PREDIKSI KERUGIAN PADA KECELAKAAN KENDARAAN MENGGUNAKAN METODE *BOOTSTRAP AGGREGATION*

(Studi Kasus : Kecelakaan Pembalap MotoGP 2022)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program Studi Statistika



Disusun Oleh:

Muhammad Rafy Ryana

20611038

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Prediksi Kerugian Pada Kecelakaan Kendaraan Menggunakan Metode *Bootstrap Aggregation* (Studi kasus: Kecelakaan Pembalap MotoGP 2022)
Nama Mahasiswa	:	Muhammad Rafy Ryana
NIM	:	20611038

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Mengetahui	Yogyakarta, 14 Mei 2024
Ketua Prodi Statistika	Dosen Pembimbing
Dr. Atina Ahdika, S.Si., M.Si.	Dr. Atina Ahdika, S.Si., M.Si.

Dipindai dengan CamScanner

HALAMAN PENGESAHAN
TUGAS AKHIR

PREDIKSI KERUGIAN PADA KECELAKAAN KENDARAAN MENGGUNAKAN METODE *BOOTSTRAP AGGREGATION*

(Studi Kasus : Kecelakaan Pembalap MotoGP 2022)

Nama Mahasiswa : Muhammad Rafy Ryana

NIM : 20611038

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL : 22 MEI 2024

Nama Penguji	**Tanda Tangan**
1. Abdullah Ahmad Dzikrullah, S.Si., M.Sc.	.....
2. Rahmadi Yotenka, S.Si., M.Sc.	.....
3. Dr. Atina Ahdika, S.Si., M.Si.	.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Riyanto, S.Pd., M.Si., Ph.D.

Dipindai dengan CamScanner

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Assalamu'alaikum Wr.Wb

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkat, rahmat, dan hidayah-Nya berupa keimanan, kekuatan, kesabaran, kelancaran, dan kesehatan sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "Prediksi Kerugian Pada Kecelakaan Kendaraan Menggunakan Metode Bootstrap Aggregation" ini dengan baik dan tepat waktu. Shalawat serta salam penulis panjatkan kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan umatnya.

Tugas Akhir ini dibuat dengan tujuan untuk memperoleh gelar sarjana Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam di Universitas Islam Indonesia.

Dalam menyusun Proposal Tugas Akhir ini, tak lepas dari rintangan serta halangan yang penulis hadapi. Akan tetapi hal tersebut dapat terselesaikan karena adanya bantuan dari pihak-pihak yang mendukung serta membimbing penulis baik secara moral maupun material. Oleh karena itu, penulis ingin mengucapkan rasa hormat dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Ilmu Pengetahuan Alam, Universitas Islam Indonesia beserta seluruh jajarannya.
2. Bapak Dr. Edy Widodo, S.Si., M.Si. selaku Ketua Jurusan Statistika beserta seluruh jajarannya
3. Ibu Dr. Atina Ahdika, S.Si., M.Si selaku Ketua Program Studi Statistika beserta seluruh jajarannya dan selaku Dosen Pembimbing Tugas Akhir dengan kesabarannya telah membekali dan membimbing penulis dari awal hingga selesainya Proposal Tugas Akhir ini
4. Ayah, Ibu, dan Dede yang telah memberikan dukungan, doa, kasih sayang, serta motivasi kepada peneliti
5. Seluruh pihak yang sudah berpartisipasi dan memberi dukungan moral maupun materi yang tidak dapat penulis sebutkan satu persatu.

Semoga segala dukungan yang diberikan kepada penulis mendapatkan imbalan yang lebih besar dari Allah SWT. Penulis menyadari bahwa Proposal

Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis menerima segala kirtik dan saran yang sifatnya membangun untuk perbaikan kedepannya. Semoga Tugas Akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan.

Wassalamu'alaikum Wr.Wb

Yogyakarta, 14 Mei 2024

Muhammad Rafy Ryana

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR ii
HALAMAN PENGESAHAN TUGAS AKHIR iii
KATA PENGANTAR iv
DAFTAR ISI vi
DAFTAR TABEL vii
DAFTAR GAMBAR viii
DAFTAR LAMPIRAN ix
PERNYATAAN x
INTISARI xi
ABSTRACT xii
BAB I PENDAHULUAN 1
1.1. Latar Belakang Masalah 1
1.2. Rumusan Masalah 3
1.3. Batasan Masalah 4
1.4. Tujuan Penelitian 4
1.5. Manfaat Penelitian 4
BAB II TINJAUAN PUSTAKA 6
BAB III LANDASAN TEORI 12
3.1. MotoGP 12
3.2. Pengertian Asuransi 13
3.3. Asuransi Kendaraan 14
3.4. *Machine Learning* 15
3.5. *Bagging* 16
3.6. *Variable Importance* 17
3.7. *Partial Dependence Plots* 18
3.8. *Ordinary Least Square* (OLS) 19
BAB IV METODOLOGI PENELITIAN 21
4.1. Populasi dan Sampel Penelitian 21
4.2. Tempat dan Waktu Penelitian 21
4.3. Definisi Operasional Variabel Penelitian 21
4.4. Alat dan Cara Organisir Data 22
4.5. Metode Penelitian 23
BAB V PEMBAHASAN 24
5.1. Statistika Deskriptif 24
5.2. Ordinary Least Square 29
5.3. *Bootstrap Aggregation* 32
5.4. *Variable Importance* 33
5.5. *Partial Dependence Plots* 34
5.6. Prediksi Jumlah Kecelakaan 36
5.7. Perkiraan Kerugian 39
BAB VI PENUTUP 42
6.1. Kesimpulan 42
6.2. Saran 42
DAFTAR PUSTAKA 43
LAMPIRAN 47

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya 9
Tabel 4.1 Penjelasan Variabel Penelitian 21
Tabel 5.1 Statistika Deskriptif Variabel Penelitian 24
Tabel 5.2 Hasil Analisis OLS 30
Tabel 5.3 Statistik Uji dan Keputusan Uji t 31
Tabel 5.4 Nilai RMSE untuk beberapa model 33
Tabel 5.5 Nilai *Variable Importance* 34
Tabel 5.6 Statistika Deskriptif Prediksi Jumlah Kecelakaan 37
Tabel 5.7 Statistika Deskriptif Prediksi Jumlah Kecelakaan Berdasarkan Tim ... 38
Tabel 5.8 Statistika Deskriptif Perkiraan Kerugian dalam Juta Dolar Amerika .. 40
Tabel 5.9 Statistika Deskriptif Perkiraan Kerugian Berdasarkan Tim dalam Juta Dolar Amerika 40

DAFTAR GAMBAR

Gambar 4.1 Tahapan Penelitian 23
Gambar 5.1 Grafik $X1$ atau Usia 25
Gambar 5.2 Grafik $X2$ atau Asal 25
Gambar 5.3 Grafik $X3$ atau Merek 26
Gambar 5.4 Grafik $X4$ atau Harga 27
Gambar 5.5 Grafik $X5$ atau Usia Kendaraan 27
Gambar 5.6 Grafik $X6$ atau Tim 28
Gambar 5.7 Grafik Y atau Jumlah Kecelakaan 29
Gambar 5.8 *Variable Importance* Plot 34
Gambar 5.9 *Partial Dependence Plot* $X1$ dengan Prediksi Kecelakaan Berdasarkan $X3$ 35
Gambar 5.10 *Partial Dependence Plot* $X4$ dengan Prediksi Kecelakaan Berdasarkan $X3$ 36
Gambar 5.11 Histogram Prediksi Jumlah Kecelakaan 37
Gambar 5.12 Barplot Rata-Rata Prediksi Jumlah Kecelakaan Berdasarkan Tim 38
Gambar 5.13 Barplot Rata-Rata Perkiraan Kerugian Berdasarkan Tim 41

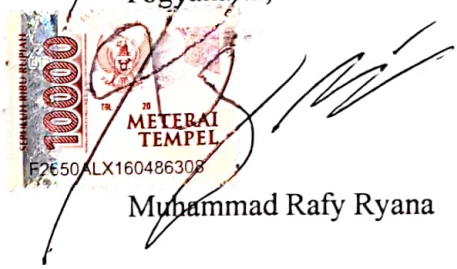
DAFTAR LAMPIRAN

Lampiran 1 Sintaks R 47

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya-karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 14 Mei 2024



Muhammad Rafy Ryana

Dipindai dengan CamScanner

INTISARI

PREDIKSI KERUGIAN PADA KECELAKAAN KENDARAAN MENGGUNAKAN METODE *BOOTSTRAP AGGREGATION*

(Studi Kasus : Kecelakaan Pembalap MotoGP 2022)

Muhammad Rafy Ryana

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

MotoGP adalah acara balap sepeda motor tertua yang diadakan di beberapa sirkuit utama. Pada tahun 2022, terdapat 335 kecelakaan yang melibatkan pembalap MotoGP. Angka ini menunjukkan bahwa acara balap sepeda motor kelas dunia ini membawa risiko yang signifikan, yang mengakibatkan kerugian besar bagi mereka yang terkena dampak. Oleh karena itu, sangat penting untuk memprediksi sejauh mana kerugiannya sehingga tim-tim MotoGP dapat mengantisipasi besarnya kerugian potensial sebelumnya. Dalam penelitian ini, penulis tertarik untuk memprediksi sejauh mana kerugian dalam kecelakaan balap MotoGP 2022 menggunakan metode *bootstrap aggregation*. Data diambil dari situs web resmi MotoGP, terdiri dari 6 variabel independen antara lain usia, asal, merek, harga, usia kendaraan, serta tim dan 1 variabel dependen, yaitu jumlah kecelakaan. Metode *bootstrap aggregation* digunakan untuk memprediksi jumlah kecelakaan. Dalam estimasi kerugian, kami mengasumsikan bahwa setiap kecelakaan mengakibatkan kerugian sebesar nilai sepeda motor yang terlibat. Hal ini dilakukan untuk memperkirakan skenario terburuk. Nilai kerugian berkisar antara 4.385 juta hingga 12.742 juta dolar Amerika, dengan kerugian rata-rata sebesar 8.379 juta dolar Amerika dan standar deviasi sebesar 2260.579.

Kata Kunci : *Bootstrap Aggregation*, Kecelakaan, Kerugian, MotoGP.

ABSTRACT

LOSS PREDICTION IN VEHICLE ACCIDENTS USING BOOTSTRAP AGGREGATION METHOD

(Case Study : Racer Accident MotoGP 2022)

Muhammad Rafy Ryana
Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

MotoGP is the oldest motorcycle racing event that held at several major circuits. In 2022, there were 335 accidents involving MotoGP riders. This figure indicates that this world-class motorcycle racing event carries significant risks, resulting in substantial losses for those affected. Therefore, it is highly necessary to predict the extent of the losses so that MotoGP teams can anticipate the magnitude of the potential losses beforehand. In this study, the author is interested in predicting the extent of losses in MotoGP 2022 racing accidents using the bootstrap aggregation method. The Data was taken from the official MotoGP website, consisting of 6 independent variables including age, origin, brand, price, vehicle age, and team, and 1 dependent variable, which is the number of accidents. The bootstrap aggregation method is used to predict the number of accidents. In loss estimation, we assume that each accident results in a loss equal to the value of the motorcycle involved. This is done to estimate the worst-case scenario. The loss values range from 4.385 million to 12.742 million US dollars, with an average loss of 8.379 million US dollars and a standard deviation of 2260.579.

Keywords *Accidents, Bootstrap Aggregation, Loss, MotoGP*

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

MotoGP merupakan ajang balap motor kelas dunia yang diadakan di beberapa sirkuit besar. Kompetisi ini diadakan dalam beberapa seri dan diikuti oleh pembalap-pembalap handal dunia (Pinch & Reimer, 2017). Selain itu sepeda motor yang digunakan menggunakan mesin *prototype* dengan teknologi yang mutakhir dan dibuat oleh pabrikan termasuk Ducati, Aprilia, Honda, KTM, dan Yamaha. Balapan MotoGP dimulai pada hari Jumat dan melibatkan serangkaian sesi yang dijadwalkan selama tiga hari. Selain itu, pergelaran MotoGP juga diikuti oleh beberapa kelas yaitu Moto3, Moto2, dan MotoGP. Pada tiap-tiap kelas dilakukan beberapa sesi antara lain *Free Practice*, *Qualifying*, *Warm Up*, hingga *Race*. Puncak acaranya terletak pada sesi *race* MotoGP yang dilaksanakan pada hari Minggu.

Kejuaraan balap motor dunia ini dimulai oleh FIM (Federation Internationale de Motocyclisme) pada tahun 1949, dan saat ini, MotoGP memasuki tahun ke-74nya. Ini adalah balap motor tertua di dunia, dengan tiga kelas balap utama yang diadakan selama akhir pekan. Awalnya dikenal dengan sebutan '500'cc, kejuaraan ini mengalami perubahan pada tahun 2002 dengan diperkenalkannya mesin empat-tak, serta peningkatan kapasitas mesin menjadi 990cc, sehingga dikenal sebagai MotoGP. Pada tahun 2007, peraturan diubah untuk membatasi kapasitas mesin menjadi 800cc, dan pada tahun 2012, batasan kapasitas mesin diperluas menjadi 1000cc. MotoGP telah dikelola oleh Dorna Sport, pemilik hak komersialnya, di bawah pengawasan FIM sejak tahun 1992 (Wulandari & Arsyad, 2023).

Pada tahun 2022 telah terjadi 335 kecelakaan yang melibatkan para pembalap MotoGP (MotoGP, 2022). Angka ini menggambarkan bahwa ajang balap motor kelas dunia ini memiliki risiko yang besar. Hal itu menghasilkan kerugian yang besar bagi pihak-pihak yang terdampak. Kerugian tersebut dapat berupa motor pembalap yang hanya tergores karena bergesekan dengan aspal sirkuit atau motor yang hancur karena kecelakaan pada kecepatan tinggi. Misalkan pada balapan MotoGP yang digelar di Austria pada 2020 terjadi insiden antara Franco Morbidelli yang menyenggol ban belakang Johan Zarco pada kecepatan kurang lebih 300

km/jam sehingga motor kedua pembalap tersebut terpental hingga rusak dan nyaris mengenai pembalap Valentino Rossi. Oleh karena itu, sangat diperlukan sebuah pertanggungan risiko untuk menggantikan kerugian yang terjadi pada motor yang mengalami kerusakan.

Asuransi merupakan salah satu sarana perlindungan dari kerugian finansial. Ada berbagai jenis asuransi seperti asuransi jiwa, asuransi kesehatan, asuransi kendaraan, dan lain-lain. Kontribusi terbesar di bidang asuransi berasal dari asuransi kendaraan. Tujuan utama asuransi kendaraan adalah untuk melindungi kendaraan dari kerusakan, cedera diri, dan tanggung jawab pihak ketiga (Ghose & Akanchha, 2020). Pihak tertanggung dapat memperoleh perlindungan dari kerugian finansial dengan cara mengajukan klaim kepada pihak asuransi.

Asuransi kendaraan tidak hanya untuk kendaraan yang biasa digunakan sehari-hari saja. Asuransi kendaraan juga digunakan dalam olahraga yang menggunakan kendaraan. Misalnya pada MotoGP, salah satu tim satelit dari merek Ducati yaitu Pramac-Ducati memiliki sponsor utama Prima Assicurazioni. Prima Assicurazioni adalah perusahaan asuransi terkemuka di Italia yang memiliki posisi kuat di segmen sepeda motor, dengan hampir satu dari sepuluh sepeda motor di Italia diasuransikan dengan Prima (Pradipta, 2022). Selain itu, tim Pramac-Ducati memiliki sponsor Aon yang merupakan perusahaan asuransi dan manajemen risiko yang sering bekerja dengan tim-tim olahraga besar. Salah satu produk asuransi Aon dalam olahraga yang menggunakan kendaraan adalah asuransi fasilitas olahraga dan asuransi kecelakaan (Aon, 2024).

Pada MotoGP, banyaknya kerugian yang disebabkan oleh kecelakaan, maka akan mengakibatkan kerugian yang signifikan yang dialami oleh tim-tim MotoGP. Terutama pada kerugian berupa motor yang rusak. Kerugian apabila motor rusak parah diperkirakan sangat besar karena harga motor MotoGP diperkirakan sangat mahal. Oleh karena itu, sangat penting untuk memprediksi sejauh mana kerugian sehingga tim-tim MotoGP dapat memperkirakan besarnya kerugian yang mungkin terjadi sebelumnya.

Salah satu cara untuk memperkirakan sejauh mana kerugian adalah menggunakan *machine learning*. Saat ini, banyak perusahaan asuransi yang beralih menggunakan *machine learning* untuk memprediksi ukuran klaim (Terefe, 2023).

