PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN ALGORITMA C5.0 PADA LAJU INFLASI DI INDONESIA

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Matematika
dalam Ilmu Matematika

Oleh :

MUHAMAD CAHYO NUGROHO

NIM : 2008046021

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI
WALISONGOSEMARANG

2024

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama	:	Muhamad Cahyo Nugroho
NIM	:	2008046021
Jurusan/Program Studi	:	Matematika / Matematika

menyatakan bahwa skripsi yang berjudul :

PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN ALGORITMA C5.0 PADA LAJU INFLASI DI INDONESIA

secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 15 Juli 2024
Pembuat pernyataan,

Muhamad Cahyo Nugroho
NIM : 2008046021

KEMENTERIAN AGAMA R.I.
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka (Kampus II) Ngaliyan Semarang
Telp. 024-7601295 Fax. 7615387

PENGESAHAN

Naskah skripsi berikut ini :

Judul : PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN ALGORITMA C5.0 PADA LAJU INFLASI DI INDONESIA
Penulis : Muhamad Cahyo Nugroho
NIM : 2008046021
Jurusan : Matematika

Telah diujikan dalam sidang *tugas akhir* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Matematika.

Semarang, 15 Juli 2024

DEWAN PENGUJI

Penguji I,

Seftina Diyah Miasary, M.Sc.
NIP : 198709212019032010

Penguji II,

Dinni Rahma Oktaviani, M.Si.
NIP : 199410092019032017

Penguji III,

Yulia Romadiastri, S.Si., M.Sc.
NIP : 198107152005012008

Penguji IV,

Agus Wayan Yulianto, M.Sc.
NIP : 198907162019031007

Pembimbing,

Ariska Kurnia Rachmawati, M.Sc
NIP : 198908112019032019

NOTA DINAS

Semarang, 3 Juli 2024

Yth. Ketua Program Studi Matematika
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan,arahan dan koreksi naskah skripsi dengan:

Judul	: PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN ALGORITMA C5.0 PADA LAJU INFLASI DI INDONESIA
Nama	: Muhamad Cahyo Nugroho
NIM	: 2008046021
Jurusan	: Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

Wassalamu'alaikum warahmatullahi wabarakatuh

Pembimbing,

Ariska Kurnia Rachmawati, M.Sc
NIP : 198908112019032019

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillah segala puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, serta shalawat dan salam kepada jujungan besar Nabi Muhammad SAW sehingga penulis dapat menyelesaikan tugas akhir yang berjudul **"Perbandingan Metode Regresi Logistik Biner dan Algoritma C5.0 Pada Laju Inflasi Di Indonesia"** dengan baik.

Skripsi ini dilakukan sebagai salah satu peryaratan yang harus dipenuhi dalam menyelesaikan Program Studi Jurusan Matematika, Fakulas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang. Selama proses penyusunan skripsi ini tidak lepas dari doa, bantuan, bimbingan, motivasi dan peran dari banyak pihak. Sehingga penulis mengucapkan terimakasih kepada :

1. Prof.Dr. H.Musahadi, M.Ag. selaku Dekan Fakultas Sains dan Teknologi.
2. Any Muanalifah, M.Si., Ph.D. selaku Ketua Program Studi Matematika.
3. Seftina Diyah Miasary, M.Sc. selaku Wali Dosen
4. Ariska Kurnia Rachmawati, M.Sc. selaku Dosen Pembimbing skripsi atas bimbingan dan kesabarannya selama menyelesaikan skripsi ini.
5. Seluruh dosen dan Staff Program Studi Matematika di Universitas Islam Negeri Walisongo Semarang yang telah memberikan ilmunya selama delapan semester.
6. Bapak Slamet dan Ibu Ning Asih selaku kedua orang tua yang selalu setia mendoakan, mendukung, menemani serta memberikan motivasi kepada penulis dalam menyelesaikan skripsi ini.

7. Semua kakak laki-laki dan perempuan penulis yang selalu mendoakan dan mendukung hingga selesai penulisan skripsi ini.
8. Semua Adik laki-laki dan perempuan penulis yang selalu mendoakan dan mendukung hingga selesai penulisan skripsi ini
9. Rizanatul Mukharomah selaku teman spesial penulis yang selalu memberikan dukungan, motivasi, kebersamaan dan kasih sayangnya.
10. UKM Saintek Sport yang selalu mendoakan dan mendukung penulis dalam keadaaan apapun.
11. Terima kasih untuk semua teman prodi matematika angkatan 2020 yang tidak bisa penulis sebutkan satu per satu yang selalu menjadi semangat bagi penulis.
12. Semua Pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya skripsi ini.

Semoga kebaikan semuanya menjadi amal ibadah yang diterima dan mendapat pahala yang berlimpah dari Allah SWT. Amiin.

Atas segala kekurangan dan kelemahan dalam skripsi ini penulis mengharapkan saran dan kritik yang membangun. Semoga karya tulis yang sederhana ini dapat menjadi bacaan yang bermanfaat dan dapat dikembangkan bagi peneliti-peneliti selanjutnya.

Wassalamu'alaikum Wr. Wb.

Semarang, 30 Mei 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

ABSTRAK

Judul : **Perbandingan Metode Regresi Logistik Biner dan Algoritma C5.0 Pada Laju Inflasi Di Indonesia**

Penulis : **Muhamad Cahyo Nugroho**

Nim : **2008046021**

Indonesia mengalami inflasi pada kategori inflasi ringan. Meskipun dalam kategori ringan, inflasi tetap memberikan dampak terhadap pertumbuhan ekonomi dan kenaikan harga. Penelitian ini menggunakan metode regresi logistik biner dan algoritma C5.0 sebagai metode klasifikasi. Dalam penelitian ini bertujuan untuk membandingkan metode regresi logistik biner dan algoritma C5.0 pada laju inflasi di Indonesia. Faktor – faktor yang digunakan adalah nilai tukar rupiah, suku bunga, dan jumlah uang beredar dari tahun 2010 hingga 2023. Data *training* dan *testing* yang digunakan adalah 80% dan 20%. Hasil penelitian dengan menggunakan teknik regresi logistik biner menunjukan bahwa variabel independen yang dapat mempengaruhi secara signifikan adalah suku bunga dengan dengan ketepatan klasifikasi 95,83% dan nilai AUC senilai 0,9722. Sedangkan teknik algoritma C5.0 menghasilkan variabel independen suku bunga sebagai variabel terpenting yang artinya mempengaruhi secara signifikan dengan ketepatan klasifikasi 91,67% dan nilai AUC senilai 0,8334. Hasil dari penelitian ini dapat disimpulkan bahwa tektik regresi logistik biner lebih unggul disbanding strategi algoritma C5.0 untuk analisis klasifikasi laju inflasi di Indonesia.

Kata kunci *: Laju Inflasi, Klasifikasi, Regresi Logistik Biner, Algoritma C5.0*

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

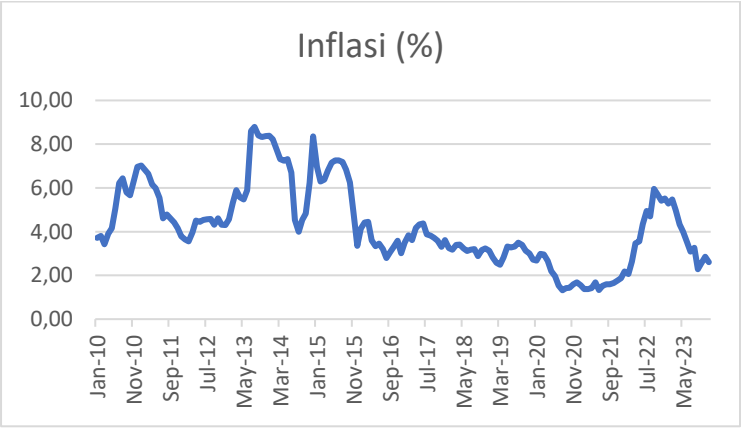
Dalam teori makroekonomi, setiap negara selalu menghadapi masalah seperti pertumbuhan ekonomi, pengangguran, inflasi, ketidakstabilan kegiatan ekonomi dan neraca perdagangan. Isu ekonomi yang selalu menjadi perhatian pemerintah negara-negara di dunia, khususnya Indonesia yaitu inflasi. Inflasi yaitu peningkatan harga produk dan jasa secara keseluruhan pada bidang ekonomi. Stabilitas ekonomi sering diukur dengan menggunakan tingkat inflasi sebagai indikator. Ketika inflasi rendah dan stabil, pertumbuhan ekonomi meningkat. Setiap kali terjadi kerusuhan sosial, politik atau ekonomi di dalam atau di luar negeri, orang selalu mengaitkannya dengan inflasi (Arjunita, C, 2016).

Saat ini, Indonesia menunjukkan perkembangan positif dalam indikator ekonomi makro. Proyeksi pertumbuhan ekonomi mencapai 5,03%, didorong terutama oleh permintaan domestik, industri pengolahan, dan perdagangan. Kontribusi signifikan juga datang dari konsumsi Lembaga Non-Profit yang melayani Rumah Tangga (LNPRT), terutama selama periode kampanye Pemilihan Umum. Tingkat inflasi umum yang mencapai 2,57% dianggap terkendali dan berada dalam kisaran yang diharapkan (Sekarsari dkk., 2024).

Indonesia memiliki inflasi dengan kenaikan yang sangat kecil.

Inflasi di Indonesia dianggap sebagai penyakit yang telah ada sejak lama dan bergantung pada sejarah. Salah satu dari cara untuk mengendalikan inflasi yaitu melalui jumlah uang beredar. Inflasi dipengaruhi oleh jumlah uang yang beredar di negara tersebut (Kalalo, 2016).

Ketika perkembangan ekonomi sangat cepat, biasanya mengarah pada kenaikan harga. Jika harga tidak dikendalikan maka akan menyebar ke seluruh barang dan jasa yang dibutuhkan masyarakat, artinya pada akhirnya akan terjadi inflasi. Tingkat inflasi di Indonesia diupayakan tidak melebihi 5,43% agar tidak menghambat pertumbuhan ekonomi (Vinayagathasan, 2013). Ketika inflasi terjadi maka harga barang naik sedangkan daya beli masyarakat turun. Dan ketika inflasi terjadi menyebabkan nilai uang domestik melemah. Inflasi juga menyebabkan ketidakpastian bagi pelaku ekonomi dalam melakukan produksi dan investasi, yang menjadi penghambat pertumbuhan ekonomi. Faktor-faktor seperti nilai tukar mata uang nasional terhadap dolar AS, jumlah uang yang beredar di masyarakat , dan tingkat suku bunga memengaruhi inflasi. Oleh karena itu, Indonesia membutuhkan pengelolaan yang tepat (Aprileven, 2017).



Gambar 1. 1 Grafik inflasi di Indonesia pada Tahun 2010 – 2023

(Sumber: https://www.bi.go.id/.)

Dalam kurun waktu sekitar empat belas tahun, laju inflasi di Indonesia masih dikategorikan sebagai inflasi ringan karena tingkat inflasi selama periode tersebut tetap di bawah 10%. Oleh karena itu, dapat dikatakan bahwa inflasi di Indonesia masih berada dalam kategori inflasi ringan. Meskipun kategori inflasi di Indonesia selama empat belas tahun ini ringan namun tetap mempengaruhi kenaikan harga BBM, sembako, biaya pendidikan, upah tenaga kerja dan lain sebagainya (Nanga, 2005).

Di Indonesia inflasi cukup berubah dari tahun ke tahun. Pada tahun 2012 inflasi di Indonesia sebesar 4,30 %, kemudian pada tahun 2013 inflasi meningkat sebesar 8,38% diikuti dengan kenaikan harga Bahan Bakar Minyak (BBM) bersubsidi, dengan BBM premium menjadi Rp 6.500/liter dan solar Rp 5.500/liter. Pada tahun 2014 inflasi menurun menjadi 8,36%. Selanjutnya, pada tahun 2015 inflasi kembali menurun menjadi 3,35%. Dari

tahun 2016 – 2020 inflasi terus menurun hingga sebesar 1,68% pada 2020. Inflasi kemudian naik kembali pada tahun 2022 sebesar 5,51% disebabkan oleh kombinasi antara faktor global, pasca pandemik dan cuaca yang berdampak pada gagal panen hingga distribusi terhambat serta menyebabkan harga BBM pertalite menjadi naik sebesar Rp 10.000/liter dari harga sebelumnya yaitu sebesar Rp 7.650/liter sedangkan BBM solar menjadi naik sebesar Rp 6.800/liter dari harga sebelumnya yaitu sebesar Rp 5.150/liter (Ningsih, dkk., 2018).

Hingga pada tahun 2023 inflasi turun kembali menjadi sebesar 2,61% namun harga BBM belum kunjung turun sampai saat ini. Akibatnya, analisis diperlukan untuk mengetahui beberapa faktor yang dapat mempengaruhi inflasi sehingga pemerintah dan penduduk dapat memprediksi kapan inflasi akan terjadi. Asumsi penelitian ini berdasarkan penelitian sebelumnya yang dilakukan oleh Aprileven (2017) dan Widiarsih, Romanda (2020) untuk faktor-faktor inflasi meliputi nilai tukar mata uang nasional terhadap dolar AS, suku bunga dan jumlah uang yang beredar di masyarakat.

Metode regresi logistik biner pernah digunakan untuk menganalisis laju inflasi di Indonesia dan memperoleh hasil yang baik berdasarkan penelitian yang dilakukan oleh Amanda dan Marhaeni (2023). Sedangkan metode algoritma c5.0 belum ada penelitian tentang analisis laju inflasi di Indonesia oleh karena itu dalam penelitian ini akan menampilkan proses dan hasil analisis

laju inflasi menggunakan metode algoritma c5.0. Kemudian membandingan antara metode regresi logistik biner dan algoritma c5.0 pada laju inflasi di Indonesia. Untuk mengetahui metode mana yang lebih unggul untuk menganalisis laju inflasi. Metode regresi logistik biner dan metode algoritma C5.0 merupakan metode klasifikasi dikarenakan metode regresi logistik merupakan metode analisis regresi dimana variabel dependennya adalah variabel biner atau kategorikal, untuk variabel dependennya biner atau dikotomis dan terdiri dari dua kategori yaitu 1 dan 0, maka faktor-faktor yang mempengaruhi inflasi diteliti dengan menggunakan analisis regresi logistik (Tampil, 2017). Sedangkan metode algoritma C5.0 sebagai bagian dari teknologi klasifikasi data mining yang mempunyai kemampuan untuk mengklasifikasikan ke dalam pohon keputusan atau buku aturan. Algoritma C5.0 adalah pengklasifikasi yang dapat mengklasifikasikan data dalam waktu yang relatif singkat dibanding pengklasifikasi lainnya dengan menggunakan memori minimal untuk membuat pohon keputusan dan meningkatkan akurasi (Pandya, 2015).

Metode regresi logistik biner dapat digunakan untuk menganalisis variabel dependen atau hubungan antara variabel dependen dengan variabel independen atau variabel independen dalam skala kategoris atau kontinu. Metode regresi logistik biner digunakan untuk menganalisis hubungan antara variabel dependen dengan beberapa variabel independen, dimana

variabel dependen berupa data kualitatif dikotomis, dengan nilai 1 menunjukkan adanya fitur dan nilai 0 menunjukan tidak adanya fitur (Tampil, 2017).

Metode algoritma C5.0 merupakan salah satu algoritma klasifikasi yang sesuai untuk kumpulan data besar. Algoritma C5.0 lebih baik daripada c4.5 dalam hal memori, kecepatan, dan efisiensi. Pemilihan atribut yang akan diproses pada algoritma C5.0 didasarkan pada ukuran *gain ratio*. Ukuran *gain ratio* digunakan untuk memilih atribut uji untuk setiap *node* di *tree*. Ukuran ini dapat digunakan dalam memilih maupun mendefinisikan *nodes* pada pohon. Atribut dengan nilai *gain ratio* terbesar dipilih sebagai kepala dari *node* berikutnya. Langkah-langkah membuat *tree* pada algoritma C5.0 mempunyai kemiripan dengan *tree* algoritma c4.5. Kesamaannya ada pada perhitungan *entropy* dan *gain*. Ketika algoritma c4.5 berhenti sampai perhitungan *gain*, algoritma C5.0 melanjutkan menghitung *gain ratio* menggunakan *gain* dan *entropy* yang ada (Pratiwi, 2020).

Pada penelitian sebelumnya yang dilakukan Permana, Siregar, Masruriyah, Juwita (2020) tentang perbandingan hasil prediksi kredit macet pada koperasi menggunakan Algoritma KNN dan C5.0 menghasilkan bahwa Algoritma C5.0 lebih baik dari KNN. Selanjutnya penelitian yang dilakukan oleh Rumaeda, Wilandari, Safitri (2016) tentang perbandingan Regresi Logistik Biner dan Algoritma C4.5 untuk mengklasifikasi penyakit

hipertensi menghasilkan bahwa Regresi Logistik Biner lebih baik daripada Algoritma C4.5 dalam mengklasifikasikan jenis penyakit hipertensi di UPT Puskesmas Ponjong I, Gunungkidul. Kemudian penelitian yang dilakukan oleh Pratiwi, Hayati, Prangga (2019) dengan membandingkan klasifikasi Algoritma C5.0 dengan *Clssification and Regression Tree* menghasilkan bahwa metode CART merupakan metode yang lebih baik dalam pengklasifikasian data rata-rata pendapatan masyarakat Desa Teluk Baru Kecamatan Muara Ancalong tahun 2019 dibandingkan dengan metode algoritma C5.0. Untuk penelitian tentang inflasi oleh Widiarsih, Romanda (2020) tentang analisis faktor-faktor yang mempengaruhi inflasi di Indonesia tahun 2015-2019 dengan pendekatan *error corection model* menghasilkan bahwa ketiga variabel Jumlah Uang Beredar, Nilai Tukar Rupiah terhadap Dolar, dan BI Rate mempunyai pengaruh terhadap inflasi.

Berdasarkan latar belakang yang telah disebutkan sebelumnya, penelitian ini bertujuan untuk mengetahui bagaimana model pengaruh nilai tukar, suku bunga dan jumlah uang beredar terhadap inflasi Indonesia dengan metode regresi logistik biner dan algoritma C5.0 dan untuk mengetahui metode mana yang lebih baik untuk menganalisis klasifikasi tingkat inflasi Indonesia.

B. Rumusan Masalah

1. Bagaimana kinerja metode regresi logistik biner dalam

memprediksi laju inflasi di Indonesia?

2. Bagaimana kinerja algoritma C5.0 dalam memprediksi laju inflasi di Indonesia?
3. Bagaimana perbandingan kinerja metode regresi logistik biner dan algoritma C5.0 dalam memprediksi laju inflasi di Indonesia?

C. Tujuan Penelitian

1. Untuk mengetahui kinerja metode regresi logistik biner dalam memprediksi laju inflasi di Indonesia.
2. Untuk mengetahui kinerja algoritma C5.0 dalam memprediksi laju inflasi di Indonesia.
3. Untuk membandingkan antara metode regresi logistik biner dan algoritma C5.0 pada prediksi laju inflasi di Indonesia.

D. Manfaat Penelitian

1. Memberikan informasi kepada masyarakat tentang faktor yang mempengaruhi laju inflasi di Indonesia dalam penerapan ilmu matematika untuk pengklasifikasian laju inflasi di Indonesia menggunakan metode regresi logistik biner dan algoritma C5.0.
2. Sebagai bentuk pengembangan dan penerapan ilmu statistika, khususnya pada metode regresi logistik biner dan algoritma C5.0.
3. Dapat meningkatkan pemahaman dan pengalaman langsung tentang cara menggunakan metode regresi logistik biner dan algoritma C5.0.

