

SKRIPSI

**IMPLEMENTASI SOFT VOTING CLASSIFIER UNTUK
PREDIKSI KINERJA MITRA
(STUDI KASUS : PT TELKOM INDONESIA)**

Disusun dan diajukan oleh:

**DEA WAHSA SAPUTRI
D121 19 1083**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**



Optimized using
trial version
www.balesio.com

LEMBAR PENGESAHAN SKRIPSI

IMPLEMENTASI SOFT VOTING CLASSIFIER UNTUK PREDIKSI KINERJA MITRA (STUDI KASUS : PT TELKOM INDONESIA)

Disusun dan diajukan oleh

DEAWAHSA SAPUTRI
D121 19 1083

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin Pada tanggal 12 Januari 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

Elly Warni S.T., M.T.
NIP 198202162008122001

Prof. Dr. Eng. Ir. Intan Sari Arni, ST., MT, IPU
NIP 197502032000122002

Ketua Program Studi,



Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 197507162002121004



PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;
Nama : Dea Wahsa Saputri
NIM : D121 19 1083
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Implementasi Soft Voting Classifier Untuk Prediksi Kinerja Mitra
(Studi Kasus : PT Telkom Indonesia)

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

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

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

Gowa, 12 Januari 2024

Menyatakan

Dea Wahsa Saputri



ABSTRAK

DEA WAHSA SAPUTRI. *Implementasi Soft Voting Classifier untuk Prediksi Kinerja Mitra* (Studi Kasus: PT Telkom Indonesia) (dibimbing oleh Elly Warni dan Intan Sari Areni)

Kinerja mitra pada Divisi Regional VII Makassar di PT Telkom Indonesia dievaluasi berdasarkan kontribusi mitra dalam pekerjaan proyek pengadaan dan pemasangan *Outside Plant Fiber Optic* (OSP-FO), yang melibatkan pemasangan jaringan *fiber optic* di luar bangunan di wilayah telekomunikasi Makassar hingga ke bagian timur Indonesia. Evaluasi ini mencakup berbagai aspek, yaitu *General Affair* (GA), *Planning Engineering Development* (PED), *Supply Management Information for Logistic Enhancement* (SMILE) dan penilaian lapangan oleh Wilayah Telekomunikasi (Witel). Untuk menghasilkan prediksi kinerja mitra, penelitian ini mengimplementasikan metode *soft voting classifier*.

Pada penelitian ini, tujuan dari implementasi metode *soft voting classifier* yang menggabungkan algoritma XGBoost, algoritma AdaBoost dan algoritma *gradient boosting* yaitu untuk meningkatkan akurasi prediksi kinerja mitra. Hasil implementasi metode tersebut digunakan dalam membuat sistem informasi untuk proses evaluasi seluruh kinerja mitra.

Dalam penggunaan model *soft voting classifier*, metode *firefly* diterapkan untuk menentukan bobot optimal pada ketiga algoritma yang digunakan. Bobot optimal ini diterapkan untuk menggabungkan hasil klasifikasi dari tiap algoritma dalam membuat prediksi akhir.

Penelitian ini menggunakan 7329 data proyek dari 39 mitra. Hasil penelitian menunjukkan bahwa prediksi kinerja mitra dengan metode *soft voting classifier* memiliki performa yang lebih baik yaitu mencapai akurasi 91,6% dibandingkan dengan akurasi algoritma XGBoost 87,1%, algoritma AdaBoost 81,2% dan algoritma *gradient boosting* 86,7%.

Kata Kunci: Kinerja Mitra, PT Telkom Indonesia, *Soft Voting Classifier*



ABSTRACT

DEA WAHSA SAPUTRI. Implementation of Soft Voting Classifier for Partner Performance Prediction (Case Study: PT Telkom Indonesia) (supervised by Elly Warni and Intan Sari Areni)

The performance of partners in the Regional Division VII Makassar at PT Telkom Indonesia is evaluated based on their contributions to the project work of procuring and installing the Outside Plant Fiber Optic (OSP-FO), which involves installing fiber optic networks outside buildings in the telecommunications area of Makassar and extending to the eastern parts of Indonesia. This evaluation encompasses various aspects, including General Affairs (GA), Planning Engineering Development (PED), Supply Management Information for Logistic Enhancement (SMILE), and field assessments by the Telecommunications Area (Witel). To generate performance predictions, this study implements the soft voting classifier method.

In this research, the purpose of implementing the soft voting classifier method, which combines the XGBoost, AdaBoost, and gradient boosting algorithms, is to improve the accuracy of partner performance predictions. The results of this method are used in creating an information system for the evaluation process of all partner performances.

In the soft voting classifier model, the firefly method is applied to determine the optimal weights for the three algorithms used. These optimal weights are applied to combine the classification results from each algorithm in making the final prediction.

This study uses data from 7329 projects from 39 partners. The results show that partner performance predictions with the soft voting classifier method have better performance, achieving an accuracy of 91.6%, compared to the accuracy of the XGBoost algorithm at 87.1%, the AdaBoost algorithm at 81.2%, and the gradient boosting algorithm at 86.7%.

Keywords: Partner Performance, PT Telkom, Soft Voting Classifier



DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
PERNYATAAN KEASLIAN.....	Error! Bookmark not defined.
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR SINGKATAN DAN ARTI SIMBOL	ix
DAFTAR LAMPIRAN.....	xi
KATA PENGANTAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 PT Telkom Indonesia (Persero) Tbk Wilayah Telekomunikasi (Witel) Makassar	4
2.2 <i>Supply Management Information for Logistic Enhancement (SMILE)</i>	5
2.3 <i>General Affair (GA)</i>	6
2.4 <i>Planning Engineering Deployment (PED)</i>	6
2.5 <i>Outside Plant Fiber Optic (OSP-FO)</i>	7
2.6 <i>Data Mining</i>	8
2.7 <i>Metode Ensemble</i>	9
2.8 <i>Extreme Gradient Boosting (XGBoost)</i>	11
2.9 <i>Adaptive Boosting (AdaBoost)</i>	14
2.10 <i>Gradient Boosting</i>	15
2.11 <i>Firefly</i> atau Kunang-Kunang	17
2.12 <i>Distribusi Chi-Square</i>	19
2.13 <i>Confusion Matrix</i>	20
BAB III METODE PENELITIAN.....	24
3.1 Waktu dan Lokasi Penelitian	24
3.2 Instrumen Penelitian	24
3.3 Tahapan Penelitian.....	24
3.4 Teknik Pengambilan Data.....	28
3.5 Pengolahan Data	30
3.6 Perancangan Sistem	31
3.7 Model Klasifikasi XGBoost.....	43
3.8 Model Klasifikasi AdaBoost.....	46
3.9 Model Klasifikasi <i>Gradient Boosting</i>	49
3.10 Implementasi Soft Voting Classifier.....	51
3.11 Implementasi Klasifikasi.....	53
3.12 Implementasi Algoritma dalam Sistem Informasi	54
HASIL DAN PEMBAHASAN	55



4.1 Pengumpulan Data dan Pelabelan Data	55
4.2 <i>Splitting Dataset</i>	56
4.3 Analisis Klasifikasi dengan XGBoost.....	56
4.4 Analisis Klasifikasi dengan AdaBoost.....	59
4.5 Analisis Klasifikasi dengan <i>Gradient Boosting</i>	61
4.6 Analisis Klasifikasi dengan <i>Soft Voting Classifier</i>	63
4.7 Perbandingan Kinerja Model	66
4.8 Prediksi Kinerja Mitra.....	68
4.9 Implementasi <i>Soft Voting Classifier</i> dalam Sistem Informasi	75
BAB V KESIMPULAN DAN SARAN.....	82
5.1 Kesimpulan	82
5.2 Saran.....	82
DAFTAR PUSTAKA	83