Namun, pemilihan model prediksi *machine learning* yang beragam bukanlah hal yang mudah.

Metode *bootstrap aggregation* atau biasa disebut dengan *bagging* digunakan untuk meningkatkan estimasi atau skema klasifikasi yang tidak stabil pada sekumpulan data. *Bagging* adalah metode *ensemble* yang terdiri dari dua langkah utama. Pertama, sampel dilakukan *bootstrap* dari *training data* dan digunakan untuk mengkalibrasi kumpulan model, di mana setiap model dikalibrasi ke dalam sampel yang berbeda. Kedua *ensemble* model diterapkan pada data di luar sampel dan estimasi setiap model dikumpulkan menjadi satu *output* (Brodeur, Herman, & Steinschneider, 2020). Metode *bagging* merupakan metode yang menggunakan *boostrap* sehingga terbilang efektif dan lebih baik untuk menyelesaikan suatu masalah (Putri, Mariani, & Sunarmi, 2021). *Bootstrap* melakukan *resampling* data asli dengan pengembalian dan membuat kesimpulan dari sampel untuk mengestimasi nilai parameter. Pada penelitian oleh Edossa Merga Terefe (2023) menghasilkan evaluasi model *bagging* lebih baik dalam memprediksi ukuran klaim asuransi kendaraan dibandingkan dengan metode *ordinary least square*.

Berdasarkan pada uraian latar belakang, penulis tertarik untuk melakukan prediksi besarnya kerugian pada kecelakaan kendaraan khususnya pada motor MotoGP 2022. Hasil prediksi besarnya kerugian motor MotoGP 2022 dengan metode *bootstrap* diharapkan dapat memberikan kontribusi terhadap perkembangan dalam pemanfaatan *machine learning* untuk mengukur kerugian dan memberikan pemahaman yang lebih mendalam bagi pihak-pihak terkait agar kerugian yang dialami bila terjadi kejadian yang tidak diinginkan dapat diperhitungkan sebelumnya. Sehingga, penelitian ini memiliki kontribusi yang signifikan terhadap proses perkembangan MotoGP agar lebih baik lagi.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini antara lain:

1. Bagaimana hubungan variabel-variabel yang berkaitan dengan pergelaran MotoGP seperti usia pembalap, asal pembalap, merk motor, dan lainnya dengan besarnya kerugian?

2. Bagaimana metode *bagging* dapat digunakan untuk memprediksi besar kerugian pada kecelakaan kendaraan dalam konteks MotoGP pada tahun 2022?

1.3. Batasan Masalah

Penulis akan membatasi pada beberapa parameter antara lain:

1. Penelitian ini membatasi periode pengumpulan data dan analisis pada tahun 2022 untuk pergelaran MotoGP.
2. Penelitian ini akan fokus pada besar kerugian pada kecelakaan kendaraan dalam pergelaran MotoGP, terutama sepeda motor balap yang digunakan dalam kompetisi tersebut.
3. Penelitian ini akan mempertimbangkan beberapa variabel pendukung yang dapat memengaruhi besarnya kerugian, termasuk usia pembalap, asal pembalap, merk motor, dan faktor-faktor lain yang relevan.
4. Penelitian ini akan menggunakan metode *bagging* sebagai alat utama untuk memprediksi kerugian pada kecelakaan sepeda motor MotoGP.
5. Data yang digunakan dalam penelitian ini akan bersumber dari sumber-sumber yang dapat diakses dan relevan untuk pergelaran MotoGP dalam periode yang ditentukan.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dalam penelitian ini antara lain:

1. Mengetahui hubungan besarnya kerugian pada kecelakaan kendaraan dalam konteks MotoGP dengan usia pembalap, asal pembalap, merk motor, dan lainnya.
2. Memprediksi besar kerugian pada kecelakaan kendaraan dalam konteks MotoGP pada tahun 2022 dengan menggunakan metode *bagging*.

1.5. Manfaat Penelitian

Beberapa manfaat dari penulisan Tugas Akhir ini antara lain:

1. Meningkatkan pengetahuan dan pemahaman penulis mengenai metode *bagging*.
2. Penelitian ini dapat bermanfaat pada perkembangan ilmu *machine learning*.

3. Mengetahui besarnya kerugian pada kecelakaan kendaraan jika diimplementasikan dalam pergelaran MotoGP.
4. Pihak-pihak terkait dalam pergelaran MotoGP dapat menambah pemahaman yang lebih mendalam agar kerugian yang dialami apabila terjadi kejadian yang tidak diinginkan dapat diperhitungkan sebelumnya.
5. Dapat digunakan sebagai referensi baru untuk penelitian selanjutnya terkait dengan penggunaan metode *bagging*.

BAB II
TINJAUAN PUSTAKA

Beberapa penelitian terdahulu diperlukan oleh penulis sebagai inspirasi dan acuan dalam penulisan tugas akhir. Dengan mengacu pada penelitian terdahulu, penulis dapat mengetahui ada atau tidaknya plagiasi. Selain itu, penulis dapat mengetahui kekurangan dan kelebihan penelitian terdahulu untuk kepentingan penulisan tugas akhir. Dengan begitu, penulis dapat mengembangkan, meningkatkan, dan mengimplementasikan hasil-hasil yang diperoleh dari penelitian terdahulu. Penyusunan tugas akhir ini bertujuan untuk mengungkapkan pentingnya penulisan terhadap kemajuan dalam ranah ilmu pengetahuan..

Penelitian pertama ditulis oleh Hong-Jun Yoon, dkk (2020) yang berjudul "*Accelerated training of bootstrap aggregation-based deep information extraction systems from cancer pathology reports*". Tujuan penelitian ini untuk menunjukkan efektifitas metode *proposed partitioned bagging* dan membagi masalah menjadi sub-masalah dapat meningkatkan kinerja tugas dan mempercepat pelatihan dalam konteks ekstraksi informasi dari laporan patologi kanker. Hasil dari penelitian ini adalah model *bagging* meningkatkan kinerja klasifikasi dibandingkan dengan pendekatan model tunggal. Metode *partitioned bagging* yang diusulkan mencapai skor akurasi klasifikasi yang lebih tinggi pada empat tugas, khususnya untuk ekstraksi data histologi kanker. Model *partitioned bagging* menunjukkan kinerja unggul dalam tugas klasifikasi histologi kanker, yang mencakup lebih dari 500 label kelas.

Penelitian kedua ditulis oleh Zachary P. Brodeur, Jonathan D. Herman, dan Scott Steinschneider (2020) yang berjudul "*Bootstrap Aggregation and Cross-Validation Methhods to Reduce Overfitting in Reservoir Control Policy Search*". Tujuan penelitian ini untuk mengembangkan perluasan teknik *bootstrap aggregating* (*bagging*) dan *cross-validation* untuk meningkatkan kinerja kebijakan pengendalian waduk, khususnya yang berkaitan dengan rangkaian kejadian banjir dan kekeringan. Hasil dari penelitian ini menunjukkan bahwa strategi validasi-kalibrasi untuk pemilihan kebijakan, ditambah dengan metode *bootstrap*

aggregation, dapat meningkatkan kinerja dalam tujuan penyediaan air dan risiko banjir dibandingkan kinerja dasar.

Penelitian ketiga ditulis oleh Jeyoung Shin dan Chan-Hwan Im (2020) yang berjudul "*Performance Improvement of Near-Infrared Spectoscopy-Based-Brain-Computer Interface Using Regularized Linear Discriminant Analysis Ensemble Classifier Based on Bootstrap Aggregating*". Tujuan penelitian ini untuk mengevaluasi kemanjuran *regularized linear discriminant analysis ensemble classifiers* berdasarkan *bootstrap aggregating* untuk meningkatkan kinerja *near-infrared spectroscopy-based brain-computer interface* dengan meningkatkan akurasi klasifikasi dan kecepatan bit. Hasil dari penelitian ini adalah *regularized linear discriminant analysis* berdasarkan *bootstrap aggregating* menunjukkan peningkatan signifikan dalam *bitrate* dan akurasi klasifikasi untuk keempat kumpulan data *near-infrared spectroscopy-based brain-computer interface* yang digunakan dalam penelitian ini.

Penelitian keempat ditulis oleh Imad Abdallah, Konstantinos Tatsis, and Eleni Chatzi (2020) yang berjudul "*Unsupervised local cluster-weighted bootstrap aggregating the output from multiple stochastic simulators*". Tujuan penelitian ini untuk mengatasi masalah penggabungan *output* dari beberapa simulator komputer stokastik untuk membuat inferensi pada kuantitas yang diinginkan, dengan tujuan mengurangi ketidakpastian bentuk model yang melekat karena tidak adanya pengukuran apa pun. Hasil dari penelitian ini adalah metode yang diusulkan, *unsupervised local cluster-weighted bootstrap aggregation*, dibandingkan dengan *classical bagging*, *bayesian model averaging*, dan *stacking of predictive distributions*. Hasilnya menunjukkan efektivitas metode yang diusulkan dalam mengurangi kesalahan generalisasi dan meningkatkan ketepatan estimator untuk interval kepercayaan sebesar 95%. Kinerja metode ini dipengaruhi oleh faktor-faktor seperti ukuran wadah masukan, ukuran sampel per simulator, dan simulasi.

Penelitian kelima ditulis oleh Dwi Liana Wella Putri, Scolastika Mariani, dan Sunarmi (2021) yang berjudul "Peningkatan Ketepatan Klasifikasi Model Regresi Logistik Biner dengan Metode Bagging (Bootstrap Aggregating)". Penelitian ini bertujuan untuk mengevaluasi ketepatan dan efektivitas penggunaan regresi logistik dan regresi logistik metode bagging (bootstrap aggregating) dalam meramalkan

status peserta keluarga berencana di Kota Tegal pada tahun 2016. Hasil penelitian menunjukkan bahwa metode bagging menghasilkan tingkat keakuratan sebesar 75,641%, dengan tingkat kesalahan klasifikasi sebesar 24,359%. Penggunaan metode bagging secara signifikan meningkatkan akurasi model regresi logistik dari 69,74% menjadi 75,641%.

Penelitian keenam ditulis oleh Edossa Merga Terefe (2023) yang berjudul "*Tree-Based Machine Learning Methods For Vehicle Insurance Claims Size Prediction*". Tujuan penelitian ini untuk mengeksplorasi penerapan metode *tree-based machine learning* seperti *bagging*, *random forest*, dan *gradient boosting* dalam memprediksi ukuran klaim asuransi kendaraan dan untuk mengevaluasi serta membandingkan kinerja model-model ini dengan *ordinary least square*. Hasil penelitian ini adalah metode *tree-based ensemble*, seperti *bagging*, *random forest*, dan *gradient boosting*, mengungguli metode *ordinary least square* dalam memprediksi ukuran klaim asuransi kendaraan.

Penelitian ketujuh ditulils oleh Yulia Resti, Chandra Irsan, Jeremy Firdaus Latif, Irsyad Yani, and Novi Rustiana Dewi (2023) yang berjudul "*A Bootstrap-Aggregating in Random Forest Model for Classification of Corn Plant Diseases and Pests*". Tujuan penelitian ini untuk mengklasifikasikan dan meningkatkan kinerja metode *random forest* berdasarkan *bootstrap-agregating* mengenai penyakit dan hama tanaman jagung dengan mengeksplorasi jumlah *decision trees* yang optimal. Hasil penelitian ini adalah kinerja model *random forest* dievaluasi berdasarkan jumlah *decision trees*. Ditemukan bahwa kinerja model tertinggi dicapai dengan enam *decision trees* dalam *bootstrap aggregating* dari *random forest*. Hanya kelas DWD dan HAP yang memiliki observasi dengan klasifikasi buruk dalam model dengan tiga *decision trees*.

Penelitian kedelapan ditulis oleh Michael Lau, Sara Kress, Tamara Schikowski, and Holger Schwender (2023) yang berjudul "*Efficient gene-environment interaction testing through bootstrap aggregating*". Tujuan penelitian ini untuk mengembangkan prosedur uji statistik yang menggunakan *bootstrap aggregating* (*bagging*) untuk interaksi gen-lingkungan (GxE) yang melibatkan paparan polusi udara dan rheumatoid arthritis. Hasil penelitian ini berfokus pada pengembangan prosedur pengujian yang efisien untuk interaksi GxE, khususnya

menggunakan *bagging* dan *random forest*. Hasilnya menunjukkan bahwa metode ini memiliki kekuatan statistik yang lebih tinggi dalam mengendalikan kesalahan tipe I, menjadikannya prosedur pengujian yang valid. Penerapan prosedur ini pada kumpulan data nyata menunjukkan adanya interaksi GxE terkait dengan rheumatoid arthritis dengan paparan polusi udara.

Tabel 2.1 Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2020	Hong-Jun Yoon, Hilda B. Klasky, John P. Gounley, Mohammed Alawad, Shang Gao, Eric B. Durbin, Xiao-Cheng Wu, Antoinette Stroup, Jennifer Doherty, Linda Coyle, Lynne Penberthy, J. Blair Christian, and Georgia D. Tourassi	*Accelerated training of bootstrap aggregation-based deep information extraction system from cancer pathology reports*	Penelitian menunjukkan bahwa model *bagging* meningkatkan kinerja klasifikasi dibandingkan dengan pendekatan model tunggal. Metode *partitioned bagging* yang diusulkan mencapai skor akurasi klasifikasi yang lebih tinggi pada empat tugas, khususnya untuk ekstraksi data histologi kanker. Model *partitioned bagging* menunjukkan kinerja unggul dalam tugas klasifikasi histologi kanker, yang mencakup lebih dari 500 label kelas.
2020	Zachary P. Brodeur, Jonathan D. Herman, and Scott Steinschneider	*Bootstrap Aggregation and Cross-Validation Methhods to Reduce Overfitting in Reservoir Control Policy Search*	Hasilnya menunjukkan bahwa strategi validasi-kalibrasi untuk pemilihan kebijakan, ditambah dengan metode *bootstrap aggregation*, dapat meningkatkan kinerja dalam tujuan penyediaan air dan risiko banjir dibandingkan kinerja dasar. Formulasi validasi-kalibrasi dan *bootstrap aggregation* memberikan kinerja terbaik untuk biaya pasokan air dan pengendalian banjir.

Tahun	Nama	Judul	Hasil Penelitian
2020	Imad Abdallah, Konstantinos Tatsis, and Eleni Chatzi	*Unsupervised local cluster-weighted bootstrap aggregating the output from multiple stochastic simulators*	Metode yang diusulkan, *unsupervised local cluster-weighted bootstrap aggregation*, dibandingkan dengan *classical bagging*, *bayesian model averaging*, dan *stacking of predictive distributions*. Hasilnya menunjukkan efektivitas metode yang diusulkan dalam mengurangi kesalahan generalisasi dan meningkatkan ketepatan estimator untuk interval kepercayaan sebesar 95%. Kinerja metode ini dipengaruhi oleh faktor-faktor seperti ukuran wadah masukan, ukuran sampel per simulator, dan simulasi.
2021	Dwi Liana Wella Putri, Scolastika Mariani, dan Sunarmi	Peningkatan Ketepatan Klasifikasi Model Regresi Logistik Biner dengan Metode *Bagging* (*Bootstrap Aggregating*)	Metode bagging menghasilkan tingkat keakuratan sebesar 75,641%, dengan tingkat kesalahan klasifikasi sebesar 24,359%. Penggunaan metode bagging secara signifikan meningkatkan akurasi model regresi logistik dari 69,74% menjadi 75,641%.
2023	**Edossa Merga Terefe**	***Tree-Based Machine Learning Methods For Vehicle Insurance Claims Size Prediction***	**Studi ini menemukan bahwa metode *tree-based ensemble*, seperti *bagging*, *random forest*, dan *gradient boosting*, mengungguli metode *classical least square* dalam memprediksi ukuran klaim asuransi kendaraan.**

Tahun	Nama	Judul	Hasil Penelitian
2023	Michael Lau, Sara Kress, Tamara Schikowski, and Holger Schwender	*Efficient gene-environment interaction testing through bootstrap aggregating*	Makalah ini berfokus pada pengembangan prosedur pengujian yang efisien untuk interaksi GxE, khususnya menggunakan *bagging* dan *random forest*. Hasilnya menunjukkan bahwa metode ini memiliki kekuatan statistik yang lebih tinggi dalam mengendalikan kesalahan tipe I, menjadikannya prosedur pengujian yang valid. Penerapan prosedur ini pada kumpulan data nyata menunjukkan adanya interaksi GxE terkait dengan rheumatoid arthritis dengan paparan polusi udara.

