Chen & Guestrin, 2016)

2.12 Tuning Hyperparameter

Algoritma pembelajaran mesin seperti *deep neural network*, *gradient Boosting models*, dan lainnya memiliki *hyperparameter* yang beragam yang perlu disesuaikan untuk mencapai kinerja optimal. Optimasi *hyperparameter* adalah proses menyetel *hyperparameter* dari algoritma untuk mencapai kinerja yang terbaik. Setiap algoritma memiliki jenis *hyperparameter* yang berbeda. Dampak dari optimasi *hyperparameter* terhadap kinerja algoritma pembelajaran mesin telah terbukti secara teoritis maupun empiris melalui banyak penelitian yang ada dalam literatur. Namun, optimasi *hyperparameter* dapat menjadi tugas yang membosankan dan memakan waktu jika dilakukan secara manual. Beberapa pendekatan umum untuk mengatasi hal ini adalah *grid search*, *random search*, dan pendekatan lain seperti optimasi Bayesian (Putatunda et al., 2018).

Salah satu algoritma populer yang memanfaatkan optimasi *hyperparameter* adalah *Extreme Gradient Boosting* (XGBoost). Beberapa parameter utama *hyperparameter* yang dapat ditingkatkan untuk meningkatkan kinerja XGBoost adalah:

1. *Max depth: hyperparameter max depth* mengontrol kedalaman maksimum dari setiap pohon dalam model. Semakin besar nilai *max depth*, semakin kompleks model pohonnya, dan semakin kuat kemampuan penyesuaiannya.
 2. *Learning rate: learning rate (eta)* adalah *hyperparameter* yang menentukan seberapa cepat model belajar dari kesalahan sebelumnya. Nilai *learning rate* yang lebih kecil akan membuat model belajar lebih lambat, tetapi dapat membantu mencapai konvergensi yang lebih baik. Di sisi lain, nilai *learning rate* yang terlalu besar dapat menyebabkan model melewati titik konvergensi yang optimal dan menyebabkan ketidakstabilan dalam pembelajaran.
- n_estimators: hyperparameter n_estimators* menentukan jumlah pohon yang akan dibangun dalam model. Semakin banyak pohon yang dibangun, semakin



kompleks dan kuat modelnya. Namun, perlu diingat bahwa penambahan $n_estimators$ yang terlalu besar juga dapat menyebabkan *overfitting* dan meningkatkan waktu pelatihan.

4. *colsample_bytree* adalah parameter dalam algoritma XGBoost yang mengontrol proporsi fitur (kolom) yang digunakan saat membangun setiap pohon dalam *Ensemble*.
5. *subsample*: parameter subsample dalam model XGBoost mengontrol fraksi sampel yang akan digunakan dalam pelatihan setiap pohon. Dalam XGBoost, setiap pohon diperoleh dengan memilih secara acak sebagian sampel dari dataset pelatihan.
6. Gamma: Gamma adalah *hyperparameter* yang menentukan ambang batas untuk pemisahan node pohon. Jika penurunan skor (*gain*) yang dihasilkan oleh pemisahan node kurang dari gamma, maka pemisahan tidak akan dilakukan. Pengaturan gamma dapat membantu mengontrol kompleksitas model dan mengurangi *overfitting*.
7. *reg_lambda* membantu mencegah model XGBoost agar terlalu mencocokkan data pelatihan dan memperkuat generalisasi pada data yang belum pernah dilihat sebelumnya.
8. *reg_alpha* membantu model untuk melakukan seleksi fitur secara otomatis. Dengan memberikan nilai *reg_alpha* yang lebih tinggi, kita memberikan sinyal kepada model bahwa kita ingin menekan beberapa koefisien yang kurang penting sehingga menjadi nol atau mendekati nol.

RandomizedSearch, seperti yang disebutkan sebelumnya, adalah salah satu pendekatan alternatif untuk optimasi *hyperparameter* dalam algoritma pembelajaran mesin. Berbeda dengan *Grid Search* yang melakukan pencarian secara eksplisit pada semua kombinasi nilai *hyperparameter* yang mungkin, *RandomizedSearch* melakukan

acak dengan memilih nilai *hyperparameter* secara acak dari distribusi yang sebelumnya.