E. Pembatasan Masalah

1. Metode yang digunakan dalam menganalisis faktor- faktor yang mempengaruhi laju inflasi adalah metode regresi logistik biner dan algoritma C5.0.
2. Pemilihan model terbaik dari metode regresi logistik biner dan algoritma C5.0 dengan melihat nilai ketepatan klasifikasi, nilai AUC (Area Under Curve) dan kurva ROC.
3. Data yang digunakan adalah data laju inflasi di Indonesia pada rentang waktu 2010 sampai dengan 2023 yang bersumber dari https://www.bps.go.id/ dan https://www.bi.go.id/

BAB II

LANDASAN PUSTAKA

A. Laju Inflasi

Definisi Inflasi yaitu ketika harga beberapa barang yang dibutuhkan orang terus meningkat, meskipun kenaikan satu kali saja cukup besar, itu bukan inflasi (Nopirin, 2014). Inflasi dihitung berlandaskan angka indeks yang diperoleh dari berbagai komoditas pada setiap tingkat harga di pasar (barang-barang umum atau kebutuhan utama masyarakat). Berdasarkan tingkat harga tersebut, dibuat suatu angka indeks yang menggabungkan seluruh barang yang dibeli pelanggan dengan harga berapa pun yang disebut Indeks Harga Konsumen (IHK). Berdasarkan IHK, mungkin telah dihitung tingkat umum kenaikan harga selama jangka waktu tertentu (Iskandar, 2018). Ada beberapa kriteria yang dapat digunakan untuk membedakan inflasi (Pujadi, 2022):

a. Kurang dari 10% per bulan, yang disebut inflasi ringan

b. 10-30 % bulan dianggap inflasi sedang

c. 30-100 % per bulan dianggap inflasi tinggi

d. Lebih dari 100% per bulan disebut hiperinflasi.

Beberapa faktor yang dapat mempengaruhi inflasi Indonesia yaitu suku bunga, nilai tukar (kurs), dan jumlah uang beredar.

1. Nilai Tukar

Nilai tukar (*kurs*) menunjukkan nilai mata uang suatu negara dalam mata uang negara lain. Nilai tukar dapat

diartikan sebagai jumlah mata uang lokal atau rupiah yang diperlukan untuk mendapatkan satu unit mata uang asing (Sadono, 2015). Bank sentral negara menetapkan nilai tukar untuk menjaga stabilitas pasar domestik, kecukupan cadangan devisa dan pengendalian inflasi (Widiarsih, 2020).

2. Suku Bunga

Suku bunga (bunga bank) adalah upah yang dibayarkan oleh bank kepada konsumen yang membeli atau menjual barangnya sesuai dengan persyaratan standar. Suku bunga bank bisa diartikan sebagai imbalan yang dibayarkan oleh nasabah yang memiliki simpanan dan juga sebagai biaya yang harus dibayar nasabah kepada bank ketika menerima kredit (Kasmir, 2014).

Komponen utama yang mempengaruhi suku bunga adalah tingkat inflasi di Indonesia. Jumlah uang yang beredar di dalam negeri, serta produksi dan permintaan masyarakat mempengaruhi tingkat inflasi. Suku bunga akan naik jika inflasi naik, dan sebaliknya, jika inflasi turun, Bank Indonesia akan menurunkan suku bunga. Selain mempengaruhi kenaikan harga, perubahan suku bunga juga mempengaruhi pertumbuhan ekonomi negara dan masyarakat di seluruh dunia. Di sisi lain, pada saat ekonomi lemah, Bank Indonesia menurunkan suku bunga untuk mendorong berkembangnya perusahaan kecil dan bidang ekonomi

lainnya. Oleh sebab itu, pemerintah berharap dapat mempertahankan stabilitas ekonomi dengan menahan inflasi. Penetapan suku bunga juga berdampak signifikan pada keadaan ekonomi sehari-hari. Misalnya, ketika panen yang sulit atau kekurangan bahan pokok tertentu membuat harga barang kebutuhan pokok meningkat, suku bunga diturunkan guna memperlancar peredaran kredit di masyarakat. Ketika ekonomi membaik dan jumlah uang beredar meningkat, ada harapan harga bahan makanan pokok ini akan menurun dan kemudian kembali stabil. Namun, untuk mencegah inflasi, suku bunga juga sangat penting dalam pengendalian uang di masyarakat. Ketika inflasi melaju kencang, Institusi perbankan lebih suka menyimpan uang mereka di Bank Indonesia akibatnya peredaran uang berangsur-angsur berkurang (Widiarsih, 2020).

3. Jumlah Uang Beredar

Uang dapat diartikan sebagai sesuatu yang dapat digunakan atau diperoleh untuk membayar barang atau jasa. Fungsi uang secara umum yaitu untuk mengukur nilai, menjadi alat tukar dan menyimpan kekayaan. Jumlah uang beredar juga sering disebut uang beredar. Jumlah uang yang disalurkan oleh masyarakat dan lembaga keuangan dikenal sebagai uang beredar. Jumlah uang beredar adalah salah satu variabel di bidang keuangan yang dianggap penting dan diperlukan. Sebuah studi menyebutkan JUB sebagai

indikator yang memberikan sinyal positif pertumbuhan ekonomi (Feri Wibowo & Rina Arifati, 2016).

B. Data Mining

Menurut Agresti (2009), data mining adalah proses ekstraksi atau penemuan informasi penting dari kumpulan data yang sangat besar. Dalam melakukannya, data mining mengekstraksi informasi berharga dengan menganalisis pola atau hubungan yang terkait dengan database besar. Sedangkan menurut Han (2004), data mining didefinisikan sebagai menambang atau mengekstraksi informasi yang sebelumnya tidak diketahui tetapi dapat dipahami dan dimanfaatkan dari basis data besar dan digunakan untuk membuat keputusan yang sangat penting bagi bisnis.

Data mining dikenal sebagai "penemuan data atau pengetahuan" untuk menemukan pola yang tersembunyi dalam data. Berdasarkan definisi tersebut, data mining merupakan pencarian informasi penting dan berharga yang tersembunyi dan tidak diketahui dalam basis data yang sangat besar untuk ditemukan. Dalam data mining, terdapat langkah-langkah kunci dalam proses pencarian informasi, yang terdiri dari tujuh tahap berikut:

1. Pembersihan Data

   Pembersihan data menghapus duplikat data, memeriksa ketidakkonsistenan data, dan memperbaiki kesalahan data termasuk kesalahan ketik. Biasanya, informasi yang

didapatkan dari database atau hasil tes perusahaan memiliki konten yang tidak lengkap, seperti informasi yang hilang dan salah, atau bahkan hanya kesalahan ketik.

2. Integrasi Data

   Integrasi data adalah proses menggabungkan data dari berbagai database menjadi data baru yang diperlukan untuk pencarian informasi, atau melengkapi data yang ada dengan data yang relevan.

3. Pemilihan Data

   Pemilihan data yang tepat dan analisis data operasional dilakukan. Informasi hasil pemilu disimpan dalam database khusus.

4. Transformasi Data

   Proses mengubah data menjadi format yang sesuai untuk digunakan dalam proses penambangan data dikenal sebagai transformasi data. Misalnya, input kategorikal hanya dapat diterima oleh metode standar seperti analisis asosiasi dan pengelompokan.

5. Data Mining

   Menemukan pola atau informasi yang menarik yang dihasilkan melalui penggunaan metode, algoritma, atau teknik tertentu.

6. Evaluasi Pola

   Mengidentifikasi pola yang sangat menarik dalam hasil data mining. Pada tahap ini, Hasil dari teknik data mining melibatkan

evaluasi model tipikal dan prediktif untuk menilai keabsahan hipotesis yang ada saat ini.

7. Presentasi Data

   Presentasi Data adalah proses menampilkan pola atau model yang dibuat pada tahap data mining. Visualisasi ini membantu menyajikan hasil dari proses penambangan data dalam format yang mudah dipahami (Han, 2014).

C. Data *Training* dan Data *Testing*

Data yang akan digunakan dalam pengujian klasifikasi dibagi menjadi dua yaitu data *training* dan *testing*. Data *training* yaitu data atau vektor yang sudah diketahui sebelumnya untuk label kelas dan digunakan untuk membangun model *classifier*. Sedangkan data *testing* yaitu data atau vektor yang belum diketahui label kelasnya menggunakan model *classifier* yang sudah dibangun (Prasetyo, 2014).

Saat membagi suatu data menjadi data *training* dan data *testing* untuk menghasilkan model klasifikasi yang akurat, biasanya digunakan kombinasi persentase pemisahan yang berbeda. Kombinasi yang umum mencakup 90%: 10%, 85%: 15%, 80%: 20%, dan 75%: 25%. Angka pertama menunjukkan persentase data *training*, dan angka kedua menunjukkan persentase data *testing*. Kombinasi terbaik dari kombinasi tertentu ditentukan oleh hasil akurasi klasifikasi yang diperoleh dari persentase optimal data pelatihan dan pengujian. Menghitung akurasi pada data *training* dan data *testing*

merupakan proses pembuatan pohon klasifikasi yang berkualitas tinggi (Nair, dkk., 2001).

Adapun beberapa metode yang digunakan untuk membagi data *training* dan *testing* yaitu:

1. *Random Split*

   Random split adalah metode yang umum digunakan yang secara acak membagi dataset menjadi set pelatihan dan pengujian. Set pelatihan digunakan untuk melatih model pembelajaran mesin, ini adalah himpunan data inti tempat model belajar memahami pola dan hubungan dalam data. Set pengujian memberikan evaluasi yang adil atas kinerja model pada data yang tidak terlihat. Ini sangat penting untuk menilai kemampuan model untuk menggeneralisasi ke data yang tidak diketahui (David, 2023).

2. *K-fold Cross Validation*

   K-fold cross validation membagi dataset menjadi lipatan berukuran sama "k", memungkinkan beberapa putaran pelatihan dan validasi. *K-fold cross validation* adalah teknik yang ampuh untuk menilai performa dan kemampuan generalisasi model pembelajaran mesin. Ini mengatasi tantangan untuk mengevaluasi kinerja model secara efektif dengan data terbatas dengan memanfaatkan sampel yang tersedia. Metode ini membagi dataset menjadi "k" subset berukuran sama, di mana "k" mewakili jumlah grup atau segmen tempat data dipartisi. Selanjutnya, setiap lipatan

digunakan sebagai set validasi, dan lipatan yang tersisa berfungsi sebagai set pelatihan dalam iterasi "k". Proses ini memastikan bahwa setiap lipatan digunakan untuk validasi tepat sekali, dan model dilatih dan dievaluasi pada subset data yang berbeda. Dengan melakukan beberapa putaran pelatihan dan validasi, penilaian kinerja model yang lebih andal dan akurat diperoleh. Teknik ini sangat menguntungkan ketika berhadapan dengan kumpulan data terbatas, karena memaksimalkan penggunaan data yang tersedia untuk pelatihan dan validasi, memberikan estimasi kemampuan prediksi model yang lebih kuat. Setelah menyelesaikan iterasi "k", hasil evaluasi dirata-ratakan untuk menawarkan perkiraan performa model yang lebih stabil dan representatif, mengurangi dampak partisi data pada evaluasi keseluruhan (David, 2023).

3. *Stratified K-fold Cross Validation*

Stratified K-Fold Cross-Validation (SKCV) merupakan perluasan dari *Cross-Validation* (CV). SKCV memastikan bahwa setiap kelas didistribusikan secara merata di seluruh *k-fold*. Dengan kata lain, kumpulan datanya tidak disebarkan secara acak ke dalam *k-fold*, namun sedemikian rupa sehingga tidak mengganggu rasio distribusi sampel antar kelas dalam SKC. Hasilnya, SKCV lebih disukai daripada CV untuk menangani masalah klasifikasi dengan distribusi kelas yang tidak merata. objek CV ini mengembalikan lipatan bersarang

dan merupakan varian dari *K-Fold*. Pelipatan dicapai dengan menjaga fraksi sampel dengan setiap kelas tetap konstan (Widodo & Dkk, 2022).

4. *Leave One Out Cross Validatio*

   Leave One Out Cross Validation adalah metode validasi silang yang komprehensif, sangat cocok untuk kumpulan data kecil. Dalam metode ini, masing-masing sampel N diambil sebagai sampel uji, sedangkan sampel N-1 yang tersisa berfungsi sebagai set pelatihan. Proses ini diulang untuk setiap sampel, menghasilkan pengklasifikasi N dan hasil uji N. Kinerja model kemudian dinilai dengan rata-rata hasil N ini (David, 2023).

5. *Time Series Split*

   Time series split adalah metode yang dirancang khusus untuk menangani data deret waktu, yang terdiri dari urutan pengamatan yang direkam pada titik waktu yang berbeda, seperti harga stok harian, catatan cuaca bulanan, atau data lalu lintas situs web per jam. Tantangan unik dari data deret waktu adalah bahwa urutan titik data sangat penting, karena pengamatan biasanya bergantung pada hasil sebelumnya. Dalam konteks ini, pemisahan deret waktu sangat berharga. Memisahkan himpunan data menjadi subset untuk pelatihan, validasi, dan pengujian memastikan pelestarian urutan temporal data. Metode pemisahan deret waktu menggunakan teknik khusus (misalnya, kelas Time Series Split di scikit-learn)

untuk membuat himpunan bagian yang menghormati urutan kronologis data. Tidak seperti pemisahan acak tradisional yang mengacak data, pemisahan deret waktu mengelompokkan data menjadi fragmen, dengan setiap segmen mewakili periode waktu yang berbeda. Misalnya, memiliki data harian selama setahun, setiap segmen sesuai dengan satu minggu atau sebulan. Jenis pemisahan deret waktu ini sangat penting untuk data deret waktu karena meniru skenario dunia nyata di mana prediksi masa depan hanya dapat didasarkan pada peristiwa masa lalu. Dengan mempertahankan urutan kronologis ini, pemisahan deret waktu memungkinkan untuk melatih model menggunakan data historis dan memvalidasi model menggunakan data yang diperbarui, mensimulasikan performa model di lingkungan dunia nyata (David, 2023).

6. *Bootstrap*

Metode *bootstrap* adalah resample data dari data asli dengan pengembalian untuk mendapatkan replika data baru dengan banyak pengulangan yang terjadi. Dikarenakan banyaknya pengulangan ini, metode bootstrap juga kadang disebut sebagai metode computer-intensive. Dimulai dari tahun 1979, penerapan metode bootstrap mengalami banyak kemajuan dari tahun ke tahun. Salah satu penerapan metode bootstrap adalah untuk menguji hipotesis non parametrik. Metode Bootstrap merupakan suatu metode resample atau pengambilan sampel-sampel baru secara acak dengan

pengembalian berdasarkan sampel asli sebanyak B kali (Agustius & Dkk, 2013).

D. Klasifikasi

Klasifikasi adalah jenis analisis data yang membantu dalam memprediksi label atau kategori dari sampel yang akan dikelompokkan. Berbagai bidang, termasuk pembelajaran mesin, pembelajaran komputer, sistem pakar, dan statistik, telah mengusulkan berbagai metode klasifikasi. Model pertama kali dilatih dengan kumpulan data historis dengan label kelas yang diketahui. Kemudian *classifier* terlatih ini diterapkan pada nama kelas contoh terbaru Liao (2007). Sedangkan klasifikasi adalah proses mencari pola yang menjelaskan dan membedakan jenis (Han, 2014). Tipe atau kategori data yang digunakan untuk memprediksi atau memperkirakan jenis kategori item dalam kasus di mana jenis kategori variabel respons tidak diketahui. Klasifikasi adalah proses mencari pola atau karakteristik yang menjelaskan atau mengidentifikasi ide dari suatu konsep atau pengetahuan, dengantujuan mengevaluasi kategori ambigu dari suatu objek (Prasetyo, 2012).

Proses pengujian klasifikasi tidak bisa mengabaikan pentingnya membagi data menjadi dua kelompok, yaitu data pelatihan (data *training*) dan data pengujian (data *testing*). Data *training* dan data *testing* memainkan peran penting dalam membangun dan menguji model klasifikasi. Setelah menggunakan data pelatihan untuk membangun model

klasifikasi, gunakan data pengujian untuk menguji kemampuan model dalam membuat prediksi yang benar berdasarkan data yang belum pernah dilihat sebelumnya (Musu, dkk., 2021).

Komposisi atau proporsi data yang dipertukarkan antara data *training* dan *testing* memiliki dampak signifikan terhadap penilaian akurasi model. Jika konfigurasi tidak seimbang, misalnya jika data *training* lebih banyak dibandingkan data *testing*, atau sebaliknya, model mungkin bias terhadap subset data. Oleh karena itu, penting untuk merencanakan distribusi data yang seimbang dan mewakili distribusi yang sebenarnya (Musu, dkk., 2021).

E. Statistika Deskriptif

Statistik deskriptif adalah metode pengumpulan, penggabungan, dan identifikasi informasi untuk menghasilkan data yang berguna dan mengkategorikannya ke dalam struktur yang disiapkan untuk penelitian. Oleh karena itu, berbagai wawasan berada pada tahap pengembangan dan pembahasan penyajian, termasuk penyajian informasi. Pada tahap ini, pengukuran aktual seperti ukuran fokus, ukuran transfer, distribusi ukuran wilayah dan alokasi informasi diperiksa. Statistik deskriptif adalah suatu sistem untuk mengumpulkan, menyajikan, menganalisis, dan menjelaskan informasi. Metode ini dibagi menjadi dua, yaitu meliputi pengukuran wawasan dan inferensi. Statistik deskriptif menggabungkan pengumpulan informasi, memutuskan kualitas dan kapasitas yang terukur,

membuat diagram, grafik, dan gambar (Walpole, 1995).

Statistik deskriptif adalah pengukuran yang menggunakan contoh dan informasi demografis untuk menggambarkan atau memberikan gambaran tentang topik yang dipelajari tanpa melakukan penelitian atau menarik kesimpulan yang relevan untuk masyarakat pada umumnya (Sugiyono, 2007).

Statistik deskriptif adalah cabang statistika yang mempelajari cara pengumpulan dan penyajian data agar mudah dipahami. Statistik deskriptif memberikan gambaran atau informasi tentang data atau kondisi tertentu (Hasan, 2004).

F. Regresi Logistik Biner

Menurut Hosmer (2000) Regresi logistik biner adalah metode analisis data yang digunakan untuk menemukan hubungan antara variabel respon (y), yang berupa variabel biner, dan variabel prediktor (x). Variabel respon (y) dalam regresi logistik biner mengambil dua nilai, yaitu 1 untuk keberhasilan dan 0 untuk kegagalan. Dalam konteks ini, variabel y mengikuti distribusi Bernoulli untuk setiap pengamatan. Fungsi kemungkinan untuk setiap pengamatan diberikan sebagai berikut:

$$f(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}; y_i = 0,1 \quad (2.1)$$

y_i : Variabel acak yang mengambil nilai 0 atau 1.

$\pi(x_i)$: Probabilitas sukses (y_i = 1).

$1-\pi(x_i)$: Probabilitas gagal ($y_i = 0$).

Jika $y_i = 0$ maka $f(y_i) = 1-\pi(x_i)$ dan jika $y_i = 1$ maka $f(y_i) = \pi(x_i)$.

Fungsi regresi logistik dapat ditulis sebagai berikut:

$$f(z) = \frac{e^z}{1+e^z} \tag{2.2}$$

Dengan $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$.

Dengan demikian, model umum persamaan regresi logistik dengan beberapa variabel bebas x dapat dirumuskan sebagai berikut:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_p X_p}} \tag{2.3}$$

Keterangan:

$\pi(x)$: Peluang peristiwa dengan nilai probabilitas $0 \le \pi(x) \le 1$.

X_i : Variabel prediktor ke - i $(i = 1, 2, 3, ... p)$.

β_0 : Konstanta.

β_i : Koefisien variabel prediktor ke - i $(i = 1, 2, 3, ... p)$.

Model umum regresi logistik dengan fungsi $\pi(x)$ merupakan fungsi non linier maka untuk membuatnya menjadi fungsi linier

agar dapat melihat hubungan antara variabel respon dan variabel prediktor harus dilakukan transformasi. Bentuk logit dari $\pi(x)$ dapat dituliskan sebagai berikut:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \quad (2.4)$$

$g(x)$: Transformasi logit dari $\pi(x)$.