Penelitian terdahulu yang ditulis tebal merupakan penelitian yang menjadi rujukan utama pada penelitian kali ini. Perbedaan yang paling signifikan dengan penelitian kali ini adalah pada bagian studi kasus. Pada penelitian ini membahas mengenai prediksi kerugian pada kecelakaan motor MotoGP 2022. Sedangkan pada penelitian sebelumnya membahas mengenai prediksi ukuran klaim asuransi kendaraan pada *Ethiopian Insurance Corporation*.

BAB III
LANDASAN TEORI

3.1. MotoGP

MotoGP atau "*Motorcycle Grand Prix*" merupakan kejuaraan balap sepeda motor yang melibatkan pembalap dari berbagai negara di seluruh dunia dengan menggunakan sepeda motor kelas tinggi dan berteknologi canggih. MotoGP dimulai pada saat musim semi tiba dan berakhir ketika musim gugur berakhir. Dalam satu musim pergelaran MotoGP dilaksanakan di berbagai lokasi di seluruh dunia. Setiap sirkuit memiliki karakteristik yang berbeda, termasuk panjang trek, tikungan, dan kondisi cuaca. Setiap balapan memberikan poin kepada pembalap, dan pada akhir musim, pembalap dengan jumlah poin tertinggi dinobatkan sebagai juara dunia.

Dorna Sports adalah penyelenggara MotoGP sejak tahun 1991 dan menjadi pemegang hak komersial dan televisi eksklusif. Berbasis di Spanyol dan anak perusahaan di Italia, Grup ini menjadi salah satu manajemen olahraga, pemasaran dan media yang terkemuka. Dorna tidak hanya menyelenggarakan MotoGP sebagai kejuaraan balap motor, tetapi juga menyelenggarakan *World Superbike Championship, Idemitsu Asia Talent Cup, Nothern Talent Cup*, dll. Mereka telah menciptakan sinergi antara kejuaraan-kejuaraan balap motor tersebut di pasar-pasar utama di seluruh dunia. Tidak hanya dalam promosi dan penyelenggaraan acara tetapi juga terus mengembangkan pengendara, pabrikan, dan merek baru.

Pada tahun 2022 MotoGP menggelar 20 balapan di beberapa sirkuit di seluruh dunia. Pada masing-masing balapan dilakukan dalam beberapa sesi, yaitu *Free Practice 1-4, Qualifying Practice 1-2, Warm Up*, dan *Race*. Selain itu, MotoGP diikuti oleh 30 pembalap berbakat dari berbagai negara di seluruh dunia. Beberapa merek produsen motor di dunia berpartisipasi dalam MotoGP 2022 sebagai tim yang bersaing satu sama lain. Produsen motor tersebut antara lain, Yamaha, Ducati, Honda, Aprilia, KTM, dan Suzuki. Setiap produsen motor tersebut mengirimkan satu atau lebih tim untuk berpartisipasi dalam MotoGP 2022. Tim tersebut terdiri dari tim pabrikan dan tim satelit. Produsen motor Suzuki dan Aprilia hanya mengirimkan satu tim pabrikan saja pada MotoGP 2022. Produsen motor Yamaha,

Honda, dan KTM mengirimkan satu tim pabrikan dan satu tim satelit. Sedangkan, produsen motor Ducati mengirimkan satu tim pabrikan dan tiga tim satelit pada MotoGP 2022. Setiap tim tersebut berisi dua pembalap yang selalu mengikuti balapan diseluruh agenda balap MotoGP 2022 dan ada juga beberapa pembalap tes yang mengikuti balapan hanya di beberapa sirkuit saja.

Sepeda motor yang digunakan pada balapan MotoGP memiliki komponen-komponen yang luar biasa canggih dan sangat mahal. Ketika terjadi kecelakaan komponen-komponen tersebut beresiko mengalami kerusakan. Oleh karena itu, biaya yang dikeluarkan untuk memperbaiki atau mengganti komponen yang rusak akibat kecelakaan diperkirakan hingga 20 juta dolar Amerika (Samudra, 2023). Biaya tersebut sepenuhnya ditanggung oleh tim atau sponsor yang bertanggung jawab. Meskipun pembalap yang sangat berpengaruh dalam terjadinya kecelakaan, mereka tidak perlu mengeluarkan biaya apapun apabila terjadi kerusakan pada motor (Banaran, 2022).

3.2. Pengertian Asuransi

Pengertian asuransi telah dikemukakan oleh beberapa ahli terdahulu menurut pandangannya masing-masing. Pengertian asuransi pertama dikemukakan oleh Mark R. Greene (1968), Asuransi adalah suatu entitas ekonomi yang berupaya mengurangi risiko dengan mengumpulkan sejumlah individu atau entitas dan mengelola risiko tersebut untuk meramalkan tingkat kerugian dengan lebih detail, terutama dalam konteks geografis dan periode waktu tertentu. Sedangkan menurut Robert I. Mehr dan Emerson Cammack (1972), Asuransi adalah sarana untuk mengurangi potensi kerugian finansial dengan mengumpulkan sejumlah unit risiko yang mencukupi sehingga kerugian pada tingkat individu dapat diestimasi dengan lebih akurat. Kerugian yang dapat diperkirakan kemudian akan ditanggung secara merata oleh para peserta. Menurut C Arthur Williams Jr. dan Richard M. Heins (1989), Asuransi adalah produk yang menawarkan tingkat keamanan risiko, di mana sekelompok individu atau perusahaan berkumpul dengan berkontribusi melalui pembayaran premi atau penyisihan dana, yang kemudian digunakan untuk menutup klaim-klaim.

Dari beberapa pengertian asuransi yang dikemukakan oleh para ahli di atas penulis dapat menyimpulkan bahwa asuransi memungkinkan para peserta atau

pemegang polis untuk melindungi diri mereka dari kerugian finansial yang dapat diperkirakan, seperti kecelakaan atau kerusakan properti. Risiko ini kemudian ditanggung secara merata oleh semua peserta dalam kelompok atau perusahaan asuransi. Dalam hal ini, asuransi berfungsi sebagai alat untuk meminimalkan risiko finansial, sehingga individu atau entitas yang diasuransikan dapat merasa lebih aman terhadap kemungkinan kerugian yang mungkin terjadi. Asuransi adalah suatu kesepakatan antara perusahaan asuransi atau pihak penanggung dengan pemegang polis atau tertanggung. Dalam kesepakatan ini, pihak penanggung berkomitmen untuk memberikan penggantian kepada tertanggung atas kerugian, kerusakan, atau kehilangan keuntungan yang mungkin timbul akibat peristiwa yang tidak pasti. Pihak tertanggung membayar premi asuransi sebagai kompensasi. Kesepakatan asuransi merupakan perjanjian timbal balik yang melibatkan penanggung yang menerima risiko terhadap benda atau objek yang diasuransikan, sementara tertanggung berkewajiban membayar premi sebagai balasannya. Perjanjian asuransi harus dibuat secara tertulis dalam bentuk polis, yang berisi persetujuan, ketentuan, dan komitmen khusus yang menjadi dasar hak dan kewajiban kedua belah pihak dalam mencapai tujuan asuransi. Dengan demikian, polis adalah dokumen tertulis yang menegaskan adanya kontrak asuransi antara tertanggung (pemegang polis) dan pihak penanggung (perusahaan asuransi).

Berdasarkan cakupan perlindungannya, asuransi dapat diklasifikasikan menjadi dua kategori, yaitu asuransi jiwa dan asuransi umum. Asuransi jiwa adalah bentuk perlindungan yang melibatkan pengalihan risiko untuk memberikan kompensasi dalam hal terjadinya kerugian finansial akibat kematian yang tak terduga dari pemegang polis. Sementara itu, asuransi umum merupakan jenis asuransi yang bertujuan untuk mengalihkan risiko terkait aset yang rentan terhadap kerusakan atau kehilangan (Fitriani & Gunardi, 2020).

3.3. Asuransi Kendaraan

Kendaraan adalah suatu peranti yang digunakan untuk mengangkut orang atau barang dari suatu lokasi ke lokasi lain, yang digerakkan oleh mesin atau sistem mekanis lainnya. (Alam & Mutaqin, 2023). Kendaraan dapat mencakup berbagai jenis, seperti mobil, sepeda motor, truk, dan lain-lain. Beberapa kendaraan

dirancang khusus untuk transportasi pribadi, sementara yang lain digunakan untuk transportasi umum, pengiriman barang, dan keperluan khusus lainnya.

Asuransi kendaraan termasuk dalam kategori asuransi umum. Jenis asuransi ini menyediakan perlindungan terhadap kerugian finansial dan kerusakan pada kendaraan yang disebabkan oleh berbagai faktor, seperti tabrakan, kecelakaan, tindakan kriminal, kebakaran, dan faktor lainnya.

Kecelakaan adalah suatu peristiwa yang tidak diinginkan atau direncanakan yang mengakibatkan kerusakan, cedera, atau kerugian. Kecelakaan kendaraan terjadi sering kali disebabkan oleh beberapa faktor seperti kelalaian, ketidakhati-hatian, cuaca buruk, dan lain-lain. Kecelakaan dapat memiliki konsekuensi serius, seperti cedera fisik, kerusakan properti, atau bahkan kehilangan nyawa.

Sama seperti dalam jenis asuransi lainnya, dalam asuransi kendaraan, pemegang polis memiliki tanggung jawab untuk membayar premi kepada perusahaan asuransi sesuai dengan kesepakatan. Besaran premi yang harus dibayarkan ditentukan oleh perusahaan asuransi dengan mempertimbangkan berbagai faktor terkait kondisi kendaraan, jenis kendaraan, usia kendaraan, tingkat perlindungan, penggunaan kendaraan, serta risiko yang mungkin terjadi pada lokasi penggunaan.

3.4. *Machine Learning*

Machine learning (ML) kini sudah banyak digunakan dalam berbagai bidang. Berbeda dengan pendekatan pemodelan statistik, algoritma ML tidak mengasumsikan struktur model tertentu untuk data. Metode ML menangkap struktur dasar data sehingga lebih efisien dalam menangani data besar dengan tingkat kompleksitas yang berubah-ubah. Salah satu tugas utama ML adalah membuat model yang baik dari kumpulan data.

Di antara algoritma ML, metode ansambel adalah salah satu pilihan yang umum untuk menganalisis data yang besar dan kompleks. Awalnya dikembangkan untuk mengurangi varians sehingga meningkatkan akurasi sistem pengambilan keputusan otomatis, metode ansambel telah berhasil digunakan untuk mengatasi berbagai masalah ML (Zhang & Ma, 2012), seperti pemilihan prediktor, data dengan kelas yang tidak seimbang, estimasi kepercayaan, koreksi kesalahan, dan lain-lain.

Ide utama di balik metode ansambel adalah menimbang beberapa pola ML individu, dan menggabungkannya untuk mendapatkan ML yang kinerjanya lebih baik dari sebagian besar ML. Faktanya, menggabungkan keluaran ML tidak serta merta menghasilkan kinerja yang dijamin lebih baik daripada ML terbaik dalam ansambel tersebut. Sebaliknya, hal ini mengurangi kemungkinan memilih ML dengan kinerja buruk. Metodologi ansambel meniru untuk mencari beberapa pendapat sebelum mengambil keputusan penting. Pendapat individu tersebut ditimbang, dan digabungkan untuk mencapai keputusan akhir (Polikar, 2006).

Prinsip umum metode ansambel adalah membangun kombinasi linier dari beberapa metode pemasangan model, alih-alih menggunakan satu metode untuk meningkatkan kinerja prediktif pembelajaran statistik atau teknik pemasangan model tertentu. Lebih tepatnya, pertimbangkan estimasi fungsi bernilai riil $f: \mathbb{R}^p \rightarrow \mathbb{R}$ berdasarkan data $\{(x_i, y_i); i = 1, \dots, n\}$ dengan x adalah variabel prediktor berdimensi p dan y merupakan respons univariat. Kemudian dapat menggeneralisasi fungsi $f(x)$ dan tipe data lainnya (Terefe, 2023).

Beberapa data dimasukkan untuk dipelajari fungsi-fungsinya $\left(\hat{f}_1, \hat{f}_2, \hat{f}_3, \dots, \hat{f}_B\right)$, atau disebut juga sebagai *learners*, dengan mengubah data yang dimasukkan berdasarkan pembobotan ulang yang berbeda, kemudian dapat membuat estimasi fungsi berbasis ansambel $\hat{f}_{ens}(x)$ dengan mengambil kombinasi linier dari masing-masing *learners* sebagai perluasan aditif dari *learners* (Elish, 2009) $\hat{f}_i(x)$:

$$\hat{f}_{ens}(x) = \sum_{i=1}^{B} w_i \hat{f}_i(x) \tag{3.1}$$

$\hat{f}_i(x)$ adalah perkiraan yang diperoleh dari kumpulan data ke-i yang ditimbang ulang dan w_i adalah koefisien kombinasi linier. Misalnya, $w_i = 1/B$, yang merupakan rata-rata untuk *bootstrap aggregating*.

3.5. *Bagging*

Bagging atau singkatan dari *bootstrap aggregating* adalah metode ansambel untuk membuktikan skema estimasi atau klasifikasi yang tidak stabil. Sesuai dengan namanya, dua unsur utama *bagging* adalah *bootstrap* dan *aggregating* (Breiman, 1996).

Bagging mengadopsi distribusi bootstrap untuk menghasilkan *learners* yang berbeda. Dengan kata lain, *bagging* menerapkan *bootstrap sampling* untuk memperoleh subset data untuk *training learners* (Efron & Tibshirani, 1994). Secara rinci, dengan kumpulan data asli, dihasilkan kumpulan data yang berisi n observasi *training* dengan pengambilan sampel dan penggantian. Beberapa observasi asli dapat muncul lebih dari satu kali, tetapi beberapa observasi asli dapat pula tidak ada dalam sampel. Apabila diterapkan proses B kali, diperoleh B sampel observasi pelatihan. Kemudian, dari setiap sampel satu *learner* dapat dilatih dengan menerapkan ML.

Bagging juga mengadopsi strategi yang paling populer untuk menggabungkan *output* dari *learners*, yaitu memilih klasifikasi dan membuat rata-rata regresi. Suatu kejadian pengujian diprediksi dengan mengambil regresi. Misalnya *bagging* memasukkan contoh tersebut ke *leaners* dan mengumpulkan semua *output*, lalu mengambil rata-rata dari *output* tersebut sebagai prediksi.

Pada regresi, metode bagging diterapkan sebagai berikut. *Learner* $\hat{f}_i(x)$ dipasang pada masing-masing sampel B yang di-*bootstrap*, dengan $\hat{f}_i(x)$ menunjukkan nilai respons yang diprediksi dari $i = 1,2, \dots, B$ *learners*. Kemudian B *learners* yang dikonstruksi kemudian digabungkan menggunakan *aggregation*, sehingga rata-rata prediksi $f_{av}(x)$ diperkirakan sebagai rata-rata prediksi *outputs* dari $\hat{f}_i(x)$ sebagai berikut:

$$\hat{f}_{av}(x) = \frac{1}{B}\sum_{i=1}^{B} \hat{f}_i(x) \tag{3.2}$$

dengan:

$\hat{f}_{av}$: hasil rata-rata nilai tunggal

$\hat{f}_i(x)$: *Bootstrap* (pengulangan) ke-i

Perhitungan ini bertujuan untuk memperoleh satu *learner* yang mempunyai variansi rendah.

3.6. *Variable Importance*

Fungsi untuk mengevaluasi *variable importance* dapat dikelompokkan menjadi dua kelompok: yang menggunakan informasi model dan yang tidak. Menggunakan pendekatan berbasis model memiliki keunggulan karena lebih erat terkait dengan kinerja model dan mungkin dapat mengintegrasikan struktur korelasi

di antara prediktor ke dalam perhitungan *variable importance*. Pada *Bagging*, untuk setiap pohon, akurasi prediksi pada data *out-of-bag* dicatat. Proses yang sama diulang pada setiap variabel prediktor. Perbedaan antara akurasi tersebut di seluruh pohon kemudian dihitung rata-rata dan dinormalisasi oleh *standard error*. Pada regresi, *Root Mean Squared Error* (RMSE) dihitung pada data *out-of-bag* untuk setiap pohon, dan kemudian dihitung ulang pada setiap variabel. Perbedaan tersebut dihitung rata-rata dan dinormalisasi oleh *standard error*. Jika kesalahan standar sama dengan 0 untuk suatu variabel, tidak ada pembagian yang dilakukan (Kuhn, 2019).