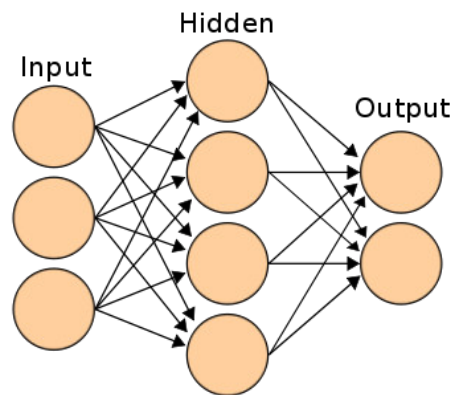
2.13 Artificial Neural Network (ANN)

2.13.1 Pengertian *Artificial Neural Network*

Pada metode *Artificial Neural Network* (ANN) memiliki alur penyelesaian masalah dengan menerapkan sistem kerja syaraf dari otak manusia (Anitescu *et al.*, 2019). Struktur jaringan terdiri dari beberapa jenis, diantaranya *Single Layer Networks*, *Multi Layer Networks*, dan *Recurrent Networks*. Diantara jenis arsitektur jaringan tersebut, yang digunakan pada penelitian ini adalah arsitektur *Multi Layer Networks*. Proses pembelajaran terjadi ketika penyesuaian bobot dan bias dilakukan. Metode yang paling umum digunakan dalam proses pembelajaran adalah *backpropagation*. Dalam metode ini, bobot diatur untuk meminimalkan perbedaan kuadrat antara keluaran model dan keluaran perkiraan, yang umumnya dikenal sebagai kesalahan kuadrat atau *sum of square error*. Istilah "buatan" digunakan karena jaringan saraf ini diimplementasikan menggunakan program komputer yang mampu melakukan sejumlah perhitungan selama proses pembelajaran (Abiodun *et al.*, 2018)

2.13.2 Komponen *Artificial Neural Network*

Menurut (Untoro, 2020) pada arsitektur *Multi Layer Networks* memiliki tiga lapisan jaringan yaitu *input layer*, *hidden layer* dan *output layer*. Posisi *hidden layer* berada di antara *input* dan *output layer*. Susunan pada arsitektur *Multi Layer Networks* terdiri dari tiga lapisan diantaranya *input layer*, *hidden layer* dan *output layer*, seperti pada gambar berikut:



Gambar 2. 11 Komponen ANN



Menurut (Izati *et al.*, 2019) diantara masing-masing *layer* tersebut memiliki fungsi yang berbeda, antara lain:

1. Lapisan Masukan (*Input layer*), difungsikan sebagai penghubung jaringan dengan data yang digunakan.
2. Lapisan Tersembunyi (*Hidden layer*), berfungsi sebagai pengubah masukan yang diterima dari *input layer* yang selanjutnya akan dihubungkan pada unit *output*. *Hidden layer* terletak diantara *input* dan *output layer*
3. Lapisan Keluaran (*Output Layer*), memiliki fungsi sebagai lapisan yang memberikan keluaran sebagai hasil dari proses yang telah dijalankan.

2.13.3 Jenis Fungsi Aktivasi

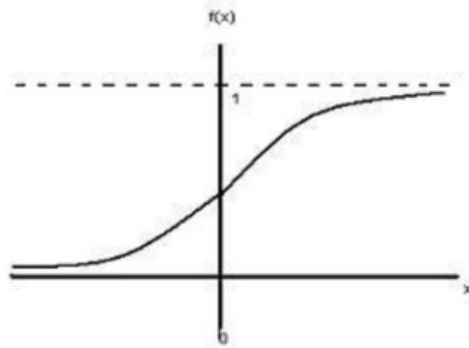
Fungsi aktivasi dalam *Artificial Neural Network* (ANN) berfungsi sebagai sinyal untuk menentukan *output* yang akan diteruskan ke neuron-neuron lainnya. Fungsi aktivasi ini memiliki peran yang sangat penting karena penggunaannya bergantung pada kebutuhan dan target yang diinginkan. Selain itu, fungsi aktivasi juga berperan dalam menentukan bobot yang akan diterapkan pada neuron-neuron (Putra *et al.*, 2022). Fungsi aktivasi digunakan untuk mengaktifkan setiap neuron yang digunakan dalam jaringan tersebut. Fungsi Aktivasi yang digunakan

