

SKRIPSI

PERBANDINGAN ALGORITMA C4.5 DAN RANDOM FOREST PADA PEMETAAN KONDISI KERUSAKAN JALAN BERBASIS WEBGIS (STUDI KASUS : LUWU RAYA)

Disusun dan diajukan oleh:

**PUTRI AYU NENGSIH
D121 19 1070**



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

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN ALGORITMA C4.5 DAN RANDOM FOREST PADA PEMETAAN KONDISI KERUSAKAN JALAN BERBASIS WEBGIS (STUDI KASUS : LUWU RAYA)

Disusun dan diajukan oleh

Putri Ayu Nengsih
D121 19 1070

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 Juni 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,


Dr. Ir. Ingrid Nurtanio, M.T.
NIP 196108131988112001


Dr. Ir. Zahir Zainuddin, M.Sc.
NIP 196404271989101002

Ketua Program Studi,


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



PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama : Putri Ayu Nengsih
NIM : D121191070
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

**PERBANDINGAN ALGORITMA C4.5 DAN RANDOM FOREST PADA PEMETAAN
KONDISI KERUSAKAN JALAN BERBASIS WEBGIS
(STUDI KASUS : LUWU RAYA)**

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 Juni 2024

Yang Menyatakan



Putri Ayu Nengsih



ABSTRAK

PUTRI AYU NENGSIH. Perbandingan Algoritma C4.5 dan Random Forest Pada Pemetaan Kondisi Kerusakan Jalan Berbasis WebGIS (Studi Kasus : Luwu Raya) (dibimbing oleh Ingrid Nurtanio dan Zahir Zainuddin)

Kerusakan jalan diwilayah Luwu Raya merupakan tanggung jawab BBPJN VI Makassar yang dimana kerusakan disetiap ruasnya memiliki tenaga penilikan jalan untuk melakukan pelaksanaan, pengamatan, pemanfaatan jalan dan kondisi jalan setiap hari dan pelaporan pengamatan serta usulan tindakan terhadap hasil pengamatan disampaikan kepada penyelenggara jalan atau PPK (Pejabat Pembuat Komitmen). Untuk menghasilkan prediksi prioritas perbaikan jalan serta persebaran wilayah kerusakan, penelitian ini mengimplementasikan perbandingan metode C4.5 dan Random Forest.

Pada penelitian ini, tujuan dari implementasi perbandingan algoritma C4.5, dan algoritma Random Forest yaitu untuk mendapatkan hasil yang konsisten dan lebih baik diantara kedua algoritma tersebut dalam prediksi prioritas perbaikan. Hasil implementasi perbandingan kedua metode tersebut digunakan dalam membuat sistem informasi pemetaan wilayah berbasis WebGIS untuk proses penentuan prioritas perbaikan jalan raya.

Penelitian ini menggunakan 6100 data kerusakan jalan pada 3 ruas ditahun 2021- 2023. Hasil klasifikasi dengan membandingkan algortima C4.5 dan Random Forest menunjukkan akurasi yang tidak jauh beda tetapi algoritma Random Forest menunjukkan hasil yang konsisten dan lebih baik. Untuk C4.5, nilai presisi sebesar 87,9% *recall* sebesar 82,6%, f1-score sebesar 87,8% dan akurasi sebesar 88%, sedangkan Random Forest memiliki nilai presisi sebesar 86,6% *recall* sebesar 86,8%, f1-score sebesar 86,6% dan akurasi sebesar 87%.



Kata Kunci: Kerusakan Jalan, BBPJN VI Makassar, Algoritma C4.5, Random Forest, WebGIS

ABSTRACT

PUTRI AYU NENGSIH. Comparison of the C4.5 and Random Forest Algorithms in WebGIS-Based Mapping of Road Damage Conditions (Case Study: Luwu Raya) (supervised by Ingrid Nurtanio and Zahir Zainuddin)

Road damage in the Luwu Raya area is the responsibility of BBPJN VI Makassar, where damage to each section has road surveillance personnel to carry out implementation, observation, use of roads and road conditions every day and report observations and proposed actions regarding the results of observations submitted to the road administrator or PPK (Official Commitment Maker). To produce predictions of road repair priorities and the distribution of damaged areas, this research implements a comparison of the C4.5 and Random Forest methods.

In this research, the aim of implementing the comparison of the C4.5 algorithm and the Random Forest algorithm is to obtain consistent and better results between the two algorithms in predicting improvement priorities. The results of the comparative implementation of the two methods were used in creating a WebGIS-based regional mapping information system for the process of determining highway repair priorities.

This research uses 6100 road damage data on 3 sections in 2021-2023. Classification results by comparing the C4.5 and Random Forest algorithms show that the accuracy is not much different, but the Random Forest algorithm shows consistent and better results. For C4.5, the precision value is 87.9%, the recall is 82.6%, the f1-score is 87.8% and the accuracy is 88%, while Random Forest has a precision value of 86.6%, the recall is 86.8%, f1-score of 86.6% and accuracy of 87%.

Keywords: Road Damage, BBPJN VI Makassar, Algorithm C4.5, Random Forest, WebGIS



DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
PERNYATAAN KEASLIAN	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR SINGKATAN DAN ARTI SIMBOL	ix
DAFTAR LAMPIRAN.....	x
KATA PENGANTAR	xi
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Balai Besar Pelaksanaan Jalan Nasional (BBPJN) VI Makassar.....	5
2.2 Kerusakan Jalan	6
2.3 <i>Data Mining</i>	8
2.4 <i>Metode Comparing</i>	10
2.5 Decision Tree	10
2.5.1 <i>Classification and Regression Trees, version 4.5 (C4.5)</i>	11
2.5.2 <i>Random Forest</i>	13
2.6 Distribusi Chi-Square.....	14
2.7 Confusion Matrix	16
2.8. Web Geographic Information System (WebGIS)	20
BAB III METODE PENELITIAN	22
3.1 Lokasi Penelitian.....	22
3.2 Benda Uji dan Alat.....	22
3.3 Tahapan Penelitian	23
3.4 Teknik Pengambilan Data	25
3.5 Pengolahan Data.....	26
3.6 Perancangan Sistem	27
3.7 Model Klasifikasi C4.5	33
3.8 Model Klasifikasi Random Forest	35
3.9 Hasil Klasifikasi	36
3.9.1 Implementasi Algoritma dalam WebGIS.....	37
3.9.2 Implementasi Algoritma dalam WebGIS.....	38
3.9.3 Implementasi Algoritma dalam WebGIS.....	38
3.9.4 Implementasi Algoritma dalam WebGIS.....	38
3.9.5 Implementasi Algoritma dalam WebGIS.....	39