DAFTAR GAMBAR

Gambar 1 Tampilan SMILE.....	5
Gambar 2 Proses dan tahap data mining	9
Gambar 3 Ilustrasi <i>soft voting</i>	11
Gambar 4 Diagram alur XGBoost.....	13
Gambar 5 Algoritma AdaBoost	15
Gambar 6 Diagram alur <i>gradient boosting</i>	16
Gambar 7 <i>Confusion matrix</i>	21
Gambar 8 Tahapan penelitian	25
Gambar 9 Dataset SMILE.....	29
Gambar 10 Dataset penilaian lapangan oleh Witel	30
Gambar 11 Pengolahan data	30
Gambar 12 <i>Flowchart</i> sistem prediksi kinerja mitra.....	31
Gambar 13 <i>Feature selection</i> dataset GA dan PED.....	32
Gambar 14 <i>Feature selection</i> dataset SMILE.....	32
Gambar 15 <i>Feature selection</i> dataset penilaian lapangan oleh witel.....	33
Gambar 16 <i>Data cleaning</i> pada <i>feature selection</i>	35
Gambar 17 Pemeriksaan nilai yang hilang pada <i>feature selection</i>	35
Gambar 18 Alur <i>data transformation</i>	36
Gambar 19 <i>Data concatenation</i>	37
Gambar 20 <i>Type transformation</i> variabel durasi kontrak	37
Gambar 21 <i>Type transformation</i> variabel durasi penyelesaian.....	38
Gambar 22 <i>Type transformation</i> variabel <i>local amount</i>	38
Gambar 23 Pembuatan fitur nilai proyek per LoP	39
Gambar 24 Proses ekstraksi fitur statistik.....	40
Gambar 25 Proses <i>categorical encoding</i>	40
Gambar 26 Proses normalisasi pada variabel nilai performansi KHS	41
Gambar 27 Proses <i>data transformation multilabel binarizer</i>	42
Gambar 28 Diagram proses XGBoost.....	44
Gambar 29 Diagram proses AdaBoost.....	46
Gambar 30 Diagram proses <i>gradient boosting</i>	49
Gambar 31 Ilustrasi <i>soft voting classifier</i>	52
Gambar 32 Grafik pembagian kelas dataset.....	55
Gambar 33 <i>Confusion matrix</i> model XGBoost.....	57
Gambar 34 <i>Confusion matrix</i> model AdaBoost.....	59
Gambar 35 <i>Confusion matrix</i> model <i>gradient boosting</i>	61
Gambar 36 <i>Confusion matrix soft voting</i>	64
Gambar 37 Grafik perbandingan kinerja model.....	68
Gambar 38 Prediksi kinerja mitra	69
Gambar 39 Proporsi <i>name of vendor</i> dengan label baik sekali.....	70
Gambar 40 Proporsi <i>name of vendor</i> dengan label baik	71
Gambar 41 Proporsi <i>name of vendor</i> dengan label cukup	72
Gambar 42 Proporsi <i>name of vendor</i> dengan label buruk.....	73



DAFTAR TABEL

Tabel 1 Contoh keputusan uji independensi <i>chi-square</i>	20
Tabel 2 <i>Data labelling</i> dalam memprediksi kinerja proyek mitra	27
Tabel 3 Keputusan uji independensi <i>chi-square</i> data kinerja mitra	33
Tabel 4 Rincian jumlah proyek pada setiap kelas	56
Tabel 5 Distribusi data train dan data test	56
Tabel 6 Evaluasi model XGBoost	57
Tabel 7 Evaluasi model AdaBoost	60
Tabel 8 Evaluasi model <i>gradient boosting</i>	62
Tabel 9 Hasil bobot <i>firefly</i>	63
Tabel 10 Evaluasi model <i>soft voting classifier</i>	64
Tabel 11 Perbandingan kinerja model	66
Tabel 12 Hasil peringkat kinerja mitra PT Telkom Regional VII Makassar	74
Tabel 13 Pengujian <i>black box</i>	77



DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
L	Kerugian
Ω	Regularisasi
$h_t(x)$	<i>Weak</i> atau <i>basic classifiers</i>
α_t	Tingkat pembelajaran (<i>learning rate</i>)
θ	Parameter XGBoost
y_i	Nilai aktual
\hat{y}_i	Hasil prediksi XGBoost
n	Jumlah iterasi XGBoost
$H(x)$	<i>Final classifier</i>
$y(x)$	Nilai target
$F_0(x)$	Model <i>baseline</i>
x_i^t	Posisi kunang-kunang I sebelum berpindah
t	Waktu
γ	Variasi daya tarik (<i>variation of attractiveness</i>)
ε_i^t	<i>Vector</i> bilangan acak yang didapatkan dari distribusi Gaussian
TP	<i>True positive</i>
TN	<i>True negative</i>
FP	<i>False positive</i>
FN	<i>False Negative</i>
XGBoost	Extreme gradient boosting
AdaBoost	Adaptive boosting
GA	<i>General Affair</i>
PED	<i>Planning Engineering Development</i>
SMILE	<i>Supply Management Information for Logistic Enchancement</i>
	Wilayah Telekomunikasi
	<i>Outside Plant Fiber Optic</i>



Lambang/Singkatan	Arti dan Keterangan
LoP	<i>List of Projects</i>
KHS	Kontrak Harga Satuan
OSP	<i>Outside Plant</i>



DAFTAR LAMPIRAN

Lampiran 1 Surat validasi instrumen penelitian TA	86
Lampiran 2 Tampilan dataset GA dan PED.....	87
Lampiran 3 Tampilan dataset SMILE.....	88
Lampiran 4 Tampilan dataset penilangan lapangan oleh Witel	89
Lampiran 5 Tampilan data salah prediksi dalam <i>confusion matrix</i> model XGBoost.....	90
Lampiran 6 Tampilan data salah prediksi dalam <i>confusion matrix</i> model AdaBoost.....	91
Lampiran 7 Tampilan data salah prediksi dalam <i>confusion matrix</i> model <i>gradient boosting</i>	92
Lampiran 8 Tampilan data salah prediksi dalam <i>confusion matrix</i> model <i>soft voting classifier</i>	93
Lampiran 9 Tampilan website.....	94
Lampiran 10 <i>Codingan soft voting classifier</i> dengan pembobotan <i>firefly</i>	97



KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir dengan judul " Implementasi Soft Voting Classifier untuk Prediksi Kinerja Mitra (Studi Kasus : PT Telkom Indonesia)". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Allah SWT atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Kedua orangtua penulis, Bapak Andri Lamade dan Ibu Sari Bulan yang selalu mendukung penulis dalam menempuh pendidikannya, selalu mendoakan penulis demi kelancaran urusan perkuliahannya, dan selalu memberi semangat penulis saat mengerjakan skripsinya. Penulis tidak akan sampai dititik yang sekarang tanpa doa dan restu kedua orangtua penulis, penulis mengucapkan banyak terima kasih kepada kedua orangtua penulis atas jasa dan kerja kerasnya untuk menfalitasi penulis dalam menjalankan dunia perkuliahan.
3. Ibu Elly Warni S.T., M.T. selaku pembimbing I dan Ibu Prof. Dr.Eng. Ir. Intan Sari Areni, ST., MT, IPU selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
4. Ibu Mukarramah Yusuf, B.Sc., MSc., Ph.D dan Bapak A. Ais Prayogi Alimuddin, S.T., M.Eng. selaku penguji, yang mengarahkan penulis untuk menyelesaikan tugas akhir.
5. Segenap Dosen dan Staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu penulis selama masa perkuliahan.
6. Kak Alamsyah Syamsuddin, Ibu Indah, Kak Seli, Kak Nining dan Kak Halkan selaku karyawan dari PT Telkom Indonesia Regional VII Makassar yang membantu penulis dalam memberikan data penelitian dan memberi masukan mengenai penelitian penulis.
7. Andi Rusmiati, Deby dan Pahrul menjadi tempat bertanya penulis saat penulis mengalami kendala dan masalah terkait tugas akhir.



, Reskita, Mirna, Farhan, Leon, Ilham, Rayyan, Arif, Ijlal, Hedar, Fatur, a, Atri dan teman-teman kelas C penulis yang selalu menghibur penulis nembuat kenangan-kenangan yang menyenangkan selama berkuliah.

9. Giga, Citra, Besse, Dita, Pia, Dila dan semua teman-teman Teknik Informatika Angkatan 2019 (S19NIFIER) selaku rekan yang telah memberi semangat selama penulis mengerjakan skripsi.
10. Mita dan Nanda yang selalu mendengarkan keluh kesah penulis dalam mengerjakan skripsi dan memberi semangat penulis untuk selalu mengerjakan skripsinya dan tidak putus semangat. Dan juga membuat kenangan-kenangan yang menyenangkan buat penulis agar selalu terhibur.
11. Diaz, Fadia, Nurhaliza, Yurni, Auly, Annisa dan Wiwi teman-teman SMA penulis yang mendukung penulis dan menghibur penulis juga selama berkuliah.
12. Teman-teman KKN Perhutanan Sosial Luwu Timur – Luwu Utara Gel.108 yang telah memberi pengalaman kepada penulis selama KKN.
13. Serta berbagai pihak atas segala dukungan dan bantuannya yang tidak dapat penulis tuliskan satu persatu.