Tujuannya bukan hanya untuk menemukan model yang paling akurat untuk respons, tetapi juga untuk menemukan variabel prediktor mana yang paling penting untuk membuat prediksi. Oleh karena itu, dalam melakukan analisis *variable importance*. Metode ensemble memperkirakan pentingnya suatu variabel prediktor, misalnya x_j, dengan melihat seberapa besar peningkatan kesalahan prediksi dibandingkan dengan kesalahan dasar ketika sampel *Out-of-Bag* (OOB) untuk x_j dipermutasi sementara yang lain tetap tidak berubah. Pengukuran *variable importance* yang paling banyak digunakan adalah perbedaan dalam akurasi prediksi sebelum dan setelah permutasi x_j, dirata-ratakan di seluruh pohon. Secara khusus, pentingnya variabel prediktor x_j didefinisikan sebagai berikut: (Breiman, 2001)

$$VI(x_j) = \frac{1}{B}\sum_{i=1}^{B} VI(X_j)_i \tag{3.3}$$

Dimana $VI(x_j)_i = \left(RMSE_i^{x_j} - RMSE_i^0\right)$, dengan $RMSE_i^{x_j}$ menunjukkan nilai RMSE untuk model ke i dalam ensemble yang dihasilkan dari dataset dengan permutasi acak yang diterapkan pada kovariat x_j, dan $RMSE_i^0$ mewakili nilai RMSE untuk model ini yang disesuaikan dengan dataset asli. Perlu diperhatikan bahwa $VI(x_j)_i = 0$ jika variabel x_j tidak ada dalam model ke i. Skor *variable importance* untuk setiap variabel kemudian dihitung sebagai rata-rata pentingnya di semua pohon.

3.7. *Partial Dependence Plots*

Partial dependence plots digunakan untuk menilai hubungan antara variabel prediktor dan variabel respon (Friedman, 2001). *Partial dependence plots*

membantu memvisualisasikan hubungan antara variabel prediktor dan variabel respon, lalu memperhitungkan efek rata-rata prediktor lain dalam model.

Misalkan $\hat{y}$ adalah fungsi prediksi dari model menggunakan kumpulan data, $D = \{(x_{i,j}, y_i)\}$ untuk $i = 1, \dots, n$ dan $j = 1, \dots, p$. Model ini menghasilkan prediksi dalam bentuk.

$$\hat{y}_i = f(x_{i,1}, x_{i,2}, \dots, x_{i,p}) \tag{3.4}$$

Misalkan x_k adalah prediktor tunggal yang memiliki nilai unik $(x_{1,k}, x_{2,k}, \dots, x_{n,k})$. Kemudian *partial dependence plots* diperoleh dengan menghitung rata-rata dan memplotnya pada rentang nilai x:

$$\bar{f}_k(x) = \frac{1}{n}\sum_{j=1}^{n} \hat{f}(x_{1,j}, \dots, x_{k-1,j}, x, x_{k+1,j}, \dots, x_{p,j}) \tag{3.5}$$

Fungsi $\bar{f}_k(x)$ mengindikasikan bagaimana nilai dari variabel x_k mempengaruhi model prediksi $\{\hat{y}_j\}$ setelah dihitung rata-rata pengaruh seluruh variabel. *Partial dependence plots* dari respon terhadap x_k dapat dibuat dengan cara

1. Untuk $j \in \{1,2, \dots, n\}$;
   a. Ganti nilai pelatihan x_k dengan konstanta $x_{k,j}$
   b. Hitung vektor nilai prediksi, $\{\hat{y}_j\}$ dari versi modifikasi data pelatihan
   c. Hitung prediksi rata-rata dengan rumus 3.5 untuk mendapatkan $\bar{f}_k(x_{k,j})$
2. Buat plot pasangan dari $\{x_{k,j}, \bar{f}_k(x_{k,j})\}$ untuk $j = 1,2, \dots, n$

3.8. *Ordinary Least Square* (OLS)

OLS adalah teknik untuk menghitung parameter yang tidak diketahui dalam model regresi linier. Metode ini meminimalkan jumlah kuadrat jarak vertikal antara respons yang diamati dalam kumpulan data dan respons yang diprediksi dengan pendekatan linier. Estimator yang diperoleh dapat dinyatakan dengan rumus sederhana, terutama pada kasus regresi tunggal (Cavdar, Aydin, & Karaaslan, 2015).

Pada dasarnya OLS membangun model sebagai formula yang menggabungkan prediktor dan parameter. Misalnya jika ada satu prediktor maka fungsi modelnya adalah (Usman, Garba, Zoramawa, & Usman, 2020)

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{3.6}$$

Dimana y adalah variabel dependen, x adalah variabel independen, β adalah parameter, dan ε adalah istilah kesalahan yang mewakili varians populasi yang tidak dapat dijelaskan dalam model. Metode OLS menghitung penduga β dengan menggunakan formula (Jang & Lee, 2022).

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{3.7}$$

BAB IV
METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Penelitian

Pada penelitian ini populasi yang digunakan adalah jumlah kecelakaan para pembalap MotoGP. Sampel penelitian yang digunakan adalah jumlah kecelakaan para pembalap MotoGP 2022.

4.2. Tempat dan Waktu Penelitian

Penelitian dilakukan di FMIPA UII dari September 2023 hingga Mei 2024. Waktu pengambilan data dilakukan pada tanggal 4 Oktober 2023.

4.3. Definisi Operasional Variabel Penelitian

Tabel 4.1 Penjelasan Variabel Penelitian

Variabel	Keterangan	Unit/Satuan	Definisi
$X1$	Usia	Tahun	Usia pembalap ketika MotoGP 2022 dilakukan
$X2$	Asal	-	Benua kelahiran pembalap
$X3$	Merek	-	Merek motor yang digunakan pembalap MotoGP 2022
$X4$	Harga	Ribu Dolar Amerika	Perkiraan harga motor yang digunakan pembalap MotoGP 2022
$X5$	Usia Kendaraan	Tahun	Usia kendaraan ketika MotoGP 2022 dilakukan
$X6$	Tim	-	Jenis tim yang terdiri dari pabrikan dan satelit
Y	Banyak Kecelakaan	Kecelakaan	Jumlah kecelakaan yang dialami pembalap MotoGP 2022

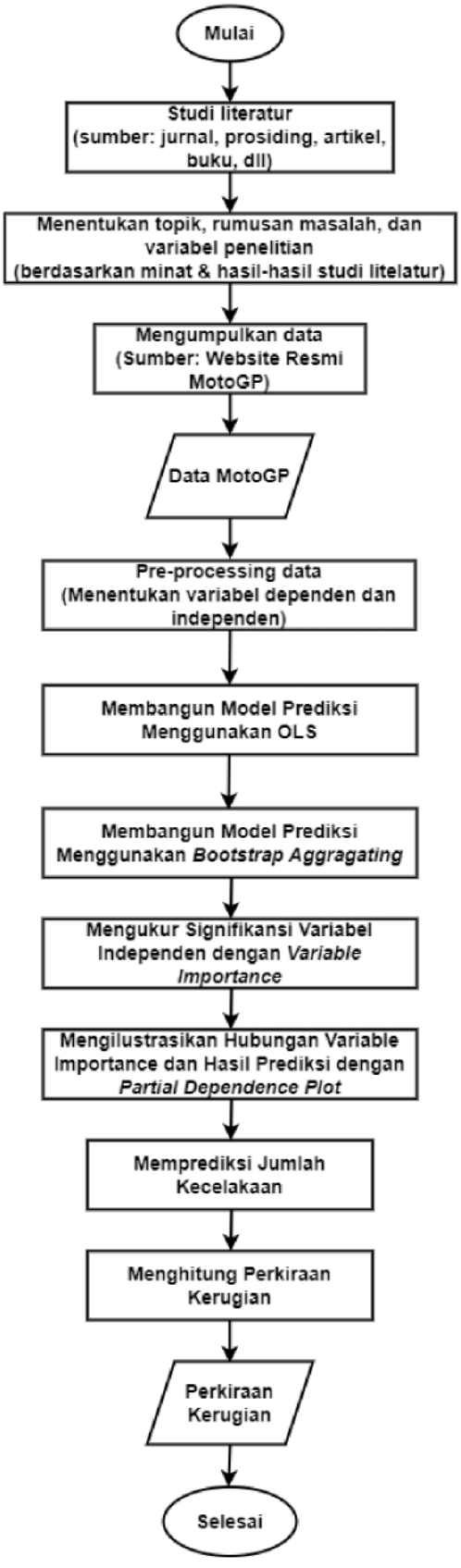
Variabel $X1$, yang dikenal sebagai usia, menunjukkan usia para pembalap pada MotoGP 2022. Variabel $X2$, atau asal, menunjukkan negara asal para pembalap MotoGP 2022 yang dikategorikan berdasarkan benua mereka, meliputi Afrika (ditandai sebagai 1), Eropa (ditandai sebagai 2), Australia (ditandai sebagai 3), dan Asia (ditandai sebagai 4). Variabel $X3$, atau merek, mencakup merek-merek sepeda motor yang berpartisipasi dalam balapan MotoGP 2022, termasuk Yamaha

(ditandai sebagai 1), Ducati (ditandai sebagai 2), Honda (ditandai sebagai 3), Aprilia (ditandai sebagai 4), KTM (ditandai sebagai 5), dan Suzuki (ditandai sebagai 6). Variabel $X4$, atau harga, berisi perkiraan harga masing-masing sepeda motor MotoGP 2022 berdasarkan merek sepeda motor dan jenis produsen atau tim satelit dalam ribuan dolar Amerika. Variabel $X5$, atau usia kendaraan, mewakili usia kendaraan selama acara MotoGP 2022, dikategorikan sebagai tahun pertama dan tahun kedua penggunaan sepeda motor. $X6$, atau tim, mengkategorikan jenis tim yang bekerja sama, termasuk tim pabrikan (ditandai sebagai 1) dan tim satelit (ditandai sebagai 2). Terakhir, variabel Y, atau jumlah kecelakaan, mencerminkan jumlah kecelakaan yang dialami oleh para pembalap MotoGP 2022.

4.4. Alat dan Cara Organisir Data

Data diperoleh dari *website* resmi MotoGP. Salah satu data dalam *website* tersebut adalah data mengenai jumlah kecelakaan yang terjadi pada setiap pembalap ketika perhelatan MotoGP 2022. Data tersebut kemudian dianalisis menggunakan aplikasi M.S. Excel dan RStudio.

4.5. Metode Penelitian



Gambar 4.1 Tahapan Penelitian

BAB V
PEMBAHASAN

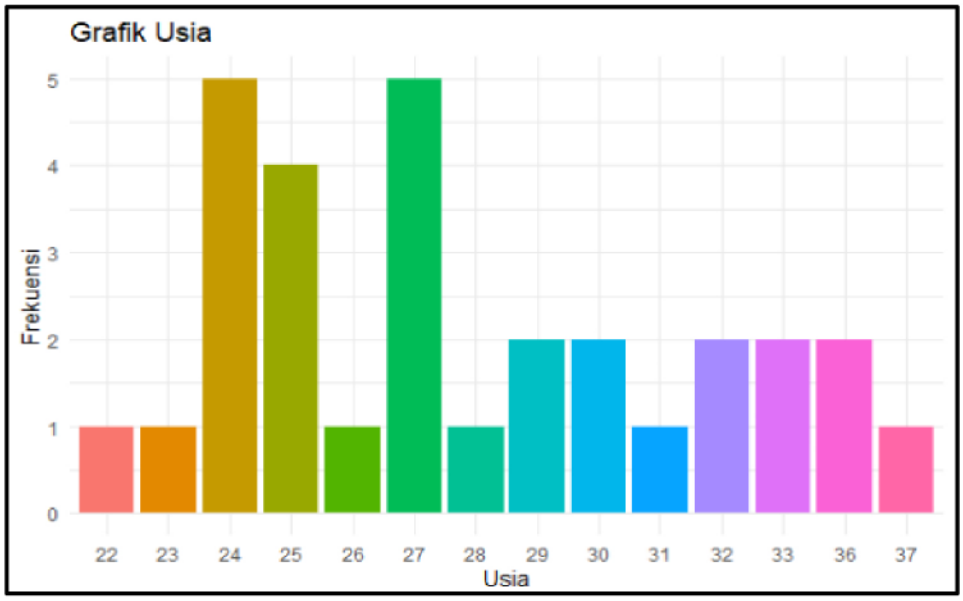
5.1. Statistika Deskriptif

Pada bagian ini, akan dijelaskan statistika deskriptif dari masing-masing variabel yang digunakan pada penelitian kali ini. Terdiri dari enam variabel independen dan satu variabel dependen. Berikut adalah statistika deskriptif untuk variabel-variabel tersebut.

Tabel 5.1 Statistika Deskriptif Variabel Penelitian

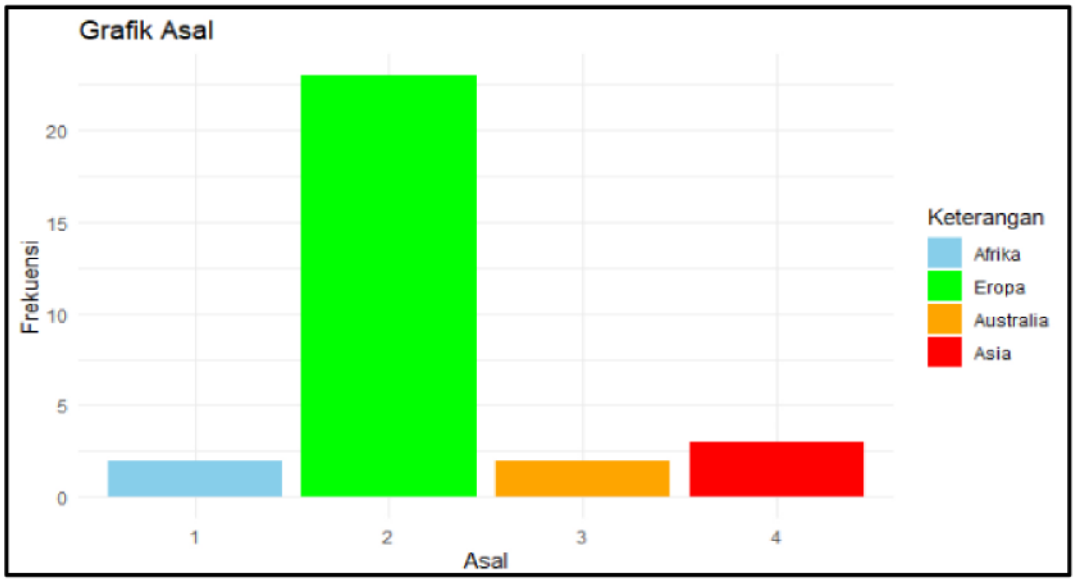
Variabel	Minimal	Rata-rata	Maksimal	Std. Deviasi
$X1$	22	28	37	4,13
$X4$	525	692,3	945	124,95
Y	1	11	27	6,9

Hasil statistika deskriptif dari masing-masing variabel dapat dilihat nilai minimal, rata-rata, maksimal, dan standar deviasinya. Pada variabel X1, atau usia, para pembalap berusia antara 22 hingga 37 tahun dengan rata-rata usia pembalap sebesar 28 tahun dan standar deviasi sebesar 4,13 artinya usia pembalap bervariasi sekitar 4,13 tahun dari nilai rata-rata usia pembalap. Pada variabel X4, atau harga, rentang harga motor MotoGP berkisar antara 525 hingga 945 ribu dolar Amerika dengan rata-rata harga motor sebesar 692,3 ribu dolar Amerika dan standar deviasi sebesar 124,95 artinya harga motor bervariasi sekitar 124,95 ribu dolar Amerika dari nilai rata-rata harga motor. Pada variabel Y, atau jumlah kecelakaan, rentang kecelakaan para pembalap berkisar antara satu hingga 27 kali dengan rata-rata kecelakaan sebanyak 11 kali dan standar deviasi sebesar 6,9 artinya jumlah kecelakaan yang dialami pembalap bervariasi sekitar 6,9 kali dari nilai rata-rata kecelakaan. Selanjutnya, akan dibuat grafik untuk masing-masing variabel agar dapat dilihat lebih detail mengenai karakteristik data setiap variabelnya.