4.3 Analisis Klasifikasi dengan C4.5	40
4.4 Analisis Klasifikasi dengan Random Forest	41
4.5 Perbandingan Kinerja Model	43
4.6 Contoh Perhitungan Manual Model	45
4.7 Prediksi Kerusakan Jalan	48
4.8 Visualisasi WebGIS	52
BAB V KESIMPULAN DAN SARAN	57
5.1 Kesimpulan	57
5.2 Saran.....	57
DAFTAR PUSTAKA	58



DAFTAR GAMBAR

Gambar 1 Tahap dan Proses <i>data mining</i>	9
Gambar 2 Cara Kerja Algoritma Random Forest.....	13
Gambar 3 Tahapan Penelitian	23
Gambar 4 Dataset Kerusakan Jalan.....	26
Gambar 5 Hasil Pengolahan Data.....	27
Gambar 6 <i>Flowchart</i> sistem prediksi kerusakan jalan.....	27
Gambar 7 <i>Feature Selection</i> Awal.....	28
Gambar 8 <i>Feature Selection</i> Akhir	30
Gambar 9 <i>Data cleaning</i> pada <i>feature selection</i>	30
Gambar 10 Pemeriksaan nilai yang hilang pada <i>feature selection</i>	30
Gambar 11 Alur <i>data transformation</i>	31
Gambar 12 Dataset Akhir dari <i>feature selection</i>	31
Gambar 13 <i>Type Transformation</i>	32
Gambar 14 Proses <i>data transformation multilabel binarizer</i>	32
Gambar 15 Diagram Proses C4.5	34
Gambar 16 Diagram Proses Random Forest	35
Gambar 17 Grafik pembagian kelas dataset.....	38
Gambar 18 <i>Confusion matrix</i> model C4.5.....	40
Gambar 19 <i>Confusion Matrix</i> model Random Forest	42
Gambar 20 Grafik Perbandingan Kinerja Model	45
Gambar 21 Contoh Kasus.....	45
Gambar 22 Pohon Keputusan C4.5	46
Gambar 23 Pohon Keputusan 1.....	47
Gambar 24 Pohon Keputusan 2.....	47
Gambar 25 Pohon Keputusan 3.....	47
Gambar 26 Prediksi Kerusakan Jalan.....	48
Gambar 27 Proporsi <i>sub kategori kerusakan</i> dengan label Rutin Biasa	49
Gambar 28 Proporsi <i>sub kategori kerusakan</i> dengan label Rutin Kondisi.....	50
Gambar 29 Proporsi <i>sub kategori kerusakan</i> dengan label Holding	51
Gambar 30 Tampilan halaman Homepage WebGIS	52
Gambar 31 Tampilan ketika klik load more pada halaman.....	52
Gambar 32 Tampilan halaman data.....	53
Gambar 33 Tampilan halaman klasifikasi	53
Gambar 34 Tampilan peta di halaman klasifikasi	53
Gambar 35 Tampilan ketika salah satu prediksi diklik pada peta	54
Gambar 36 Tampilan daftar prediksi di halaman klasifikasi.....	54
Gambar 37 Tampilan peta detail hasil klasifikasi	54
Gambar 38 Tampilan peta ketika salah satu prediksi di daftar	55
Gambar 39 Tampilan <i>button</i> diatas Daftar Alamat	55
Gambar 40 Tampilan halaman tambah data phpMyAdmin	55
Gambar 41 Tampilan halaman hapus data phpMyAdmin.....	56



DAFTAR TABEL

Tabel 1 Contoh Uji Keputusan <i>Chi-Square</i>	15
Tabel 2 Nilai <i>True Negative</i> Kelas Rutin Biasa	17
Tabel 3 Nilai <i>True Negative</i> Kelas Rutin Kondisi.....	18
Tabel 4 Nilai <i>True Negative</i> Kelas Holding	18
Tabel 5 Nilai <i>False Positive</i>	18
Tabel 6 Nilai <i>False Negative</i>	19
Tabel 7 Keputusan uji indepedensi <i>chi-square</i> data kerusakan jalan	28
Tabel 8 Rincian jumlah kerusakan jalan pada setiap kelas.....	39
Tabel 9 Distribusi Data Train dan Data Test	39
Tabel 10 Evaluasi model C4.5.....	41
Tabel 11 Evaluasi model Random Forest.....	42
Tabel 12 Perbandingan kinerja model	43



DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
BBPJN	Balai Besar Pelaksanaan Jalan Nasional
AC-WC	<i>Asphalt Concrete-Wearing Course</i>
AC-BC	<i>Asphalt Concrete-Binder Course</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
C4.5	Classification and Regression Trees, version 4.5



DAFTAR LAMPIRAN

Lampiran 1 Tampilan Dataset Kerusakan Jalan	60
Lampiran 2 Preprocessing Data.....	61
Lampiran 3 <i>Class</i> C4.5.....	68
Lampiran 4 <i>Class</i> Random Forest	75
Lampiran 5 Klasifikasi	81
Lampiran 6 Surat Penugasan Bimbingan	90
Lampiran 7 Surat Penugasan.....	91
Lampiran 8 Daftar Hadir Seminar Hasil	92
Lampiran 9 Berita Acara Seminar Hasil	93
Lampiran 10 Surat Penerbitan Panitia/Penguji Seminar Hasil.....	94
Lampiran 11 Surat Penugasan.....	95
Lampiran 12 Daftar Hadir Ujian Skripsi	96
Lampiran 13 Berita Acara Ujian Skripsi	97
Lampiran 14 Usulan Susunan Panitia/Penguji Ujian Sarjana	98
Lampiran 15 Surat Penugasan Panitia/Penguji Ujian Sarjana.....	99
Lampiran 16 Lembar Perbaikan Skripsi.....	100



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir dengan judul "Perbandingan Algoritma C4.5 dan Random Forest pada Pemetaan Kondisi Kerusakan Jalan Berbasis WebGIS (Studi Kasus : Luwu Raya)". 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 alm.Ponijan dan Ibu Semiati dan Kedua Kakak saya 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 titik yang sekarang tanpa doa dan restu kedua orangtua penulis dan keluarga, penulis mengucapkan banyak terima kasih kepada kedua orangtua penulis atas jasa dan kerja kerasnya untuk memfasilitasi penulis dalam menjalankan dunia perkuliahan.
3. Ibu Dr. Ir. Ingrid Nurtanio, M.T. selaku pembimbing I dan Bapak Dr. Ir. Zahir Zainuddin, M.Sc. selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir.