Penulis berharap semoga Tuhan membalas segala kebaikan yang telah diterima oleh penulis dari berbagai pihak yang telah membantu mempermudah penulis dalam mengerjakan tugas akhir ini. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu penulis mengharapkan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tugas akhir ini dapat memberikan pengetahuan dan manfaat bagi penulis dan pembaca.

Gowa, 12 Januari 2024

Penulis,
Dea Wahsa Saputri



BAB I PENDAHULUAN

1.1 Latar Belakang

PT Telkom Indonesia (Persero) Tbk adalah perusahaan informasi dan komunikasi serta penyedia jasa dan jaringan telekomunikasi secara lengkap di Indonesia. Telkom mengklaim sebagai perusahaan telekomunikasi terbesar di Indonesia, dengan jumlah pelanggan telepon tetap sebanyak 15 juta dan pelanggan telepon seluler sebanyak 104 juta. Telkom merupakan salah satu BUMN yang 52,09% sahamnya saat ini dimiliki oleh publik. Perusahaan ini menyediakan berbagai layanan komunikasi lainnya termasuk interkoneksi jaringan telepon, multimedia, data dan layanan terkait komunikasi internet, sewa transponder satelit, sirkuit langganan, televisi berbayar dan layanan *Voice over Internet Protocol* (VoIP) (PT Telkom Indonesia, 2020). Dalam upaya mempertahankan posisi sebagai penyedia layanan telekomunikasi terbesar di Indonesia dan menjamin kualitas layanan yang tinggi kepada jutaan pelanggannya, PT Telkom Indonesia menerapkan sistem evaluasi untuk kinerja mitra yang bekerja sama dalam berbagai proyek.

Performansi mitra dinilai setiap tahun berdasarkan berbagai aspek proyek yang dikerjakan, termasuk jumlah proyek, durasi pengerjaan, anggaran proyek dan aspek di lapangan seperti kelengkapan peralatan dan sarana kerja, ketersediaan material, jumlah anggota tim yang lengkap, serta tingkat kerapihan yang dicapai di berbagai lokasi di wilayah Timur Indonesia. Target dari Evaluasi ini adalah efisiensi proyek yang diukur melalui penyelesaian jumlah proyek dalam waktu yang ditentukan, pengelolaan anggaran yang efektif untuk memastikan penggunaan dana secara efisien dan memastikan bahwa seluruh aspek operasional di lapangan berjalan dengan baik dan sesuai dengan standar yang ditetapkan.

Sebelumnya performansi mitra dinilai menggunakan *Supply Management Information for Logistic Enhancement* (SMILE), dan memberikan rating bintang



untuk apresiasi terhadap kinerja administrasi pengadaan yang telah dilakukan oleh mitra berdasarkan laporan kinerja performansi mitra dalam pelaksanaan proyek pengadaan dan pemasangan *Outside Plant Fiber Optic* atau biasa

disebut dengan OSP-FO. Namun, penilaian SMILE memiliki kekurangan dalam hal mengevaluasi kinerja mitra sehingga adanya penambahan tiga aspek penilaian baru, yaitu *General Affair (GA)*, *Planning Engineering Development (PED)*, dan penilaian lapangan oleh Wilayah Telekomunikasi (Witel). Parameter penilaian dari ketiga penilai tersebut digabungkan dengan penilaian oleh SMILE untuk menghitung nilai performansi mitra. Dengan penggabungan aspek parameter penilaian dari 4 penilai, performansi mitra menjadi lebih menyeluruh karena mempertimbangkan beberapa aspek penilaian yang berbeda. Hasil penggabungan ini kemudian diterapkan dalam proses evaluasi kinerja 39 mitra yang terlibat dalam proyek OSP-FO di area Witel Makassar hingga witel Indonesia Timur.

Penelitian yang dilakukan oleh Nova Agustina pada tahun 2023 menggunakan metode pendekatan *ensemble* yaitu *soft voting* untuk melakukan analisis sentimen Covid19, bertujuan untuk membandingkan dan menerapkan algoritma *logistic regression*, *naïve bayes*, dan *support vector machine* menggunakan *soft voting*. Pada penelitian ini, semua algoritma mendapatkan akurasi yang sama untuk analisis sentimen, yaitu sebesar 89%. Selanjutnya, penerapan metode *ensemble* meningkatkan akurasi model untuk prediksi sentimen menjadi 91% (Agustina & Ihsan, 2023).

Maka dari itu, penelitian ini mengembangkan penelitian yang telah dilakukan sebelumnya dengan menggunakan metode *soft voting classifier* untuk meningkatkan akurasi. Hasil dari pemanfaatan dari *soft voting classifier* ini dapat menjadi solusi dalam memprediksi kinerja mitra PT Telkom Indonesia berdasarkan parameter-parameter yang digunakan berdasarkan penilaian GA, PED, SMILE dan penilaian lapangan oleh Witel.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana implementasi *soft voting classifier* dalam mengevaluasi kinerja mitra PT Telkom Indonesia?
bagaimana unjuk kerja sistem evaluasi kinerja mitra dengan *soft voting assifier*?



1.3 Tujuan Penelitian

Tujuan dalam penelitian ini adalah:

1. Mengimplementasikan *soft voting classifier* dalam mengevaluasi kinerja mitra PT Telkom Indonesia.
2. Menganalisa unjuk kerja sistem evaluasi kinerja mitra dengan *soft voting classifier*

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini diuraikan sebagai berikut:

1. Bagi institusi pendidikan, dapat digunakan sebagai referensi dalam pengembangan penelitian topik terkait penelitian prediksi kinerja mitra dengan memanfaatkan metode *soft voting classifier*.
2. Bagi PT Telkom Indonesia, sistem evaluasi kinerja mitra dengan *soft voting classifier* yang terintegrasi dalam sebuah sistem informasi dapat digunakan untuk memilih mitra terbaik untuk bekerja sama dengan performansi terbaiknya.

1.5 Ruang Lingkup

Ruang lingkup dalam penelitian ini adalah:

1. Penelitian ini menggunakan studi kasus pada proyek mitra di PT Telkom Indonesia, Divisi Regional VII Makassar yang melibatkan 7329 proyek dari 39 mitra pada tahun 2022.
2. Data yang digunakan yaitu *name of vendor* (nama mitra), *short text* (ringkasan proyek), *PO date* (tanggal pesanan pembelian), *delivery date* (tanggal pengiriman barang), *Doc date GR* (tanggal pencatatan penerimaan barang), *local amount* (nilai transaksi), jumlah proyek, nilai performansi KHS, alker/salker, stok material, jumlah team, dan kerapihan.
3. Metode klasifikasi yang digunakan yaitu XGBoost, AdaBoost dan *gradient boosting*.



BAB II TINJAUAN PUSTAKA

2.1 PT Telkom Indonesia (Persero) Tbk Wilayah Telekomunikasi (Witel) Makassar

PT Telkom Indonesia sebagai perusahaan telekomunikasi terkemuka di Indonesia menyediakan jasa telepon tetap kabel (*fixed wireline*), jasa telepon nirkabel (*fixed wireless*), jasa telepon bergerak (*mobile service*), data internet serta jasa multimedia lainnya. Tahun 2001 Telkom membeli 35% saham Telkomsel dari PT Indosat sebagai bagian dari implementasi restrukturisasi industri jasa telekomunikasi di Indonesia yang ditandai dengan penghapusan kepemilikan bersama dan kepemilikan silang antara Telkom dan Indosat. PT Telkom Indonesia juga mempunyai anak perusahaan seperti Telkomsel, Telkom Akses, Graha Sarana Duta, Patrakom, Telkom Metra, Mitratel, Telin, Jalin, Metranet, Telkom Infra dan PINS (Haslina, 2022).