1. Fungsi Sigmoid Biner

Fungsi sigmoid biner memiliki rentang nilai antara 0 hingga 1, sehingga sangat cocok digunakan pada jaringan yang menghasilkan *output* dalam rentang tersebut. Fungsi sigmoid biner sering digunakan dan efektif diterapkan pada jaringan saraf tiruan ketika algoritma pembelajaran yang digunakan adalah metode *backpropagation* (Putra *et al.*, 2022). Rumus matematis dari fungsi sigmoid biner adalah sebagai berikut:

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (2.3)$$





Gambar 2. 12 Fungsi Aktivasi Sigmoid Biner

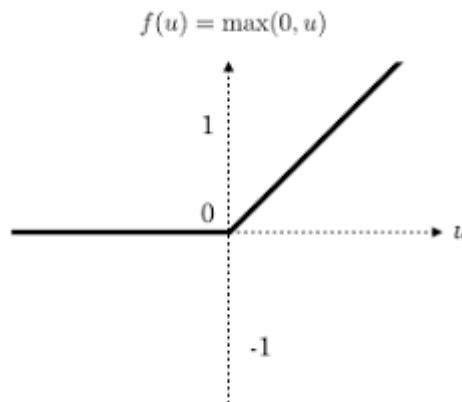
2. Fungsi ReLu

Fungsi aktivasi ReLU (Rectified Linear Unit) adalah salah satu jenis fungsi aktivasi yang umum digunakan dalam jaringan saraf tiruan. ReLU mengubah *input* negatif menjadi nol, sementara *input* positif dipertahankan tanpa perubahan (Putra *et al.*, 2022). Rumus matematis dari fungsi ReLU adalah sebagai berikut:

$$y(u) = \max(0, u)$$

atau

$$y(u) = \begin{cases} u, & \text{if } u \geq 0 \\ 0 & \text{if } u < 0 \end{cases} \quad (2.4)$$



Gambar 2. 13 Fungsi Aktivasi ReLu



2.13.4 Tahapan pada Metode *Artificial Neural Network* (ANN)

Menurut (Cynthia & Ismanto, 2017) terdapat tahapan-tahapan pada metode *Artificial Neural Network* (ANN):

1. Melakukan *import library* yang diperlukan dan membaca data dari file Excel.
2. Melakukan data *cleaning* terhadap dataset.
3. Memisahkan kolom dan target dengan variabel yang berbeda dari data.
4. melakukan *scaling* pada fitur-fitur menggunakan *Standard Scaler* untuk memastikan fitur-fitur memiliki skala yang serupa.
5. Memisahkan data menjadi data *train* dan data *test*
6. Membangun model ANN:
 - a. Definisikan arsitektur model: Tentukan jumlah *layer*, jumlah neuron dalam setiap *layer*, dan fungsi aktivasi yang akan digunakan.
 - b. Inisialisasi model: Buat objek model ANN menggunakan *library* PyTorch.
 - c. Konfigurasi model: Tentukan optimizer dan fungsi *loss* yang akan digunakan
7. Pelatihan model:
 - a. Melakukan *forward propagation* untuk menghasilkan *output* prediksi:
 - Menghitung nilai *input* pada *layer* pertama dengan mengalikan *weight* dan data *input* yang dimiliki lalu menambahkannya dengan bias, berikut rumusnya:

$$Z_j = \sum_{i=1}^n x_i \cdot w_i + b_j$$

- Saat menggunakan *library* PyTorch sehingga pengaturan *weight* dan bias diatur secara otomatis menggunakan metode inisialisasi disebut *Kaiming initialization* atau "*He initialization*". Metode inisialisasi ini mempertimbangkan jumlah *input* dan *output* dari setiap lapisan linier untuk mengatur skala bobot secara cerdas.