lk Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng. dan
lk Dr. Eng. Zulkifli Tahir, S.T., M.Sc. selaku penguji, yang mengarahkan
lis 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. Bapak Joko Pribatin selaku Pegawai Dinas PU dan Tata Ruang, Bapak Said selaku Kepala PPK 2.5 dan Kak Yatno, Kak Fadli, Kak Ari selaku Penilik Jalan Balai Besar Pelaksanaan Jalan Nasional Sulawesi Selatan yang membantu penulis dalam memberikan data penelitian dan memberi masukan mengenai penelitian penulis.
7. Andi Rusmiati, Dea Wahsa dan Pahrul menjadi tempat bertanya penulis saat penulis mengalami kendala dan masalah terkait tugas akhir.
8. Reskita, Arif, Yusuf dan teman-teman kelas C penulis yang selalu menghibur penulis dan membuat kenangan-kenangan yang menyenangkan selama berkuliah.
9. Besse, Dila, Deby, dan semua teman-teman Teknik Informatika Angkatan 2019 (S19NIFIER) selaku rekan yang telah memberi semangat selama penulis mengerjakan skripsi.
10. Yudhistira Murti yang menjadi menyemangat penulis dalam mengerjakan skripsi dan memberi dorongan penulis untuk selalu mengerjakan skripsinya.
11. Tresa dan teman-teman penulis yang mendukung penulis dan menghibur penulis juga selama berkuliah.
12. 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, 21 April 2024

Putri Ayu Nengsih



BAB I

PENDAHULUAN

1.1 Latar Belakang

Jalan raya adalah jalan utama yang menghubungkan antara suatu wilayah dengan wilayah lainnya dalam sektor perhubungan terutama untuk kesinambungan distribusi barang dan jasa. Sehingga untuk wilayah Sulawesi Selatan, jalan raya Nasional dipegang oleh Balai Besar Pelaksanaan Jalan Nasional VI Makassar yang merupakan instansi pemerintah di Indonesia yang bertanggung jawab atas penyelenggaraan dan pengelolaan jalan nasional di wilayah Makassar. Lembaga tersebut biasa dikenal dengan nama "BBPJN VI Makassar" dan berada di bawah kewenangan Kementerian Pekerjaan Umum dan Tata Ruang Indonesia. Peran utamanya adalah memastikan pembangunan, pemeliharaan, dan rehabilitasi jalan raya nasional yang tepat di dalam yurisdiksinya. BBPJN VI Makassar mengemban berbagai tanggung jawab seperti perencanaan, konstruksi dan pemeliharaan, manajemen proyek, dan manajemen lalu lintas. Penggunaan jalan raya sendiri juga telah diatur dalam Undang-Undang yang disepakati. Berdasarkan UU RI No 38 Tahun 2004 tentang Jalan, disebutkan jalan adalah prasarana transportasi darat. Hal ini tidak dipungkiri dengan kondisi jalannya banyak yang rusak dikarenakan beberapa jalan utama yang sering dilalui di wilayah Luwu Raya, seperti Ruas Bts. Kota Palopo-Bts. Kab. Luwu, Ruas Bts. Kab. Luwu Utara-Wotu- Tarengge-Malili-Bts. Prov. Sultra dan Ruas Tarengge-Kayulangi-Bts. Prov.Sulteng. Sebagian besar sudah diaspal dan terawat dengan baik. Namun, terdapat juga beberapa jalan alternatif yang lebih kecil dan jarang dilalui yang kondisinya kurang baik. Seperti halnya daerah-daerah lain di Indonesia, wilayah Luwu Raya juga mengalami kerusakan jalan pada beberapa bagian. Dari data yang ada jalanan rusak parah di Sulsel sepanjang 2.009 kilometer. Jumlah ini terus bertambah, di mana sebelumnya pada 2019, jalanan

sepanjang 1.500 kilometer, kemudian terus bertambah hingga tahun 2022 menjadi 2.009 KM (Kompas.com, 2023).

Beberapa faktor yang menyebabkan kerusakan jalan di wilayah ini adalah cuaca ekstrem, wilayah Luwu Raya memiliki iklim tropis dengan curah hujan yang tinggi, terutama pada musim hujan. Hujan yang terus-menerus dapat



menyebabkan kerusakan jalan, seperti tergenang air, tergerus, atau amblas, kepadatan lalu lintas, jalan-jalan utama di wilayah Luwu Raya sering dilalui oleh kendaraan berat, seperti truk dan bus. Hal ini dapat menyebabkan jalan cepat rusak dan berlubang, kurangnya perawatan. Beberapa jalan di wilayah Luwu Raya mungkin kurang mendapatkan perawatan secara teratur. Hal ini dapat menyebabkan kerusakan pada jalan semakin parah dan konstruksi jalan yang tidak baik. Beberapa jalan di wilayah Luwu Raya mungkin dirancang atau dibangun dengan konstruksi yang kurang baik. Ini dapat menyebabkan jalan cepat rusak dan tidak tahan lama (Infootomotif, 2021).

Pemerintah dan otoritas setempat di wilayah Luwu Raya berusaha untuk memperbaiki jalan yang rusak dan meningkatkan kondisi infrastruktur jalan. Namun, proses perbaikan jalan tidak selalu bisa dilakukan dengan cepat dan mudah, terutama jika masalah kerusakan jalan tersebut sangat kompleks. Selain itu, sumber daya yang diperlukan untuk melakukan perbaikan jalan mungkin terbatas, sehingga diperlukan pengalokasian anggaran yang cukup besar. Oleh karena itu, ada kemungkinan bahwa beberapa jalan yang rusak tidak segera diperbaiki karena prioritas perbaikan yang lebih tinggi atau karena keterbatasan sumber daya.

Namun, pemerintah dan otoritas setempat biasanya memiliki program perbaikan jalan yang terencana dan berkelanjutan untuk mengatasi kerusakan jalan. Program ini biasanya melibatkan perbaikan rutin dan pemeliharaan jalan serta peningkatan kapasitas dan kualitas jalan raya yang rusak atau berbahaya. Dalam situasi seperti ini, sebagai warga masyarakat, kita dapat memberikan masukan dan laporan kepada pihak terkait jika ada jalan yang rusak dan memerlukan perbaikan. Dengan begitu, pihak terkait dapat mempertimbangkan prioritas perbaikan yang sesuai dengan kebutuhan dan kemampuan sumber daya yang ada.