Dalam struktur organisasi, PT Telkom Indonesia mengintegrasikan unit Wilayah Telekomunikasi (Witel) yang mengoperasikan dan mengelola kegiatan bisnis telekomunikasi di wilayah-wilayah tertentu. Witel merupakan wilayah usaha telekomunikasi, dimana kantor Witel Makassar membawahi beberapa usaha cabang yang berada di beberapa kabupaten Sulawesi Selatan, yaitu diantaranya, Kabupaten Maros, Takalar, Sinjai, Bone, Jeneponto dan lain-lain. Hasil dari setiap kegiatan yang dilakukan oleh beberapa cabang Witel di beberapa Kabupaten akan di laporkan pada kantor Witel yang berada di Makassar untuk diinput dan hasil penginputan segala kegiatan Witel tersebut nantinya akan dikirim melalui monitor menuju kantor pusat yang berada di Makassar, yakni kantor Regional (Haslina, 2022).

Kantor Regional yang terletak di Jalan A.P Pettarani No. 2, tidak hanya berfungsi sebagai pusat koordinasi untuk Witel yang berada di bawahnya tetapi juga sebagai pusat kegiatan utama PT Telkom Indonesia Tbk di kota Makassar (Haslina, 2011). Di bawah kepemimpinan seorang *general manager*, Witel bertanggung jawab pelaksanaan berbagai program, termasuk pembangunan infrastruktur telekomunikasi, aktivitas pemasaran atau penjualan, serta penyediaan layanan



2.2 Supply Management Information for Logistic Enhancement (SMILE)

Supply Management Information for Logistic Enhancement (SMILE) adalah aplikasi *web online* manajemen logistik yang dapat diakses melalui <http://smile.telkom.co.id>. SMILE dapat memberikan solusi kepada mitra atau masyarakat yang ingin menjadi *supplier* Telkom. Aplikasi ini berguna untuk mengelola program kerja, proyek, maupun aktivitas lainnya secara detail dan sederhana (Astuti, 2018). Tampilan awal SMILE ditunjukkan pada Gambar 1 berikut.



Gambar 1 Tampilan SMILE

Sumber: <https://smile.telkom.co.id> (2005)

SMILE dibangun berdasarkan *Project Manajemen Office*, yang bertujuan untuk membantu mitra lebih mudah mengendalikan rencana proyek beserta implementasinya, secara resmi sistem ini diresmikan oleh Bapak Abdul Haris selaku Direktur *Netword Solution* di Surabaya pada tanggal 10 November 2005 (Astuti, 2018).

SMILE digunakan oleh mitra beserta timnya untuk berperan secara aktif dalam pengambilan keputusan, diskusi, penetapan *milestone*, laporan pekerjaan dan kolaborasi seluruh proses kegiatan. SMILE menyediakan suatu nilai tambah yang tinggi dalam layanan manajemen proyek yang dibangun berdasarkan pengalaman berbagai tipe pengguna dan organisai. Pengalaman dalam pengelolaan di diimplementasikan kedalam *template* dan aplikasi yang siap digunakan LE user (Astuti, 2018).



2.3 *General Affair (GA)*

General Affairs (GA) adalah posisi dalam perusahaan yang biasanya berada di bawah kepala divisi umum atau kepala operasi. GA bertanggung jawab untuk mengelola berbagai aspek operasional perusahaan, termasuk urusan yang tidak terduga. Posisi ini penting dalam menyediakan kejelasan mengenai tujuan, prospek, dan rencana masa depan perusahaan. Dalam struktur organisasi, GA sering terlibat dalam departemen pembelian, mengurus pengadaan barang dan jasa, serta pemeliharaan asset (Mudabicara, 2021). Selain itu, GA juga berperan penting dalam evaluasi dan pengelolaan hubungan kontrak dengan mitra kerja perusahaan.

Syarat-syarat yang harus dipenuhi sebelum melakukan kontrak mitra kerja sama yaitu dokumen justifikasi ijin prinsip/mengikuti tender, dokumen evaluasi kebutuhan barang dan atau jasa, dokumen *indicative offering request* (surat permintaan penawaran harga), dokumen berita acara rapat penjelasan pekerjaan, dokumen berita acara klasifikasi dan negoisasi, dokumen penetapan calon mitra pelaksana dan dokumen penetapan mitra pelaksana (Hakim, 2020).

2.4 *Planning Engineering Deployment (PED)*

Unit *Planning Engineering Deployment (PED)* di PT Telkom Indonesia memegang peran dalam merencanakan, menjadwalkan, mengkoordinasikan dan mengawasi proyek-proyek yang ditangani oleh unit tersebut. Selain itu, unit ini secara aktif berkomunikasi dengan mitra untuk memahami kebutuhan dan persyaratan yang spesifik. PED juga berkolaborasi erat dengan manajer proyek dan anggota tim lainnya dalam menyediakan dukungan teknis dan bantuan sesuai kebutuhan. PED secara konsisten mengevaluasi dan memperbaiki aspek teknis dari pekerjaan, mengambil tindakan korektif yang diperlukan sepanjang durasi proyek (Glassdoor, 2008).

PED berfokus pada pemeliharaan dan pengawasan yang efektif atas jaringan, dengan tujuan utama untuk menyampaikan layanan produk Telkom secara optimal.

PED ini mengambil langkah-langkah strategis untuk menjaga kestabilan dan kualitas layanan. Ini mencakup pemantauan berkelanjutan atas kinerja pelaksanaan perawatan yang terjadwal dan responsif, serta penerapan teknologi yang bertujuan meningkatkan layanan (Ananda, 2022).



2.5 Outside Plant Fiber Optic (OSP-FO)

Dalam bidang telekomunikasi, *Outside Plant Fiber Optic* (OSP) merujuk pada seluruh komponen infrastruktur yang ditempatkan di luar gedung, mencakup peralatan dan kabel yang memungkinkan operasional jaringan komunikasi seperti telepon atau kabel (Vetro, 2021). Proses desain jaringan serat optik atau *Fiber Optic* (FO) merupakan faktor penentu keberhasilan instalasi dan operasi dari jaringan ini. Melalui proses ini, ditentukan jenis sistem komunikasi yang akan ditransmisikan, tata letak geografis termasuk gedung dan kampus, serta peralatan transmisi yang dibutuhkan (The Fiber Optic Association, 2018).

Fiber optic adalah komponen utama dalam jaringan transmisi data, dibuat dari bahan seperti kaca atau plastik yang halus. Sehingga, *fiber optic* mampu mentransmisikan sinyal cahaya dengan kecepatan sangat tinggi dari satu titik ke titik lain. Kemampuan dalam transmisi cahaya, dikombinasikan dengan penggunaan LED atau laser, menjadikan *fiber optic* solusi andalan dalam komunikasi data, memberikan kecepatan transmisi yang memungkinkan komunikasi cepat dan efisien (Selular, 2020).

Dalam proyek *fiber optic*, tahap implementasi meliputi seleksi komponen yang tepat, penempatan strategis, dan proses instalasi yang dilakukan dengan cermat untuk memastikan integrasi teknis yang efisien dan kepatuhan terhadap standar regulasi. Proses ini tidak hanya terbatas pada aspek teknis, tetapi juga melibatkan perizinan, inspeksi, dan koordinasi dengan sistem lain. Aspek-aspek ini sangat penting untuk memastikan pemeliharaan yang efektif, kemampuan pemulihan jaringan dari gangguan, serta membantu dalam mengestimasi biaya dan membuat penawaran kontraktor yang akurat (The Fiber Optic Association, 2018). Selama pemasangan OSP, berbagai tantangan dihadapi, termasuk penempatan infrastruktur yang rumit mulai dari bawah tanah sampai di gedung dan pada tiang, yang menuntut strategi instalasi yang rinci dan fleksibel (Zhu, 2016). Sehingga dari tantangan tersebut, kabel FO dirancang untuk ketahanan lingkungan, esensial untuk daya tahan infrastruktur terhadap kondisi ekstrem, menekankan pentingnya desain

proteksi ekstra (Perveen3, 2022).



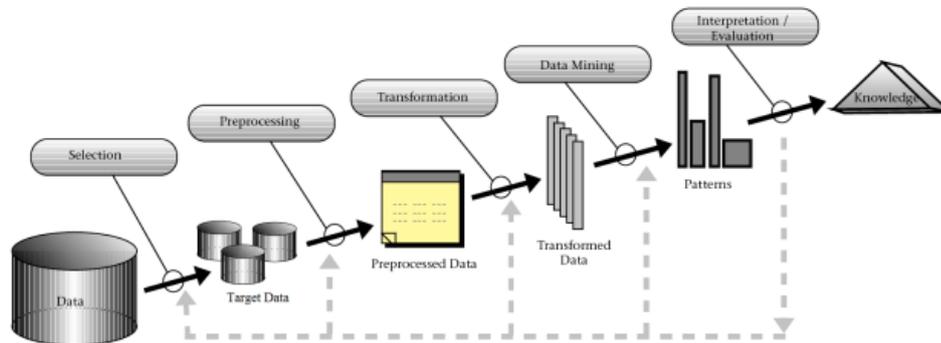
2.6 Data Mining

Data Mining (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Oleh karena itu *data mining* sebenarnya memiliki akar yang panjang dari ilmu bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*. Beberapa teknik yang sering disebut dalam literatur *data mining* antara lain yaitu *clustering*, *classification*, *association rule mining*, *neural network*, *genetic algorithm* dan lain-lain (Arif, 2019).