$\frac{\pi(x)}{1-\pi(x)}$: Rasio peluang *odds* yang menunjukan probabilitas sukses dan gagal.

1. Estimasi Parameter

Metode estimasi parameter dalam model regresi logistik biner yang digunakan adalah estimasi likehood maksimum. Metode esimasi parameter digunakan metode kuadrat terkecil seperti regresi linier, sering memberikan estimasi parameter tanpa karakteristik yang tidak bias dan dengan varian yang minimal. Hal ini terjadi karena dalam regresi logistik biner, variabel respon (y) mengikuti distribusi Bernoulli, yang menyebabkan varians berubah seiring dengan probabilitas keberhasilan. Fungsi kemungkinan yang diperoleh digunakan untuk mendapatkan estimasi parameter yang tidak diketahui. Estimasi ini dilakukan dengan memaksimalkan nilai probabilitas dari data yang diamati. Dengan memaksimalkan fungsi likelihood, metode MLE dapat mengestimasi koefisien (Hosmer, 2000).

Variabel respon y terdiri dari dua jenis yaitu sukses dan gagal yang direpresentasikan oleh y = 1 (sukses) dan y = 0 (gagal). Dalam konteks ini, variabel y mengikuti distribusi Bernoulli untuk setiap pengamatan. Fungsi likelihood untuk setiap pengamatan diberikan sebagai berikut:

$$f(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}; y_i = 0,1 \quad (2.5)$$

y_i : Variabel acak yang mengambil nilai 0 atau 1.

$\pi(x_i)$: Probabilitas sukses (y_i = 1).

$1-\pi(x_i)$: Probabilitas gagal (y_i = 0).

Dimana jika y_i = 0 maka $f(y_i)$ = $1-\pi(x_i)$ dan jika y_i = 1 maka $f(y_i)$ = $\pi(x_i)$. Jika diasumsikan bahwa y_i independen, maka persamaan fungsi probabilitas metode regresi logistik biner sebagai berikut:

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \quad (2.6)$$

$l(\beta)$: Fungsi *likelihood*.

$\prod$: Simbol produk menunjukan hasil kali dari semua elemen dari 1 hingga n.

n : Jumlah total pengamatan dalam dataset.

y_i : Variabel acak yang mengambil nilai 0 atau 1.

$\pi(x_i)$: Probabilitas sukses (y_i = 1).

$1-\pi(x_i)$: Probabilitas gagal (y_i = 0).

Variabel dependen dibagi menjadi dua kategori yaitu sukses (1) dan gagal (0).

2. Uji Rasio Likelihood

Menurut Hosmer (2000) Uji rasio likelihood dapat digunakan untuk menguji lompatan dengan mengevaluasi kepraktisan model yang diperoleh dan memutuskan apakah variabel independen model berpengaruh signifikan terhadap seluruh populasi. Pengujian ini memungkinkan model penuh (menggunakan variabel independen) dianalisis dengan model konsisten (tanpa variabel independen) untuk memeriksa apakah model konsisten keseluruhan lebih baik daripada model lengkap menggunakan persamaan berikut. Statistik Uji :

$$G^2 = -2\ln\frac{Likelihood\ tanpa\ variabel\ bebas}{Likelihood\ dengan\ variabel\ bebas} \tag{2.7}$$

Untuk kasus variabel independen tunggal, untuk menunjukan bahwa ketika variabel tersebut tidak ada dalam model, estimasi kemungkinan maksimum β_0 adalah

$\ln\left(\frac{n_1}{n_0}\right)$ dengan $n_1 = \sum_{y_i=1} y_i$, dan $n_0 = \sum_{y_i=0}(1-y_i)$ dan nilai prediksinya konstan $\frac{n_1}{n}$. Dalam hal ini maka nilai G adalah:

$$G^2 = -2\ln \frac{\left(\frac{n_1}{n}\right)n_1\left(\frac{n_0}{n}\right)n_0}{\prod_{i=1}^{n}\pi^{yi}(1-\pi_i)^{(1-y_i)}} \tag{2.8}$$

Dimana:

n_1: Besaran pengamatan pada kategori sukses (1).

n_0 : Besaran pengamatan pada kategori gagal (0).

n : Jumlah pengamatan $(n_1 + n_0)$.

Hipotesa:

H_0: Variabel terikat tidak dipengaruhi oleh variabel bebas ketika keduanya dilakukan secara bersamaan.

H_1: Ada setidaknya satu variabel bebas yang mempengaruhi variabel terikat secara bersamaan.

Kriteria Uji : Tolak H_0 jika $G > X^2_{(\alpha.p)}$ atau $p-value < \alpha$ (Hosmer, D.W. and Lemeshow, 2000).

3. Uji Wald

Uji Wald digunakan untuk menilai kekuatan hubungan antara variabel independen dengan variabel respons dalam model. Pengukuran uji Wald dihitung dengan membagi

estimasi parameter dengan *standar error* dari parameter tersebut. Statistik Uji:

$$W_j = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\}^2 \tag{2.9}$$

Dimana :

$\hat{\beta}_j$: Nilai estimasi parameter β_j.

SE : *Standar error* untuk koefisien β_j.

Hipotesis:

H_0: Variabel bebas tidak memiliki pengaruh terhadap variabel terikat.

H_1: Variabel bebas memiliki pengaruh yang signifikan terhadap variabel terikat.

Kriteria Uji : Tolak H_0 jika $W_j > Z_{\alpha/2}$ atau $p-value < \alpha$ (Hosmer, D.W. and Lemeshow, 2000).

4. Uji Multikolinieritas

Multikolinearitas adalah suatu kondisi ketika setidaknya dua faktor otonom yang digunakan dalam kekambuhan berkorespondensi. Uji multikolinearitas berarti memeriksa apakah model regresi yang disiapkan memiliki hubungan yang ideal atau tinggi antara faktor bebasnya. Efek samping multikolinearitas dapat ditemukan dalam model regresi yang memiliki nilai hubungan yang tinggi atau luar

biasa. Dalam regresi logistik diasumsikan bahwa multikolonieritas harus dihindari, karena multikolinearitas kesalahan standar dari koefisien regresi, dapat memungkinkan bahwa uji Wald menghasilkan indikator yang tidak relevan. Salah satu metode untuk mendeteksi adanya multikolinearitas adalah dengan menggunakan VIF (*Variance Inflation Error Factor*) yang menjadi suatu insentif bagi setiap indikator. Dengan asumsi nilai VIF ≥ 10, hal ini menunjukkan adanya multikolinearitas pada data yang digunakan. Sehingga, uji multikolinearitas ini harus memiliki nilai VIF < 10.

$$VIF_i = \frac{1}{1-R_i^2} \tag{2.10}$$

VIF_i : *Variance Inflation Factor* untuk variabel independen ke-i.

R_i^2 : Koefisien determinasi berganda variabel bebas x_i dan semua variabel bebas lainnya (Suliyanto, 2011).

5. Uji Kesesuaian Model

Uji kesesuaian model menggunakan uji Hosmer dan Lemeshow untuk mengevaluasi kesesuaian model..

Statistik uji :

$$\hat{C} = \sum_{f=1}^{g} \left[\frac{(O_f - n\bar{\pi}_f)^2}{(n_f \bar{\pi}_f (1-\bar{\pi}_f))} \right]$$

(2.11)

Dimana:

g : Jumlah grup.

O_f : Jumlah nilai variabel respon pada grup ke-f.

$\bar{\pi}_f$: Rata-rata taksiran peluang.

n_f : Banyak observasi pada grup ke-f.

Hipotesis:

H_0 = Model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi)

H_1 = Model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi)

Kriteria uji : Tolak H_0 jika $\hat{C} > X^2_{(g-2,\alpha)}$ atau $p-value > \alpha$ (Hosmer, D.W. and Lemeshow, 2000).

6. Odds Ratio

Odds Ratio didefinisikan sebagai Ukuran tren yang diinterpretasikan sebagai perbandingan antara jumlah individu yang mengalami peristiwa tertentu dengan jumlah individu yang tidak mengalami peristiwa tersebut, baik dalam sampel maupun populasi umum (Agresti, 2009). Rasio kemungkinan adalah representasi probabilitas yang menggambarkan peluang suatu peristiwa dibagi peluang bahwa peristiwa itu tidak akan terjadi. *Odds Ratio* adalah perbandingan antara peluang keberhasilan dan peluang

kegagalan. *Odds Ratio* untuk kemungkinan sukses didefinisikan sebagai berikut:

$$\Omega_i = \frac{\pi_i}{1-\pi_i} \tag{2.12}$$

Ω_i : Peluang (*odds*) untuk kejadian pada grup ke-i.

π_i : Probabilitas kejadian pada grup ke-i.

Rasio kemungkinan Ω_1 dan Ω_2 yang kemudian disebut odds rasio adalah sebagai berikut:

$$\psi = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1 / 1-\pi_1}{\pi_2 / 1-\pi_2} \tag{2.13}$$

ψ : Perbandingan antara peluang kejadian pada dua grup yang berbeda.

Untuk distribusi peluang bersama π_{ij} nilai *odds* dalam baris ke-i adalah $\Omega_i = \frac{\pi_{1i}}{\pi_{2i}}$ dengan i = 1,2. Sehingga persamaan *Odds Ratio* adalah sebagai berikut :

$$\psi = \frac{\pi_{11} / \pi_{12}}{\pi_{21} / \pi_{22}} = \frac{\pi_{11} / \pi_{22}}{\pi_{12} / \pi_{21}} \tag{2.14}$$

ψ : Perbandingan antara peluang kejadian pada dua grup yang berbeda.

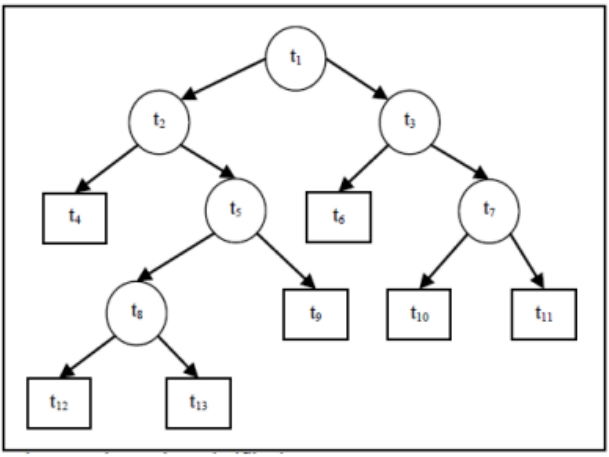
G. Decision Tree

Pohon keputusan *(Decision Tree)* merupakan salah satu metode yang sering digunakan untuk analisis data dan pengambilan keputusan. Metode ini suatu aliran keputusan berdasarkan aturan yang berasal dari data yang ada. Pohon keputusan memiliki struktur seperti pohon, dimana akar mewakili *node* awal yang mewakili fitur atau atribut, cabang mewakili pilihan atau nilai fitur, dan daun mewakili hasil atau kelas keputusan. Pembagian populasi data menjadi kelompok yang lebih kecil dan seragam secara karakteristik adalah konsep dasar pohon keputusan. Proses ini dilakukan dengan memilih atribut terbaik pada setiap langkah berdasarkan kriteria tertentu seperti information *gain* atau *gain ratio*. Atribut yang dipilih menjadi cabang di pohon, dan proses ini diulangi hingga data di setiap cabang cukup homogen atau kondisi terminasi lainnya tercapai. Faktor-faktor berikut membentuk karakteristik pohon keputusan (Prasetyo, dkk., 2013).

1. Sebuah *node* yang menunjukkan variabel. Ini dapat berupa variabel akar (juga dikenal sebagai *root node*), variabel cabang, atau kelas.
2. Nilai hasil pengujian pada node non-daun diwakili oleh setiap cabang.
3. Node akar tidak memiliki *edge* (sisi) atau lengan masukan dan tidak memiliki lengan keluar, yaitu nol atau lebih.
4. *Node internal* yang bukan *node non*-daun (*non-terminal*) dengan satu lengan masukan dan dua atau lebih lengan

keluaran. *Node* ini mewakili pengujian berdasarkan nilai fitur.

5. Node daun, juga dikenal sebagai node terminal, adalah simpul dengan satu lengan masukan dan tidak ada lengan keluaran. Nama kelas (penentuan) ditampilkan di node ini.



Gambar 2. 1 Struktur Pohon Klasifikasi

Simpul utama (*root node*) disebut t_1, dan simpul dalam (*internal nodes*) disebut t_2, t_3, t_5, t_7, dan t_8. Simpul akhir, juga dikenal sebagai simpul terminal (*terminal nodes*), adalah $t_4, t_6, t_9, t_{10}, t_{11}, t_{12}$, dan t_{13}, di mana tidak ada lagi pemilahan. Dimulai dengan simpul utama atau t_1, yang berada pada kedalaman 1, dan simpul t_2 dan t_3, yang berada pada kedalaman 2, dan seterusnya sampai simpul terminal, atau t_{12} dan t_{13}, yang berada pada kedalaman 5. Simpul-simpul ini ditunjukkan pada Gambar 2.1 gambar ◯ sebagai *node* utama dan *node* dalam sedangkan gambar □ sebagai simpul terminal (Pratiwi dan Zain, 2014).

H. Algoritma C5.0

Algoritma C5.0 adalah sebuah algoritma dalam data mining yang digunakan khususnya untuk membuat pohon keputusan. Algoritma C5.0 merupakan pengembangan dari algoritma ID3 dan C4.5 yang diciptakan oleh Ross Quinlan pada tahun 1987. Algoritma ini menggunakan rasio gain untuk menangani pemilihan atribut. Algoritma ini membuat pohon dengan jumlah cabang berbeda untuk setiap node (Putri, dkk., 2013).

Algoritma C5.0 membangun pohon dengan jumlah cabang berbeda pada setiap node. Algoritma ini memperlakukan variabel kontinu dengan cara yang sama seperti CART, namun untuk variabel kategori, algoritma C5.0 memperlakukan nilai variabel kategori sebagai pemisah. Bagian sampel dari cabang yang terbentuk kemudian dibelah lagi. Proses ini berlanjut hingga sampel subset tidak dapat lagi dibagi. Pada akhirnya, subsampel yang berkontribusi lebih sedikit pada model akan ditolak (Yusuf, 2007).

Langkah-langkah membangun pohon dengan algoritma C5.0 hampir sama dengan langkah-langkah membangun pohon dengan algoritma C4.5. Kesamaan ini juga mencakup perhitungan gain dan entropy. Jika algoritma C4.5 berhenti menghitung gain, algoritma C5.0 menggunakan gain dan entropy yang telah ada untuk menghitung rasio gain. Cara untuk menghitung nilai entropy adalah:

$$Entropy(S) = -\sum_{j=1}^{k} p_j \log_2 p_j$$

(2.15)

Dimana:

S : Himpunan kasus.

k : Jumlah kelas pada variabel A.

p_j : Proporsi dari S.

Kemudian untuk mencari nilai gain digunakan persamaan di bawah ini:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{m} \frac{|S_i|}{|S|} \times Entropy(S_i)$$

(2.16)

Dimana:

S : Himpunan kasus.

S_i : Himpunan kasus pada kategori ke- i.

A : Variabel.

m : Jumlah kategori pada variabel A.

$|S_i|$: Jumlah kasus pada kategori ke-i.

$|S|$: Jumlah kasus dalam S.

Setelah mengetahui nilai entropy dan gain, langkah berikutnya adalah menentukan nilai gain ratio. Rumus untuk melakukan ini adalah sebagai berikut.:

$$Gain\ Ratio = \frac{Gain(S, A)}{\sum_{i=1}^{m} Entropy(S_i)} \quad (2.17)$$

Dimana:

$Gain(S, A)$: Nilai gain dari suatu variabel.

$\sum_{i=1}^{m} Entropy(S_i)$: Jumlah nilai entropy dalam suatu variabel.

Proses diulang untuk setiap cabang sampai setiap kelas di cabang memiliki kelasnya sendiri (Putri, dkk, 2013).

I. Variabel *Dummy*

Variabel *dummy* disebut faktor multilabel, faktor penjelas, faktor subyektif, atau faktor dikotomis. Modus berulang mencakup faktor langsung sebagai faktor otonom, tetapi juga dapat digabungkan dengan faktor matematis independen lainnya (Supranto, 2004). Variabel dummy dapat digunakan pada variabel dependen atau variabel dependen, sehingga variabel dependennya adalah 0 atau 1 yang berarti ya atau tidak (dikotomis). Menurut Nachowi da Usman (2002), variabel dummy adalah faktor data subyektif yang digunakan untuk klasifikasi atau karakterisasi.

Berdasarkan pemahaman di atas, dapat disimpulkan bahwa variabel dummy adalah data yang diubah menjadi informasi kualitatif untuk tujuan membuat kategori.

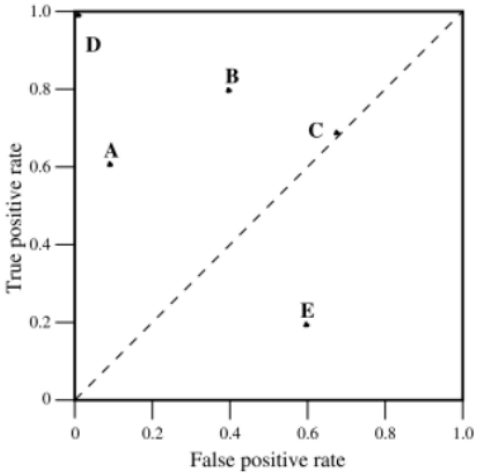
J. Pengujian Akurasi dengan Metode Analisis ROC (*Receiver Operating Characteristics*)

ROC (*Receiver Operating Characteristic*) adalah teknik untuk menggambarkan, mengatur dan menempatkan hal-hal yang tidak sepenuhnya diatur bagaimana direpresentasikan dalam model fakta. Selama Perang Dunia II, teknik ini digunakan untuk menganalisis keakuratan sinyal tertentu yang terdeteksi oleh radar. Sekarang, analisis ROC telah diperluas untuk karakterisasi dan analisis kerangka kerja analitik. Teknik ROC juga digunakan dalam studi dinamis yang menampilkan kurva ROC dalam uji bukti di bidang medis (Fawcett, 2006).

1. Grafik dan Kurva ROC (*Receiver Operating Characteristics*)

Kurva ROC adalah representasi grafik dua dimensi yang mengilustrasikan hubungan antara *True Positive Rate* (TPR) atau sensitivitas (Y) dan *False Positive Rate* (FPR) atau 1-spesifitas (X). Penting untuk dicatat dalam kurva ROC bahwa jika kurva dimulai dari titik dasar kiri (0,0), ini menunjukkan probabilitas yang tidak menghasilkan kondisi positif. Ini menandakan bahwa pengelompokan tidak menghasilkan *True Positive* atau *False Positive*. Sebaliknya, skor titik kanan atas (1,1) pada kurva menunjukkan nilai peluang kondisi yang positif. Nilai TPR dan FPR terkait satu sama lain, jika TPR membesar maka FPR akan berkurang begitu juga sebaliknya. Kurva ROC dapat membuat garis sudut ke sudut dengan menunjuk pengklasifikasi secara acak yang disebut kinerja acak. Setiap kali data kategori yang lengkap

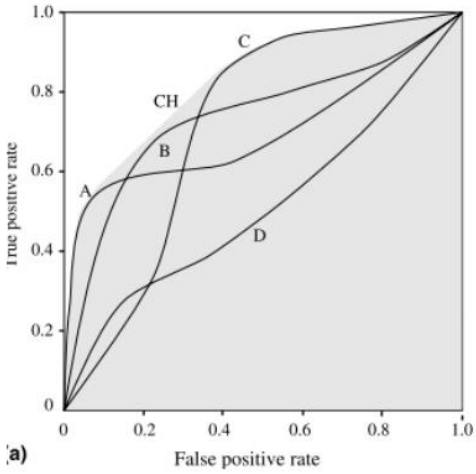
digabungkan dengan TPR dan FPR, data dapat digambarkan dalam diagram ROC. Setiap titik data yang diproses langsung dapat terhubung, menghasilkan perubahan dalam kurva ROC. Kurva ini menunjukkan tingkat kemungkinan atau presisi dari model tersebut. Nilai akurasi dikatakan sempurna atau teratur tanpa cela, dengan asumsi bahwa kurva dekat titik kiri atas (0,1) (Fawcett, 2006).