Penelitian yang dilakukan oleh Nahot Frastian, Senna Hendrian, dan V.H.Valentino pada tahun 2018 menggunakan metode pendekatan komparasi C4.5, Naïve Bayes, dan Random Forest, bertujuan melakukan lingkaran nilai akurasi untuk menentukan kelulusan mata kuliah. Hasil yang ketika menggunakan metode C4.5 sebesar 98.89%, Naïve Bayes sebesar , sedangkan nilai akurasi Random Forest sebesar 95.56%. (Frastian et al.,



Maka dari itu banyaknya kerusakan jalan raya menimbulkan pihak terkait mengalami kesulitan untuk pengambilan keputusan mengenai urutan penanganan sedangkan kasus atau masalah tersebut harus ditangani dengan bijak dan merata. Maka dari itu, penelitian ini akan mengembangkan penelitian yang telah dilakukan sebelumnya dengan melakukan klasifikasi terhadap parameter yang berpengaruh sehingga diperoleh prediksi prioritas kerusakan jalan raya. Pemanfaatan dari klasifikasi *data mining* menggunakan Algoritma C4.5 dan Random Forest, kemudian divisualisasikan ke dalam *Website Geographic Information System* (WebGIS) ini yaitu dapat menjadi solusi dalam membantu pihak terkait dalam melakukan monitoring terhadap wilayah yang mengalami kerusakan jalan raya berdasarkan kategori penelitian. Dengan adanya penelitian ini, sistem WeGIS yang dibangun diharapkan menjadi solusi untuk penentuan prioritas perbaikan jalan raya untuk mengambil keputusan ruas jalan yang harus dilakukan perbaikan terlebih dahulu.

1.2 Rumusan Masalah

- a) Bagaimana cara mengimplementasikan algoritma C4.5 dan Random Forest pada klasifikasi kerusakan jalan raya?
- b) Bagaimana cara memvisualisasikan hasil prediksi menjadi peta digital (WebGIS)?

1.3 Tujuan Penelitian

- a) Mengimplementasikan algoritma C4.5 dan Random Forest untuk klasifikasi data kerusakan jalan raya dan prediksi penentuan prioritas perbaikan pada data kerusakan jalan raya.
- b) Membangun sistem berbasis WebGIS untuk memvisualisasikan hasil prediksi penentuan prioritas perbaikan kerusakan jalan raya.

1.4 Manfaat Penelitian

- a) Untuk pemerintah, dapat memberikan informasi sebagai dasar dalam penentuan prioritas perbaikan disetiap jalan yang rusak dan tindakan untuk neliharaan kedepannya bagi pihak terkait.



- i para peneliti, penelitian ini memiliki potensi menjadi sumber referensi untuk studi-studi atau penelitian mendatang.

1.5 Ruang Lingkup

- a. Penelitian ini menggunakan studi kasus pada kerusakan jalan di Luwu Raya yang dibawah naungan Badan Pengelola Jalan Nasional (BBPJJN) Wilayah VI Makassar yang melibatkan 6100 data kerusakan pada tahun 2021- 2023
- b. Data yang digunakan yaitu PPK, provinsi, satker, nomor ruas jalan, nama ruas jalan, tahun, nomor identitas, lokasi/km, sub kategori (kategori kerusakan), rusak ruas kiri, rusak ruas kanan, waktu pengamatan, nama wilayah, nama penilik jalan, panjang (M), lebar (M), tebal/tinggi (M), volume, balai dan keterangan.
- c. Metode klasifikasi yang digunakan yaitu : C4.5 dan Random Forest dan divisualisasikan kedalam WebGIS.



BAB II TINJAUAN PUSTAKA

2.1 Balai Besar Pelaksanaan Jalan Nasional (BBPJN) VI Makassar

Jalan nasional merupakan salah satu infrastruktur transportasi darat yang kewenangan pengelolaannya adalah pemerintah. Jalan mempunyai bagian-bagian jalan yang pemanfaatannya tidak hanya untuk menghubungkan daerah satu dengan daerah lain, namun juga dapat dimanfaatkan untuk keperluan pihak-pihak lain. Pemanfaatan bagian-bagian jalan dapat dilakukan oleh pihak perorangan, perusahaan maupun pemerintah dengan mendapat izin dari pengelola jalan nasional yaitu Balai Besar Pelaksanaan Jalan Nasional/Balai Pelaksanaan Jalan Nasional (Bagian-Bagian Jalan Nasional et al., 2020).

Balai Besar Pelaksanaan Jalan Nasional (BBPJN) adalah institusi yang dibentuk oleh Kementerian Pekerjaan Umum dan Perumahan Rakyat yang membidangi infrastruktur jalan. Balai Besar Pelaksanaan Jalan Nasional berfungsi untuk melaksanakan dan mengendalikan Penggunaan dana APBN pada infrastruktur Jalan Nasional. Pada Pasal 103 Peraturan Menteri Pekerjaan Umum dan Perumahan rakyat, BBPJN bertanggung jawab kepada Direktur Jenderal Bina Marga (BBPJN, 2024).

Menurut Peraturan Menteri Pekerjaan Umum nomor 11/PRT/M/2010 tentang Tata Cara dan Persyaratan Layak Fungsi Jalan, layak fungsi jalan adalah kondisi suatu ruas jalan yang memenuhi persyaratan teknis kelaikan jalan untuk memberikan keselamatan bagi penggunaannya, dan persyaratan administratif yang memberikan kepastian hukum bagi penyelenggara jalan dan pengguna jalan, sehingga jalan tersebut dapat dioperasikan untuk umum. Pasal 102 Peraturan Pemerintah RI nomor 34 ayat 4 menyebutkan bahwa suatu ruas jalan umum dinyatakan layak fungsi secara teknis apabila memenuhi persyaratan dari aspek teknis struktur perkerasan jalan, teknis struktur bangunan pelengkap jalan, teknis geometri jalan, teknis pemanfaatan bagian-bagian jalan, teknis penyelenggaraan manajemen dan rekayasa lalu lintas, s perlengkapan jalan (Adwang, 2020).



dasarkan Pasal 30 Undang-Undang RI nomor 38 tahun 2004 tentang Jalan tkan bahwa jalan umum dioperasikan setelah ditetapkan memenuhi an layak fungsi jalan secara teknis administratif. Jalan adalah prasarana

transportasi darat yang meliputi segala bagian jalan, termasuk bangunan pelengkap dan perlengkapannya yang diperuntukkan bagi lalu lintas, yang berada pada permukaan tanah, di atas permukaan tanah, di bawah permukaan tanah dan air, serta di permukaan air, kecuali jalan kereta api, jalan lori dan jalan kabel. Jalan umum adalah jalan yang diperuntukkan bagi lalu lintas. Jalan umum menurut statusnya dikelompokkan atas : a) jalan nasional; b) jalan provinsi; c) jalan kabupaten; d) jalan kota; dan e) jalan desa (Adwang, 2020).