Alasan utama mengapa *data mining* sangat menarik perhatian industri informasi dalam beberapa tahun belakangan ini adalah karena tersedianya data dalam jumlah yang besar dan semakin besarnya kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna. *Data mining* adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan. Kegunaan *data mining* yaitu untuk menspesifikasikan pola yang harus ditemukan dalam tugas *data mining*. Secara umum tugas *data mining* dapat diklasifikasikan ke dalam dua kategori yaitu deskriptif dan prediktif. Tugas menambang secara deskriptif adalah untuk mengklasifikasikan sifat umum suatu data di dalam *database*. Tugas *data mining* secara prediktif adalah mengambil kesimpulan terhadap data terakhir untuk membuat prediksi (Arif, 2019).

Data mining merupakan bagian yang terintegrasi dari *Knowledge Discovery in Databases* (KDD) yang bertujuan untuk proses transformasi data mentah menjadi informasi berguna. Jika di gambarkan secara detail tahapan KDD menjadi 5 tahap. Tampilan proses dan tahap data mining ditunjukkan pada Gambar 2 berikut.





Gambar 2 Proses dan tahap data mining

Sumber: Arif (2019)

Tahapan-tahapan pada Gambar 2 diuraikan sebagai berikut.

a. Seleksi

Bertujuan mentransformasikan data mentah ke format yang sesuai untuk analisis. Terdiri atas proses seleksi fitur, reduksi dimensionalitas, normalisasi dan subsetting data. Proses penyeleksian atau segmentasi data menurut beberapa kriteria.

b. *Preprocessing*

Bertujuan untuk menjamin bahwa hasil proses data mining yang diintegrasikan pada sistem penunjang keputusan, benar-benar hasil yang valid. Proses pembersihan data, dimana informasi yang tidak dibutuhkan dibuang.

c. Transformasi

Proses transformasi agar data menjadi berguna dan dapat ditelusuri.

d. *Data Mining*

Proses yang berfokus pada ekstraksi pola-pola data.

e. Interpretasi dan Evaluasi

Pola diidentifikasi sistem, lalu diinterpretasikan sebagai pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan manusia.

2.7 Metode *Ensemble*



Ensemble methods adalah teknik *machine learning* yang menggabungkan model untuk mencapai performa lebih baik daripada hanya mengandalkan model saja. Teknik ini sangat populer dan banyak digunakan di industri karena

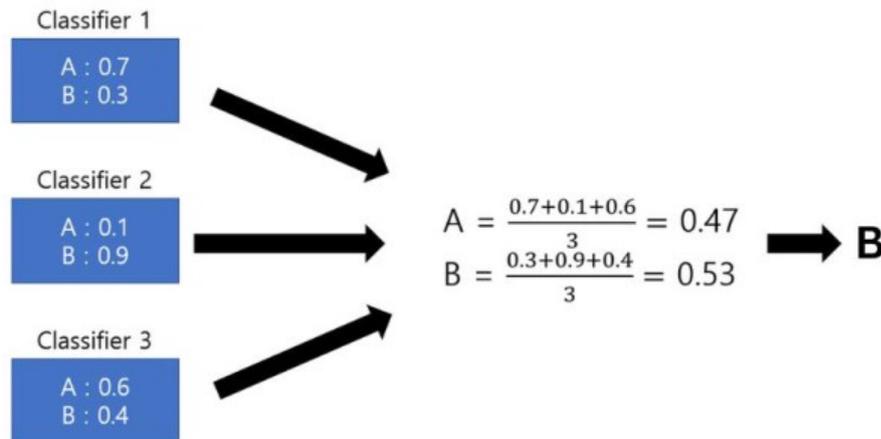
dapat meningkatkan akurasi prediksi dan mengurangi risiko *overfitting* (Kantinit, 2021).

Dalam teori pembelajaran *ensemble* yang biasa disebut model pembelajar lemah atau model dasar dapat digunakan sebagai blok bangunan untuk merancang model yang lebih kompleks dengan menggabungkan beberapa diantaranya. Biasanya, model-model dasar ini tidak berkinerja optimal secara sendiri, sering kali karena model-model tersebut memiliki bias yang tinggi atau *varians* yang berlebihan, sehingga kurang baik. Kemudian, konsep utama metode *ensemble* adalah mengatasi bias atau *varians* dari pelajar lemah tersebut dengan menggabungkan beberapa di antaranya untuk membentuk pelajar yang lebih kuat (Ichi.pro, 2023).

Salah satu contoh metode *ensemble* adalah *voting classifier*, dimana model ini dilatih dengan sekumpulan model yang berbeda dan memprediksi hasil (kelas) berdasarkan probabilitas tertinggi dari kelas yang dipilih sebagai *output*. Model ini membuat prediksi kelas *output* dengan mengkombinasikan hasil dari setiap pengklasifikasi yang terintegrasi dalam *voting classifier*, berdasarkan suara mayoritas yang paling tinggi. *Voting classifier* mendukung dua jenis pemungutan suara, yaitu pemungutan suara keras atau *hard voting* dan pemungutan suara lunak atau *soft voting*, yang masing-masing berfokus pada mayoritas suara tertinggi dan rata-rata probabilitas (Shaju, 2022).

Metode *soft voting* adalah teknik pemungutan suara berdasarkan probabilitas yang dikenal dengan pendekatannya yang efektif dan sederhana. Prinsipnya yaitu melibatkan penggabungan probabilitas yang dihasilkan oleh berbagai model untuk setiap kelas, dan kemudian memilih kelas yang memiliki probabilitas tertinggi di antara semuanya (Devkor, 2019).



Gambar 3 Ilustrasi *soft voting*

Sumber: Devkor (2019)

Gambar 3 menunjukkan ilustrasi *soft voting*, dimana setiap model komponen memberikan prediksi dalam bentuk probabilitas untuk setiap kelas yang ada. Misalnya, jika terdapat dua kelas yang mungkin, seperti 'A' dan 'B', setiap model memberikan probabilitas untuk keduanya. Prediksi akhir dihasilkan dengan cara menjumlahkan probabilitas dari semua model untuk setiap kelas dan memilih kelas dengan probabilitas tertinggi sebagai hasil akhir. Metode ini bermanfaat untuk mendapatkan informasi yang lebih terperinci dari model-model komponen, seperti probabilitas kelas untuk menghasilkan prediksi yang lebih tepat.

Dalam contoh konkret, jika terdapat beberapa model komponen dengan probabilitas prediksi yang berbeda-beda, *soft voting* akan menjumlahkan probabilitas dari setiap model untuk setiap kelas dan memutuskan kelas akhir berdasarkan total probabilitas tertinggi. Metode ini efektif terutama ketika model-model komponen memiliki karakteristik yang berbeda-beda, dan juga cocok untuk situasi, dimana model-model memberikan probabilitas yang mencerminkan tingkat keyakinan dalam prediksi yang dihasilkan.

2.8 *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) merupakan pengembangan dari *Gradient Boosting Machine (GBM)* dengan beberapa fitur tambahan guna dalam mempercepat proses komputasi dan mencegah terjadinya *overfitting*. Kunci kecepatan komputasi XGBoost terletak pada optimasi



penggunaan memori dan *cache* di komputer sehingga dapat bekerja secara efisien, meskipun berhadapan dengan data berukuran besar. Fitur itu menyebabkan XGBoost bisa dijalankan dengan waktu yang lebih cepat dibandingkan model-model canggih lainnya saat ini, seperti *deep learning* dan *random forest*. Sementara, pencegahan kejadian *overfitting* dilakukan dengan teknik regularisasi dengan cara memberikan komponen penalti pada fungsi kerugian. Dengan cara tersebut algoritma akan terhindar dari model yang terlalu kompleks, tetapi berkinerja buruk dalam memprediksi kejadian dengan data baru (Aldy, 2023).