Gambar 2. 2 Contoh Kurva ROC pada Kinerja Acak

(Sumber: Fawcett, 2006)

Gambar 2.2 menunjukkan bahwa pada grafik ROC, garis diagonal dapat dibuat jika nilai pada sumbu y = x.



Gambar 2. 3 Contoh Kurva ROC

(Sumber: Fawcett, 2006)

Gambar 2. 3 menunjukkan contoh kurva bentuk ROC dari dua model yang dievaluasi.

2. *Area Under Curve* (AUC)

Daerah di bawah kurva merujuk pada area yang menggambarkan akurasi model yang diharapkan, dan diukur menggunkan teknik estimasi yang dikenal sebagai *Area Under Curve* (AUC). AUC adalah daerah persegi penting yang nilainya biasanya dimulai dari 0 hingga 1. Kinerja acak menghasilkan AUC senilai 0,5 kurva yang dibuat menghasilkan kemiringan antara titik (0,0) dan titik (1,1). Jika AUC berikutnya < 0,5, evaluasi model berikutnya memiliki presisi yang sangat rendah dan menunjukkan bahwa model tersebut sangat buruk saat digunakan (Fawcett, 2006). Berikut ini penilaian AUC untuk

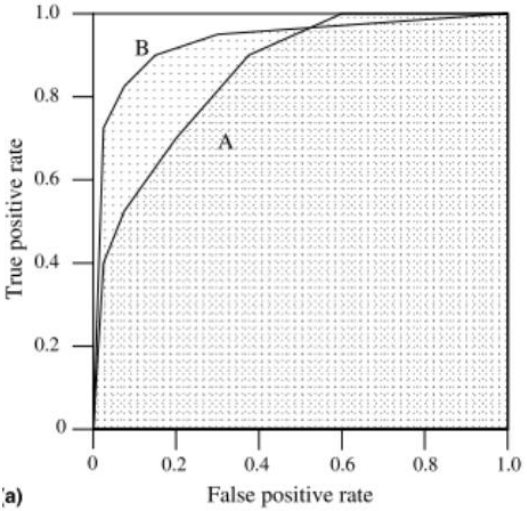
mengklasifikasikan model prediksi yang dihasilkan:

Tabel 2. 1 Klasifikasi Nilai AUC

Nilai AUC	Keterangan
0.9 ≤ AUC ≤ 1	Sangat Baik
0.8 ≤ AUC < 0.9	Baik
0.7 ≤ AUC < 0.8	Cukup Baik
0.6 ≤ AUC < 0.7	Cukup
0.5 ≤ AUC < 0.6	Tidak Baik

(Sumber: Fawcett, 2006)

Tabel 2. 1 merupakan cara untuk mengevaluasi AUC (*Area Under the Curve*) dalam klasifikasi model prediksi.



Gambar 2. 4 Contoh Perbandingan Nilai AUC

(Sumber: Fawcett, 2006)

Gambar 2. 4 menampilkan contoh perbandingan nilai AUC dari kedua metode yaitu dari metode A dan B.

K. Confusion Matriks

Confusion Matrix merupakan tabel yang menunjukkan seberapa banyak informasi pengujian yang diurutkan secara tepat dan seberapa banyak informasi yang dicoba tidak dikelompokkan secara tepat. Estimasi pengukuran karakterisasi pada informasi pertama dan hasil informasi dari model pengelompokan dilakukan dengan memanfaatkan tabulasi silang (tab cross) yang berisi data tentang kelas informasi pertama yang dialamatkan pada tabel dan kelas informasi yang diprediksi dengan perhitungan dimasuk kan kedalam kolom pengelompokkan. Berikutnya adalah ilustrasi dari confusion matrix:

1. *True Positive* (TP), merupakan jumlah data yang benar dari kelas 1 dan dikategorikan sebagai kelas 1.
2. *True Negative* (TN), merupakan jumlah data yang benar dari kelas 0 dan dikategorikan sebagai kelas 0.
3. *False Positive* (FP), merupakan jumlah data dari yang salah kelas 0 dan dikategorikan sebagai kelas 1.
4. *False Negative* (FN), merupakan jumlah data yang salah dari kelas 1 dan dikelompokkan sebagai kelas 0.

Ketepatan klasifikasi harus dilihat dari nilai akuraasi yang disusun. Akurasi klasifikasi menunjukkan penyajian model klasifikasi secara umum, dengan asumsi jika semakin besar nilai akurasi semakin bagus performa model pengelompokkan model yang digunakan.

$$Akurasi\ klasifikasi = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%$$

(2.18)

Kemudian, pada saat itu juga dapat menghitung nilai error classification atau tingkat dugaan error yang merupakan ukuran penilaian yang digunakan untuk mengenali kemungkinan dari pengelompokkan yang dibuat oleh model klasifikasi. Semakin rendah nilai error classification, semakin baik hasil klasifikasinya. Rumus untuk menghitung error classification sebagai berikut:

$$Error\ klasifikasi = \frac{FP + FN}{TP + FN + TN + FP} \times 100\%$$

(2.19)

Semua pengklasifikasi mecoba untuk membuat model dengan nilai akurasi tinggi dan tingkat kesalahan yang rendah (Prasetyo, 2012).

L. Penelitian Terdahulu

Penelitian terdahulu penting bagi penulis sebagai acuan untuk menuju ke penelitian yang terbaru, agar dapat memilih Untuk memahami keterkaitan antara penelitian yang akan dilakukan penulis dengan penelitian sebelumnya, untuk menghindari pencurian sastra dan untuk membantu penelitian ini dalam kemajuan ilmu pengetahuan. Beberapa peneliti sebelumnya telah menginvestigasi topik ini dengan menggunakan regresi logistik biner dan algoritma klasifikasi C5.0. Berikut adalah beberapa studi terkait dengan penelitian ini:

Penelitian yang dilakukan oleh Ramdani (2022) mengenai perbandingan Metode Regresi Logistik Biner dan *Naive Bayes* pada klasifikasi status pengguna KB di Kelurahan Mangunharjo menghasilkan bahwa metode Regresi Logistik Biner adalah

pilihan yang lebih unggul untuk mengklasifikasikan status pengguna KB di Kelurahan Mangunharjo. Hal ini disebabkan, karena dalam regresi logistik biner, kurva hasilnya mendekati titik (0,1) dan memiliki nilai AUC yang lebih besar dari naïve bayes yaitu sebesar 0,9424. Perbedaan antara penelitian oleh Ramdani (2022) dengan penelitian ini adalah bahwa penelitian ini menggunakan metode Regresi Logistik Biner dan Algoritma C5.0 dengan data inflasi di Indonesia periode 2010 – 2023 dan dalam analisis data menggunakan aplikasi *R Studio*.

Penelitian oleh Pratiwi (2020) membandingkan klasifikasi Algoritma C5.0 dan *Classification And Regression Tree* dalam konteks data sosial Kepala Keluarga Masyarakat Desa Teluk Baru yang menghasilkan metode CART terbukti lebih efektif dalam mengklasifikasikan data rata-rata pendapatan masyarakat Desa Teluk Baru, dengan tingkat akurasi rata-rata mencapai 84,63%. Perbedaan penelitian oleh Pratiwi (2020) dengan penelitian ini yaitu dalam penelitian tersebut menggunakan Algoritma C5.0 dan *Classification And Regression Tree* dengan konteks data sosial Kepala Keluarga Masyarakat Desa Teluk Baru sedangkan paada penelitian ini menggunakan metode Regresi Logistik Biner dan Algoritm C5.0 dengan data inflasi di Indonesia periode 2010 – 2023 dan dalam analisis data menggunakan aplikasi *R Studio*.

Penelitian oleh Amanda dan Marhaeni (2023) yang berisi tentang analisis data inflasi menggunakan metode Regresi Logistik yang menghasilkan bahwa menganalisis data inflasi menggunakan

metode Regresi Logistik memperoleh nilai akurasi yang cukup tinggi yaitu sebesar 75%. Faktor-faktor yang memengaruhi inflasi meliputi indeks harga konsumen, nilai tukar, dan jumlah uang yang beredar. Pada dasarnya, inflasi dipengaruhi oleh uang dalam konteks yang bervariasi. Perbedaan antara penelitian tersebut dan penelitian saat ini adalah bahwa penelitian tersebut hanya menggunakan satu metode yaitu metode Regresi Logistik sedangkan pada penelitian ini membandingkan antara metode Regresi Logistik Biner dan Algoritma C5.0 dengan data inflasi di Indonesia periode 2010 – 2023 dan dalam analisis data menggunakan aplikasi *R Studio*.

Penelitian oleh Umaroh (2020) tentang perbandingan metode Regresi Logistik Biner dan CART untuk mengklasifikasikan status kesejahteraan rumah tangga di Kota Batu. Hasilnya menunjukkan bahwa metode Regresi Logistik Biner memiliki tingkat akurasi klasifikasi sebesar 94,69%, sementara metode CART memiliki tingkat akurasi klasifikasi sebesar 89,36% jadi kesimpulannya metode Regresi Logistik Biner adalah model terbaik untuk memprediksi faktor kesejahteraan rumah tangga di Kota Batu. Perbedaan antara penelitian tersebut dengan penelitian saat ini adalah bahwa penelitian tersebut membandingkan metode Regresi Logistik Biner dan metode CART dalam mengklasifikasikan status kesejahteraan rumah tangga di Kota Batu, sementara penelitian ini menggunakan metode Regresi Logistik Biner dan Algoritma

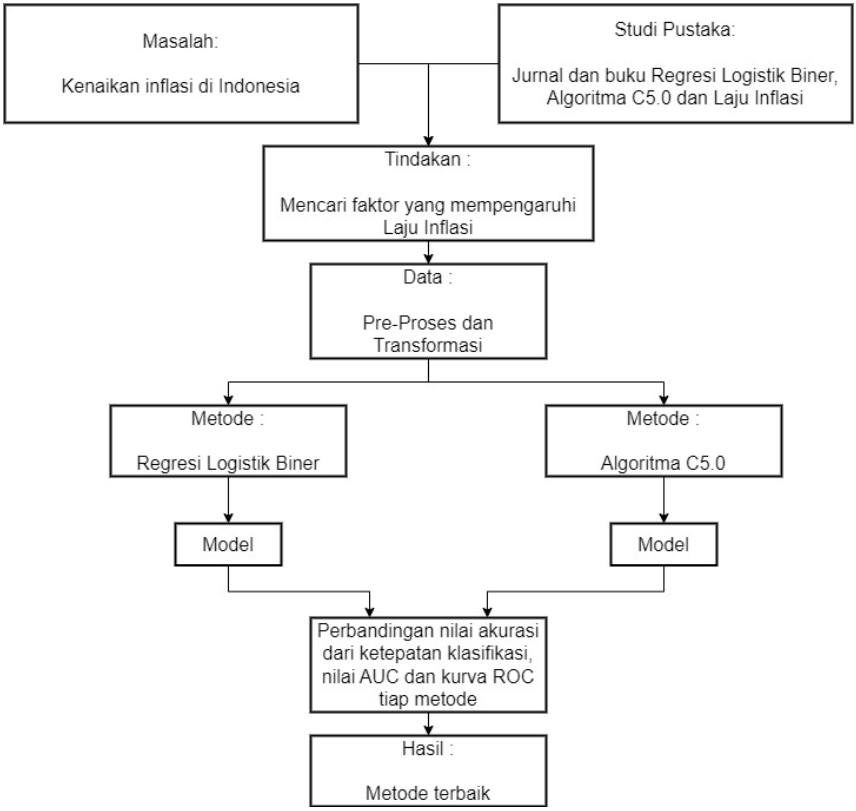
C5.0 untuk menentukan metode yang lebih efektif dalam mengklasifikasikan laju inflasi di Indonesia dari tahun 2010 hingga 2023. Analisis data dilakukan menggunakan aplikasi *R Studio*.

Penelitian oleh Hanin (2023) membahas penerapan Algoritma C5.0 dan metode SMOTE dalam klasifikasi status kesejahteraan rumah tangga yang menghasilkan nilai accuracy sebesar 76,32% dengan nilai sensitivity sebesar 79,75% dan nilai specificity sebesar 72,6%. Tingkat ketepatan klasifikasi juga diukur dengan metode AUC-ROC, nilai AUC yaitu sebesar 0,7617. Nilai AUC masuk dalam kategori fair classification. Perbedaan antara penelitian sebelumnya dan penelitian ini adalah bahwa penelitian sebelumnya menggunakan Algoritma C5.0 dan metode SMOTE untuk mengklasifikasikan status kesejahteraan rumah tangga, sedangkan penelitian ini membandingkan metode Regresi Logistik Biner dengan metode Algoritma C5.0 untuk menentukan metode mana yang lebih efektif dalam mengklasifikasikan laju inflasi di Indonesia periode 2010 – 2023. Analisis data dilakukan dengan menggunakan aplikasi *R Studio*.

M. Kerangka Berfikir

Kerangka berpikir penelitian sangat penting dalam kelangsungan penelitian karena dengan mengulas studi sebelumnya atau ulasan literatur yang relevan dengan penelitian, peneliti dapat lebih memahami gambaran besar dari topik atau pertanyaan yang sedang diteliti. Peneliti harus memberikan

gambaran singkat tentang proses persiapan penelitian ini dengan kerangka berpikir sebagai berikut:



Gambar 2. 5 Kerangka Berfikir

BAB III

METODE PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini berasal dari sumber data sekunder yaitu dokumen-dokumen yang diperoleh dari https://www.bps.go.id/dan https://www.bi.go.id/.

B. Metode Pengumpulan Data

Dalam penelitian ini, data dikumpulkan dengan cara dokumentasi. Salah satu metode untuk menemukan data (informasi) yang diperlukan untuk penelitian adalah analisis dokumentasi. Walaupun data (informasi) yang diperoleh dari telaah dokumentasi ini tidak tergolong sebagai data primer, Namun, meskipun termasuk dalam kategori data sekunder, dokumen ini memiliki signifikansi penting, terutama untuk data yang digunakan dalam penelitian (Abdullah, 2015). Dokumen ini mencakup data mengenai laju inflasi di Indonesia dari tahun 2010 hingga 2023.

C. Populasi dan Sampel

Populasi dalam penelitian ini adalah total jumlah data inflasi yang diklasifikasikan dalam laju inflasi di Indonesia dari tahun 2010 hingga 2023, yaitu sebanyak 168 data Inflasi.

Ukuran sampel akan dihitung menggunakan Rumus Slovin, yang mengakomodasi toleransi terhadap kesalahan pengambilan sampel yang masih dapat diterima, seperti misalnya 5%. Slovin menyatakan nilai toleransi ini sebagai persentase. Rumus Slovin

yang digunakan adalah sebagai berikut (Arieska, 2018):

$$n = \frac{N}{1+(N\alpha^2)} \tag{3.1}$$

Dimana:

n : ukuran sampel.

N : ukuran populasi.

α : toleransi ketidak telitian dalam persen (%).

$$n = \frac{168}{1+(168\times 0,05^2)}$$

$$n = \frac{168}{1+(168\times 0,0025)}$$

$$n = \frac{168}{1+(0,42)}$$

$$n = 118$$

Berdasarkan perhitungan menggunakan Rumus Slovin, jumlah sampel dalam penelitian ini adalah sebanyak 118 sampel.

Teknik pengambilan sampel yang digunakan dalam penelitian ini adalah *Simple Random Sampling*, karena populasi yang ada bersifat homogen, sehingga sampel yang dipilih secara acak dapat mewakili populasi tersebut. *Simple Random Sampling* merupakan salah satu teknik sampling yang paling sederhana dan banyak digunakan. Pemilihan responden dilakukan berdasarkan angka acak, dan jumlah responden yang terpilih sesuai dengan jumlah sampel yang ditentukan (Arieska, 2018). Berdasarkan metode *Simple Random Sampling*, maka sampel yang akan digunakan berjumlah 118 untuk mengukur tingkat inflasi.

D. Variabel Penelitian

Variabel penelitian bisa menyebabkan perbedaan atau variasi pada nilai tertentu. Dalam penelitian ini, variabel terbagi menjadi variabel bebas atau independen dan variabel terikat atau dependen.

1. Variabel Independen

   Variabel independen atau variabel bebas adalah variabel yang peneliti percaya akan memengaruhi variabel dependen atau variabel terikat dalam penelitian ini. Variabel independen atau variabel bebas yang digunakan dalam penelitian ini adalah:

Tabel 3. 1 Variabel Independen

No	Jenis Variabel	Variabel	Keterangan
1.	Independen	Nilai Tukar (X_1)	Nilai Tukar uang domestik terhadap uang dolar Amerika
2.	Independen	Suku Bunga (X_2)	Suku Bunga yang ditetapkan oleh pemerintah
3.	Independen	Jumlah Uang Beredar (X_3)	Jumlah Uang Kartal yang beredar di Indonesia

2. Variabel dependen atau variabel terikat adalah variabel yang diperkirakan oleh peneliti akan terpengaruh oleh variabel lain dalam suatu studi (Ahyar, dkk., 2020). Pada penelitian ini merujuk pada penelitian yang dilakukan oleh Amanda dan Marhaeini (2023) untuk menentukan kategori inflasi naik dan turun yaitu dengan melihat batas bawah dari total populasi

inflasi. Sesuai perhitungan tersebut menghasilkan batas bawah dari populasi inflasi yaitu sebesar 2,42%. Oleh karena itu, variabel dependen dalam penelitian ini terdiri dari 2 kategori yaitu:

a. Kategori Inflasi naik jika x ≥ 2,42% (1)
b. Kategori Inflasi turun jika x < 2,42% (0)

Tabel 3. 2 Variabel Dependen

Jenis Variabel	Variabel	Keterangan
Dependen	Laju Inflasi (x)	1 = Inflasi ≥ 2,42 % 0 = inflasi < 2,42 %

E. Metode Analisis Data

Teknik analisis data yang diterapkan dalam penelitian ini adalah menggunakan metode Regresi Logistik Biner serta Algoritma C5.0. untuk mengklasifikasikan laju inflasi bulanan dengan menggunakan variabel bebas seperti Nilai Kurs, Tingkat Suku Bunga, dan Jumlah Uang Beredar. *Microsoft Excel* dan *R Studio* adalah program komputer yang digunakan dalam penelitian ini.

Langkah-langkah dalam penelitian ini adalah sebagai berikut:

1. Analisis Statistika Deskriptif

   Analisis statistika deskriptif mengumpulkan dan menyajikan kumpulan data untuk memberikan informasi bermanfaat tentang subjek penelitian. Penyajian data

dilakukan dengan membuat *cross tabulation* dan diagram lingkaran menggunakan bantuan dari *Software Microsoft Excel*.