2.2 Kerusakan Jalan

Jalan raya umumnya terbuat dari aspal beton. Aspal beton adalah suatu lapisan pada konstruksi jalan yang terdiri dari campuran aspal keras dan agregat, dicampur dan dihampar dalam keadaan panas serta dipadatkan dengan suhu tertentu (Sukirman, 1992).

Lapisan yang terdiri dari campuran aspal keras (AC) dan agregat yang mempunyai gradasi menerus, dicampur dan dipadatkan pada suhu tertentu. Lapis ini digunakan sebagai lapis permukaan struktural dan lapis pondasi (*Asphalt Concrete Base/Asphalt Treated Base*) (Andi Tenrisukki, 2002).

Sesuai fungsi Laston mempunyai macam campuran yaitu :

- Lapisan *Sub-Grade*, lapisan ini terletak di atas permukaan tanah dasar atau *sub-grade* yang telah dipadatkan sesuai dengan persyaratan.
- Lapisan pondasi bawah adalah bagian perkerasan yang terletak antara lapisan pondasi dengan tanah dasar (Teori, n.d.).
- Lapisan sebagai lapisan aus, dikenal dengan nama AC-WC (*Asphalt Concrete-Wearing Course*) dengan tebal minimum 4 cm.
- Lapisan sebagai lapisan antara dikenal dengan nama AC-BC (*Asphalt Concrete-Binder Course*) dengan tebal minimum 5 cm.

Lapisan aspal beton (Laston) secara umum digunakan di berbagai negara adalah direncanakan untuk mendapatkan kepadatan yang tinggi, nilai struktural yang tinggi, kadar aspal yang rendah. Hal ini biasanya menjadikan suatu bahan relatif kaku hingga konsekuensi ketahanan rendah dan keawetan yang terjadi juga rendah (Sugeng dkk, 2021).



Kerusakan pada perkerasan jalan dapat dikelompokkan dalam dua jenis:

- Yang pertama, kerusakan struktural ditandai dengan adanya kerusakan pada satu atau lebih bagian dari struktur perkerasan jalan yang disebabkan lapisan tanah dasar tidak stabil, beban lalu lintas, keausan permukaan, dan pengaruh kondisi lingkungan sekitar.
- Yang kedua, kerusakan fungsional adalah kerusakan pada permukaan jalan yang dapat menyebabkan terganggunya fungsi jalan tersebut. Pada kerusakan fungsional, perkerasan jalan masih mampu menahan beban yang bekerja namun tidak memberikan tingkat kenyamanan dan keamanan yang diinginkan. Untuk lapisan permukaan perkerasan harus dirawat agar permukaan kembali baik.

Pada wilayah Luwu Raya, pihak yang bertugas mengamati dan melakukan pelaporan di masing-masing ruas yakni Penilik Jalan. Penilik Jalan adalah tenaga pelaksana pengamatan pemanfaatan kondisi jalan dan pemeliharaan jalan, termasuk di dalamnya laporan pengamatan, dan usulan penanganan gangguan jalan. Yang dimana tugas Penilik Jalan yaitu :

- Pengamatan atas pemanfaatan dan kondisi bagian-bagian jalan, dilakukan setiap hari.
- Pelaporan atas hasil pengamatan, disampaikan secara tertulis kepada penyelenggara jalan atau instansi yang berwenang paling lambat tanggal 10 (sepuluh) setiap bulan.
- Pengusulan tindakan yang diperlukan terhadap hasil pengamatan.
- Menerima keluhan/masukan/informasi dari masyarakat/pengguna jalan.
- Dari hasil pelaporan pengamatan, selanjutnya akan diberikan kepada PPK (Pejabat Pembuat Komitmen) untuk mengambil keputusan dan/ atau melakukan tindakan yang dapat mengakibatkan pengeluaran anggaran belanja negara/anggaran belanja daerah. (Pekerjaan et al., 2021).

Menurut Peraturan Direktorat Jenderal Bina Marga No.03/MN/B/1983 tentang Kual Pemeliharaan Jalan dan Jenis Kerusakan Jalan, menyatakan bahwa kerusakan jalan dapat diklasifikasikan dalam beberapa jenis, yaitu: retak (*cracking*), deformasi, distorsi, cacat permukaan, kerusakan di pinggir perkerasan, kerusakan tekstur permukaan jalan. Selain itu, kerusakan pada konstruksi



perkerasan jalan dapat disebabkan oleh beberapa faktor, yaitu faktor lalu lintas, faktor air yang berasal dari air hujan, sistem drainase jalan yang tidak baik serta naiknya air akibat sifat kapilaritas, faktor material konstruksi perkerasan, faktor ini dapat disebabkan oleh sifat material itu sendiri atau karena sistem pengolahan yang tidak baik, faktor iklim, faktor kondisi tanah dasar yang tidak stabil, biasanya disebabkan oleh sistem pelaksanaan yang kurang baik ataupun karena sifat tanah dasarnya tidak bagus, faktor proses pemadatan yang kurang baik. Kerusakan jalan yang terjadi di beberapa ruas jalan dapat menimbulkan kerugian terutama bagi pengguna jalan seperti ketidak-kenyamanan, waktu tempuh yang lama, kemacetan, kecelakaan, dan lain-lain.

2.3 Data Mining

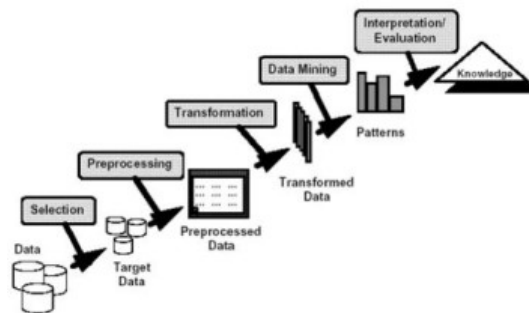
Data Mining (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Oleh karena itu data mining sebenarnya memiliki akar yang panjang dari ilmu bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *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 disklasifikasikan ke dalam dua kategori yaitu deskriptif dan prediktif. Tugas menambang secara deskriptif adalah untuk mengklasifikasikan sifat umum suatu data dalam *database*. Tugas *data mining* secara prediktif adalah mengambil an terhadap data terakhir untuk membuat prediksi (Arif, 2019).

ata mining merupakan bagian yang terintegrasi dari *Knowledge Discovery* *ises* (KDD) yang bertujuan untuk proses transformasi data mentah menjadi berguna. Jika di gambarkan secara detail tahapan KDD menjadi 5



tahap. Tampilan proses dan tahap *data mining* ditunjukkan pada Gambar 1 berikut.