- Setelah mendapatkan nilai *input*, memasukkan nilai tersebut ke dalam fungsi aktivasi ReLU untuk menghasilkan *output* neuron. Fungsi aktivasi diterapkan untuk memperkenalkan non-linearitas pada *output layer* dan memungkinkan model untuk mempelajari pola yang lebih kompleks dalam data. Aktivasi ini kemudian akan menjadi *input* untuk *layer* berikutnya, yaitu *layer* kedua. Langkah ini akan terus berlanjut hingga mencapai *layer output*. Fungsi aktivasi sigmoid diterapkan setelah *output* dari model untuk menghasilkan probabilitas *output* yang berada dalam rentang (0, 1) pada masalah klasifikasi biner.

b. Melakukan perhitungan *loss*:

- Membandingkan *output* prediksi dengan label biner yang sebenarnya.
- Fungsi *loss* yang digunakan yaitu BCELoss:

$$BCELoss = \frac{1}{N} \sum_{i=1}^N -(y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i))$$

c. *Backpropagation*

Setelah mendapatkan nilai *error*, kemudian dilakukan dengan *backpropagation* untuk memperbaiki nilai *weight* dan bias. Rumus utama untuk memperbaiki suatu bobot *w* berdasarkan *error* *E* adalah:

$$w_{new} = w_{old} - \alpha \frac{\partial E}{\partial w}$$

Rumus untuk memperbaiki nilai bias:

$$b_{new} = b_{old} - \alpha \frac{\partial E}{\partial b}$$

menunjukkan *learning rate*, sebuah konstanta (biasanya antara 0 - 1) yang menentukan seberapa cepat proses pembelajaran model dilakukan



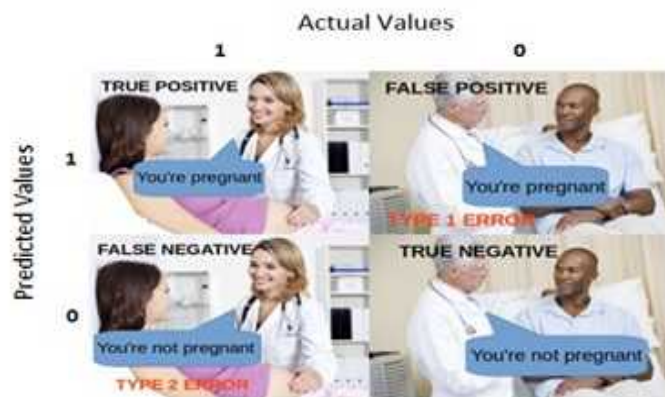
8. Melakukan evaluasi model menggunakan data uji dan periksa metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.

2.14 Metrik Klasifikasi pada *Machine Learning*

Metrik adalah suatu alat pengukuran atau petunjuk yang digunakan untuk menilai prestasi dari model *machine learning*. Metrik ini berguna untuk mengevaluasi sejauh mana model *machine learning* mampu mencapai tujuannya, seperti menghasilkan prediksi yang akurat atau memisahkan data ke dalam kelas-kelas yang tepat. Dengan menggunakan metrik, kita dapat membandingkan kinerja model *machine learning* dan memahami cara kerjanya dalam berbagai situasi. Selain itu, metrik juga membantu menentukan apakah model *machine learning* perlu ditingkatkan atau diperbaiki, serta memberikan petunjuk tentang bagaimana mencapai hasil yang lebih baik.

2.14.1 *Confusion matrix*

Confusion matrix adalah metode evaluasi yang digunakan untuk menilai performa model *machine learning* dalam masalah klasifikasi dengan dua kelas atau lebih. *confusion matrix* dapat diwakili dalam bentuk tabel yang terdiri dari empat kombinasi nilai prediksi dan nilai aktual yang berbeda. Terdapat empat istilah yang digunakan dalam *confusion matrix* untuk mewakili hasil klasifikasi, yaitu *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif*.