2. Proses Data

a. Pembersihan data

Sebelum melakukan klasifikasi menggunakan Regresi Logistik Biner dan Algoritma C5.0, langkah pertama adalah membersihkan data menggunakan *R Studio*.

b. Membagi Data *Training* dan Data *Testing*

Langkah kedua adalah membagi data training dan data testing dengan menggunakan metode random split diikuti dengan melakukan pengacakan untuk memastikan bahwa setiap data memiliki kesempatan yang sama untuk digunakan baik sebagai data pelatihan maupun data pengujian. Ini dilakukan dengan menggunakan program *Microsoft Excel.*

3. Analisis Regresi Logistik Biner

Digunakan untuk mengumpulkan data tentang variabel independen atau variabel bebas mana pun yang secara signifikan memengaruhi variabel dependen, dan juga diklasifikasikan menggunakan metode Regresi Logistik Biner pada aplikasi *R Studio*, aplikasi yang membantu untuk menentukan tingkat akurasidari klasifikasi tersebut. Langkah-langkah analisis menggunakan regresi logistik biner berikut:

a. Memasukkan data *training* atau data pelatihan

b. Melakukan uji multikolinieritas

c. Menghitung estimasi parameter

d. Uji Wald

e. Uji Likehood

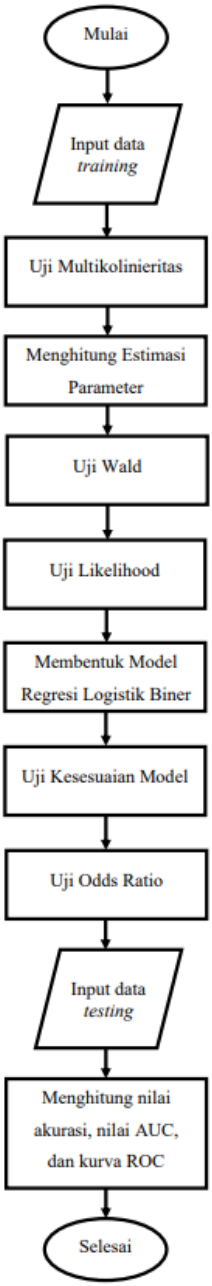
f. Membentuk model Regresi Logistik Biner

g. Melakukan uji kesesuaian model

h. Menghitung nilai Odds Ratio

i. Memasukan data *testing* atau data uji

j. Menghitung nilai akurasi, nilai AUC dan kurva ROC.

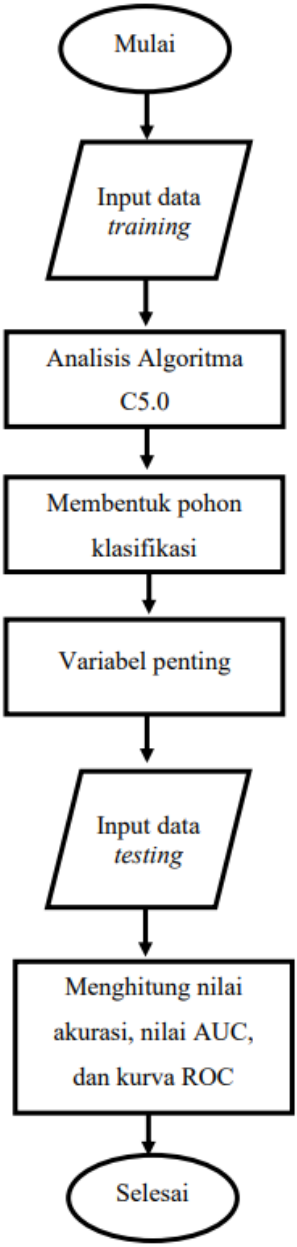


Gambar 3. 1 *Flowchart* Uji Regresi Logistik Biner

4. Analisis Algoritma C5.0

Proses analisis metode Algoritma C5.0 menggunakan aplikasi *R Studio*, untuk membantu menentukan tingkat akurasi dari klasifikasi laju inflasi. Berikut langkah- langkah menganalisis dengan menggunakan metode algoritma C5.0:

a) Input data *training*
b) Analisis Algoritma C5.0
c) Membentuk Pohon Klasifikasi
d) Mencari Variabel Penting
e) input data *testing*
f) Menghitung nilai akurasi, nilai AUC, dan kurva ROC.



Gambar 3. 2 *Flowchart* Uji Algoritma C5.0

5. Perbandingan Model Metode Regresi Logistik Biner dan Algoritma C5.0

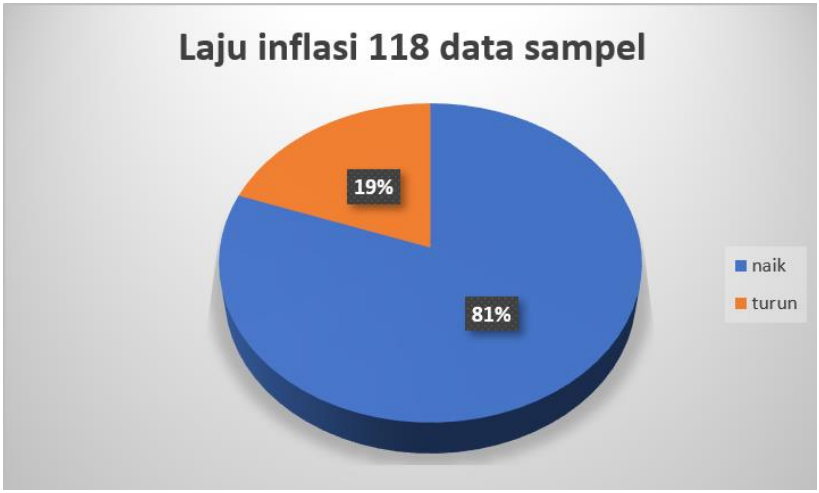
Setelah langkah-langkah dilakukan menggunakan metode regresi logistik biner dan algoritma C5.0, hasil klasifikasi yang tepat, nilai akurasi, dan kurva ROC dibandingkan dan diputuskan metode mana yang lebih baik untuk pengklasifikasian laju inflasi.

BAB IV

HASIL DAN PEMBAHASAN

A. Analisis Statistika Deskriptif

Inflasi di Indonesia bisa diklasifikasikan menjadi kategori inflasi naik dan inflasi turun. Gambaran laju inflasi di Indonesia disajikan dalam gambar berikut:



Gambar 4. 1 Laju Inflasi di Indonesia

Dari gambar 4.1 dapat dilihat bahwa data laju inflasi di Indonesia dari tahun 2010 hingga 2023 mencakup 168 data dengan sampel sebesar 118 yang telah diteliti , bisa diketahui bahwa inflasi naik di Indonesia lebih mendominasi yang artinya Indonesia sering mengalami kenaikan inflasi. Dengan data sampel yang telah di teliti maka diperoleh sebanyak 81% atau sejumlah 95 bulan dalam klasifikasi inflasi naik. Sedangkan 19% atau sejumlah 23 bulan dalam klasifikasi inflasi turun.

1. Deskripsi Nilai Tukar

Nilai tukar adalah suatu variabel independen yang diasumsikan dapat mempengaruhi laju inflasi di Indonesia. Dikarenakan data nilai tukar merupakan data numerik, maka dari itu menggunakan rata-rata dari variabel nilai tukar untuk melakukan analisis deskriptif terhadap inflasi. Berdasarkan data nilai tukar yang berjumlah 118 sampel mendapatkan nilai rata-rata sebesar Rp. 12836,62.

Selanjutnya deskripsi laju inflasi berdasarkan rata-rata nilai tukar ditunjukan pada tabel berikut:

Tabel 4. 1 Nilai Tukar

No	Nilai Tukar	Inflasi Naik	Infllasi Turun	Total
1	X_1 ≤ Rp. 12.837	42 (44%)	0 (0%)	42 (36%)
2	X_1 > Rp. 12.837	53 (56%)	23 (100%)	76 (64%)
	Total	95 (100%)	23 (100%)	118 (100%)

Berdasarkan Tabel 4.1 menunjukan bahwa nilai tukar berdasarkan laju inflasi, sebanyak 95 bulan merupakan kategori inflasi naik dengan 44% atau sebanyak 42 bulan dengan nilai tukar X_1 ≤ Rp. 12.837, dan 53% atau sebanyak 53 bulan dengan nilai tukar X_1 > Rp. 12.837. Sementara itu, pada kategori inflasi turun terdapat sebanyak 23 bulan yang termasuk inflasi turun.

Sebanyak 0% atau sebanyak 0 bulan dengan nilai tukar X_1 ≤ Rp. 12.837, dan 100% atau sebanyak 23 bulan dengan nilai tukar X_1 > Rp. 12.837. Sehingga dapat disimpulkan bahwa sebagian besar inflasi naik di Indonesia memiliki nilai tukar X_1 > Rp. 12.837.

2. Deskripsi Suku Bunga

Suku bunga adalah suatu variabel independen yang diasumsikan dapat mempenngaruhi laju inflasi. Dikarenakan data suku bunga merupakan data numerik, maka dari itu menggunakan rata-rata dari variabel suku bunga untuk melakukan analisis deskriptif terhadap inflasi. Berdasarkan data suku bunga yang berjumlah 118 sampel mendapatkan nilai rata-rata sebesar 5,63%.

Selanjutnya deskripsi laju inflasi berdasarkan suku bunga ditunjukan pada tabel berikut:

Tabel 4. 2 Suku Bunga

No	Suku Bunga	Inflasi Naik	Infllasi Turun	Total
1	$X_2 \leq 5{,}63\%$	28 (29%)	22 (96%)	50 (42%)
2	$X_2 > 5{,}63\%$	67 (71%)	1 (4%)	68 (58%)
	Total	95 (100%)	23 (100%)	118 (100%)

Dari Tabel 4.2 menunjukan bahwa suku bunga berdasarkan laju inflasi, sebanyak 95 bulan merupakan kategori inflasi naik

dengan 29% atau sebanyak 28 bulan dengan suku bunga X_2 ≤ 5,63%, dan 71% atau sebanyak 67 bulan dengan suku bunga X_2 > 5,63%. Sementara itu, pada kategori inflasi turun terdapat sebanyak 23 bulan yang termasuk inflasi turun. Sebanyak 96% atau sebanyak 22 bulan dengan suku bunga X_2 ≤ 5,63%, dan 4% atau sebanyak 1 bulan dengan suku bunga X_2 > 5,63%. Sehingga dapat disimpulkan bahwa sebagian besar inflasi naik di Indonesia memiliki suku bunga X_2 > 5,63%.

3. Deskripsi Jumlah Uang Beredar

Jumlah uang beredar dianggap sebagai variabel independen yang dapat mempengaruhi tingkat inflasi. Karena data jumlah uang beredar berupa data numerik, analisis deskriptif terhadap inflasi dilakukan dengan menggunakan nilai rata-rata dari variabel jumlah uang beredar. Berdasarkan data jumlah uang beredar yang berjumlah 118 sampel mendapatkan nilai rata-rata sebesar 532471,14 milyar rupiah.

Selanjutnya deskripsi laju inflasi berdasarkan jumlah uang beredar ditunjukan pada tabel berikut:

Tabel 4. 3 Jumlah Uang Beredar

No	Jumlah Uang Beredar	Inflasi Naik	Infllasi Turun	Total
1	X_3 ≤ 532471,14 milyar rupiah	61 (64%)	0 (0%)	61 (52%)

2	X_3 > 532471,14 milyar rupiah	34 (36%)	23 (100%)	57 (48%)
	Total	95 (100%)	23 (100%)	118 (100%)

Berdasarkan Tabel 4.3 menunjukan bahwa jumlah uang beredar berdasarkan laju inflasi, sebanyak 95 bulan merupakan kategori inflasi naik dengan 64% atau sebanyak 61 bulan dengan jumlah uang beredar X_3 ≤ 532471,14 milyar rupiah, dan 36% atau sebanyak 34 bulan dengan jumlah uang beredar X_3 > 532471,14 milyar rupiah. Sementara itu, pada kategori inflasi turun terdapat sebanyak 23 bulan yang termasuk inflasi turun. Sebanyak 0% atau sebanyak 0 bulan dengan jumlah uang beredar X_3 ≤ 532471,14 milyar rupiah, dan 100% atau sebanyak 23 bulan dengan jumlah uang beredar X_3 > 532471,14 milyar rupiah. Sehingga dapat disimpulkan bahwa sebagian besar inflasi naik di Indonesia memiliki jumlah uang beredar X_3 ≤ 532471,14 milyar rupiah.

B. *Preprocessing* Data (Persiapan Data)

1. Pembersihan data

Sebelum melakukan proses klasifikasi, langkah pertama yang perlu dilakukan yaitu melakukan pembersihan data. Tujuan dari pembersihan data yaitu untuk mengurangi munculnya misinformasi akibat data yang belum dibersihkan.

Pembersihan data dilakukan dengan bantuin *R Studio*. Berdasarkan hasil analisis *R Studio* dengan sintaks di lampiran 4. Maka didapatkan hasil seperti berikut:

Tabel 4. 4 Pembersihan Data

Variabel	*Missing Value*
Y (inflasi)	0
X_1 (nilai tukar)	0
X_2 (suku bunga)	0
X_3 (jumlah uang beredar)	0

terlihat pada hasil analisis diatas bahwa data sampel yang digunakan pada penelitian ini tidak terdapat *missing value* dan duplikat data.

2. Pembagian data *training* dan *testing*

Setelah melakukan pembersihan data lanjut membagi data *training* dan *testing,* kemudian dilakukan pengacakan terlebih dulu supaya setiap data mempunyai kesempatan yang sama untuk menjadi data *training* dan *testing.* Penelitian ini menggunakan perbandingan data *training* dan *testing* dengan proporsi 80:20 dikarenakan untuk menggunakan perbandingan data *training* dan *testing* harus menggunakan komposisi yang pas supaya menghindari *overfitting* (Musu, dkk., 2021). Berikut merupakan contoh perhitungan cara

menentukan banyaknya data *training* dan *testing* dengan menggunakan perbandingan 80:20.

$$\text{Jumlah data } training = 80\% \times 118$$
$$= 94$$
$$\text{Jumlah data } testing = 20\% \times 118$$
$$= 24$$

Berdasarkan perhitungan tersebut dapat dilihat bahwa data yang masuk dalam data *training* sebanyak 94 data dan yang masuk dalam data *testing* sebanyak 24 data.

C. Analisis Regresi Logistik Biner

1. Estimasi Parameter

Prinsip dari MLE (*Maximum Likelihood Estimator*) menyatakan bahwa MLE dapat digunakan dengan memaksimalkan nilai β dari persamaan (2.6). Namun, lebih mudah dan matematis dalam bekerja dengan membuat log dari persamaan (2.6). *Log Likelihood* dapat didefinisikan sebagai berikut:

$$L(\beta) = \ln\left[l(\beta)\right] = \Sigma_{i=1}^{n}\left\{y_i \ln\left[\pi(x_i)\right] + (1-y_i)\ln\left[1-\pi(x_i)\right]\right\} \quad (4.1)$$

$$= \Sigma_{i=1}^{n}\left\{y_i \ln\left[\pi(x_i)\right] + (1-y_i)\ln\left[1-\pi(x_i)\right]\right\}$$
$$= \Sigma_{i=1}^{n} y_i \ln\left(\frac{e^{x_i}}{1+e^{x_i}}\right) + (1-y_i)\ln\left(1-\left(\frac{e^{x_i}}{1+e^{x_i}}\right)\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] + \left(1-y_i\right) \ln\left(1 - \left(\frac{e^{x_i}}{1+e^{x_i}}\right)\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] + \left(1-y_i\right) \ln\left(\frac{1+e^{x_i}-e^{x_i}}{1+e^{x_i}}\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] + \left(1-y_i\right) \ln\left(\frac{1}{1+e^{x_i}}\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] + \left(1-y_i\right) \left[\ln 1 - \ln\left(1+e^{x_i}\right) \right]$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] - \left(1-y_i\right) \ln\left(1+e^{x_i}\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right] - \ln\left(1+e^{x_i}\right) + y_i \ln\left(1+e^{x_i}\right)$$

$$= \Sigma_{i=1}^{n} y_i \left[\ln\left(e^{x_i}\right) - \ln\left(1+e^{x_i}\right) \right]$$

$$= \Sigma_{i=1}^{n} \left(y_i x_i\right) - \ln\left(1+e^{x_i}\right)$$

Karena $g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = x_i\beta$, dengan $\beta = \left(\beta_1, \beta_2, \beta_3, ..., \beta_p\right)$

dan $x_i = \left(x_1, x_2, x_3, ..., x_p\right)$ adalah nilai dari rediktor ke-1 sampai ke-*p* dari pengamatan ke-*i* sedemikian rupa sehingga $g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \Sigma_{j=0}^{p} x_{ij}\beta_j$, maka:

$$L(\beta) = \Sigma_{i=1}^{n} y_i \left(\Sigma_{j=0}^{p} x_{ij}\beta_j\right) - \Sigma_{i=1}^{n} \ln\left(1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}\right) \qquad (4.2)$$

Selanjutnya persamaan (4.2) diturunkan terhadap $(\beta_0, \beta_1, \beta_2, ..., \beta_p)$. Turunan itu diakhiri dengan menyamakan dengan nol dengan tujuan memproleh hasil sebagai berikut:

Turunan fungsi log likelihood terhadap β_0 :

$$\frac{\partial L(\beta)}{\partial \beta_0} = \frac{\partial\left(\Sigma_{i=1}^{n} y_i x_{i0}\beta_0 - \Sigma_{i=1}^{n} \ln\left(1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}\right)\right)}{\partial \beta_0} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i0} - \Sigma_{i=1}^{n} x_{i0} \frac{e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}}{1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i0} - \Sigma_{i=1}^{n} x_{i0}\pi(x_i) = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i - \Sigma_{i=1}^{n} \pi(x_i) = 0$$

Turunan fungsi log likelihood terhadap β_1 :

$$\frac{\partial L(\beta)}{\partial \beta_1} = \frac{\partial\left(\Sigma_{i=1}^{n} y_i x_{i1}\beta_1 - \Sigma_{i=1}^{n} \ln\left(1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}\right)\right)}{\partial \beta_1} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i1} - \Sigma_{i=1}^{n} x_{i1} \frac{e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}}{1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i1} - \Sigma_{i=1}^{n} x_{i1}\pi(x_i) = 0$$

Turunan fungsi log likelihood terhadap β_2 :

$$\frac{\partial L(\beta)}{\partial \beta_2} = \frac{\partial\left(\Sigma_{i=1}^{n} y_i x_{i2}\beta_2 - \Sigma_{i=1}^{n} \ln\left(1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}\right)\right)}{\partial \beta_2} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i2} - \Sigma_{i=1}^{n} x_{i2} \frac{e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}}{1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{i2} - \Sigma_{i=1}^{n} x_{i2}\pi(x_i) = 0$$

Turunan fungsi log likelihood terhadap β_p :

$$\frac{\partial L(\beta)}{\partial \beta_p} = \frac{\partial\left(\Sigma_{i=1}^{n} y_i x_{ip}\beta_p - \Sigma_{i=1}^{n} \ln\left(1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}\right)\right)}{\partial \beta_p} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{ip} - \Sigma_{i=1}^{n} x_{ip} \frac{e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}}{1+e^{\Sigma_{j=0}^{p} x_{ij}\beta_j}} = 0$$

$$\Leftrightarrow \Sigma_{i=1}^{n} y_i x_{ip} - \Sigma_{i=1}^{n} x_{ip}\pi(x_i) = 0$$

Berdasarkan turunan pertama dari $L(\beta)$ terhadap $\left(\beta_0, \beta_1, \beta_2, ..., \beta_p\right)$, maka dapat diperoleh persamaan umum sebagai berikut:

$$\frac{\partial L(\beta)}{\partial \beta_p} = \Sigma_{i=1}^{n} y_i x_{ip} - \Sigma_{i=1}^{n} x_{ip}\pi(x_i)$$

$$\Leftrightarrow \frac{\partial L(\beta)}{\partial \beta_p} = \Sigma_{i=1}^{n} x_{ip}\left[y_i - \pi(x_i)\right] \qquad (4.3)$$

Estimasi β diperoleh dengan menuntaskan turunan dari fungsi log-likelihood untuk setiap batas β dalam kondisi (4.3). Kondisi (4.3) masih dalam struktur implisit, sehingga diperlukan teknik matematis untuk mengatasinya. Susunan diselesaikannya harus dilakukan dengan memanfaatkan teknik Newton Raphson. Teknik ini menangani kondisi nonlinier dalam perhitungan berulang untuk menemukan maksimalisasi suatu fungsi, seperti probabilitas log likelihood (Agresti, 2009).