Gambar 1 Tahap dan Proses *data mining*

Tahapan-tahapan pada Gambar 1 diuraikan sebagai berikut.

a) Seleksi

Tahapan pertama dalam proses data mining adalah memilih data yang relevan dan sesuai dengan tujuan analisis. Ini melibatkan identifikasi sumber data yang diperlukan dan mengumpulkan data yang diperlukan untuk analisis.

b) *Preprocessing*

Data yang dikumpulkan seringkali tidak sempurna dan mungkin mengandung noise, outlier, atau kekosongan. Tahapan ini mencakup aktivitas pembersihan dan pemrosesan data untuk menghilangkan anomali, mengisi nilai yang hilang, atau menangani noise yang mungkin memengaruhi hasil analisis.

c) Transformasi

Data seringkali perlu ditransformasikan ke dalam format atau representasi yang lebih sesuai untuk analisis data mining. Ini bisa melibatkan konversi data ke dalam bentuk yang lebih standar, normalisasi data, reduksi dimensi, atau pengkodean kategori.

d) *Data Mining*

Tahapan utama di mana teknik-teknik data mining diterapkan untuk mengekstraksi pola, tren, atau pengetahuan yang tersembunyi dari data yang telah diproses sebelumnya. Berbagai algoritma dan teknik, seperti *clustering*, klasifikasi, asosiasi, dan regresi, dapat digunakan di sini, tergantung pada tujuan analisis.



Setelah proses *data mining* selesai, langkah selanjutnya adalah menginterpretasikan hasil yang ditemukan. Ini melibatkan pemahaman terhadap pola atau hubungan yang teridentifikasi oleh model *data mining* dan menafsirkan arti atau implikasi dari hasil tersebut. Tahap akhir dari proses *data mining* adalah mengevaluasi kualitas dan keefektifan model yang telah dibuat. Evaluasi dilakukan dengan menggunakan metrik dan teknik evaluasi yang relevan untuk memastikan bahwa model yang dihasilkan sudah baik.

2.4 Metode Comparing

Comparing methods adalah teknik untuk mengetahui dan atau menguji perbedaan dua kelompok atau lebih. Metode komparasi juga untuk membandingkan suatu variabel (objek penelitian), antara subjek yang berbeda atau waktu yang berbeda dan menemukan hubungan sebab-akibatnya. Metode komparasi adalah suatu metode yang digunakan untuk membandingkan data-data yang ditarik ke dalam konklusi baru. Komparasi sendiri dari bahasa inggris, yaitu *compare*, yang artinya membandingkan untuk menemukan persamaan dari kedua konsep atau lebih (R Meikalyan.2017).

2.5 Decision Tree

Decision Tree merupakan metode penelitian yang paling sering digunakan untuk masalah klasifikasi. decision tree adalah sebuah struktur yang bisa digunakan untuk membagi kumpulan data besar menjadi himpunan record yang lebih kecil melalui serangkaian aturan keputusan. Setiap simpul pada daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari simpul internal dan akar yang terdiri dari kondisi tes atribut pada sebagian record yang memiliki karakteristik yang berbeda. Simpul internal dan akar ditandai dengan bentuk oval dan simpul daun ditandai dengan segi empat (Supriyadi et al., 2020).

Decision tree adalah model dasar yang digunakan dalam algoritma C4.5 dan juga digunakan dalam pembangunan setiap pohon dalam Random Forest. C4.5 adalah spesifik untuk membangun pohon keputusan, sementara Random Forest metode *ensemble learning* yang menggunakan pohon keputusan sebagai dasarnya.



2.5.1 Classification and Regression Trees, version 4.5 (C4.5)

Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decision tree*) adalah algoritma C4.5. Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul *The Top Ten Algorithms in Data Mining*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang di ciptakan oleh J. Rose Quinlan (Yuli Mardi.2016).

Algoritma Pohon Keputusan C4.5 atau Classification version 4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut, algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3 (Novianti et al., 2016).

Secara umum, proses algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- a) Pilih atribut sebagai akar
- b) Buat cabang untuk tiap-tiap nilai
- c) Bagi kasus dalam cabang
- d) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Secara khusus, algoritma C4.5 Decision Tree menggunakan kriteria split yang telah dimodifikasi yang dinamakan Gain Ratio dalam proses pemilihan split atribut. Split atribut merupakan proses utama dalam pembentukan pohon keputusan (Decision Tree) di C4.5 (Novianti et al., 2016). Tahapan dari algoritma C4.5 adalah sebagai berikut.

- 1) Menghitung nilai *Entropy*,
- 2) Menghitung nilai *Gain Ratio* untuk masing-masing atribut,
- 3) Atribut yang memiliki Gain Ratio tertinggi dipilih menjadi akar (root) dan atribut yang memiliki nilai Gain Ratio lebih rendah dari akar (root) dipilih menjadi cabang (branches),
- 4) Menghitung lagi nilai Gain Ratio tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (root) di tahap sebelumnya,
- 5) Atribut yang memiliki Gain Ratio tertinggi dipilih menjadi cabang (branches),



- 6) Mengulangi langkah ke-4 dan ke-5 sampai dengan dihasilkan nilai Gain = 0 untuk semua atribut yang tersisa.

Untuk menghitung nilai Entropy dapat dihitung dengan persamaan

$$\text{Entropy}(S) = \sum_{i=1}^n - \rho_i * \log_2 \rho_i \quad (1)$$

dimana:

S = himpunan kasus
A = fitur
n = jumlah partisi S
 ρ_i = proporsi dari S_1 terhadap S

Sementara itu nilai *information gain* (Gain) dapat dihitung menggunakan persamaan:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

dimana :

S = himpunan kasus
A = atribut
n = jumlah partisi atribut A
 $|S_i|$ = jumlah kasus pada partisi ke-i
 $|S|$ = jumlah kasus dalam S

Selanjutnya nilai Split Info dapat dihitung dengan persamaan :

$$\text{SplitInfo}(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

dimana :

S = himpunan kasus
A = atribut
 S_i = jumlah sampel untuk atribut i

Maka nilai Gain Ratio yang menentukan sebuah atribut dapat dijadikan akar maupun cabang suatu pohon keputusan dapat dihitung dengan persamaan :

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)} \quad (4)$$

dimana :