Gambar 2. 14 Contoh *Confusion matrix*



1. *True Positif* mengacu pada kasus ketika model dengan benar mengklasifikasikan seorang wanita sebagai hamil.
2. *False Positif* terjadi ketika model salah mengklasifikasikan seorang wanita sebagai hamil, padahal sebenarnya tidak.
3. *False Negatif* terjadi ketika model salah mengklasifikasikan seorang wanita sebagai tidak hamil, padahal sebenarnya dia hamil.
4. *True Negatif* mengindikasikan kasus ketika model dengan benar mengklasifikasikan seorang wanita sebagai tidak hamil.

Dengan menggunakan *confusion matrix* dan istilah-istilah ini, kita dapat memperoleh wawasan yang lebih rinci tentang bagaimana model *machine learning* melakukan klasifikasi dan seberapa baik performanya. Hal ini membantu kita dalam mengevaluasi apakah model tersebut berhasil mencapai tujuan yang diinginkan, serta memberikan pemahaman yang lebih baik tentang langkah-langkah perbaikan atau peningkatan yang dapat dilakukan untuk mencapai hasil yang lebih baik. Dengan menggunakan *confusion matrix*, kita dapat menghitung berbagai metrik kinerja yang digunakan untuk mengevaluasi performa model yang telah dibuat.

Pada bagian ini, mari kita memahami beberapa metrik kinerja populer yang sering digunakan, yaitu akurasi (*accuracy*), presisi (*precision*), dan *recall*.

1. Akurasi adalah metrik yang mengukur sejauh mana model mampu mengklasifikasikan data dengan benar secara keseluruhan. Metrik ini diperoleh dengan membagi jumlah prediksi yang benar (*true positive* dan *true negative*) dengan jumlah total data yang dievaluasi.

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

2. Presisi mengukur seberapa baik model dalam mengklasifikasikan data positif dengan benar. Metrik ini menghitung persentase prediksi benar positif (*true positive*) dari semua prediksi positif (*true positive* dan *false positive*). Presisi



berguna dalam mengidentifikasi seberapa akurat model dalam mengklasifikasikan data positif memiliki target FP sekecil mungkin.

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

Contoh:

- klasifikasi email *spam* vs *non spam*
 - kebanyakan email kita adalah *non spam*, beberapa email yang sebenarnya adalah *spam*
 - Akan fatal kalau email biasa tapi kita anggap *spam* (*False Positive*)
3. *Recall*, juga dikenal sebagai sensitivitas, mengukur seberapa baik model dalam menemukan kembali semua data positif yang seharusnya. Metrik ini menghitung persentase prediksi benar positif (*true positive*) dari semua data yang sebenarnya positif (*true positive* dan *false negative*). *Recall* penting dalam mengidentifikasi seberapa baik model dalam menangkap atau mendeteksi data positif memiliki target FN sekecil mungkin.

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

Contoh:

- Klasifikasi kanker vs non kanker,
 - Kebanyakan label kita tidak kanker,
 - Akan fatal kalau kanker tapi kita bilang tidak kanker (*False Negative*)
4. *F1-Score* adalah salah satu metrik evaluasi yang umum digunakan dalam masalah klasifikasi. Metrik ini menggabungkan presisi (*precision*) dan *recall* untuk memberikan gambaran keseluruhan tentang performa model.



$$\frac{1}{F_1} = \frac{1}{2} \left(\frac{1}{Precision} + \frac{1}{Recall} \right)$$

Dengan memahami dan menghitung metrik-metrik ini menggunakan *confusion matrix*, kita dapat mendapatkan wawasan yang lebih dalam tentang kinerja model yang telah dibuat. Metrik-metrik ini membantu kita mengevaluasi sejauh mana model mampu memenuhi tujuan yang diinginkan dan memberikan informasi yang berguna dalam mengidentifikasi kekuatan dan kelemahan model tersebut (Heydarian *et al.*, 2022)