Tianqi Chen mengklaim bahwa fitur pengontrolan *overfitting* menyebabkan XGBoost memiliki performa yang lebih baik. Klaim ini antara lain ditunjukkan oleh hasil yang didapatkan para data *scientist* pengguna XGBoost di beberapa kompetisi, misalnya KDD Cup dan CERN LHCb *Experiment Flavor of Physics*. Dalam banyak kesempatan, XGBoost hampir selalu mampu menempati posisi tiga besar kompetisi-kompetisi itu. Salah satu yang bisa disebutkan adalah raihan juara pertama kompetisi Avito Context Ad Click oleh Owen Zhang yang menggunakan XGBoost dalam pemodelannya. Kompetisi mengharuskan peserta untuk mengembangkan model yang mampu memprediksi apakah pengunjung situs Avito akan mengklik iklan yang sesuai dengan minat mereka. Menurut hasil wawancara yang dilakukan oleh Kaggle, Owen menyebutkan bahwa ada dua kunci yang membuatnya berhasil memprediksi dengan tepat. Yang pertama adalah melakukan *feature engineering* untuk memperoleh prediktor, dan yang kedua adalah menerapkan algoritma XGBoost disertai proses optimasi *hiperparameter* (Aldy, 2023).

Konsep pada XGBoost yang mengubah parameter pembelajaran secara berulang merupakan bagian dari proses optimasi untuk mengurangi nilai kerugian pada fungsi *loss*. XGBoost memanfaatkan model yang lebih terstruktur untuk membangun pohon *regresi*. Model ini menghasilkan performa yang unggul serta mampu mengurangi kompleksitas model guna menghindari *overfitting*. Prediksi akhir dari XGBoost diperoleh dengan menjumlahkan hasil prediksi dari tiap pohon



Algoritma berbasis pohon keputusan ini efektif dalam menangani data fitur kategorikal dan tidak terlalu terpengaruh oleh distribusi kelas yang tidak seimbang (Herni Yulianti et al., 2022).

Dalam pembelajaran mesin, terutama pada algoritma penguatan *gradient* seperti XGBoost, peningkatan akurasi prediksi merupakan salah satu tujuan utama dan aspek terpenting. Algoritma ini dirancang untuk mengelola data kompleks secara efektif dan efisien, memberikan manfaat yang besar dalam berbagai aplikasi analisis data. Untuk mencapai peningkatan akurasi prediksi, XGBoost menggunakan rumus pada persamaan (1) (Aws, 2023).

$$F_m(X) = F_{m-1}(X) + \alpha_m h_m(X, r_{m-1}) \quad (1)$$

dimana,

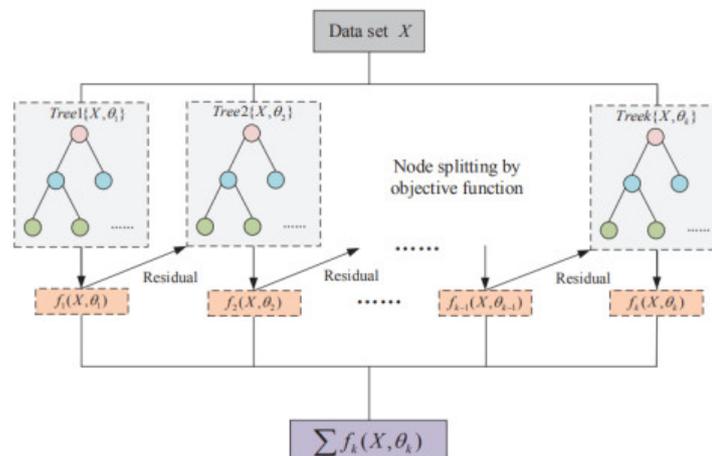
F_m = model yang telah diperbarui setelah menambahkan pohon ke-m,

F_{m-1} = model diawal,

α_m = learning rate,

h_m = pohon regresi baru yang dibangun pada iterasi ke-m,

r_{m-1} = residu dari model sebelumnya.



Gambar 4 Diagram alir XGBoost

Sumber: Guo et al. (2020)

Gambar 4 menunjukkan alur dari proses komputasi XGBoost. Selama proses pelatihan, model menghitung kerugian pada simpul-simpul secara berkelanjutan



emilih simpul daun dengan peningkatan kerugian terbesar. XGBoost tuhkan pohon-pohon baru dengan terus membagi fitur-fitur. Setiap han pohon pada setiap *iterasi* sebenarnya adalah pembelajaran fungsi baru

$f_k(X, \theta_k)$ untuk menyesuaikan sisa prediksi sebelumnya. Setelah pelatihan dan mendapatkan K pohon, atribut-atribut prediksi pada sampel akan sesuai dengan simpul daun tertentu di setiap pohon, dan masing-masing simpul daun ini sesuai dengan sebuah skor. Pada akhirnya, skor-skoranya dari setiap pohon dijumlahkan untuk mendapatkan nilai prediksi yang menggambarkan sampel tersebut.

2.9 Adaptive Boosting (AdaBoost)

Algoritma *Adaptive Boosting* (AdaBoost) adalah sebuah teknik *boosting* yang digunakan sebagai metode *ensemble* dalam *machine learning*. Algoritma ini disebut *adaptive boosting* karena bobot diberikan ulang pada setiap *instance*, dengan bobot yang lebih tinggi diberikan pada *instance* yang salah diklasifikasikan. *Boosting* sendiri bertujuan untuk mengurangi bias serta variasi dalam *supervised learning*. Algoritma ini banyak digunakan dalam berbagai aplikasi, termasuk *computer vision*, pengolahan bahasa alami (*natural language processing/NLP*) dan deteksi penipuan (*fraud detection*). Cara kerja AdaBoost adalah dengan cara *iterative* melatih *weak learners*, seperti *decision tree* atau model *linear*, pada sebuah dataset dan memberikan bobot pada setiap *instance training* berdasarkan kesalahan klasifikasinya (Trivusi, 2023).

Dalam AdaBoost, setiap sampel pelatihan mendapatkan bobot untuk menentukan peluangnya dipilih dalam set data latihan. Klasifikasi akhir didasarkan pada akumulasi bobot dari pengklasifikasi lemah, sehingga rumus umum algoritma AdaBoost ditunjukkan pada persamaan (3) sebagai berikut (Ihsan & Yusuf, 2013).

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (3)$$

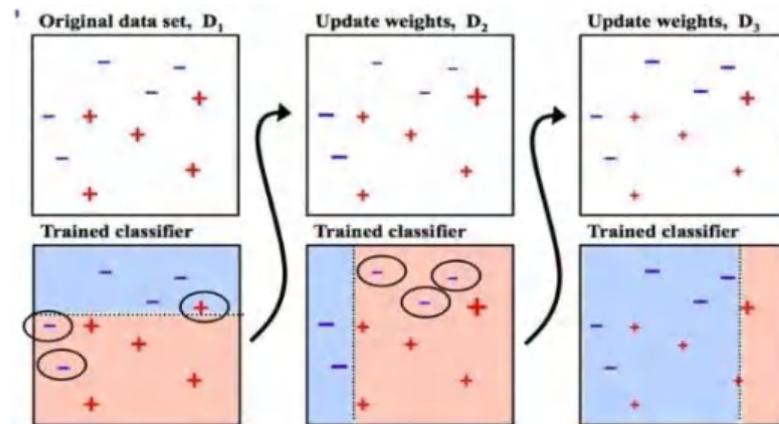
dimana,

$h_t(x)$ = *weak* atau *basic classifiers*,

α_t = bobot *weak learner*,

$f(x)$ = *final classifier*.





Gambar 5 Algoritma AdaBoost

Sumber: Trivusi (2023)

Gambar 5 menunjukkan alur dari proses AdaBoost, algoritma AdaBoost bersifat *iterative* atau berulang. Cara kerja algoritma ini dimulai dengan melatih sebuah *weak classifier* pada data pelatihan. *Weak classifier* kemudian diberi bobot berdasarkan performanya. Selanjutnya, algoritma melatih *weak classifier* kedua menggunakan data yang telah diberi bobot. *Weak classifier* kedua kemudian diberi bobot berdasarkan performanya. Proses ini diulang sejumlah *iterasi* tertentu atau hingga tingkat kesalahan berada di bawah ambang batas yang ditentukan. *Classifier* akhir adalah rata-rata terbobot dari semua *weak classifiers*. Bobot ditentukan berdasarkan tingkat kesalahan dari masing-masing *weak classifier*. Semakin rendah tingkat kesalahan, semakin tinggi bobotnya (Trivusi, 2023).