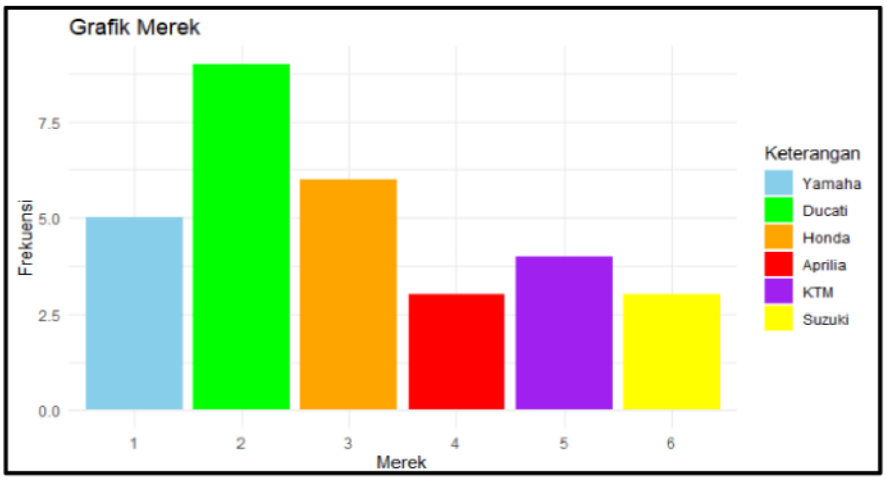
Gambar 5.1 Grafik $X1$ atau Usia

Pada grafik $X1$, atau usia, terlihat bahwa pembalap termuda berusia 22 tahun dan pembalap tertua berusia 37 tahun. Selain itu, paling banyak pembalap berusia 24 dan 27 tahun masing-masing sebanyak lima orang. Usia pembalap yang paling sedikit berada pada usia 22, 23, 26, 28, 31, dan 37 tahun sebanyak satu orang untuk masing-masing usia tersebut. Setengah dari jumlah pembalap berusia 24 hingga 27 tahun. Artinya pada usia-usia tersebut para pembalap sedang berada pada masa keemasan mereka. Sehingga banyak tim yang mempercayai pembalap pada usia 24 hingga 27 tahun.



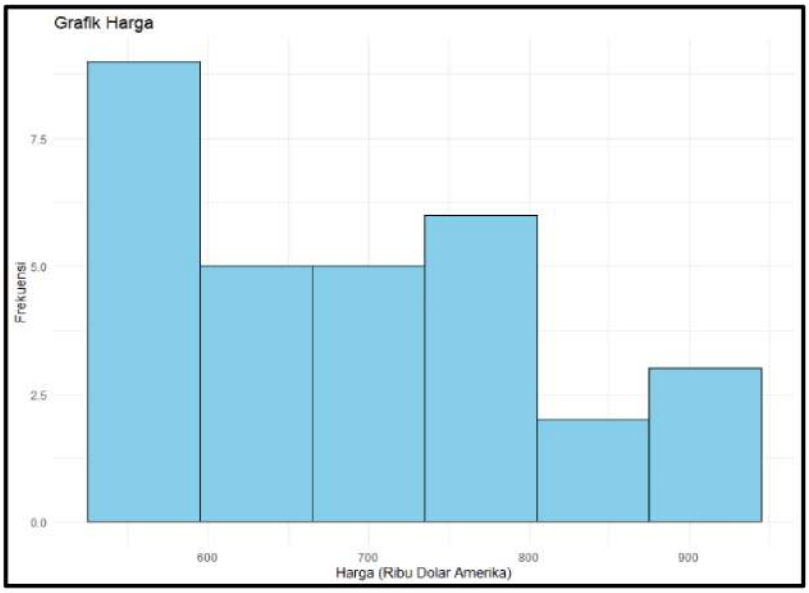
Gambar 5.2 Grafik $X2$ atau Asal

Pada grafik $X2$, atau asal, terlihat bahwa para pembalap berasal dari empat benua, yaitu Afrika, Eropa, Australia, dan Asia. Dari ke empat benua tersebut, benua Eropa memiliki pembalap terbanyak yaitu sebanyak 23 orang. Mayoritas pembalap berasal dari benua Eropa karena di Eropa industri MotoGP sangat berkembang. Hal tersebut ditunjukkan dengan 11 dari 20 kali balapan MotoGP dilaksanakan di benua Eropa. Sehingga kebanyakan tim MotoGP lebih banyak mempercayakan balapan pada pembalap yang berasal dari benua Eropa.



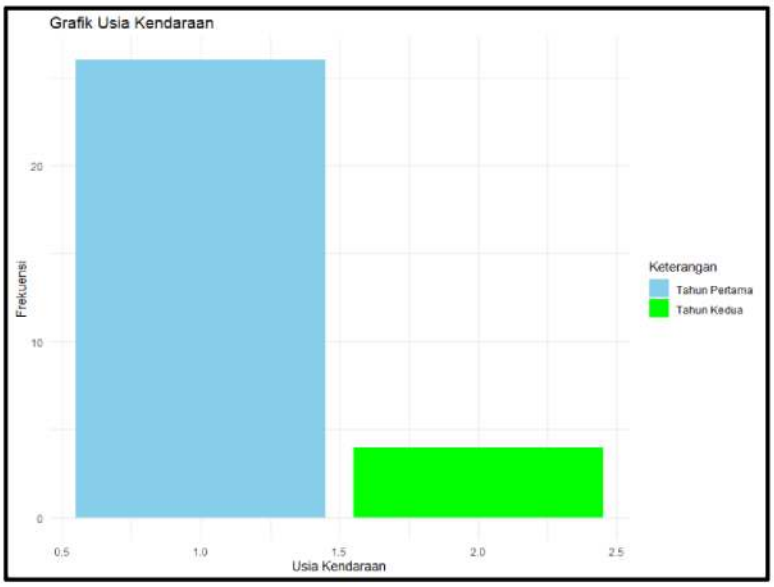
Gambar 5.3 Grafik $X3$ atau Merek

Pada grafik $X3$, atau merek, terlihat bahwa balapan MotoGP diikuti oleh enam merek motor yaitu Yamaha, Ducati, Honda, Aprilia, KTM, dan Suzuki. Mayoritas pembalap menggunakan motor dari merek Ducati yaitu sebanyak sembilan orang. Sedangkan minoritas pembalap menggunakan motor dari merek Aprilia dan Suzuki yaitu masing-masing sebanyak tiga orang. Para pembalap cenderung lebih tertarik mengendarai motor dengan merek Ducati karena pada tahun 2021 tiga dari lima pembalap dengan poin teratas menggunakan merek Ducati. Sehingga motor dengan merek Ducati cenderung lebih kompetitif dibandingkan motor dengan merek yang lain. Selain itu, tim yang memilih menggunakan motor dengan merek Ducati lebih banyak dibandingkan tim yang memilih menggunakan motor dengan merek yang lain. Terdapat lima tim yang menggunakan motor dengan merek Ducati, terdiri dari satu tim pabrikan dan empat tim satelit.



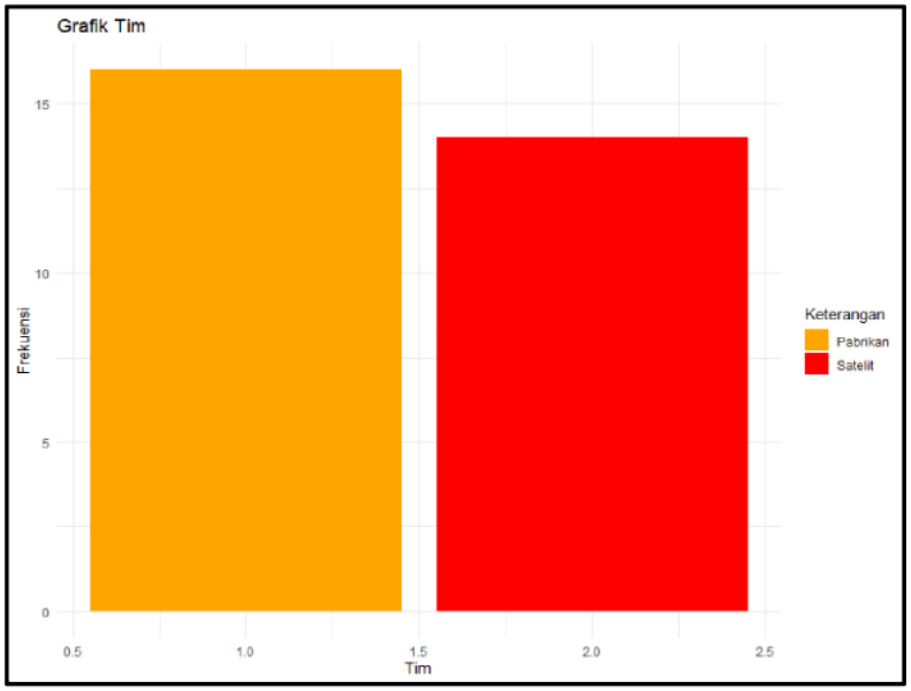
Gambar 5.4 Grafik $X4$ atau Harga

Pada grafik $X4$, atau harga, terlihat bahwa motor MotoGP memiliki harga berkisar antara 525 hingga 945 ribu dolar Amerika. Paling banyak harga motor berada pada kisaran 525 hingga 595 ribu dolar Amerika sebanyak sembilan motor. Sedangkan paling sedikit harga motor berada pada kisaran harga 805 hingga 875 ribu dolar Amerika sebanyak dua motor. Mayoritas motor berada pada kisaran harga 525 hingga 805 ribu dolar Amerika karena harga motor di atas 800 ribu dolar Amerika hanya tim-tim dari merek motor yang terhitung mahal yang dapat mendanai motor dengan harga lebih dari 800 ribu dolar Amerika.



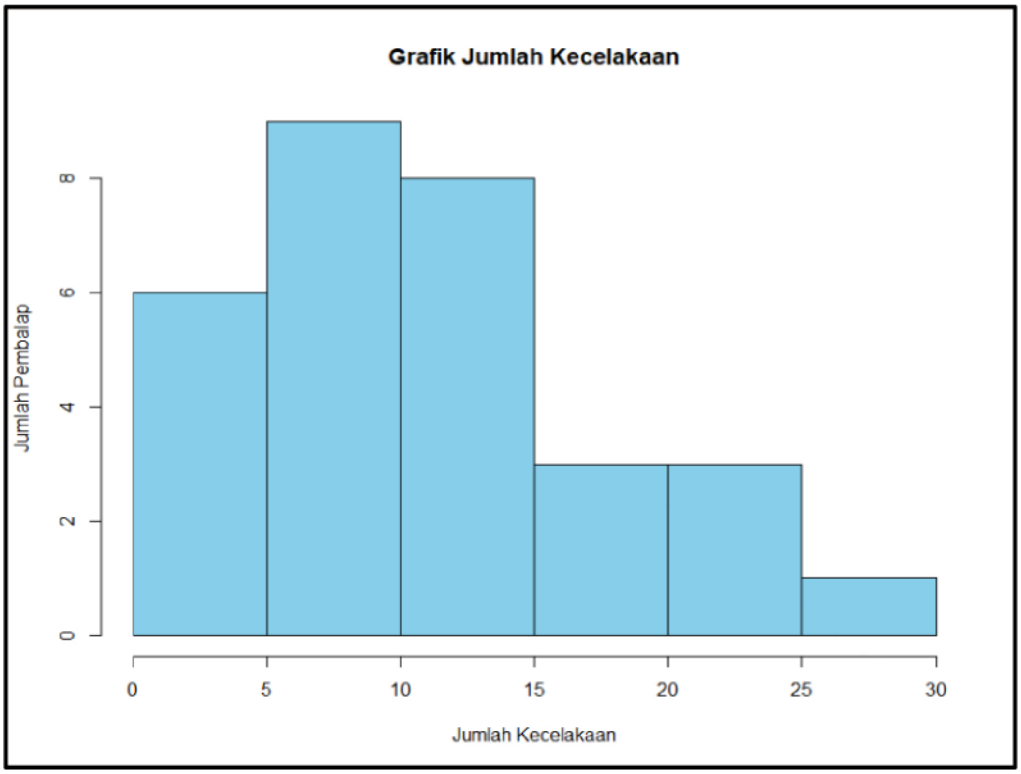
Gambar 5.5 Grafik $X5$ atau Usia Kendaraan

Pada grafik $X5$, atau usia kendaraan, terlihat bahwa motor MotoGP berusia dua tahun saja yaitu tahun pertamanya dan tahun keduanya motor tersebut berlaga dalam MotoGP 2022. Mayoritas merupakan motor tahun pertama sebanyak 26 motor dan motor tahun kedua hanya sebanyak empat motor saja. Hal ini terjadi karena tim satelit biasanya diberikan motor tahun lalu atau motor yang berusia dua tahun, tetapi beberapa merek motor mempercayakan tim satelit untuk menggunakan motor terbaru atau motor yang berusia satu tahun. Alasan merek motor yang mempercayakan tim satelit untuk menggunakan motor terbaru karena pada tahun 2021 tim satelit atau pembalap yang ditunjuk menggunakan motor terbaru memperoleh prestasi yang bagus. Sehingga mereka memberikan motor terbaru kepada pembalap tersebut meskipun tim mereka adalah tim satelit.



Gambar 5.6 Grafik $X6$ atau Tim

Pada grafik $X6$, atau tim, berisi dua tim yang mengikuti balapan MotoGP 2022 yaitu tim pabrikan dan tim satelit. Tim pabrikan dan tim satelit masing-masing berjumlah 16 dan 14 tim. Biasanya satu merek motor yang mengikuti MotoGP mempunyai dua tim yaitu tim pabrikan dan tim satelit. Tetapi, merek motor Suzuki hanya mempunyai satu tim pabrikan saja. Merek motor Suzuki hanya mempunyai satu tim pabrikan saja karena dana mereka untuk pengembangan MotoGP sudah dibatasi. Terlihat pada tahun-tahun selanjutnya, merek motor Suzuki tidak lagi mengikuti balapan MotoGP.



Gambar 5.7 Grafik *Y* atau Jumlah Kecelakaan

Pada grafik *Y*, atau jumlah kecelakaan, pembalap MotoGP paling sedikit mengalami kecelakaan sekali dan paling banyak mengalami kecelakaan sebanyak 27 kali. Terdapat sebanyak sembilan pembalap mengalami kecelakaan enam sampai sepuluh kali. Sedangkan pembalap yang mengalami kecelakaan 26 hingga 30 kali hanya terdapat satu pembalap saja. Dalam satu musim, MotoGP 2022 melakukan balapan hingga 20 kali balapan. Selain itu, dalam sekali balapan terdapat beberapa sesi yang harus dilakukan pembalap yaitu sesi latihan bebas sebanyak empat kali, sesi kualifikasi sebanyak dua kali, sesi *warm up* sebanyak sekali, dan sesi balapan sebanyak sekali. Sehingga wajar apabila mayoritas pembalap terjatuh hingga 10 kali dalam satu musim balapan MotoGP.

5.2. Ordinary Least Square

Regresi *Ordinary Least Square* (OLS) akan mencoba menemukan garis yang secara akurat mewakili hubungan antara variabel prediktor dan respons. Semua variabel prediktor dimasukkan ke dalam model untuk melihat apakah ada pengaruh dari ke-enam variabel prediktor tersebut. Berikut adalah hasil OLS dari ke-enam variabel dengan satu response.

Tabel 5.2 Hasil Analisis OLS

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	16,65316	17,96758	0,927	0,364
$X1$	-0,41623	0,325792	-1,278	0,214
$X2$	-1,63009	1,782754	-0,914	0,37
$X3$	-0,41851	0,893258	-0,469	0,644
$X4$	0,002606	0,010408	0,25	0,804
$X5$	5,729788	4,139753	1,384	0,18
$X6$	1,877291	2,927571	0,641	0,528

Persamaan regresi yang dihasilkan adalah sebagai berikut

$$Y = 16{,}65316 - 0{,}41623X1 - 1{,}63009X2 - 0{,}41851X3 + 0{,}002606X4 + 5{,}729788X5 + 1{,}877291X6 \quad (5.1)$$

Nilai *intercept* sebesar 16,65316 mewakili skor Y ketika variabel independen yang lain bernilai nol. Selanjutnya, *root mean square error* (RMSE) dalam hal ini menunjukkan bahwa, secara rata-rata, nilai Y yang diamati menyimpang sebesar 5,39675 poin dari skor yang diprediksi oleh garis regresi.

Hasil OLS tersebut lalu dilakukan uji F untuk menguji kelayakan model dan menguji parameter regresi secara keseluruhan. Berikut adalah hipotesis uji F.

Hipotesis

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_1: minimal ada satu i di mana $\beta_i = 0$, $i = 1.2{,}3{,}4{,}5{,}6$

Tingkat Signifikansi

$\alpha = 0{,}05$

Daerah Kritis

Tolak H_0 jika $F_{stat} > F_{(6{,}23;0.05)}$

Statistik Uji

Diperoleh $F_{stat} = 2{,}23$

Keputusan

Karena nilai $F_{stat} < F_{(6{,}23;0{,}05)}$ maka keputusannya adalah gagal tolak H_0

Kesimpulan

Dengan menggunakan tingkat kepercayaan 95% maka berdasarkan uji F dapat disimpulkan data gagal tolak H_0 atau tidak ada pengaruh keenam variabel independen terhadap variabel Y.