Algoritma iterasi Newton Raphson untuk memperoleh nilai $\hat{\beta}$ melalui langkah- langkah berikut:

a. Menghitung nilai estimasi awal $(\beta_{(t)})$ untuk β dengan menggunakan metode kuadrat terkecil (OLS)

$$\beta_{(t)} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}$$

b. Menghitung $X^t(Y-\pi)$ dan X^tVX, selanjutnya menentukan invers dari X^tVX

c. Menentukan estimasi baru pada iterasi ke- t

$$\hat{\beta}_{(t+1)} = \beta_{(t)} + (X^tVX)^{-1}(X^t(Y-\pi))$$

d. Jika nilai estimasi konvergen, iterasi dapat dihentikan.

$$\hat{\beta}_{(t+1)} \cong \beta_{(t)} \text{ atau } \left\| \hat{\beta}_{(t+1)} - \beta_{(t)} \right\| < \varepsilon$$

dengan ε merupakan tingkat ketelitian yang ditunjukkan oleh hasil hitungan yang positif.

2. Penerapan Metode Regresi Logistik Biner Pada Laju Inflasi di Indonesia

Analisis klasifikasi laju inflasi di Indonesia ini yang pertama menggunakan Metode Regresi Logistik Biner dengan bantuan *R Studio*. Dengan data *training* dan data *testing* yang sudah ditentukan. Pengelompokan data *training* dan data *testing* yaitu 80% dan 20% yang berisi 94 data dan 24 data. Berikut adalah hasil analisis regresi logistik biner untuk mengidentifikasi variabel – variabel indepeden yang dapat mempengaruhi laju inflasi di Indonesia menggunakan *R Studio*.

1) Uji Multikolinieritas

Uji multikolinieritas digunakan untuk mengevaluasi apakah terdapat hubungan korelasi antara variabel independen dalam model regresi dengan menguji nilai VIF. Berikut adalah hasil evaluasi multikolinieritas.

H_0 = Ada gejala multikolinieritas antar variabel independent atau variabel bebas apabila (VIF > 10)

H_1 = Tidak ada gejala multikolinieritas antar variabel independent atau variabel bebas apabila (VIF < 10)

Tingkat signifikansi: α = 0,05

Daerah keputusan: Jika nilai VIF < 10, maka H_0 ditolak

Berdasarkan hasil analisis menggunakan *R Studio* dengan sintaks yang ada di lampiran 5. Dibawah ini adalah hasil uji nilai VIF sesuai di lampiran 9.

Tabel 4. 5 Nilai VIF

Variabel Response	Nilai VIF
X_1 (nilai tukar)	1,717434
X_2 (suku bunga)	1,096447
X_3 (jumlah uang beredar)	1,692692

Berdasarkan hasil pengujian VIF, ditemukan bahwa nilai VIF untuk setiap variabel independen lebih rendah dari 10 yang artinya H_0 ditolak, maka H_1 diterima. Maka dari itu, dapat disimpulkan bahwa tidak terdapat indikasi multikolinearitas antara variabel independen dalam dataset yang digunakan.

2) Estimasi Parameter

Berdasarkan hasil analisis menggunakan *R Studio* dengan sintaks yang ada di lampiran 5. Di bawah ini adalah hasil dari estimasi parameter sesuai di lampiran 9, yang kemudian akan dievaluasi untuk mengetahui seberapa signifikan.

Tabel 4. 6 Estimasi Parameter

Variabel	**Estimasi** β_i
Konstanta	-3,512e+00
X_1 (nilai tukar)	-1,379e-04
X_2 (suku bunga)	2,223e+00
X_3 (jumlah uang beredar)	-5,492e-06

3) Uji Wald

Uji Wald dilakukan untuk mengevaluasi signifikansi parsial parameter yang dilakukan terhadap setiap variabel independen secara terpisah. Pengaruh setiap variabel independen terhadap variabel dependen ditentukan dengan menggunakan uji parsial ini. Hasil uji signifikansi parsial parameter ditunjukkan di bawah ini.

Pengujian Hipotesis:

H_0 = Variabel independen tidak memiliki pengaruh yang signifikan terhadap variabel dependen.

H_1 = Variabel independen memiliki pengaruh yang signifikan terhadap variabel dependen.

Tingkat signifikansi: α = 0,05

Domain keputusan: jika $|W_j| > Z_{\alpha/2} = 1,96$ atau $p - value < \alpha$ maka H_0 ditolak

Statistik uji: $|W_i| = \frac{\beta_i}{SE\beta_i}$

Tabel 4. 7 Uji Wald

Variabel	**Estimasi** β_i	**SE**	**Wald**	***p - value***	**Keputusan**
Konstanta	-3,512e+00	9,292e+00	-0,378	0,705448	
X_1 (nilai tukar)	-1,379e-04	7,964e-04	-0,173	0,862536	Terima H_0
X_2 (suku bunga)	2,223e+00	6,286e-01	3,536	0,000407	Tolak H_0
X_3 (jumlah uang beredar)	-5,492e-06	5,803e-06	-0,946	0,343914	Terima H_0

Berdasarkan Tabel 4.7 hasil uji signifikansi parsial menunjukkan bahwa satu variabel independen berpengaruh signifikan terhadap laju inflasi di Indonesia yaitu variabel X_2 atau suku bunga dengan hasil nlai |2,223|> $Z_{\alpha/2} = 1,96$ atau 0,000407 < α maka H_0 ditolak. Artinya, variabel independen X_2 (suku bunga) berpengaruh signifikan secara parsial terhadap laju inflasi.

Sedangkan variabel independen lain seperti X_1 (nilai tukar) dan X_3 (jumlah uang beredar) tidak memiliki pengaruh secara parsial terhadap variabel dependen, yaitu

laju inflasi. Hal ini terlihat dari hasil uji wald yang lebih kecil dari nilai $Z_{\alpha/2}$ dan *p-value* lebih besar dari nilai α .

4) Uji Rasio *Likelihood*

Uji rasio *likelihood* digunakan untuk mengevaluasi secara menyeluruh signifikansi masing-masing parameter terhadap variabel independen. Pengujian ini juga dikenal sebagai uji G atau uji rasio likelihood maksimum. Berikut adalah hasil dari pengujian signifikansi parameter secara komprehensif.

H_0 = Secara simultan, variabel independen tidak memiliki pengaruh yang signifikan terhadap variabel dependen.

H_1 = Minimal satu variabel independen memiliki pengaruh yang signifikan secara simultan terhadap variabel dependen.

Tingkat signifikansi: α = 0,05

Domain keputusan: jika $G > X^2_{(\alpha;db)}$ atau $p-value < \alpha$ maka H_0 ditolak

Berdasarkan hasil analisis menggunakan aplikasi *R Studio* dengan sintaks yang ada di lampiran 5. Di bawah ini adalah hasil uji yang diperoleh sesuai di lampiran 9.

Tabel 4. 8 Uji Rasio Likelihood

X^2	db	*p-value*

54,481	3	8,858e-12

Berdasarkan Tabel 4.8 hasil uji rasio *likelihood* menunjukkan bahwa nilai $p-value = 8,858\times10^{-12}$ lebih kecil dari tingkat signifikansi α = 0,05. Hal ini mengindikasikan bahwa H_0 ditolak sehingga dapat disimpulkan bahwa setidaknya satu variabel independen memiliki pengaruh signifikan secara bersama-sama terhadap variabel dependen.

5) Model Regresi Logistik Biner

Berdasarkan hasil uji signifikansi parsial parameter yang tercantum dalam Tabel 4.6, dapat disimpulkan bahwa dari semua parameter yang diuji, hanya X_2 (suku bunga) yang menunjukkan signifikansi. Oleh karena itu, nilai koefisien regresi yang diperoleh adalah sebagai berikut:

Tabel 4. 9 Model RLB

Variabel	**Estimasi** β_i
Konstanta	-3,512e+00
X_2 (suku bnga)	2,223e+00

Di bawah ini adalah model regresi logistik biner yang dimasukan ke dalam bentuk logit.

$g(x) = -3,512 + 2,223X_2$

Sehingga diperoleh model akhir regresi logistik biner sebagai berikut:

$$\pi(x_i) = \frac{e^{(-3,512+2,223X_2)}}{1+e^{(-3,512+2,223X_2)}}$$

Berdasarkan model yang telah dibentuk, diketahui bahwa variabel yang mempengaruhi terhadap laju inflasi adalah variabel X_2 (suku bunga). Dapat dipahami bahwa tinggi rendahnya suku bunga yang ditetapkan oleh pemerintah membuat kecenderungan membuat laju inflasi naik. Dengan demikian dapat berdasarkan model Regresi Logistik Biner didapatkan suku bunga menjadi variabel atau faktor yang dapat berpengaruh terhadap laju inflasi di Indonesia.

6) Uji Kesesuaian Model

Uji kesesuaian model dilaksanakan untuk mengevaluasi apakah terdapat perbedaan yang signifikan antara hasil yang diamati dengan nilai yang diprediksi oleh model. Uji ini menggunakan uji Hosmer dan Lemeshow. Di bawah ini adalah hasil dari pengujian kesesuaian model yang telah diperoleh.

Pengujian Hipotesis:

H_0 = Tidak ada perbedaan yang signifikan antara hasil yang diamati dan kemungkinan hasil prediksi model (model yang diperoleh fit)

H_1 = Ada perbedaan yang signifikan antara hasil yang diamati pengamat dan kemungkinan hasil prediksi

model (model yang diperoleh tidak fit)

Tingkat signifikansi: α = 0,05

Domain keputusan: jika $\hat{C} > X^2_{(g-2,\alpha)}$ atau *p-value* > α maka H_0 diterima

Berdasarkan hasil analisis menggunakan *R Studio* dengan sintaks yang ada di lampiran 5. Di bawah ini adalah hasil uji yang diproleh sesuai di lampiran 9.

Tabel 4. 10 Uji Hosmer dan Lemeshow

X^2	db	*p-value*
10,178	8	0,2528

Bedasarkan hasil uji Hosmer dan Lemeshow dapat nilai $p-value = 0,2528$ melebihi taraf signifikansi α = 0,05. Ini menunjukan bahwa H_0 diterima. Dengan demikian, dapat disimpulkan bahwa tidak ada perbedaan yang signifikan antara hasil yang diamati dan prediksi atau model yang diperoleh.

7) *Odds Ratio*

Odds Ratio adalah ukuran kecenderungan antara kategori-kategori berbeda dari variabel independen yang bersifat kualitatif. Berikut adalah hasil *Odds Ratio* dari variabel independen yang signifikan dalam model.

Berdasarkan hasil analisis menggunakan *R Studio* dengan sintaks yang ada di lampiran 5. Berikut adalah hasil yang diperoleh sesuai di lampiran 9.

Tabel 4. 11 *Odds Ratio*

Variabel	Estimasi β_i	*Odds Ratio*
X_2	2,223	$\psi = e^{2,223}$ = 9,23499

Berdasarkan Tabel 4.11, nilai *Odds Ratio* digunakan untuk menilai pengaruh variabel independen terhadap variabel dependen, yakni laju inflasi. Hasil interpretasinya adalah sebagai berikut:

Variabel X_2 (suku bunga) mempunyai nilai *Odds Ratio* sebesar 9,23499. Ini menunjukan bahwa suku bunga mempunyai cenderung untuk membuat laju inflasi naik.

8) Ketepatan Klasifikasi Regresi Logistik Biner

Akurasi klasifikasi model dapat dievaluasi menggunakan data pengujian. Berikut ini adalah hasil akurasi klasifikasi dari analisis pengujian yang dilakukan menggunakan *R Studio* dengan sintaks yang terlampir di Lampiran 5 dan hasil sesuai di lampiran 9.

Tabel 4. 12 Ketepatan Klasifikasi

Hasil Observasi	Taksiran Klasifikasi		Total
	Naik	Turun	
Naik	17	1	18
Turun	0	6	6
Total	17	7	24

Nilai akurasi klasifikasi menggunakan metode Regresi Logistik Biner dapat dihitung dengan menggunakan

perhitungan sebagai berikut:

$$Akurasi\ klasifikasi = \frac{TP + TN}{TP + FP + FN + TN}$$

$$= \frac{17 + 6}{17 + 0 + 1 + 6}$$

$$= 0,9583$$

$$Error\ klasifikasi = \frac{FP + FN}{TP + FP + FN + TN}$$

$$= \frac{0 + 1}{17 + 0 + 1 + 6}$$

$$= 0,0417$$

$$sensitifity = \frac{TP}{TP + FN}$$

$$= \frac{17}{17 + 1}$$

$$= 0,9444$$

$$specificity = \frac{TN}{TN + FP}$$

$$= \frac{6}{6 + 0}$$

$$= 1$$

Dari hasil perhitungan di atas memperoleh nilai akurasi sebesar 95,83% dan nilai error sebesar 4,17% sedangkan nilai sensitifity sebesar 94,44% dan specificity sebesar 100%. Artinya bahwa metode Regresi Logistik Biner dapat memprediksi hasil pengamata secara akurat hingga sebesar

95,83%.

9) Nilai AUC

Pengujian menggunakan AUC-ROC memberikan gambaran yang solid tentang kapabilitas model dalam membedakan antara kelas-kelas yang berbeda.

Nilai AUC

$$AUC = \frac{1}{2}(specificity + sensitivity)$$

$$= \frac{1}{2}(1 + 0,9444)$$

$$= 0,9722$$

Berdasarkan perhitungan diatas, dapat diamati bahwa nilai AUC sebesar 0,9722 atau 97,22%. Artinya nilai AUC menghasilkan nilai yang sangat baik.

D. Analisis Algoritma C5.0

Analisis klasifikasi pada laju inflasi di Indonesia yang kedua menggunakan Metode Algoritma C5.0 dengan menggunakan pembagian data *training* dan data *testing* yang sudah ditentukan yaitu diperoleh data *training* 80% berjumlah 94 data dan data *testing* 20% berjumlah 24 data. Berikut adalah hasil analisis algoritma C5.0 untuk mengidentifikasi variabel – variabel indepeden yang dapat mempengaruhi laju inflasi di Indonesia.

Dari persamaan (2.15) akan didapatkan nilai *entropy* sebagai berikut:

1) Menghitung *entropy* total:

$$Entropy(total) = \left(\left(-\frac{77}{94}\right)\times \log_2\left(\frac{77}{94}\right)+\left(-\frac{17}{94}\right)\times \log_2\left(\frac{17}{94}\right)\right) = 0,682$$

2) Menghitung *entropy* tiap kategori dari variabel Nilai Tukar (X_1):

a. x ≤ 12837

$$Entropy(x \le 12837) = \left(\left(-\frac{36}{36}\right)\times \log_2\left(\frac{36}{36}\right)+\left(-\frac{0}{36}\right)\times \log_2\left(\frac{0}{36}\right)\right) = 0$$

b. x > 12837

$$Entropy(x > 12837) = \left(\left(-\frac{41}{58}\right)\times \log_2\left(\frac{41}{58}\right)+\left(-\frac{17}{58}\right)\times \log_2\left(\frac{17}{58}\right)\right) = 0,873$$

3) Menghitung *entropy* tiap kategori dari variabel Suku Bunga (X_2):

a. x ≤ 5,63%

$$Entropy(x \le 5,63\%) = \left(\left(-\frac{22}{39}\right)\times \log_2\left(\frac{22}{39}\right)+\left(-\frac{17}{39}\right)\times \log_2\left(\frac{17}{39}\right)\right) = 0,988$$

b. x > 5,63%

$$Entropy(x > 5,63\%) = \left(\left(-\frac{55}{55}\right)\times \log_2\left(\frac{55}{55}\right)+\left(-\frac{0}{55}\right)\times \log_2\left(\frac{0}{55}\right)\right) = 0$$

4) Menghitung *entropy* tiap kategori dari variabel Jumlah Uang Beredar (X_3):

a. x ≤ 532471,14

$$Entropy(x \le 532471,14) = \left(\left(-\frac{47}{47}\right)\times \log_2\left(\frac{47}{47}\right)+\left(-\frac{0}{47}\right)\times \log_2\left(\frac{0}{47}\right)\right) = 0$$

b. x > 532471,14

$$Entropy(x > 532471,14) = \left(\left(-\frac{30}{47}\right)\times \log_2\left(\frac{30}{47}\right)+\left(-\frac{17}{47}\right)\times \log_2\left(\frac{17}{47}\right)\right) = 0,944$$

Dari persamaan (2.16) akan didapatkan nilai *gain* sebagai

berikut:

- $$Gain(X_1) = Entropy(total) - \left(\begin{array}{l} \left(\left(\frac{36}{94} \right) \times Entropy\left(x \le 12837 \right) \right) + \\ \left(\left(\frac{58}{94} \right) \times Entropy\left(x > 12837 \right) \right) \end{array} \right)$$

$$= 0,682 - \left(\left(\left(\frac{36}{94} \right) \times 0 \right) + \left(\left(\frac{58}{94} \right) \times 0,873 \right) \right) = 0,143$$

- $$Gain(X_2) = Entropy(total) - \left(\begin{array}{l} \left(\left(\frac{39}{94} \right) \times Entropy\left(x \le 5,63\% \right) \right) + \\ \left(\left(\frac{55}{94} \right) \times Entropy\left(x > 5,63\% \right) \right) \end{array} \right)$$

$$= 0,682 - \left(\left(\left(\frac{39}{94} \right) \times 0,988 \right) + \left(\left(\frac{55}{94} \right) \times 0 \right) \right) = 0,272$$

- $$Gain(X_3) = Entropy(total) - \left(\begin{array}{l} \left(\left(\frac{47}{94} \right) \times Entropy\left(x \le 532471,14 \right) \right) + \\ \left(\left(\frac{47}{94} \right) \times Entropy\left(x > 532471,14 \right) \right) \end{array} \right)$$

$$= 0,682 - \left(\left(\left(\frac{47}{94} \right) \times 0 \right) + \left(\left(\frac{47}{94} \right) \times 0,944 \right) \right) = 0,210$$

Dari persamaan (2.17) akan didapatkan nilai *gain ratio* sebagai berikut:

- $$Gain\ Ratio(X_1) = \frac{Gain(X_1)}{\left(Entropy\left(x \le 12837 \right) + Entropy\left(x > 12837 \right) \right)}$$

$$= \frac{0,143}{\left(0 + 0,873 \right)} = 0,164$$

- $$Gain\ Ratio(X_2) = \frac{Gain(X_2)}{\left(Entropy\left(x \leq 5{,}63\%\right) + Entropy\left(x > 5{,}63\%\right)\right)}$$

$$= \frac{0{,}272}{\left(0{,}988 + 0\right)} = 0{,}275$$

- $$Gain\ Ratio(X_3) = \frac{Gain(X_3)}{\left(Entropy\left(x \leq 532471{,}14\right) + Entropy\left(x > 532471{,}14\right)\right)}$$

$$= \frac{0{,}210}{\left(0 + 0{,}944\right)} = 0{,}222$$

Berdasarkan perhitungan diatas didapatkan bahwa variabel yang memiliki nilai *gain ratio* tertinggi adalah variabel suku bunga (X_2) yang memperoleh nilai sebesar 0,275. Oleh karena itu variabel suku bunga (X_2) ditetapkan sebagai *node akar* (*node* 1). Maka cabang untuk *node* akar ada 2 yaitu x ≤ 5,63% sebagai *node* 2 (22/17) dan x > 5,63% sebagai *node* 3 (55/0). *Node* 3 sebagai *node terminal* karena data sampel hanya ada di salah satu kelas. Langkah selanjutnya adalah menentukan *node internal* / *node* dalam dengan melihat nilai *gain ratio* terbesar dari variabel X_2 yaitu x ≤ 5,63% dihitung seperti langkah awal mencari *node* akar namun data yang digunakan adalah sisa data terhadap komposisi kelas x ≤ 5,63% yaitu sebanyak 39 data. Adapun hasil perhitungan *entropy, gain, dan gain ratio* disajikan sebagai berikut:

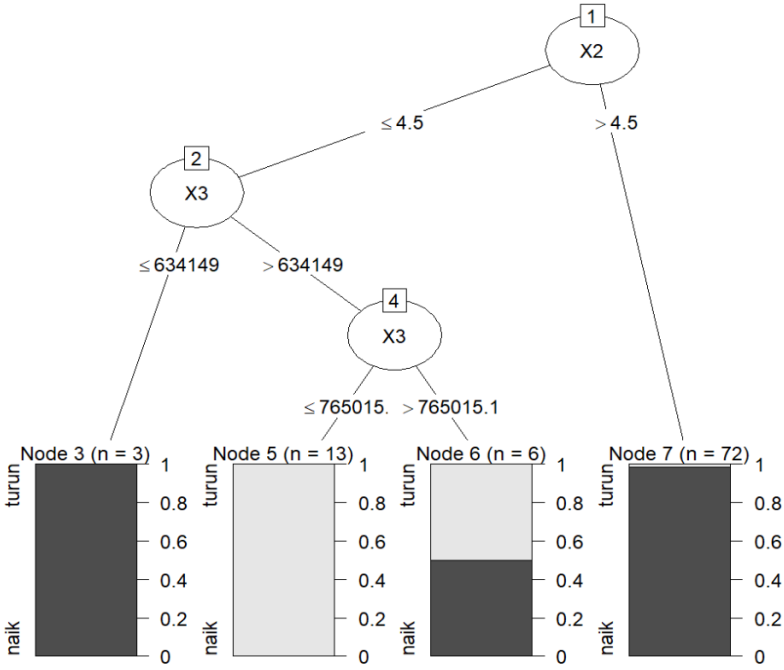
Tabel 4. 13 Perhitungan *Node Internal*

Variabel		Jumlah (S)	1	0	*Entropy*	*Gain*	*Gain Ratio*
X_2 (suku bunga)	≤5,63	39	22	17	0,988		
X_1 (nilai tukar)	≤12837	16	16	0	0	0,549	0,589
	>12837	23	15	8	0,932		
X_3 (jumlah uang beredar)	≤532471	15	15	0	0	0,559	0,615
	>532471	24	16	7	0,908		

Dari Tabel 4.13 diketahui bahwa yang memiliki nilai *gain ratio* tertinggi yaitu variabel X_3 sehingga ditetapkan sebagai *node internal*. Selanjutnya pembentukan pohon klasifikasi berdasarkan perhitungan nilai *gain ratio* tertinggi menggunakan *Software R Studio*.