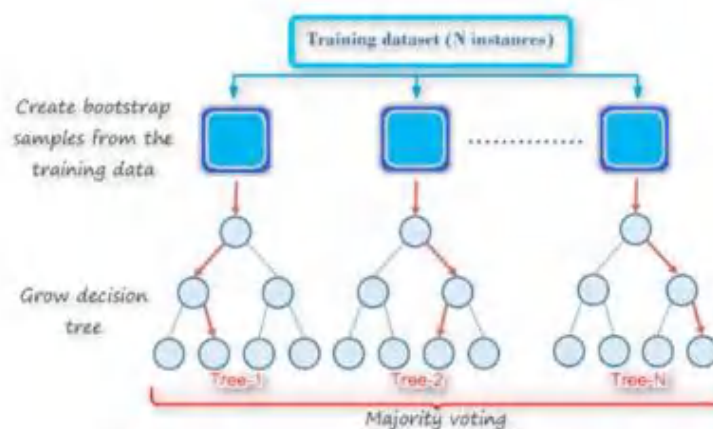
S = himpunan kasus
A = atribut
Gain (S,A) = info gain pada atribut A
SplitInfo (S,A) = split info pada atribut A



2.5.2 Random Forest

Random Forest (RF) adalah algoritma yang menggunakan metode pemisahan biner rekursif untuk mencapai node akhir dalam struktur pohon berdasarkan pohon klasifikasi dan regresi. Algoritma Random Forest menunjukkan beberapa kelebihan diantaranya mampu menghasilkan error yang relatif rendah, performa yang baik dalam klasifikasi, dapat mengatasi data pelatihan dalam jumlah besar secara efisien, serta metode yang efektif untuk mengestimasi missing data. Random Forest menghasilkan banyak pohon independen dengan subset yang dipilih secara acak melalui *bootstrap* dari sampel pelatihan dan dari variable input disetiap node (Pamuji & Ramadhan, 2021).

Random Forest adalah salah satu jenis ensemble learning yang menggabungkan banyak pohon keputusan (Decision Tree) untuk memperbaiki akurasi prediksi. Konsep Random Forest sendiri terinspirasi oleh algoritma Breiman sebelumnya, yaitu bagging, yang juga merupakan teknik ensemble learning yang menggabungkan banyak model untuk meningkatkan akurasi. Random Forest adalah algoritma klasifikasi yang terdiri dari beberapa pohon keputusan atau Decision Tree yang dibangun dengan menggunakan vektor acak (Pahrul, 2023). Random Forest adalah algoritma yang termasuk ke dalam teknik supervised learning yang pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 dengan menggabungkan teknik *bootstrap aggregating* dengan resampling. Ada beberapa kelebihan dari metode Random Forest yaitu hasil akurasi yang bagus. Dan juga mampu mengatasi *missing value* dan *noise* yang ada pada data, serta algoritma ini cocok untuk mengklasifikasikan data dalam jumlah yang besar (Pahrul, 2023).



Gambar 2 Cara Kerja Algoritma Random Forest



Gambar 2 adalah ilustrasi bagaimana Random Forest bisa menghasilkan sebuah hasil prediksi dengan menggunakan *majority voting*. Dari gambar tersebut dapat dilihat bahwa algoritma Random Forest terdiri dari kombinasi beberapa Decision Tree dimana setiap tree bergantung pada nilai random vector yang dijadikan sampel secara merata pada semua tree yang ada dalam forest tersebut. Metode Random Forest memiliki 2 tahapan. Tahap pertama adalah pembentukan forest dan tahap kedua adalah voting hasil klasifikasi (Pahrul, 2023).

$$forest = \{h(x, \theta_k), k = 1, \dots\}$$

dimana :

h = Hipotesa atau klasifikasi

x = Input vector

θ_k = Independent and identically distributed random vectors

Persamaan pertama menyatakan bahwa setiap forest terbentuk dari sekumpulan klasifikasi atau hipotesa yang berjumlah k . Input tiap hipotesa adalah x yang kemudian dilakukan resampling dengan random vector dari x itu sendiri (Pahrul, 2023).

$$Crf = majority\ vote\ \{C_n(x)\} = 1$$

dimana :

Crf = Class hasil klasifikasi Random Forest

x = Input vector

C_n = Class prediksi dari tree ke-n pada Random Forest

Setelah dilakukan pembuatan forest, maka langkah selanjutnya adalah melakukan voting untuk klasifikasi dan mengukur performansi dari Random Forest dengan menggunakan persamaan kedua (Pahrul, 2023).

2.6 Distribusi Chi-Square

Uji *Chi-Square* merupakan uji statistik non-parametrik yang paling banyak digunakan dalam penelitian bidang kesehatan masyarakat, karena uji ini memiliki kemampuan membandingkan dua kelompok atau lebih pada data-data yang telah disajikan. Meski demikian, uji *chi-square* dapat pula dipakai pada pengujian kelompok dan berskala interval/rasio (Heryana, 2020).



distribusi *Chi-square* (dibaca “khai square” atau khai kuadrat dengan simbol χ^2) adalah distribusi probabilitas teoritis yang asimetrik dan kontinyu. 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 (Heryana, 2020). Nilai statistik χ^2 dihitung dengan rumus sebagai berikut:

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

dimana,

f_o = banyaknya frekuensi yang diobservasi,

f_e = banyaknya frekuensi yang diharapkan.

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-value* < α , maka uji hipotesis menunjukkan menolak hipotesis nol, dan apabila *p-value* > α , 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 indepedensi *chi-square* ditunjukkan pada Tabel 1.

Tabel 1 Contoh Uji Keputusan *Chi-Square*

No	Fitur	Pearson Chi-Square	Signifikasi
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%.

2.7 Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengukur atau melakukan perhitungan akurasi pada konsep data mining. *Confusion matrix* terdiri dari baris data uji yang diprediksi benar dan tidak benar suatu data oleh model klasifikasi (Kiding, 2019).

Berikut tabel dibawah ini sebuah confusion matrix untuk pengklasifikasikan ke dalam tiga kelas.

Tabel 2 *Confusion Matrix Multiclass*

		Predicted		
Actual		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa	TR_{BRB}	FR_{KR_B}	FHR_{B_B}
	Rutin Kondisi	FR_{BR_K}	TR_{KR_K}	FHR_{K_K}
	Holding	FR_{BH}	FR_{KH}	THH

Tabel 2 menunjukkan *confusion matrix* yang terdiri dari 9 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix* (Pahrul.2023).