Setelah melakukan uji F, dilanjutkan dengan uji t untuk menguji parameter estimasi model regresi secara parsial, yaitu untuk mengetahui apakah variabel independen berpengaruh secara signifikan terhadap variabel dependen.

Hipotesis

$H_0: \beta_i = 0;\ i = 1,2,3,4,5,6$

$H_1: \beta_i \neq 0;\ i = 1,2,3,4,5,6$

Tingkat Signifikansi

$\alpha = 0{,}05$

Daerah Kritis

Tolak H_0 jika $t_{value} > t_{(23;0.05)}$

Statistik Uji dan Keputusan

Tabel 5.3 Statistik Uji dan Keputusan Uji t

Koefisien	t_{value}	Tanda	$t_{23;0.05}$	Keputusan
(Intercept)	0,927	<	1,714	Gagal Tolak H_0
$X1$	-1,278	<	1,714	Gagal Tolak H_0
$X2$	-0,914	<	1,714	Gagal Tolak H_0
$X3$	-0,469	<	1,714	Gagal Tolak H_0
$X4$	0,25	<	1,714	Gagal Tolak H_0
$X5$	1,384	<	1,714	Gagal Tolak H_0
$X6$	0,641	<	1,714	Gagal Tolak H_0

Kesimpulan

Dengan menggunakan tingkat kepercayaan 95% maka berdasarkan uji-t dapat disimpulkan data gagal tolak H_0 atau tidak ada pengaruh masing-masing variabel independen terhadap variabel Y.

Berdasarkan uji F dan uji t menghasilkan ke enam variabel independen tidak berpengaruh terhadap variabel dependen. Hal ini dapat disebabkan oleh data yang hanya berjumlah 30 atau tidak cukup banyak untuk melihat pengaruh variabel independen terhadap variabel dependen. Hal ini dapat disebabkan oleh pemilihan

variabel independen yang tidak terlalu berpengaruh terhadap variabel dependen. Sehingga menyebabkan model tidak terlalu signifikan untuk melihat pengaruh variabel independen terhadap variabel dependen.

5.3. *Bootstrap Aggregation*

Karena data yang kurang dan variabel independen yang tidak terlalu berpengaruh terhadap variabel dependen maka salah satu solusinya adalah dengan melakukan *bootstrap aggregating*. Kelebihan *bootstrap aggregating* dalam mengatasi masalah-masalah tersebut karena *bootstrap aggregating* melakukan *resampling bootstrap* agar data menjadi lebih banyak dan *bootstrap aggregating* dapat menghitung nilai *variable importance* untuk memilih variabel independen yang berpengaruh terhadap variabel dependen. Dalam melakukan *bagging* di R Studio, dibutuhkan beberapa *library* untuk membuat model bagging, yang meliputi *caret* untuk *fitting* model secara keseluruhan, *rpart* untuk membuat pohon keputusan, dan *ipred* untuk menghasilkan *bagging decision tree*. Awalnya, peneliti harus membuat model *bagging*. Pada R Studio, ini dapat dicapai dengan menggunakan fungsi `bagging()` dari *library ipred*. Perlu dicatat bahwa peneliti memilih untuk menggunakan 235 sampel bootstrap untuk membangun model bagging dan mengatur coob ke TRUE untuk mendapatkan estimasi kesalahan *out-of-bag*. Selain itu, peneliti menggunakan konfigurasi tertentu dalam fungsi `rpart.control()`.

- `minsplit = 2`. Artinya mengindikasikan bahwa model membutuhkan setidaknya 2 pengamatan di dalam sebuah node untuk melakukan split
- `cp = 0`. Parameter ini, yang dikenal sebagai *complexity parameter*, diatur ke nilai 0, menunjukkan tidak mewajibkan model untuk meningkatkan keseluruhan kecocokan oleh berapapun sebelum melakukan split

Pada dasarnya, kedua parameter ini memungkinkan pohon-pohon individual untuk tumbuh sangat dalam, menghasilkan pohon-pohon dengan variabilitas tinggi namun bias yang minimal. Selanjutnya, ketika menerapkan metode *bagging*, peneliti dapat mengurangi variabilitas dari model akhir sambil tetap mempertahankan bias yang rendah.

Berdasarkan output model, estimasi RMSE out-of-bag adalah 7.1954, yang mewakili perbedaan rata-rata antara nilai yang diprediksi dan diamati dari jumlah kecelakaan.

Pada pembentukan model kali ini, berdasarkan pada hasil nilai RMSE terkecil. Berikut adalah beberapa percobaan model yang dilakukan peneliti untuk memperoleh model terbaik kali ini

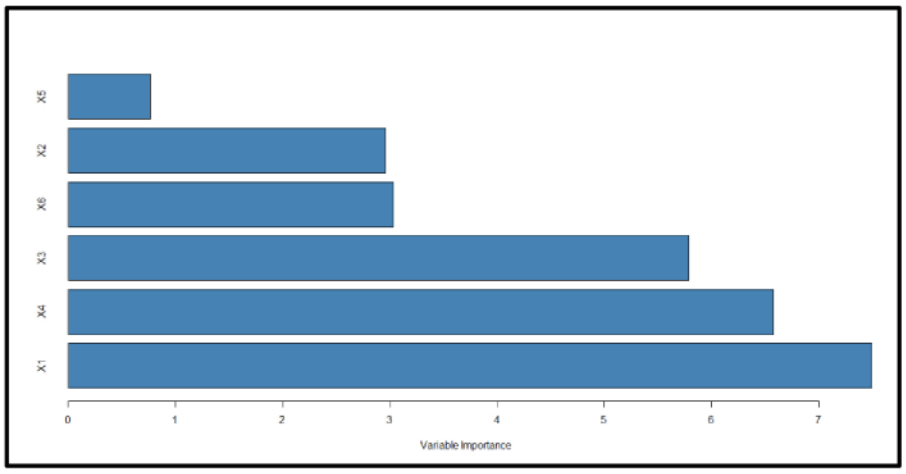
Tabel 5.4 Nilai RMSE untuk beberapa model

Bootstrapped samples	RMSE
220	7,233
230	7,2179
235	**7,1954**
240	7,1965
245	7,2067

Dari beberapa model tersebut dapat dilihat bahwa model dengan 235 *bootstrappped sample* menghasilkan nilai RMSE terkecil. Oleh sebab itu, peneliti memilih model dengan 235 *bootstrapped sample* untuk melanjutkan penelitian ke tahap-tahap selanjutnya. Meskipun *bootstrapped sample* lebih banyak tidak berarti menghasilkan nilai RMSE yang lebih kecil dari *bootstrapped sample* yang lebih sedikit. Mengapa semakin banyak *bootstrapped sample* tidak menjadikan nilai RMSE yang lebih rendah dapat dicari penyebabnya oleh penelitian selanjutnya.

5.4. *Variable Importance*

Meskipun *bagging* model secara umum memberikan prediksi yang lebih akurat dibandingkan dengan pohon keputusan individual, menginterpretasi dan memvisualisasikan hasil dari *bagging* model bisa menjadi tantangan. Namun, dapat divisualisasikan signifikansi variabel prediktor dengan menghitung total pengurangan RMSE akibat pembagian berdasarkan suatu prediktor tertentu, yang dirata-ratakan di semua pohon. Nilai yang lebih tinggi menunjukkan kepentingan yang lebih besar dari prediktor tersebut. Penulis bertujuan untuk membuat plot *variable importance* untuk model *bagging* menggunakan fungsi `varImp()` dari *library caret*.



Gambar 5.8 *Variable Importance* Plot

Terlihat bahwa $X1$, atau usia, menjadi variabel prediktor paling signifikan dalam model, diikuti oleh $X4$, atau harga, sebagai variabel prediktor kedua paling penting, dan $X3$, atau merek, sebagai variabel prediktor ketiga yang paling signifikan dalam model. Nilai *variable importance* untuk setiap variabel ditunjukkan dalam tabel berikut.

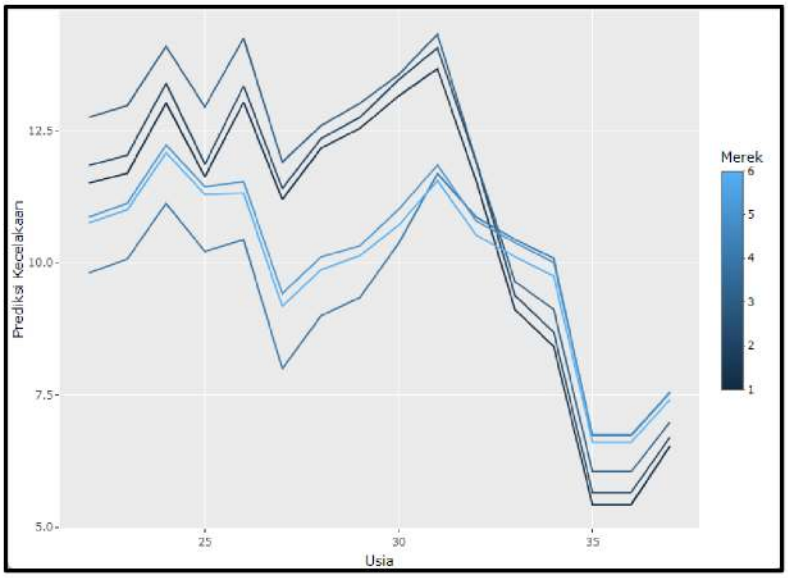
Tabel 5.5 Nilai *Variable Importance*

Variabel	Nilai
$X1$	7,493278
$X2$	2,96043
$X3$	5,787732
$X4$	6,577742
$X5$	0,766145
$X6$	3,028378

5.5. *Partial Dependence Plots*

Partial dependence plots merupakan alat untuk mengilustrasikan hubungan antara fitur tertentu dan hasil prediksi dari model pembelajaran mesin, tetapi dengan menjaga fitur lain tetap konstan. Berdasarkan analisis *variable importance*, akan dilihat hubungan antara variabel $X1$ dan $X4$ dengan hasil prediksi dari model pembelajaran mesin, serta mempertimbangkan kategori $X3$. Variabel $X1$ dan $X4$ dipilih karena keduanya memiliki nilai *variable importance* terbesar yaitu sebesar 7,493278 dan 6,577742. Selain itu, Pembuatan *partial dependence plots* berdasarkan $X3$ karena variabel karegorik yang memiliki nilai *variabel importance* terbesar adalah variabel $X3$ yaitu sebesar 5,787732. *Partial dependence plots*

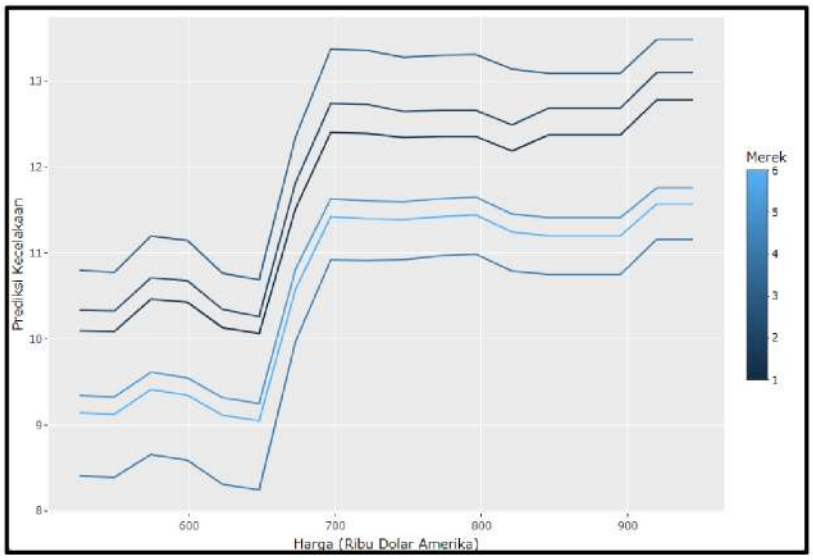
dibentuk untuk variabel $X1$ dan $X4$ dengan hasil prediksi dari model *bagging*, juga mempertimbangkan kategori $X3$, dengan menggunakan fungsi `partial()` dari *library pdp*.



Gambar 5.9 *Partial Dependence Plot* $X1$ dengan Prediksi Kecelakaan Berdasarkan $X3$

Korelasi antara variabel $X1$ dengan prediksi kecelakaan mengalami fluktuasi. Fluktuasi dalam prediksi kecelakaan muncul ketika $X1$, atau usia pembalap, berada dalam rentang usia 22 hingga 31 tahun. Fluktuasi ini dapat disebabkan oleh ketidakstabilan emosional para pembalap yang mengakibatkan terkadang mereka berada pada fase emosional yang tinggi sehingga terjadi kecelakaan akibat tidak berhati-hati dan terkadang mereka berada pada fase emosional yang rendah sehingga mereka lebih berhati-hati untuk mencegah kecelakaan. Selanjutnya, terjadi penurunan signifikan dalam prediksi kecelakaan hingga usia 36 tahun. Penurunan ini dapat dikaitkan dengan kematangan emosional para pembalap antara usia 32 dan 36 tahun, sehingga para pembalap yang berusia pada rentang tersebut lebih hati-hati untuk menghindari kecelakaan. Terakhir, terjadi sedikit peningkatan dalam prediksi kecelakaan pada usia 37 tahun. Pada usia tersebut, para pembalap menuju masa pensiun akibat berkurangnya konsentrasi dan stamina fisik, yang dapat menyebabkan peningkatan prediksi kecelakaan pada usia ini. Selain itu, mulai dari usia 33 hingga 37 tahun, para pembalap dengan motor merek 1, 2, dan 3 mengalami prediksi kecelakaan yang lebih rendah dibandingkan dengan pembalap dengan motor merek 4, 5, dan 6. Hal ini menunjukkan bahwa motor merek 1, 2, dan

3 lebih sedikit terjadi kecelakaan ketika digunakan oleh pembalap yang berusia sekitar antara 33 dan 37 tahun. Sebaliknya, motor merek 4, 5, dan 6 lebih sedikit terjadi kecelakaan ketika digunakan oleh pembalap di bawah usia 33 tahun.



Gambar 5.10 *Partial Dependence Plot* $X4$ dengan Prediksi Kecelakaan Berdasarkan $X3$

Korelasi antara variabel $X4$ dan prediksi kecelakaan cenderung stabil pada interval-interval tertentu. Terutama, ketika harga motor ($X4$) berada dalam kisaran 525 hingga 665 ribu dolar Amerika dan 693 hingga 945 ribu dolar Amerika, prediksi kecelakaan stabil dengan memperhatikan merek motor ($X3$). Konsistensi dalam prediksi kecelakaan dalam rentang harga ini dapat disebabkan oleh performa motor yang relatif seragam, menghasilkan karakteristik yang serupa ketika dikendarai. Namun, terjadi lonjakan signifikan dalam prediksi kecelakaan ketika harga motor berkisar antara 665 hingga 693 ribu dolar Amerika. Lonjakan ini disebabkan oleh perbedaan yang signifikan dalam spesifikasi motor, yang berpotensi membuat motor lebih bertenaga dan sulit dikendalikan oleh para pembalap. Selain itu, hubungan antara variabel $X3$ (merek motor) dan prediksi kecelakaan menghasilkan variasi yang berbeda di antara merek motor. Namun, ketika variabel $X3$ dan $X4$ dianalisis bersama, mereka menghasilkan pola prediksi kecelakaan yang hampir sama di berbagai merek motor.

5.6. Prediksi Jumlah Kecelakaan

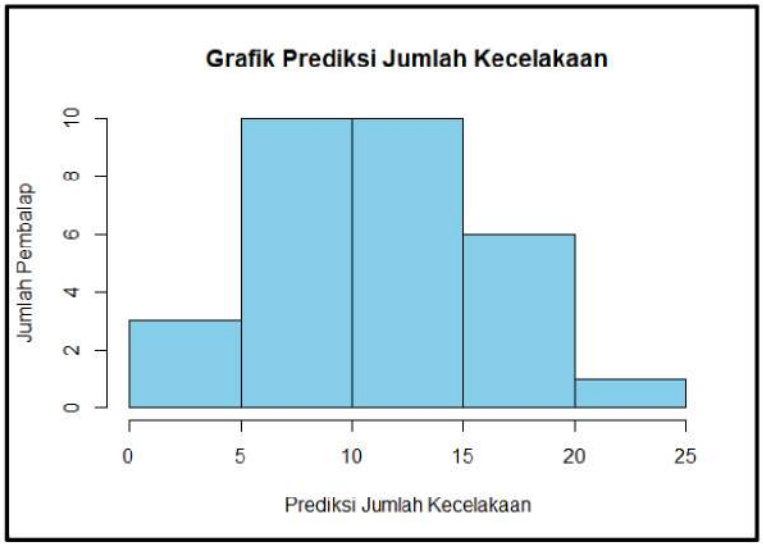
Prediksi jumlah kecelakaan dilakukan untuk melanjutkan penelitian pada tahap perkiraan kerugian. Setelah didapatkan model *bagging* terbaik, maka jumlah

kecelakaan dapat diprediksi dengan model tersebut.. Berikut adalah ringkasan prediksi jumlah kecelakaan.