Salah satu keunggulan utama dari AdaBoost adalah kemampuannya dalam mengatasi data yang kompleks dan interaksi antar fitur. Dan juga dapat mencegah *overfitting* dengan menggunakan kombinasi terbobot dari *weak learners*. Selain itu, AdaBoost telah terbukti memiliki akurasi yang tinggi dan relatif mudah untuk diimplementasikan (Trivusi, 2023).

2.10 Gradient Boosting

Gradient boosting adalah salah satu teknik dalam *machine learning* yang digunakan untuk meningkatkan performa model prediksi. Teknik ini bekerja dengan menggabungkan beberapa model prediksi sederhana menjadi satu model yang kompleks dan akurat. *Gradient boosting* juga dapat digunakan untuk mengatasi masalah *overfitting* pada model prediksi (Mulyawan, 2023).



Menurut Friedman (2001) dalam bukunya yang berjudul “Greedy Function Approximation”. *Gradient boosting* adalah sebuah metode pembelajaran mesin yang menggunakan teknik *boosting* untuk meningkatkan performa model prediksi. Teknik ini bekerja dengan cara menggabungkan beberapa model prediksi sederhana, seperti *decision tree*, menjadi satu model yang lebih kompleks dan akurat. *Gradient boosting* juga menggunakan *gradient* dari fungsi kerugian sebagai sinyal untuk mengoptimalkan model prediksi (Mulyawan, 2023).

Gradient boosting yaitu salah satu teknik *machine learning* yang digunakan untuk memprediksi nilai target berdasarkan beberapa fitur atau variabel. Rumus atau formula *Gradient boosting* ditunjukkan pada persamaan (4) sebagai berikut (Mulyawan, 2023).

$$y(x) = F_0(x) + \alpha F_1(x) + \alpha_2 F_2(x) + \dots + \alpha_m F_m(x) \quad (4)$$

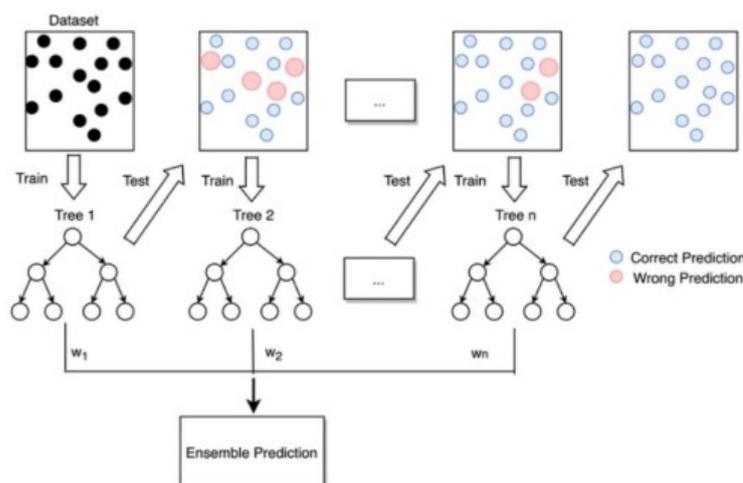
dimana,

$y(x)$ = nilai target yang ingin diprediksi,

$F_0(x)$ = model *baseline* yang digunakan sebagai acuan awal,

α = *learning rate* atau tingkat pembelajaran yang digunakan untuk mengontrol seberapa besar kontribusi setiap model pada prediksi akhir,

$F_m(x)$ = model-model tambahan yang dibangun secara bertahap untuk meningkatkan performa prediksi.



Gambar 6 Diagram alur *gradient boosting*

Sumber: Zhang et al. (2021)



Gambar 6 menunjukkan proses dimana model prediktif keseluruhan dibangun dari rangkaian *classifier* yang secara individual memiliki kekuatan prediktif yang rendah, yang disebut sebagai *weak classifiers*. Dalam metode ini, setiap titik data yang diprediksi salah oleh *classifier* sebelumnya akan mendapatkan peningkatan bobot, sehingga *classifier* yang akan dilatih selanjutnya memberikan perhatian lebih terhadap titik-titik tersebut untuk memperbaiki kesalahan prediksi yang terjadi. Pada akhirnya, keputusan atau prediksi final model dibuat dengan menghitung rata-rata tertimbang dari semua prediksi yang dihasilkan oleh setiap *classifier*, dimana *classifier* yang memiliki tingkat akurasi lebih tinggi akan memberikan kontribusi yang lebih besar terhadap prediksi akhir. Ini menghasilkan sebuah model yang kuat yang merupakan agregasi dari sejumlah *classifier* lemah yang telah diadaptasi dan dikoreksi secara berurutan (T. Zhang et al., 2021).

Konsep dari model *gradient boosting* yaitu melakukan perbaikan pada prediksi pohon pertama. Dengan kata lain, model baru yang digunakan adalah gabungan dari pohon pertama dan pohon kedua. Selanjutnya, menghitung kesalahan klasifikasi dari model *ensemble* baru dan mengembangkan pohon ketiga untuk memprediksi *residual* yang telah diubah. Proses ini diulang dalam jumlah *iterasi* tertentu. Pohon-pohon berikutnya akan membantu mengidentifikasi observasi, dimana pohon-pohon sebelumnya gagal atau menunjukkan kesalahan. Akhirnya, prediksi *ensemble* model terakhir akan berupa hasil prediksi terbobot yang disediakan oleh model pohon sebelumnya (Data Science Team, 2020).

2.11 Firefly atau Kunang-Kunang

Algoritma kunang-kunang atau *firefly algorithm* dikembangkan pertama kali oleh Yang mulai akhir tahun 2007 hingga tahun 2008 di Universitas Cambridge. Algoritma kunang-kunang terinspirasi oleh perilaku dan pola berkedip kunang-kunang. Namun, beberapa karakteristik kedipan dari kunang-kunang telah diidealkan untuk menyusun algoritma kunang-kunang, yaitu: (1) kunang-kunang dianggap *unisex*, sehingga seekor kunang-kunang akan tertarik pada kunang-kunang yang an tidak mempedulikan jenis kelaminnya, (2) daya tarik sebanding dengan eterangan, dan sama-sama menurun ketika jarak semakin besar. Oleh 1, untuk sepasang kunang-kunang yang sedang berkedip, kunang-kunang



yang lebih tidak terang akan bergerak menuju kunang-kunang yang lebih terang. Jika tidak ada yang lebih terang, kunang-kunang akan bergerak secara acak dan (3) tingkat keterangan kunang-kunang ditentukan oleh kondisi dari fungsi objektif (Setiawan et al., 2015).

Berdasarkan konsep algoritma kunang-kunang, hubungan antara daya tarik kunang-kunang dapat dirumuskan sebagai 'attractiveness' (β) dengan jarak r , seperti yang ditunjukkan pada persamaan (5). Jarak antara dua kunang-kunang, yaitu kunang-kunang i dan j yang diwakili oleh x_i dan x_j ditunjukkan pada persamaan (6). Pergerakan kunang-kunang i yang tertarik kepada kunang-kunang j yang lebih terang ditunjukkan pada persamaan (7) sebagai berikut (Kwiecień & Filipowicz, 2012).

$$\beta = \beta_0 e^{-\gamma r^2} \quad (5)$$

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (6)$$

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha_t \varepsilon^t \quad (7)$$

dimana,

β_0 = daya Tarik pada $r = 0$,

γ = koefisien penyerapan cahaya yang mengontrol penurunan intensitas cahaya,

$X_{i,k}$ = komponen ke k dari koordinat spasial x_i dari kunang-kunang,

i = jumlah dimensi,

d = jumlah dimensi,

x_i^t = posisi kunang-kunang i sebelum berpindah,

$(\alpha_t \varepsilon^t)$ = pergerakan acak kunang-kunang,

α = bilangan random antara 0 dan 1,

β_0 = bervariasi antara 0 sampai 1,

γ = bervariasi 0.1 sampai 10

ε^t = *vector* bilangan acak yang didapatkan dari distribusi Gaussian atau distribusi Uniform pada waktu t .