- TR_{BRB} : Variabel ini mewakili jumlah prediksi dimana kategori Rutin Biasa diklasifikasikan dengan benar. Ini juga merupakan *True Positive* untuk kelas Rutin Biasa

FR_{KR_B} : Variabel ini mewakili jumlah prediksi dimana kategori Rutin Biasa salah diklasifikasikan sebagai Rutin Kondisi



- FHR_B : Variabel ini mewakili jumlah prediksi dimana kategori Rutin Biasa salah diklasifikasikan sebagai Holding
- FR_{BR_K} : Variabel ini mewakili jumlah prediksi dimana Rutin Kondisi salah diklasifikasikan sebagai Rutin Biasa
- TR_{KR_K} : Variabel ini mewakili jumlah prediksi dimana Rutin Kondisi diklasifikasikan dengan benar. Ini juga merupakan *True Positive* untuk kelas Rutin Kondisi
- FHR_K : Variabel ini mewakili jumlah prediksi dimana Rutin Kondisi salah diklasifikasikan sebagai Holding
- FR_{BH} : Variabel ini mewakili jumlah prediksi dimana Holding salah diklasifikasikan sebagai Rutin Biasa
- FR_{KH} : Variabel ini mewakili jumlah prediksi dimana Holding salah diklasifikasikan sebagai Rutin Kondisi
- THH : Variabel ini mewakili jumlah prediksi dimana Holding diklasifikasikan dengan benar. Ini juga merupakan *True Positive* untuk kelas Holding

Pada Confusion Matrix, ada 4 istilah yang digunakan untuk merepresentasikan hasil klasifikasi. Keempat istilah tersebut adalah:

1) *True positive* (TP) merupakan data positif yang diprediksi benar.

- $TP_{\text{rutinbiasa}} = TR_B R_B$
- $TP_{\text{rutinkondisi}} = TR_K R_K$
- $TP_{\text{holding}} = THH$

2) *True negative* (TN) merupakan data negatif yang diprediksi benar.

Tabel 2 Nilai *True Negative* Kelas Rutin Biasa

		Predicted		
Actual		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa			
	Rutin Kondisi		$TR_K R_K$	FHR_K
	Holding		$FR_K H$	THH



Melalui tabel dapat kita lihat bahwa, *True Negative* untuk kelas rutin biasa adalah:

$$TN_{\text{rutinbiasa}} = TR_K R_K + FHR_K + FR_K H + THH$$

Tabel 3 Nilai *True Negative* Kelas Rutin Kondisi

		Predicted		
Actual		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa	$TR_B R_B$		FHR_B
	Rutin Kondisi			
	Holding	$FR_B H$		THH

Melalui tabel dapat kita lihat bahwa, True Negative untuk kelas rutin kondisi adalah:

$$TN_{\text{rutinkondisi}} = TR_B R_B + FHR_B + FR_B H + THH$$

Tabel 4 Nilai *True Negative* Kelas Holding

		Predicted		
Actual		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa	$TR_B R_B$	$FR_K R_B$	
	Rutin Kondisi	$FR_B R_K$	$TR_K R_K$	
	Holding			

Melalui tabel dapat kita lihat bahwa, True Negative untuk kelas holding adalah:

$$TN_{\text{holding}} = TR_B R_B + FR_K R_B + FR_B R_K + TR_K R_K$$

- 3) *False positive* (FP) atau type 1 error merupakan data negative namun diprediksi sebagai data positif.

Tabel 5 Nilai *False Positive*

		Predicted		
1		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa		$FR_K R_B$	FHR_B
	Rutin Kondisi	$FR_B R_K$		FHR_K
	Holding	$FR_B H$	$FR_K H$	



False Positive pada kelas rutin biasa ditandai dengan warna biru, pada kelas rutin kondisi ditandai dengan warna pink dan pada kelas holding ditandai dengan warna coklat.

- $F_{Prutinbiasa} = FR_K R_B + FHR_B$
- $F_{Prutinkondisi} = FR_B R_K + FR_B H$
- $F_{Pholding} = FHR_B + FHR_K$

4) *False Negative* (FN) atau type 2 error merupakan data positif namun diprediksi sebagai data negatif.

Tabel 6 Nilai *False Negative*

		Predicted		
Actual		Rutin Biasa	Rutin Kondisi	Holding
	Rutin Biasa		$FR_K R_B$	FHR_B
	Rutin Kondisi	$FR_B R_K$		FHR_K
	Holding	$FR_B H$	$FR_K H$	

False Negative pada kelas rutin biasa ditandai dengan warna biru, pada kelas rutin kondisi ditandai dengan warna pink dan pada kelas holding ditandai dengan warna coklat.

- $F_{Prutinbiasa} = FR_K R_B + FHR_B$
- $F_{Prutinkondisi} = FR_B R_K + FHR_K$
- $F_{Pholding} = FR_B H + FR_K H$

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 (Nugroho.2019).

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 (sebenarnya). 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).

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

4. 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).

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



Geographic Information System (WebGIS)

WebGIS adalah aplikasi GIS atau pemetaan digital yang memanfaatkan internet sebagai media komunikasi yang berfungsi mendistribusikan, publikasikan, mengintegrasikan, mengkomunikasikan dan menyediakan

informasi dalam bentuk teks, peta digital serta menjalankan fungsi–fungsi analisis dan *query* yang terkait dengan GIS melalui jaringan internet (Prahasta (2007)).

Sedangkan menurut Setiawan (1998), penggunaan data spasial dirasakan semakin diperlukan untuk berbagai keperluan seperti penelitian, pengembangan dan perencanaan wilayah, serta manajemen sumber daya alam. Pengguna data spasial merasakan minimnya informasi mengenai keberadaan dan ketersediaan data spasial yang dibutuhkan. Penyebaran (diseminasi) data spasial yang selama ini dilakukan dengan menggunakan media yang telah ada yang meliputi media cetak (peta), cd-rom, dan media penyimpanan lainnya dirasakan kurang mencukupi kebutuhan pengguna. Pengguna diharuskan datang dan melihat langsung data tersebut pada tempatnya (data provider). Hal ini mengurangi mobilitas dan kecepatan dalam memperoleh informasi mengenai data tersebut. Karena itu dirasakan perlu adanya WebGIS.

Arsitektur aplikasi pemetaan di web dibagi menjadi dua pendekatan sebagai berikut :

- **Pendekatan Thin Server**

Pendekatan ini memfokuskan diri pada sisi server. Hampir semua proses dan analisis data dilakukan berdasarkan request di sisi server. Data hasil pemrosesan kemudian dikirim ke client dalam format standar.

- **Pendekatan Thick Client**

Pada pendekatan ini, pemrosesan data dilakukan di sisi client menggunakan beberapa teknologi (Nuryadin, 2005). Secara umum pengembangan dan implementasi WebGIS akan menunjang penyebaran informasi data spasial. Sehingga orang awam pun akan dapat memiliki akses terhadap data dan hasil analisis GIS.