Tabel 5.6 Statistika Deskriptif Prediksi Jumlah Kecelakaan

Minimum	Rata-rata	Maksimum	Deviasi Standar
3,79	10,95	21,42	4,7

Pada prediksi jumlah kecelakaan, pembalap mengalami kecelakaan sebanyak 3 hingga 21 kali dalam satu musim balapan. Selain itu, selama satu musim balap MotoGP, prediksi jumlah kecelakaan rata-rata adalah sekitar 11 kecelakaan dengan standar deviasi sebesar 4,7 artinya prediksi jumlah kecelakaan bervariasi sekitar 4,7 kali kecelakaan dari nilai rata-rata prediksi jumlah kecelakaan. Selanjutnya, akan dibuat grafik untuk prediksi jumlah kecelakaan agar dapat dilihat lebih detail mengenai karakteristik data prediksi jumlah kecelakaan.



Gambar 5.11 Histogram Prediksi Jumlah Kecelakaan

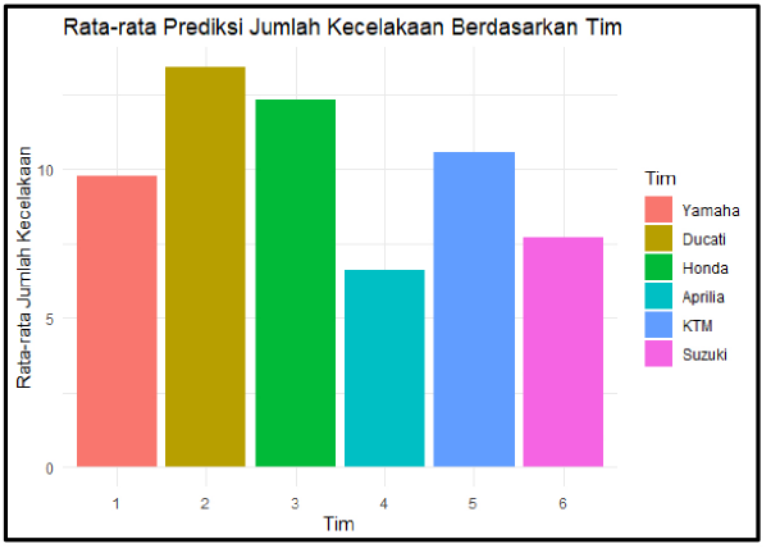
Pada histogram prediksi jumlah kecelakaan terlihat bahwa diperkirakan jumlah kecelakaan yang dialami oleh pembalap berkisar antara 3 hingga 21 kali dalam satu musimnya. Prediksi jumlah kecelakaan paling sering terjadi berkisar antara 5 hingga 15 kali yaitu sebanyak lebih dari 20 pembalap. Prediksi jumlah kecelakaan paling jarang berkisar antara 20 hingga 21 kali, yaitu hanya 1 pembalap saja. Bila dihubungkan kembali dengan karakteristik data jumlah kecelakaan, atau Y, rata-rata jumlah kecelakaan terjadi sebanyak 11 kali. Sehingga wajar apabila prediksi jumlah kecelakaan paling sering terjadi berkisar antara 5 hingga 15 kali dalam satu musim.

Selanjutnya akan dibuat prediksi jumlah kecelakaan pada masing-masing tim MotoGP. Prediksi jumlah kecelakaan tersebut berguna untuk tahap selanjutnya pada penelitian ini yaitu perhitungan perkiraan kerugian pada masing-masing tim MotoGP. Selain itu, prediksi jumlah kecelakaan tersebut dapat berguna pada masing-masing tim untuk mengembangkan teknologi dan penyesuaian performa pada motor agar dapat mengurangi risiko kecelakaan. Berikut adalah statistika deskriptif prediksi jumlah kecelakaan pada masing-masing tim MotoGP.

Tabel 5.7 Statistika Deskriptif Prediksi Jumlah Kecelakaan Berdasarkan Tim

Tim	Minimum	Rata-rata	Maksimum	Deviasi Standar
Yamaha	3,79	9,75	21,4	6,92
Ducati	4,49	13,4	17,7	4,21
Honda	7,89	12,3	17,6	4,46
Aprilia	4,37	6,6	10,4	3,31
KTM	9,27	10,6	12,7	1,51
Suzuki	5,21	7,7	10,8	2,84

Tim yang diprediksi mengalami kecelakaan paling sedikit dan paling banyak adalah tim Yamaha yaitu sebanyak 4 kali hingga 21 kali. Sedangkan rata-rata kecelakaan masing-masing tim diprediksi berkisar antara 7 hingga 13 kali dalam satu musimnya. Selanjutnya, akan dibuat grafik untuk rata-rata prediksi jumlah kecelakaan pada masing-masing tim MotoGP agar dapat dilihat lebih detail mengenai karakteristik datanya.



Gambar 5.12 Barplot Rata-Rata Prediksi Jumlah Kecelakaan Berdasarkan Tim

Pada barplot rata-rata prediksi jumlah kecelakaan berdasarkan tim diketahui bahwa tim Ducati memiliki rata-rata jumlah kecelakaan paling sering yaitu sebanyak 13 kali. Hal ini dapat disebabkan oleh kebanyakan tim Ducati menggunakan motor yang memiliki harga lebih dari 693 ribu dolar Amerika. Sesuai dengan penjelasan dalam *partial dependence plots* mengenai motor yang memiliki harga lebih dari 693 ribu dolar Amerika berpotensi membuat motor lebih bertenaga dan sulit dikendalikan oleh para pembalap. Sehingga tim Ducati lebih sering mengalami kecelakaan dibandingkan dengan tim lainnya.

5.7. Perkiraan Kerugian

Dalam perkiraan kerugian ini, diasumsikan bahwa setiap kecelakaan mengakibatkan kerugian yang sama dengan nilai sepeda motor yang terlibat. Artinya pada perhitungan perkiraan kerugian kali ini dilakukan dengan cara mengkalikan jumlah kecelakaan dengan harga motor yang terlibat dalam kecelakaan tersebut. Hal ini dilakukan untuk memperkirakan skenario terburuk. Selain itu, hal ini memberikan indikasi tentang potensi kewajiban kerugian yang harus disiapkan oleh sebuah tim jika skenario terburuk ini terjadi. Sedangkan untuk harga kendaraan yang digunakan berupa harga yang telah diakumulasikan berdasarkan bagian-bagian dari motor tersebut. Terdiri dari harga mesin, ban, elektronik, kaki-kaki, dan *body*. Pada harga elektronik, kaki-kaki, dan *body* memiliki harga yang sama untuk setiap motornya karena setiap motor memiliki spesifikasi yang sama untuk bagian-bagian tersebut. Harga yang sama untuk bagian-bagian tersebut dikarenakan peraturan yang membuat mereka terpaksa untuk memakai produk yang sama untuk bagian-bagian tersebut. Sedangkan harga mesin dihitung berdasarkan merek motor, usia kendaraan motor, dan tim. Tim pabrikan memiliki dana untuk mesin motor lebih besar dibandingkan dengan tim satelit. Karena tim pabrikan memperoleh dana langsung dari perusahaan merek motor untuk mesin motor, sedangkan tim satelit hanya memperoleh sebagian dana yang diberikan oleh perusahaan merek motor. Selain itu, untuk motor yang telah berusia 2 tahun memiliki harga mesin yang lebih murah karena mesin motor yang digunakan untuk motor yang berusia 2 tahun adalah mesin motor yang sama yang digunakan tahun lalu. Sehingga biaya pengembangan atau pembuatan mesin tersebut lebih murah dibandingkan dengan motor yang berusia 1 tahun karena

menggunakan mesin motor terbaru yang masih perlu pengembangan. Berikut adalah statistik deskriptif dari perkiraan kerugian.

Tabel 5.8 Statistika Deskriptif Perkiraan Kerugian dalam Juta Dolar Amerika

Minimum	Rata-rata	Maksimum	Deviasi Standar
2,07	7,68	13,19	3,6

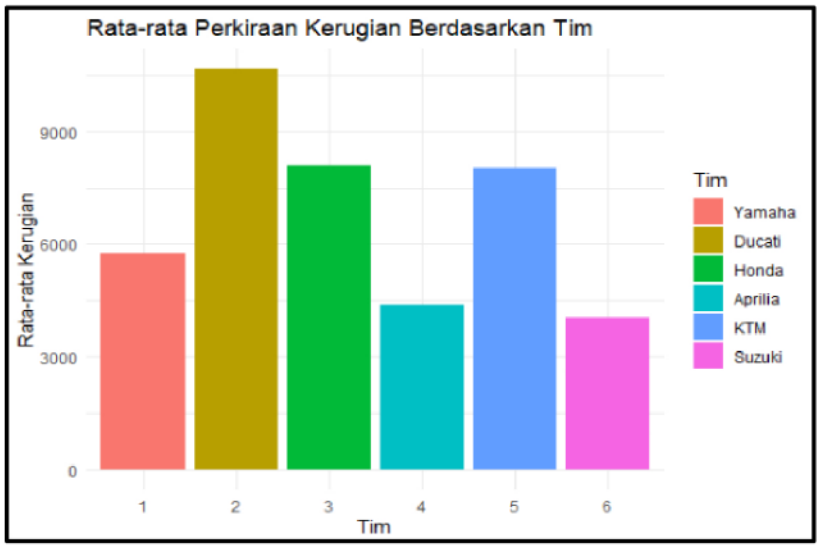
Perkiraan kerugian menghasilkan nilai kerugian potensial minimum dan maksimum sebesar 2,07 hingga 13,19 juta dolar Amerika. Hal ini menunjukkan bahwa sebuah tim harus memiliki minimal 2,07 juta dolar Amerika untuk menutupi kerugian potensial. Kerugian maksimum atau skenario terburuk yang mungkin dihadapi oleh sebuah tim adalah sebesar 13,19 juta dolar Amerika. Selain itu, selama satu musim balap MotoGP, perkiraan kerugian rata-rata adalah sekitar 7,68 juta dolar Amerika dengan standar deviasi sebesar 3,6 artinya perkiraan kerugian bervariasi sekitar 3,6 juta dolar Amerika dari nilai rata-rata perkiraan kerugian. Bila dihubungkan kembali dengan karakteristik data jumlah kecelakaan, atau Y, dan harga kendaraan, atau X4, rata-rata jumlah kecelakaan terjadi sebanyak 11 kali dan rata-rata harga kendaraan sebesar 692,3 ribu dolar Amerika. Sehingga wajar apabila rata-rata perkiraan kerugian adalah sebesar 7,68 juta dolar Amerika.

Selanjutnya untuk memudahkan interpretasi pada perkiraan kerugian kali ini, akan dihitung perkiraan kerugian yang mungkin dialami tim-tim MotoGP. Perhitungan tersebut berguna untuk memudahkan masing-masing tim mempersiapkan kerugian potensial. Karena kerugian yang timbul akibat kecelakaan akan ditanggung oleh masing-masing tim. Berikut adalah statistika deskriptif perkiraan kerugian untuk masing-masing tim MotoGP.

Tabel 5.9 Statistika Deskriptif Perkiraan Kerugian Berdasarkan Tim dalam Juta Dolar Amerika

Tim	Minimum	Rata-rata	Maksimum	Deviasi Standar
Yamaha	2,07	5,75	11,67	3,76
Ducati	4,24	10,66	13,19	3
Honda	4,56	8,08	12,54	3,42
Aprilia	2,91	4,39	6,92	2,2
KTM	6,82	8,01	8,66	0,87
Suzuki	2,73	4,04	5,66	1,49

Tim yang diperkirakan mengalami kerugian paling sedikit adalah tim Yamaha yaitu sebesar 2,07 juta dolar Amerika. Sementara tim yang diperkirakan mengalami kerugian paling banyak adalah tim Ducati yaitu sebesar 13,19 juta dolar Amerika. Hal tersebut terjadi karena tim Ducati menggunakan motor yang memiliki harga yang lebih mahal dibandingkan dengan motor dari tim lainnya. Oleh karena itu, ketika motor tim Ducati mengalami kecelakaan maka kerugian yang diperoleh pasti lebih banyak dibandingkan dengan tim lainnya. Sedangkan rata-rata kerugian masing-masing tim diperkirakan berkisar antara 4,04 hingga 10,66 juta dolar Amerika. Selanjutnya, akan dibuat grafik untuk rata-rata prediksi jumlah kecelakaan pada masing-masing tim MotoGP agar dapat dilihat lebih detail mengenai karakteristik datanya.



Gambar 5.13 Barplot Rata-Rata Perkiraan Kerugian Berdasarkan Tim

Pada barplot rata-rata perkiraan kerugian berdasarkan tim diketahui bahwa tim Ducati memiliki rata-rata perkirakan kerugian paling banyak yaitu sebesar 10,66 juta dollar Amerika. Hal tersebut terjadi karena tim Ducati menjadi tim yang mengalami rata-rata kecelakaan paling sering dibandingkan dengan tim yang lainnya. Selain itu, harga motor tim Ducati juga lebih mahal dibandingkan dengan motor dari tim lainnya. Dengan harga motor tim Ducati yang mahal serta menjadi tim dengan rata-rata kecelakaan paling sering sehingga tim Ducati menjadi tim dengan rata-rata perkiraan kerugian paling besar.

BAB VI
PENUTUP

6.1. Kesimpulan

1. $X1$, atau usia, menjadi variabel prediktor paling signifikan dalam model, diikuti oleh $X4$, atau harga, sebagai variabel prediktor kedua paling penting, $X3$, atau merek, sebagai variabel prediktor ketiga yang paling signifikan dalam model, $X6$, atau tim, sebagai variabel prediktor keempat yang paling signifikan dalam model, $X2$, atau asal, sebagai variabel prediktor kelima yang paling signifikan dalam model, $X5$, atau usia kendaraan, sebagai variabel prediktor terakhir yang paling signifikan.
2. Model bagging dibentuk dengan menggunakan 235 sampel bootstrap, model membutuhkan setidaknya 2 pengamatan di dalam sebuah node untuk melakukan split, dan tidak mewajibkan model untuk meningkatkan keseluruhan kecocokan oleh berapapun sebelum melakukan split. Berdasarkan output model, estimasi RMSE out-of-bag adalah 7.1954, yang mewakili perbedaan rata-rata antara nilai yang diprediksi dan diamati dari jumlah kecelakaan. Dalam perkiraan kerugian ini, diasumsikan bahwa setiap kecelakaan mengakibatkan kerugian yang sama dengan nilai sepeda motor yang terlibat. Perkiraan kerugian menghasilkan nilai kerugian berikisar antara 2,07 hingga 13,19 juta dolar Amerika dengan perkiraan kerugian rata-rata adalah sekitar 7,68 juta dolar Amerika dengan standar deviasi sebesar 3,6 artinya perkiraan kerugian bervariasi sekitar 3,6 juta dolar Amerika dari nilai rata-rata perkiraan kerugian.

6.2. Saran

Pada penelitian kali ini, pengaturan *bootstrapped sample yang* lebih banyak pada model *bagging* tidak berarti menghasilkan nilai RMSE yang lebih kecil dari *bootstrapped sample* yang lebih sedikit. Semakin banyak *bootstrapped sample* tidak menjadikan nilai RMSE yang lebih rendah tersebut dapat dicari penyebabnya oleh penelitian selanjutnya.

DAFTAR PUSTAKA

Abdallah, I., Tatsis, K., & Chatzi, E. (2020). Unsupervised local cluster-weighted bootstrap aggregating the output from multiple stochastic simulators. *Realibility Engineering and System Safety*, 1-14.

Alam, M. R., & Mutaqin, A. K. (2023). Pemodelan Distribusi Poisson-Sujatha pada Data Frekuensi Asuransi Kendaraan Bermotor di Indonesia. *Jurnal Riset Statistika (JRS)*, 71-78.

Alfaridzi, S. I., & Prabowo, A. (2023). Penentuan Tarif Premi pada Asuransi Kendaraan dengan Besar Klaim Berdistribusi Eksponensial dan Gamma. *Premium Insurance Business Journal*, 29-41.