2.12 Distribusi *Chi-Square*

Distribusi *chi-square* (dibaca “khai square” atau “khai kuadrat” dengan simbol χ^2 adalah distribusi probabilitas teoritis yang asimetrik dan kontinu. Nilai sebuah χ^2 selalu positif antara 0 sampai dengan ∞ (tak hingga) atau $0 \leq \chi^2 \leq \infty$, tidak seperti distribusi normal atau distribusi t yang dapat bernilai negatif. Penggunaan distribusi *chi-square* ini untuk sebagai uji kecocokan atau pengujian hipotesis model terhadap data yang diamati. Nilai statistik χ^2 ditunjukkan pada persamaan 8 (Heryana, 2020).

$$\chi^2 = \frac{(f_o - f_e)^2}{f_e} \quad (8)$$

dimana,

f_o = banyaknya frekuensi yang diobservasi,

f_e = banyaknya frekuensi yang diharapkan.

Selain distribusi *chi-square*, konsep *significant level* (α) atau tingkat signifikansi juga bertujuan dalam pengujian hipotesis dimana probabilitas tertentu menolak hipotesis nol ketika hipotesis tersebut benar. Tingkat signifikansi ini mengacu pada probabilitas kesalahan dalam menolak hipotesis nol ketika sebenarnya hipotesis tersebut adalah benar, yang dikenal sebagai kesalahan tipe I. Tingkat signifikansi yang paling sering digunakan adalah 0,05 (5%) (Putriana, 2014).

Setelah menguji nilai *chi-square*, hasil nilai probabilitas (*p-value*) digunakan untuk menentukan apakah uji hipotesis nol akan ditolak atau tidak. Dalam hal ini apabila $p\text{-value} < \alpha$, maka uji hipotesis menunjukkan menolak hipotesis nol, dan apabila $p\text{-value} > \alpha$, maka uji hipotesis menunjukkan gagal menolak hipotesis nol. Hal ini menunjukkan, *p-value* yang semakin mendekati nilai nol, maka akan semakin cenderung menolak hipotesis nol. Demikian pula sebaliknya, *p-value* yang semakin mendekati angka satu, maka semakin cenderung gagal menolak hipotesis nol (Putriana, 2014).

Adapun contoh tabel keputusan uji independensi *chi-square* ditunjukkan pada



Tabel 1 Contoh keputusan uji indepedensi *chi-square*

No	Fitur	Pearson Chi-Square	Signifikansi
1	Akses	1,396	0,319
2	Akun	40,701	9×10^{-4}
3	Aman	2,301	0,210
4	Belanja	5,938	0,018
5	Buruk	18,477	0,001
⋮	⋮	⋮	⋮
769	Voucher	1,144	0,342

Sumber: Ernayanti et al. (2023)

Tabel 1 menunjukkan contoh nilai *chi-square* dan signifikansi pada setiap fitur dimana fitur “akses” menghasilkan nilai *chi-square* sebesar 1,396 dan signifikansi sebesar $0,319 > 0,05$ sehingga gagal tolak H_0 yang berarti bahwa fitur “akses” tidak memiliki ketergantungan kuat dengan kelas (positif dan negatif). Sedangkan fitur “akun” menghasilkan nilai *chi-square* sebesar 40,701 dan signifikansi sebesar $9 \times 10^{-4} < 0,05$, sehingga H_0 ditolak yang berarti bahwa fitur “akun” memiliki ketergantungan kuat dengan kelas (positif dan negatif). Seleksi fitur dengan menggunakan *chi-square* mengakibatkan pengurangan jumlah fitur yang diperoleh. Pada taraf signifikansi 0,05, jumlah fitur yang diperoleh yaitu 160 dari jumlah fitur awal yaitu 769, berkurang sebanyak 79% (Ernayanti et al., 2023).

2.13 Confusion Matrix

Confusion matrix juga sering disebut *error matrix*. Pada dasarnya *confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui (Nugroho, 2019). Tampilan *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda ditunjukkan pada

7 berikut.



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

Gambar 7 Confusion matrix

Sumber: Nugroho (2019)

Gambar 7 menunjukkan *confusion matrix* yang terdiri dari 4 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) sebagai berikut.

- *True positive* (TP) merupakan data positif yang diprediksi benar.
- *True negative* (TN) merupakan data negatif yang diprediksi benar.
- *False positive* (FP) atau type 1 error merupakan data negative namun diprediksi sebagai data positif.
- *False Negative* (FN) atau type 2 error merupakan data positif namun diprediksi sebagai data negatif.

Confusion matrix dapat digunakan untuk menghitung berbagai *performance metrics* untuk mengukur kinerja model yang telah dibuat. Dan ada beberapa *performance metrics* populer yang umum dan sering digunakan yaitu *accuracy*, *precision*, *recall*, *F1-Score*, *macro average* dan *weighted average*.

1. Accuracy

Accuracy menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Maka, *accuracy* merupakan rasio prediksi benar (*positive* dan *negative*) dengan keseluruhan data. Dengan kata lain, *accuracy* merupakan tingkat kedekatan nilai prediksi dengan nilai *actual* arnya). Nilai *accuracy* dapat diperoleh dengan persamaan (8).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$



2. Precision (Positive Predictive Value)

Precision atau presisi menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Maka, *precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah di prediksi dengan benar, berapa banyak data yang benar-benar positif. Nilai *precision* dapat diperoleh dengan persamaan (9).

$$precision = \frac{TP}{TP+FP} \quad (9)$$

3. Recall atau sensitivity (true positive rate)

Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Maka, *recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Nilai *recall* dapat diperoleh dengan persamaan (10).

$$recall = \frac{TP}{TP+FN} \quad (10)$$

6. F1-Score

F1-Score merupakan rata-rata tertimbang antara *precision* dan *recall*. Seperti yang telah diketahui, dalam *precision* dan *recall* terdapat *false positive* dan *false negative*, sehingga *F1-Score* juga mempertimbangkan keduanya. *F1-Score* umumnya lebih berguna daripada *accuracy*, terutama jika memiliki distribusi kelas yang tidak merata. *Accuracy* bekerja dengan baik jika *false positive* dan *false negative* memiliki nilai yang mirip. Namun, jika nilai *false positive* dan *false negative* sangat berbeda, lebih baik untuk mempertimbangkan baik *precision* maupun *recall*. Nilai *F1-Score* dapat diperoleh dengan persamaan (11).

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right) \quad (11)$$

7. Macro average

Macro average merupakan *metric* yang menghitung secara terpisah untuk kelas, lalu mengambil rata-rata dari semua kelas ini dengan memberikan yang sama, tanpa mempertimbangkan seberapa sering masing-masing muncul. *Macro average* ini baik digunakan untuk distribusi kelas



seimbang (L. Zhang, 2023). Nilai *macro average* dari *precision*, *recall* dan *F1-Score* dapat diperoleh dengan *precision* dihitung menggunakan persamaan (12), *recall* pada persamaan (13), dan *F1-Score* pada persamaan (14).

$$\text{macro precision} = \sum \frac{\text{presisi per kelas}}{\text{jumlah kelas}} \quad (12)$$

$$\text{macro recall} = \sum \frac{\text{recall per kelas}}{\text{jumlah kelas}} \quad (13)$$

$$\text{macro f1 - score} = \sum \frac{\text{f1-score per kelas}}{\text{jumlah kelas}} \quad (14)$$

8. *Weighted average*

Weighted average merupakan *metric* yang menghitung dengan cara bobot yang berbeda diberikan kepada setiap kelas berdasarkan frekuensi munculnya kelas tersebut dalam dataset. *Weighted average* ini baik digunakan untuk dataset dengan ketidakseimbangan kelas yang signifikan (L. Zhang, 2023). Nilai *weighted average* dari *precision*, *recall* dan *F1-Score* dapat diperoleh dengan *precision* dihitung menggunakan persamaan (15), *recall* pada persamaan (16), dan *F1-Score* pada persamaan (17).

$$\text{weighted precision} = \frac{\sum(\text{presisi per kelas} \times \text{jumlah sampel per kelas})}{\text{total jumlah sampel}} \quad (15)$$

$$\text{weighted recall} = \frac{\sum(\text{recall per kelas} \times \text{jumlah sampel per kelas})}{\text{total jumlah sampel}} \quad (16)$$

$$\text{weighted f1 - score} = \frac{\sum(\text{f1-score per kelas} \times \text{jumlah sampel per kelas})}{\text{total jumlah sampel}} \quad (17)$$