1. Pohon Klasifikasi

Berdasarkan analisis menggunakan *R Studio* maka diperoleh pohon klasifikasi dibawah ini dengan variabel suku bunga (X_2) sebagai *node akar* (*node* 1). Dengan sintaks yang terlampir pada lampiran 7 dan hail sesuai di lampiran 10.



Gambar 4. 2 Pohon Klasifikasi

Pohon klasifikasi dapat dilihat pada Gambar 4.2. yang menunjukkan pohon klasifikasi dengan simpul utama (*root node*) yang ditandai dengan t_1 mewakili variabel X_2 (suku bunga). Pohon ini memiliki 2 simpul dalam (*internal nodes*) yaitu t_2 dan t_4. Serta 4 simpul terminal (*terminal nodes*) yaitu t_3, t_5, t_6, t_7. Dari pohon klasifikasi tersebut pada Gambar 4.2, dapat disimpulkan bahwa:

1) Apabila Kedalaman Rata - rata Terminal yang X_2 (suku bunga) > 4,5 inflasi cenderung naik.

2) Apabila Kedalaman Rata - rata Terminal yang X_2 (suku bunga) $\leq$ 4,5 dan X_3 (jumlah uang beredar) $\leq$ 634149 inflasi cenderung naik.

3) Apabila Kedalaman Rata - rata Terminal yang X_2 (suku bunga) $\leq$ 4,5 dan X_3 (jumlah uang beredar) >634149 yang X_3 (jumlah uang beredar) $\leq$ 765015 inflasi cenderung turun.

4) Apabila Kedalaman Rata - rata Terminal yang X_2 (suku bunga) $\leq$ 4,5 dan X_3 (jumlah uang beredar) >634149 yang X_3 (jumlah uang beredar) > 765015 inflasi cenderung turun.

2. Tingkat Kepentingan Variabel

Berdasarkan analisis menggunakan *R Studio* dengan sintaks yang terlampir pada lampiran 7. Diperoleh klasifikasi laju inflasi penelitian ini melibatkan dua variabel prediksi. Dalam pohon klasifikasi, kita dapat mengidentifikasi variabel yang memiliki peran paling signifikan dalam klasifikasi tingkat inflasi di Indonesia. Perhitungan tingkat kepentigan variabel pada klasifikasi berdasarkan hasil analisis menggunakan *R Studio* dengan sintaks yang ada di lampiran 7 dan hasil sesuai di lampiran 10 sebagai berikut:

$$Varlmp = \frac{nilai\ overall\ tiap\ variabel}{total\ overall} x100\%$$

1) X_1 (nilai tukar)

$$Varlmp = \frac{0}{123,40} x100\% = 0\%$$

2) X_2 (suku bunga)

$$Varlmp = \frac{100}{123,40} x100\% = 81,04\%$$

3) X_3 (jumlah uang beredar)

$$Varlmp = \frac{23,40}{123,40} x100\% = 18,96\%$$

Tabel 4. 14 Tingkat Kepentingan Variabel

No	Variabel	Overall	Tingkat Kepentingan
1.	X_2	100	81,04%
2.	X_3	23,40	18,96%
3.	X_1	0	0%

Dari Tabel 4.14, dapat disimpulkan bahwa variabel yang paling signifikan yaitu X_2 (suku bunga) dengan tingkat kepentingan 81,04%, kemudian variabel X_3 (jumlah uang beredar) dengan presentase 18,96%. Sedangkan variabel X_1 (nilai tukar) memperoleh presentase 0%. Dapat disimpulkan bahwa variabel X_2 sangat berpengaruh terhadap kenaikan laju inflasi di Indonesia. Dengan tinggi rendahnya suku bunga

yang ditetapkan oleh pemerintah dapat membuat kecenderungan terhadap kenaikan laju inflasi. Maka dari itu, dapat dikatakan bahwa model Algoritma C5.0 mendapatkan variabel atau faktor yang dapat berpengaruh terhadap laju inflasi di Indonesia yaitu suku bunga.

3. Tingkat Akurasi Ketepatan Klasifikasi

Akurasi klasifikasi laju inflasi dapat dinilai dari nilai *accuracy*. Selain itu, nilai *sensitivity* mengukur kemampuan model dalam mengidentifikasi inflasi turun, sementara nilai *specificity* mengukur kemampuan model dalam mengidentifikasi inflasi naik. Pengujian akurasi klasifikasi menggunakan 24 data *testing*.

Tabel 4. 15 *Confussion Matrix*

Actual Class	*Predicted Class*		Total
	Turun	Naik	
Turun	4	2	6
Naik	0	18	18
Total	4	20	24

Nilai akurasi klasifikasi menggunakan metode Algoritma C5.0 dapat dihitung dengan menggunakan perhitungan sebagai berikut:

$$Akurasi\ klasifikasi = \frac{TP + TN}{TP + FP + FN + TN}$$

$$= \frac{18 + 4}{18 + 2 + 0 + 4}$$

$$= 0,9167$$

$$Error\ klasifikasi = \frac{FP + FN}{TP + FP + FN + TN}$$

$$= \frac{2 + 0}{18 + 2 + 0 + 4}$$

$$= 0,0833$$

$$sensitifity = \frac{TP}{TP + FN}$$

$$= \frac{18}{18 + 0}$$

$$= 1$$

$$specificity = \frac{TN}{TN + FP}$$

$$= \frac{4}{4 + 2}$$

$$= 0,6667$$

Dari hasil perhitungan di atas memperoleh nilai akurasi sebesar 91,67% dan nilai error sebesar 8,33% sedangkan nilai sensitifity sebesar 100% dan specificity sebesar 66,67%. Artinya bahwa metode Algoritma C5.0 dapat memprediksi hasil pengamatan secara akurat hingga sebesar 91,67%.

4. Nilai AUC

Pengujian menggunakan AUC-ROC memberikan indikasi yang jelas tentang kemampuan model dalam membedakan antara kelas-kelas.

Nilai AUC

$$AUC = \frac{1}{2}(specificity + sensitivity)$$

$$= \frac{1}{2}(1 + 0,6667)$$

$$= 0,8334$$

Berdasarkan perhitungan yang telah dilakukan, dapat diamati bahwa nilai AUC sebesar 0,8334 atau 83,34%. Artinya nilai AUC menghasilkan nilai yang baik.

E. Perbandingan Metode Regresi Logistik Biner dan Algoritma C5.0

Dari hasil analisis regresi logistik biner dan algoritma C5.0 diiperoleh hasil akurasi klasifikasi, nilai AUC dan Kurva ROC pada klasifikasi laju inflasi di Indonesia. Berikut adalah perbandingan dari dua metode yang telah dipergunakan.

1. Perbandingan Ketepatan Klasifikasi

Berikut adalah perbandingan keakuratan klasifikasi antara regresi logistik biner dan algoritma C5.0.

Tabel 4. 16 Perbandingan Ketepatan Klasifikasi

Metode	Ketepatan Klasifikasi
Regresi Logistik Biner	95,83%
Algoritma C5.0	91,67%

Berdasarkan Tabel 4.16, terlihat bahwa metode Regresi Logistik Biner memiliki tingkat akurasi klasifikasi yang lebih tinggi daripada metode Algoritma C5.0 dalam memprediksi laju inflasi di Indonesia. Secara khusus, regresi logistik biner mencatat akurasi klasifikasi sebesar 95,83%.

2. Perbandingan Nilai AUC

Berikut adalah hasil perbandingan nilai AUC dari kedua metode berdasarkan hasil analisis perhitungan nilai AUC menggunakan nilai *specificity* dan *sensitifity*:

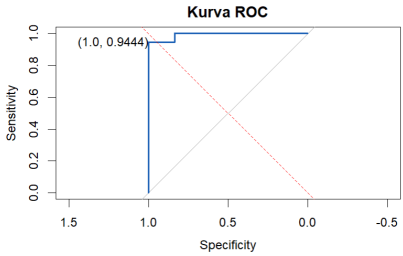
Tabel 4. 17 Perbandingan Nilai AUC

Metode	Nilai AUC
Regresi Logistik Biner	0,9722
Algoritma C5.0	0,8334

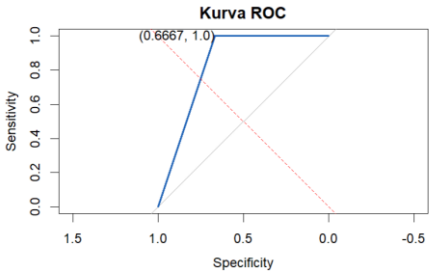
Berdasarkan nilai AUC dari kedua metode diatas dapat dikatakan bahwa metode Regresi Logistik Biner lebih efektif dalam mengklasifikasikan tingkat inflasi di Indonesia daripada metode Algoritma C5.0 karena pada regresi logistik biner menghasilkan nilai AUC yang lebih besar dibanding algoritma C5.0 yaitu sebesar 0,9722.

3. Perbandingan Kurva ROC

Berikut adalah hasil analisis kurva ROC menggunakan *R Studio* dari metode regresi logstik biner dan algoritma C5.0.



Gambar 4. 3 Kurva ROC Metode Regresi logistik Biner



Gambar 4. 4 Kurva ROC Metode Algoritma C5.0

Dari Gambar 4.3 dan Gambar 4.4 dapat dilihat bahwa metode regresi logistik biner mendapatkan hasil kurva yang mendekati titik kiri atas (0,1) yang berarti memiliki kinerja yang sangat baik dengan nilai AUC sebesar 0,9722. Sedangkan pada kurva algoritma C5.0 hampir mendekati titik kiri atas (0,1) yang memiliki kinerja yang baik dengan nilai AUC sebesar 0,8334. Berdasarkan teori kurva ROC bahwa nilai akurasi dikatakan sempurna atau teratur tanpa cela apabila kurva dekat dengan titik kiri atas (0,1). Sedangkan berdasarkan teori AUC pada penilaian untuk mengklasifikasikan model prediksi yang dihasilkan bahwa nilai AUC 0,8 - 0,9 termasuk dalam kategori baik dan nilai AUC 0,9 – 1 termasuk dalam kategori sangat baik.

Dapat ditarik kesimpulan bahwa Regresi Logistik Biner lebih efektif dalam menganalisis laju inflasi di Indonesia karena pada regresi logistik biner menghasilkan kurva ROC lebih mendekati titik kiri atas (0,1) dan memiliki nilai AUC yang lebih besar yaitu sebesar 0,9722.

4. Kelebihan dan Kekurangan Metode Regresi Logistik Biner dan Algoritma C5.0

1. Regresi Logistik Biner memiliki kelebihan yaitu memberikan probabilitas langsung untuk kejadian biner, yang memungkinkan interpretasi langsung tentang probabilitas suatu kejadian terjadi atau tidak terjadi. Regresi logistik dapat menangani variabel prediktor yang bersifat kategorikal, kontinu, dan numerik dengan baik. Regresi Logistik Biner memiliki kelemahan yaitu regresi logistik biner hanya cocok untuk variabel dependen biner tunggal. Jika variabel dependen memiliki lebih dari dua kategori, metode ini tidak cocok.
2. Algoritma C5.0 memiliki kelebihan yaitu dapat menangani baik data numerik maupun kategorikal, serta keduanya dalam kombinasi. Ini membuatnya cocok untuk berbagai jenis data. Serta pohon keputusan yang dihasilkan oleh algoritma C5.0 mudah diinterpretasikan, dengan melihat struktur pohon untuk memahami bagaimana keputusan dibuat dan mana fitur yang paling penting. Algoritma C5.0 memiliki kelemahan yaitu algoritma C5.0 rentan terhadap overfitting, terutama jika model terlalu kompleks atau jika ada noise yang signifikan dalam data atau data yang tidak teratur.

BAB V

PENUTUP

A. Kesimpulan

Dari hasil analisis yang telah dilakukan, Kesimpulan yang dicapai dalam menjawab rumusan masalah yang ada adalah:

1. Dari hasil analisis regresi logistik biner, ditemukan bahwa variabel independen yang signifikan dalam mempengaruhi laju inflasi di Indonesia adalah suku bunga. Selain itu, model regresi logistik biner untuk analisis laju inflasi di Indonesia memiliki tingkat akurasi sebesar 95,83% dan nilai AUC sebesar 0,9722.
2. Dari analisis algoritma C5.0 didapatkan variabel terpenting yaitu X2 (suku bunga) dengan tingkat kepentingan 81,04% yang berarti variabel X2 signifikan berpengaruh terhadap laju inflasi di Indonesia. Selanjutnya dapat diketahui bahwa ketepatan klasifikasi dari model algoritma C5.0 untuk analisis laju inflasi di Indonesia memberikan nilai akurasi sebesar 91,67% dan mendapatkan nilai AUC sebesar 0,8334.
3. Metode regresi logistik biner terbukti lebih efektif dalam menganalisis laju inflasi di Indonesia dibandingkan dengan algoritma C5.0. Hal ini didasarkan pada hasil yang menunjukkan metode regresi logistik biner mencapai akurasi klasifikasi sebesar 95,83% dan AUC sebesar 0,9722, yang lebih tinggi dibandingkan dengan algoritma C5.0 yang mencatat akurasi klasifikasi sebesar 91,67% dan AUC

sebesar 0,8334.

B. Saran

Setelah memperoleh kesimpulan dari hasil penelitian ini, diharapkan penelitian tentang laju inflasi di Indonesia dapat dikembangkan oleh peneliti lain dengan metode yang berbeda. Supaya mendapat hasil metode mana yang lebih cocok untuk melakukan analisis laju inflasi di Indonesia.

DAFTAR PUSTAKA

Abdullah, A. Z., Saputro, D. R. S., & Winarno, B. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C5.0 untuk Atribut Kontinu dan Diskrit. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 72–76.

Agresti, A. (2009). An Introduction to Categorical Data Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *170*(4), 1178–1178. https://doi.org/10.1111/j.1467-985x.2007.00506_2.x

Agustius, Y., Setiawan, A., & Susanto, B. (2013). Penerapan Metode Bootstrap Pada Uji Komparatif Non Parametrik Lebih Dari 2 Sampel. *Prosiding Seminar Nasional Sains Dan Pendidikan Sains VIII, Fakultas Sains Dan Matematika, UKSW*, *4*(1), Vol 4, No.1, hal : 505-512.

Ahyar, H., Maret, U. S., Andriani, H., Sukmana, D. J., & Mada, U. G. 2020. *Buku Metode Penelitian Kualitatif & Kuantitatif*. Yogyakarta: Pustaka Ilmu.

Amanda, N. A., & Marhaeni. (2023). Machine Learning Pada Analisis Data Inflasi Indonesia Menggunakan Metode Regresi Logistik. *Incomtech*, *12*(2), 37–46.

Aprileven, P. harda. (2017). Pengaruh Faktor Ekonomi Terhadap Inflasi Yang Dimediasi Oleh Jumlah Uang Beredar. *Economics Development Analysis Journal*, *4*(1), 32–41.

Arieska, P. K., & Herdiani, N. (2018). Pemilihan Teknik Sampling Berdasarkan Perhitungan Efisiensi Relatif. *Jurnal Statistika*, *6*(2), 166–171. https://jurnal.unimus.ac.id/index.php/statistik/article/view/4322/4001

Arjunita, C. 2016. Faktor-Faktor Yang Mempengaruhi Inflasi Di Indonesia. Ecosains. *Jurnal Ilmiah Ekonomi Dan Pembangunan* https://doi.org/10.24036/ecosains.11065357.00

Fawcett, T. (2006). *An introduction to ROC analysis*. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/J.PATREC.2005.10.010

Feri Wibowo, Rina Arifati, K. R. (2016). Analisis Pengaruh

Tingkat Inflasi, Suku Bunga SBI, Nilai Tukar US Dollar Pada Rupiah, Jumlah Uang Beredar, Indeks Dow Jones, Indeks Nikkei 225, dan Indeks Hangseng Terhadap Pergerakan Indeks HargaSaham Gabungan (IHSG) Periode Tahun 2010-2014. *Journal Of Accounting*, 2(2), 1–19.

Hanin, A. N. Y. (2023). *Penerapan algoritma c5.0 dan metode smote pada klasifikasi status kesejahteraan rumah tangga*. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Han, J., dan Kamber, M. (2014). *Data mining: Data mining concepts and techniques*. San Francisco, CA, itd: Morgan Kaufmann.

Harahap, A. N., & Sitompul, H. (2022). Nilai Tukar Mata Uang Inflasi dan Pengaruhnya terhadap Harga Pasar di Indonesia pada Masa Pandemi Covid 19. *JEKKP (Jurnal Ekonomi Keuangan Dan Kebijakan Publik)*, *4*(1), 29–38.

Hasan, I. (2004). *Analisis Data Penelitian Dengan Statistik*. Bumi Aksara.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley. https://doi.org/10.1002/9781118548387

Hou, S., Hou, R., Shi, X., Wang, J., dan Yuan, C. (2014). *Research on C5.0 Algorithm Improvement and the Test in Lightning Disaster Statistics. International Journal of Control and Automation*, 7(1), 181- 190.

Jayusman, I., dan Shavab, O. A. K. (2020). Aktivitas Belajar Mahasiswa Dengan Menggunakan Media Pembelajaran Learning Management System (Lms) Berbasis Edmodo Dalam Pembelajaran Sejarah. *Jurnal Artefak*, 7(1), 13. https://doi.org/10.25157/ja.v7i1.3180

Kalalo, H. Y. T., Rotinsulu, T. O., & Maramis, M. T. B. (2016). Analisis Faktor-Faktor Yang MempKalalo, Harjunata Y.T. Rotinsulu, Tri Oldy Maramis, Mauna Th. Bengaruhi Inflasi. *Jurnal Berkala Ilmiah Efisiensi*, *16*(01), 706–717.