Aon. (2024, June 9). *Sports, Recreation & Entertainment: Aon*. Diambil kembali dari Aon: https://www.aon.com/middle-east/risk-solutions/sports-recreation-entertainment

Banaran, I. (2022, July 30). *MOTOGP RUSAK, SIAPA YANG BAYAR: Youtube*. Diambil kembali dari YouTube: https://www.youtube.com/watch?v=PFjs9qySG-c

Bowers, N., Gerber, H., Hickman, J., Jones, D., & Nesbitt, C. (1987). *Actuarial Mathematics.* Illinois: The Society of Actuaries.

Breiman, L. (1996, 08 01). Bagging predictors. *Machine Learning, 24*(2), 123-140.

Breiman, L. (2001). Random Forests. *Machine Learning*, 5-32.

Brodeur, Z. P., Herman, J. D., & Steinschneider, S. (2020). Bootstrap Aggregation and Cross-Validation Methods to Reduce Overfitting in Reservoir Control Policy Search. *Water Resources Research*, 1-9.

Cavdar, S. C., Aydin, A. D., & Karaaslan, M. F. (2015). A Geometric Interpretation and Comparison of the Methods Of Ordinary Least Square (OLS) and Bivariate Lagrange Interpolation. *Mathematical Theory and Modeling*, 1-17.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap.* New York: Chapman and Hall/CRC.

Elish, M. O. (2009). Improved estimation of software project effort using multiple additive regression trees. *Expert Systems with Applications*, 10774-10778.

Fitriani, R., & Gunardi. (2020). Implementasi Metode Bayes pada Perhitungan Premi Asuransi Kendaraan Bermotor. *Journal of Fundamental Mathematics*, 112-123.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 1189-1232.

Ghose, S. B., & Akanchha, A. (2020, August). A Study on Customer's Preference for Personal Vehicle Insurance Provided by Different Insurance Companies in Bengaluru. *International Journal of Management, 5*, 69-80. doi:10.5281/zenodo

Greene, M. R. (1968). *Risk and insurance.* Cincinnati: South-western Pub. Co.

Jang, D., & Lee, B. (2022). When Machine Learning Meets Social Science: A Comparative Study of Ordinary Least Square, Stochastic Gradient Descent, and Support Vector Regression for Exploring the Determinants of Behavioral Intentions to Tuberculosis Screening. *Asian Communication Research*, 101-118.

Kania, D. S., & Mutaqin, A. K. (2022). Perhitungan Premi Risiko Asuransi Kendaraan Bermotor Berdasarkan Data Frekuensi dan Besar Klaim. *Journal Riset Statistika (JRS)*, 111-118.

Kuhn, M. (2019, 03 27). *Caret : Variable Importance.* Diambil kembali dari Topepo Github io: https://topepo.github.io/caret/variable-importance.html

Lau, M., Kress, S., Schikowski, T., & Schwender, H. (2023). Efficient gene-environment interaction testing through bootstrap aggregating. *scientific reports*, 1-15.

Maindonald, J., & Braun, W. J. (2010). Data Analysis and Graphics Using R: An Example-Based Approach. Dalam J. Maindonald, & W. J. Braun, *Cambridge Series in Statistical and Probabilistic Mathematics* (hal. 43-76). Cambridge: Cambridge University Press.

Mehr, R. I., & Cammack, E. (1972). *Principles of insurance.* Homewood: Irwin.

MotoGP. (2022, December 1). *2022 crash report: the finding*. Diambil kembali dari MotoGP: https://www.motogp.com/en/news/2022/12/01/2022-crash-report-the-findings/183009

Pinch, P., & Reimer, S. (2017). MotoGP and heterogeneous design. Dalam J. Spinney, S. Reimer, & P. Pinch, *Mobilising Design* (hal. 117-129). New York: Routledge. Dipetik 10 3, 2023

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 21-45.

Pradipta, R. S. (2022, May 27). *Resmi: Perusahaan Asuransi Prima Menjadi Sponsor Utama Baru Pramac: ridertua.com*. Diambil kembali dari ridertua.com: https://ridertua.com/2022/05/27/perusahaan-asuransi-prima-menjadi-sponsor-utama-baru-pramac/

Putri, D. L., Mariani, S., & Sunarmi. (2021). Peningkatan Ketepatan Klasifikasi Model Regresi Logistik Biner dengan Metode Bagging (Bootstrap Aggregating). *Indonesia Journal of Mathematifcs and Natural Sciences*, 61-72.

Rahmah, M. Z., & Mutaqin, A. K. (2021). Penerapan Metode Limited-Fluctuation Credibility dalam Menentukan Premi Murni pada Asuransi Kendaraan Bermotor di PT XYZ. *Indonesian Journal of Applied Statistics*, 126-139.

Resti, Y., Irsan, C., Latif, J. F., Yani, I., & Dewi, N. R. (2023, April). A Bootstrap-Aggregatin in Random Forest Model for Classification of Corn Plant Diseases and Pests. *Science and Technology Indonesia, 8*(2), 288-297. Dipetik 10 17, 2023

Samudra, M. A. (2023, September 11). *Terungkap, Ternata Segini Biaya Perbaikan Motor MotoGP Ketika Kecelakaan: gridoto.com*. Diambil kembali dari gridoto.com: https://www.gridoto.com/read/223888901/terungkap-ternyata-segini-biaya-perbaikan-motor-motogp-ketika-kecelakaan?page=all

Shin, J., & Im, C.-H. (2020). Performance Improvement of Near-Infrared Spectoscopy-Based-Brain-Computer Interface Using Regularized Linear Discriminant Analysis Ensemble Classifier Based on Bootstrap Aggregating. *Neuroscience*, 1-11.

Terefe, E. M. (2023, February 21). Tree-Based Machine Learning Methods For Vehicle Insurance Claims Size Prediction. *arXiv*, 1-20. Dipetik 10 15, 2023

Usman, U., Garba, N., Zoramawa, A. B., & Usman, H. (2020). Assesing the Performance of Ordinary Least Square and Kernel Regression. *Continental J. Applied Sciences*, 14-23.

Williams, C. A., & Heins, R. M. (1989). *Risk management and insurance* (6th ed.). New York: McGraw-Hill.

Wulandari, A., & Arsyad, R. (2023, Desember). Wonderful Indonesia dalam Pelaksanaan Moto Grand Prix Mandalika Tahun 2022. *Jurnal Hubungan Internasional Peradaban, 2*(2), 91-104. Dipetik 10 12, 2023

Yoon, H.-J., Klasky, H. B., Gounley, J. P., Alawad, M., Gao, S., Durbin, E. B., . . . Tourassi, G. D. (2020). Accelerated training of bootstrap aggregation-based deep information extraction systems from cancer pathology reports. *Journal of Biomedical Informatics*, 1-9.

Zhang, C., & Ma, Y. (2012). *Ensemble machine larning: methods and applications.* New York: Springer.

LAMPIRAN

Lampiran 1 Sintaks R

```
#Load Library
library(dplyr)       #for data wrangling
library(e1071)       #for calculating variable importance
library(caret)       #for general model fitting
library(rpart)       #for fitting decision trees
library(ipred)       #for fitting bagged decision trees
library(plotly)      #for create plots
library(pdp)         #for create partial dependence plots

#Input Data
data <- read.delim("clipboard")
data

#Statistika Deskriptif
summary(data)
sqrt(var(data$X1))
sqrt(var(data$X4))
sqrt(var(data$Y))

#Grafik
#Membuat grafik batang X1
ggplot(data, aes(x = factor(X1), fill = factor(X1))) +
  geom_bar() +
  labs(title = "Grafik Usia", x = "Usia", y = "Frekuensi") +
  theme_minimal()

# Mengubah nilai X2 sesuai dengan label yang diberikan
data$X2_label <- factor(data$X2, levels = c(1, 2, 3, 4),
                        labels = c("Afrika", "Eropa", "Australia",
"Asia"))
# Membuat plot bar dengan label yang telah ditentukan
ggplot(data, aes(x = factor(X2), fill = X2_label)) +
  geom_bar() +
  labs(title = "Grafik Asal", x = "Asal", y = "Frekuensi", fill
="Keterangan") +
  scale_fill_manual(values = c("Afrika" = "skyblue", "Eropa" =
"green", "Australia" = "orange", "Asia" = "red")) +
```

```
  theme_minimal()

# Mengubah nilai X3 sesuai dengan label yang diberikan
data$X3_label <- factor(data$X3, levels = c(1, 2, 3, 4, 5, 6),
                        labels = c("Yamaha", "Ducati", "Honda",
"Aprilia", "KTM", "Suzuki"))
# Membuat plot bar dengan label yang telah ditentukan
ggplot(data, aes(x = factor(X3), fill = X3_label)) +
  geom_bar() +
  labs(title = "Grafik Merek", x = "Merek", y = "Frekuensi", fill
= "Keterangan") +
  scale_fill_manual(values = c("Yamaha" = "skyblue", "Ducati" =
"green", "Honda" = "orange", "Aprilia" = "red", "KTM" = "purple",
"Suzuki" = "yellow")) +
  theme_minimal()

# Membuat histogram untuk X4
ggplot(data, aes(x = X4)) +
  geom_histogram(binwidth = 50, fill = "skyblue", color = "black")
+
  labs(title = "Grafik Harga", x = " Harga (Ribu Dolar Amerika)",
y = "Frekuensi") +
  theme_minimal()

# Mengubah nilai X5 sesuai dengan label yang diberikan
data$X5_label <- factor(data$X5, levels = c(1, 2),
                        labels = c("Tahun Pertama", "Tahun Kedua"))
# Membuat plot bar untuk X5 dengan label yang telah ditentukan
ggplot(data, aes(x = X5, fill = X5_label)) +
  geom_bar() +
  labs(title = " Grafik Usia Kendaraan", x = "Usia Kendaraan", y =
"Frekuensi", fill = "Keterangan") +
  scale_fill_manual(values = c("Tahun Pertama" = "skyblue", "Tahun
Kedua" = "green")) +
  theme_minimal()

# Mengubah nilai X6 sesuai dengan label yang diberikan
data$X6_label <- factor(data$X6, levels = c(1, 2),
                        labels = c("Pabrikan", "Satelit"))
```

```
# Membuat plot bar untuk X6 dengan label yang telah ditentukan
ggplot(data, aes(x = X6, fill = X6_label)) +
  geom_bar() +
  labs(title = " Grafik Tim", x = "Tim", y = "Frekuensi", fill =
"Keterangan") +
  scale_fill_manual(values = c("Pabrikan" = "orange", "Satelit" =
"red")) +
  theme_minimal()

# Histogram untuk variabel Y
hist(data$Y,
     main = " Grafik Jumlah Kecelakaan",
     xlab = "Jumlah Kecelakaan",
     ylab = "Jumlah Pembalap",
     col = "skyblue",
     border = "black")

#fit simple linear regression model
modelOLS <- lm(Y ~ X1+X2+X3+X4+X5+X6, data=data)
#view model summary
summary(modelOLS)

#make this example reproducible
set.seed(1)
#fit the bagged model
bag <- bagging(
  formula = Y ~ X1+X2+X3+X4+X5+X6,
  data = data,
  nbagg = 235,
  coob = TRUE,
  control = rpart.control(minsplit = 2, cp = 0)
)
#see the bagged model
bag

#calculate variable importance
VI         <-         data.frame(var=names(data[,c(1,2,3,4,5,6)]),
imp=varImp(bag))
#sort variable importance descending
```

```
VI_plot <- VI[order(VI$Overall, decreasing=TRUE),]
#visualize variable importance with horizontal bar plot
barplot(VI_plot$Overall,
        names.arg=rownames(VI_plot),
        horiz=TRUE,
        col='steelblue',
        xlab='Variable Importance')

# Create a PDP for Age, considering the Brand factor
PDP_Usia_Merek_Prediksi_Kecelakaan <- pdp::partial(bag, pred.var =
c("X1","X3"), grid.resolution = 18)
Merek <- PDP_Usia_Merek_Prediksi_Kecelakaan$X3
pdp_usia_merek_prediksi_kecelakaan <-
  ggplot(PDP_Usia_Merek_Prediksi_Kecelakaan, aes(X1, yhat, colour
= Merek)) +
  ylab("Prediksi Kecelakaan")+
  xlab("Usia")+
  geom_line()
plotly::ggplotly(pdp_usia_merek_prediksi_kecelakaan)

# Create a PDP for Price, considering the Brand factor
PDP_Harga_Merek_Prediksi_Kecelakaan <- pdp::partial(bag, pred.var
= c("X4", "X3"), grid.resolution = 18)
Merek <- PDP_Harga_Merek_Prediksi_Kecelakaan$X3
pdp_harga_merek_prediksi_kecelakaan <-
  ggplot(PDP_Harga_Merek_Prediksi_Kecelakaan, aes(X4, yhat, colour
= Merek)) +
  ylab("Prediksi Kecelakaan")+
  xlab("Harga (Ribu Dolar Amerika)")+
  geom_line()
plotly::ggplotly(pdp_harga_merek_prediksi_kecelakaan)

#Prediksi Jumlah Kecelakaan
data$prediksi_kecelakaan <- predict(bag, newdata = data)
View(data)
summary(data$prediksi_kecelakaan)
sqrt(var(data$prediksi_kecelakaan))
# Histogram untuk Prediksi Jumlah Kecelakaan
hist(data$prediksi_kecelakaan,
```

```
     main = "Grafik Prediksi Jumlah Kecelakaan",
     xlab = "Prediksi Jumlah Kecelakaan",
     ylab = "Jumlah Pembalap",
     col = "skyblue",
     border = "black")
#Statistika deskriptif prediksi jumlah kecelakaan per tim
prediksi_jumlah_kecelakaan_per_tim <- data %>%
  group_by(X3) %>%
  summarise(jumlah_kecelakaan_minimal = min(prediksi_kecelakaan),
            rata_rata_jumlah_kecelakaan                          =
mean(prediksi_kecelakaan),
            jumlah_kecelakaan_maksimal = max(prediksi_kecelakaan),
            std_dev_jumlah_kecelakaan                            =
sqrt(var(prediksi_kecelakaan)))
print(prediksi_jumlah_kecelakaan_per_tim)
#Grafik Prediksi Jumlah Kecelakaan Per Tim
# Mengubah nilai X3 sesuai dengan label yang diberikan
prediksi_jumlah_kecelakaan_per_tim$Tim                           <-
factor(prediksi_jumlah_kecelakaan_per_tim$X3, levels = c(1, 2, 3,
4, 5, 6),
                        labels = c("Yamaha", "Ducati", "Honda",
"Aprilia", "KTM", "Suzuki"))
# Buat barplot menggunakan ggplot2
ggplot(prediksi_jumlah_kecelakaan_per_tim, aes(x = factor(X3), y =
rata_rata_jumlah_kecelakaan, fill = Tim)) +
  geom_bar(stat = "identity") +
  labs(title = "Rata-rata Prediksi Jumlah Kecelakaan Berdasarkan
Tim",
       x = "Tim",
       y = "Rata-rata Jumlah Kecelakaan") +
  theme_minimal()

#Perkiraan Kerugian
data$perkiraan_kerugian <- data$prediksi_kecelakaan*data$X4
View(data)
summary(data$perkiraan_kerugian)
sqrt(var(data$perkiraan_kerugian))
#Statistika deskriptif perkiraan kerugian per tim
perkiraan_kerugian_per_tim <- data %>%
```

```
  group_by(X3) %>%
  summarise(kerugian_minimal = min(perkiraan_kerugian),
            rata_rata_kerugian = mean(perkiraan_kerugian),
            kerugian_maksimal = max(perkiraan_kerugian),
            std_dev_kerugian = sqrt(var(perkiraan_kerugian)))
print(perkiraan_kerugian_per_tim)
#Grafik Perkiraan Kerugian Per Tim
# Mengubah nilai X3 sesuai dengan label yang diberikan
perkiraan_kerugian_per_tim$Tim <-
factor(perkiraan_kerugian_per_tim$X3, levels = c(1, 2, 3, 4, 5, 6),
                                                    labels =
c("Yamaha", "Ducati", "Honda", "Aprilia", "KTM", "Suzuki"))
# Buat barplot menggunakan ggplot2
ggplot(perkiraan_kerugian_per_tim, aes(x = factor(X3), y =
rata_rata_kerugian, fill = Tim)) +
  geom_bar(stat = "identity") +
  labs(title = "Rata-rata Perkiraan Kerugian Berdasarkan Tim",
       x = "Tim",
       y = "Rata-rata Kerugian") +
  theme_minimal()
```