David, G. (2023). *Five Methods for Data Splitting in Machine Learning*. https://medium.com/@tubelwj/five-methods-for-data-splitting-in-machine-learning-27baa50908ed

Liao, W., & Triantaphyllou, E. (2008). Recent Advances in Data Mining of Enterprise Data: Algorithms and

Applications. *Series on Computers and Operations Research*.
Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, *X*(1), 186–195.
Nair, A., Kuban, B. D., Obuchowski, N., & Vince, D. G. (2001). Assessing spectral algorithms to predict atherosclerotic plaque composition with normalized and raw intravascular ultrasound data. *Ultrasound in Medicine & Biology*, *27*(10), 1319–1331. https://doi.org/10.1016/S0301-5629(01)00436-7
Nanga, Muana. (2005). *Makro Ekonomi Teori Masalah dan Kebijakan*. Jakarta: PT. Raja Grafindo Persada.
Ningsih, D., & Andiny, P. (2018). Analisis pengaruh inflasi dan pertumbuhan ekonomi terhadap kemiskinan di Indonesia*. Jurnal samudra ekonomika*, *2*(1), 53-61.
Pandya, R., & Pandya, J. (2015). C5. 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. *International Journal of Computer Applications*, *117*(16), 18–21. https://doi.org/10.5120/20639-3318
Prasetyo, E., Rahajoe, R. A. D., Agustin, S., & Arizal, A. (2013). Uji Kinerja Dan Analisis K-Support Vector Nearest Neighbor Terhadap Decision Tree dan Naive Bayes. *Jurnal Eksplora Informatika*, *3*(1), 1-6.
Prasetyo, E. (2012). *Data mining konsep dan aplikasi menggunakan matlab.* Yogyakarta: Andi, 1.
Prasetyo, E. (2014). *Data Mining-Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi.
Pratiwi, F. E. dan Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains dan Seni Pomits*, 3(1), 2337-3520.
Pratiwi, R. (2020). *Perbandingan Klasifikasi Algoritma C5.0 And Classification And Regression Tree* (Vol. 21, Issue 1). Universitas Mulawarman Samarinda.
Pratiwi, R., Hayati, M. N., dan Prangga, S. 2020. *Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification and*

Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019). BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 14(2), 273–284. https://doi.org/10.30598/barekengvol14iss2pp273-284

Pujadi, A. (2022). Inflasi: Teori Dan Kebijakan. *Jurnal Manajemen Diversitas*, 2(2), 73–77. https://www.febjayabaya.ac.id/

Putri, Y. R., Mukhlash, I. dan Hidayat, N. (2013). Prediksi Pola Kecelakaan Kerja pada Perusahaan Non Ekstraktif Menggunakan Algoritma Decision Tree: C4.5 dan C5.0. *Jurnal Sains dan Seni Pomits*, 2(1), 2337-3520.

Ramdani, M. N. (2022). *Perbandingan Metode Regresi Logistik Biner dan Naïve Bayes Pada Klasifikasi Status Pengguna KB di Kelurahan Mangunharjo* (Vol. 33, Issue 1).

Rulequest. (2019). *Data Mining Tools See5 and C5.0. (online)*. Quinlan, J. R. (1993). C4.5*: Programs for Machine Learning.* Morgan Kaufmann Publishers. San Mateo, California.

Sekarsari, D., Az Zahra, F. A., Ayuningtyas, F. R., & Fadilla, A. (2024). Analisis Dinamika Inflasi dan Implikasinya terhadap Stabilitas Ekonomi di Indonesia. *Journal of Macroeconomics and Social Development*, *1*(3), 1–9. https://doi.org/10.47134/jmsd.v1i3.194

Sugiyono. (2007). *Metode Penelitian Pendidikan, Pendekatan, Kuantitatif, Kualitatif, dan R&D*. Bandung: ALFABETA.

Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D'CARTESIAN*, *6*(2), 56. https://doi.org/10.35799/dc.6.2.2017.17023

Umaroh, A. (2020). *Perbandingan Metode Regresi Logistik Biner dan Classification And Regression Tree pada Klasifikasi Status Kesejahteraan Rumah Tangga di Kota Batu menyatakan*. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Vinayagathasan, T. (2013). Inflation and economic growth: A dynamic panel threshold analysis for Asian economies.

Journal of Asian Economics, Volume 26, p. 31–41.
Walpole, R. E. (1995). *Pengantar Statistik Edisi Ke-4*. Jakarta : PT. Gramedia.
Widiarsih, D., dan Romanda, R. (2020). Analisis Faktor-Faktor yang Mempengaruhi Inflasi di Indonesia Tahun 2015-2019 dengan Pendekatan Error Corection Model (ECM). *Jurnal Akuntansi Dan Ekonomika*, 10(1), 119–128. https://doi.org/10.37859/jae.v10i1.1917
Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron*, *7*(4), 2407–2414. https://doi.org/10.33395/sinkron.v7i4.11792
Yusuf, Y. (2007). Perbandingan Performansi Algoritma Decision Tree C5 . 0 , Cart dan Chaid. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, *2007*(Snati), 0–3.

LAMPIRAN

Lampiran 1. Data Populasi Penelitian Hasil Studi Literatur Terhadap Laju Infasi di Indonesia

No	Inflasi (%)	Nilai Kurs (NK) (Rp)	Tingkat Suku Bunga (TSB) (%)	Jumlah Uang Beredar (JUB) (Rp)
1	2,61	15513	6	975918,7
2	2,9	15617	6	893156,14
3	2,56	15741	6	863101,9
4	2,28	15354	5,75	865391,95
5	3,27	15245	5,75	851722,76
6	3,08	15040	5,75	853335,77
7	3,52	14932	5,75	879805,3
8	4,00	14811	5,75	859516,58
9	4,33	14867	5,75	895719,17
10	4,97	15301	5,75	832816,67
11	5,47	15126	5,75	813834,01
12	5,28	15295	5,75	830372,89
13	5,51	15615	5,5	897798,61
14	5,42	15658	5,25	840492,4
15	5,71	15417	4,75	808648,98
16	5,95	14971	4,25	807817,51
17	4,69	14850	3,75	805459,05
18	4,94	14984	3,5	822042,79
19	4,35	14688	3,5	815316,13
20	3,55	14608	3,5	820154,68
21	3,47	14368	3,5	896317,73
22	2,64	14347	3,5	792518,36
23	2,06	14351	3,5	795951,08
24	2,18	14335	3,75	765015,11

25	1,87	14328	3,5	831233,71
26	1,75	14263	3,5	775051,34
27	1,66	14198	3,5	766703,91
28	1,6	14256	3,5	748616,24
29	1,59	14397	3,5	750510,11
30	1,52	14511	3,5	758702,97
31	1,33	14338	3,5	739006,37
32	1,68	14323	3,5	743534,2
33	1,42	14558	3,5	732643,55
34	1,37	14417	3,5	692478,03
35	1,38	14042	3,5	698226,62
36	1,55	14061	3,75	712529,24
37	1,68	14173	3,75	760044,64
38	1,59	14236	3,75	712635,88
39	1,44	14749	4	707853,82
40	1,42	14847	4	674441,24
41	1,32	14724	4	661167,53
42	1,54	14582	4	668108,23
43	1,96	14195	4,25	651818
44	2,19	14906	4,5	685044,47
45	2,67	15867	4,5	634149
46	2,96	15194	4,5	620353,24
47	2,98	13776	4,75	607961
48	2,68	13732	5	616129
49	2,72	14017	5	654683
50	3	14068	5	622384
51	3,13	14105	5	611081
52	3,39	14107	5,25	614231
53	3,49	14251	5,5	622452
54	3,32	14029	5,75	619652

55	3,28	14226	6	625354
56	3,32	14400	6	675635
57	2,83	14147	6	592935
58	2,48	14211	6	585579
59	2,57	14035	6	570435
60	2,82	14132	6	579294
61	3,13	14496	6	625370,48
62	3,23	14650	6	586235,77
63	3,16	15210	5,75	581591,77
64	2,88	14868	5,75	590804,87
65	3,2	14572	5,5	587788,32
66	3,18	14422	5,25	583305,87
67	3,12	14036	5,25	605972,86
68	3,23	14066	4,75	580625,14
69	3,41	13807	4,25	549587,19
70	3,4	13757	4,25	549216,35
71	3,18	13590	4,25	531209,28
72	3,25	13360	4,25	532131,48
73	3,61	13556	4,25	586576,33
74	3,3	13527	4,25	537297,62
75	3,58	13526	4,25	519861,42
76	3,72	13303	4,25	523359,53
77	3,82	13342	4,5	527097,79
78	3,88	13342	4,75	517809,72
79	4,37	13298	4,75	561820,83
80	4,33	13323	4,75	485123,85
81	4,17	13306	4,75	483042,03
82	3,61	13346	4,75	468941,88
83	3,83	13341	4,75	462412,91
84	3,49	13359	4,75	470250,25

85	3,02	13418	4,75	508123,74
86	3,58	13311	4,75	476850,39
87	3,31	13017	4,75	467318,21
88	3,07	13118	5	469541,7
89	2,79	13165	5,25	466501,59
90	3,21	13119	6,5	474245,9
91	3,45	13355	6,5	511294,54
92	3,33	13420	6,75	440659,81
93	3,6	13180	6,75	435295,81
94	4,45	13193	6,75	420213,6
95	4,42	13516	7	422149,44
96	4,14	13889	7,25	439871,75
97	3,35	13855	7,5	469534,21
98	4,89	13673	7,5	437756,2
99	6,25	13796	7,5	435065,11
100	6,83	14396	7,5	428860,24
101	7,18	13782	7,5	423101,29
102	7,26	13375	7,5	431459,9
103	7,26	13313	7,5	409713,13
104	7,15	13141	7,5	406499,02
105	6,79	12948	7,5	395686,64
106	6,38	13067	7,5	382004,92
107	6,29	12750	7,5	387889,28
108	6,96	12579	7,75	391255,5
109	8,36	12438	7,75	419261,84
110	6,23	12158	7,75	405694,05
111	4,83	12145	7,5	396112,97
112	4,53	11891	7,5	395229,5
113	3,99	11707	7,5	399270,22
114	4,53	11689	7,5	452787,99

115	6,7	11893	7,5	381637,54
116	7,32	11526	7,5	380473,75
117	7,25	11436	7,5	372341,57
118	7,32	11427	7,5	377437,65
119	7,75	11935	7,5	367651,74
120	8,22	12180	7,5	380070,16
121	8,38	12087	7,5	399606,17
122	8,37	11613	7,5	375784,44
123	8,32	11367	7,25	363797,12
124	8,4	11346	7,25	360078,55
125	8,79	10573	7	359417,43
126	8,61	10073	6,5	383931,57
127	5,9	9882	6	347146,05
128	5,47	9761	5,75	334033,38
129	5,57	9724	5,75	324333,2
130	5,9	9709	5,75	331168,76
131	5,31	9687	5,75	321483,32
132	4,57	9687	5,75	326828,94
133	4,3	9646	5,75	361966,71
134	4,32	9628	5,75	327069,16
135	4,61	9597	5,75	326119,07
136	4,31	9566	5,75	325565,66
137	4,58	9500	5,75	327059,39
138	4,56	9457	5,75	315374,93
139	4,53	9451	5,75	314669,88
140	4,45	9290	5,75	294768,07
141	4,5	9176	5,75	290860,56
142	3,97	9165	5,75	287046,49
143	3,56	9026	5,75	280102,64
144	3,65	9109	6	286241,88

145	3,79	9088	6	307759,79
146	4,15	9015	6	279066,19
147	4,42	8895	6,5	281340,94
148	4,61	8766	6,75	279223,58
149	4,79	8532	6,75	324724,88
150	4,61	8533	6,75	275436,75
151	5,54	8564	6,75	261503,76
152	5,98	8556	6,75	254065,51
153	6,16	8651	6,75	252012,69
154	6,65	8761	6,75	241617,68
155	6,84	8913	6,75	245326,65
156	7,02	9037	6,5	247480,97
157	6,96	9023	6,5	307759,79
158	6,33	8938	6,5	260226,78
159	5,67	8928	6,5	238500,01
160	5,8	8976	6,5	235709
161	6,44	8972	6,5	229824,67
162	6,22	9049	6,5	241166,31
163	5,05	9148	6,5	228238,83
164	4,16	9183	6,5	222828,06
165	3,91	9027	6,5	214694,87
166	3,43	9174	6,5	211390,29
167	3,81	9348	6,5	205083,05
168	3,72	9275	6,5	211708,21

Lampiran 2. Data Sampel Penelitian Hasil Studi Literatur Terhadap Laju Infasi di Indonesia

No	Inflasi (%)	Nilai Kurs (NK) (Rp)	Tingkat Suku Bunga (TSB) (%)	Jumlah Uang Beredar (JUB) (Milyar Rp)
1	2,61	15513	6	975918,70
2	2,86	15617	6	893156,14
3	2,56	15741	6	863101,90
4	2,28	15354	5,75	865391,95
5	3,27	15245	5,75	851722,76
6	5,47	15126	5,75	813834,01
7	5,28	15295	5,75	830372,89
8	5,51	15615	5,5	897798,61
9	5,42	15658	5,25	84049240
10	5,71	15417	4,75	808648,98
11	5,95	14971	4,25	807817,51
12	4,69	14850	3,75	805459,05
13	4,94	14984	3,5	822042,79
14	2,64	14347	3,5	792518,36
15	2,06	14351	3,5	795951,08
16	2,18	14335	3,75	765015,11
17	1,87	14328	3,5	831233,71
18	1,75	14263	3,5	775051,34
19	1,66	14198	3,5	766703,91
20	1,6	14256	3,5	748616,24
21	1,59	14397	3,5	750510,11
22	1,52	14511	3,5	758702,97
23	1,33	14338	3,5	739006,37
24	1,68	14323	3,5	743534,20
25	1,42	14558	3,5	732643,55

26	1,37	14417	3,5	692478,03
27	1,38	14042	3,5	698226,62
28	1,55	14061	3,75	712529,24
29	1,68	14173	3,75	760044,64
30	1,59	14236	3,75	712635,88
31	1,44	14749	4	707853,82
32	1,42	14847	4	674441,24
33	1,32	14724	4	661167,53
34	1,54	14582	4	668108,23
35	1,96	14195	4,25	651818,00
36	2,19	14906	4,5	685044,47
37	2,67	15867	4,5	634149,00
38	2,96	15194	4,5	620353,24
39	2,98	13776	4,75	607961,00
40	2,68	13732	5	616129,00
41	2,72	14017	5	654683,00
42	3,0	14068	5	622384,00
43	3,13	14105	5	611081,00
44	3,39	14107	5,25	614231,00
45	3,49	14251	5,5	622452,00
46	3,32	14029	5,75	619652,00
47	3,28	14226	6	625354,00
48	3,32	14400	6	675635,00
49	3,23	14650	6	586235,77
50	3,16	15210	5,75	581591,77
51	2,88	14868	5,75	590804,87
52	3,2	14572	5,5	587788,32
53	3,18	14422	5,25	583305,87
54	3,12	14036	5,25	605972,86
55	3,23	14066	4,75	580625,14

56	3,41	13807	4,25	549587,19
57	3,72	13303	4,25	523359,53
58	3,82	13342	4,5	527097,79
59	3,88	13342	4,75	517809,72
60	4,37	13298	4,75	561820,83
61	4,33	13323	4,75	485123,85
62	3,07	13118	5	469541,70
63	2,79	13165	5,25	466501,59
64	3,21	13119	6,5	474245,90
65	3,45	13355	6,5	511294,54
66	3,33	13420	6,75	440659,81
67	3,6	13180	6,75	435295,81
68	4,45	13193	6,75	420213,60
69	4,42	13516	7	422149,44
70	4,14	13889	7,25	439871,75
71	3,35	13855	7,5	469534,21
72	4,89	13673	7,5	437756,20
73	7,26	13313	7,5	409713,13
74	7,15	13141	7,5	406499,02
75	6,79	12948	7,5	395686,64
76	6,38	13067	7,5	382004,92
77	6,29	12750	7,5	387889,28
78	6,96	12579	7,75	391255,50
79	8,36	12438	7,75	419261,84
80	6,23	12158	7,75	405694,05
81	4,83	12145	7,5	396112,97
82	6,7	11893	7,5	381637,54
83	7,32	11526	7,5	380473,75
84	7,25	11436	7,5	372341,57
85	7,32	11427	7,5	377437,65

86	7,75	11935	7,5	367651,74
87	8,22	12180	7,5	380070,16
88	8,38	12087	7,5	399606,17
89	8,37	11613	7,5	375784,44
90	8,32	11367	7,25	363797,12
91	8,4	11346	7,25	360078,55
92	8,79	10573	7	359417,43
93	8,61	10073	6,5	383931,57
94	5,9	9882	6	347146,05
95	5,47	9761	5,75	334033,38
96	5,57	9724	5,75	324333,20
97	5,9	9709	5,75	331168,76
98	5,31	9687	5,75	321483,32
99	4,57	9687	5,75	326828,94
100	4,3	9646	5,75	361966,71
101	4,5	9176	5,75	290860,56
102	3,97	9165	5,75	287046,49
103	3,56	9026	5,75	280102,64
104	3,65	9109	6	286241,88
105	3,79	9088	6	307759,79
106	4,15	9015	6	279066,19
107	4,42	8895	6,5	281340,94
108	4,61	8766	6,75	279223,58
109	6,65	8761	6,75	241617,68
110	6,84	8913	6,75	245326,65
111	7,02	9037	6,5	247480,97
112	6,96	9023	6,5	307759,79
113	6,33	8938	6,5	260226,78
114	5,67	8928	6,5	238500,01
115	5,8	8976	6,5	235709,00

116	3,43	9174	6,5	211390,29
117	3,81	9348	6,5	205083,05
118	3,72	9275	6,5	211708,21

Lampiran 3. Konversi Data Hasil Studi Literatur Terhadap Laju Inflasi di Indonesia

no	Y	X1	X2	X3
1	1	15513	6	975918,7
2	1	15617	6	893156,14
3	1	15741	6	863101,9
4	0	15354	5,75	865391,95
5	1	15245	5,75	851722,76
6	1	15126	5,75	813834,01
7	1	15295	5,75	830372,89
8	1	15615	5,5	897798,61
9	1	15658	5,25	840492,4
10	1	15417	4,75	808648,98
11	1	14971	4,25	807817,51
12	1	14850	3,75	805459,05
13	1	14984	3,5	822042,79
14	1	14347	3,5	792518,36
15	0	14351	3,5	795951,08
16	0	14335	3,75	765015,11
17	0	14328	3,5	831233,71
18	0	14263	3,5	775051,34
19	0	14198	3,5	766703,91
20	0	14256	3,5	748616,24
21	0	14397	3,5	750510,11
22	0	14511	3,5	758702,97
23	0	14338	3,5	739006,37
24	0	14323	3,5	743534,2
25	0	14558	3,5	732643,55
26	0	14417	3,5	692478,03
27	0	14042	3,5	698226,62

28	0	14061	3,75	712529,24
29	0	14173	3,75	760044,64
30	0	14236	3,75	712635,88
31	0	14749	4	707853,82
32	0	14847	4	674441,24
33	0	14724	4	661167,53
34	0	14582	4	668108,23
35	0	14195	4,25	651818
36	0	14906	4,5	685044,47
37	1	15867	4,5	634149
38	1	15194	4,5	620353,24
39	1	13776	4,75	607961
40	1	13732	5	616129
41	1	14017	5	654683
42	1	14068	5	622384
43	1	14105	5	611081
44	1	14107	5,25	614231
45	1	14251	5,5	622452
46	1	14029	5,75	619652
47	1	14226	6	625354
48	1	14400	6	675635
49	1	14650	6	586235,77
50	1	15210	5,75	581591,77
51	1	14868	5,75	590804,87
52	1	14572	5,5	587788,32
53	1	14422	5,25	583305,87
54	1	14036	5,25	605972,86
55	1	14066	4,75	580625,14
56	1	13807	4,25	549587,19
57	1	13303	4,25	523359,53

58	1	13342	4,5	527097,79
59	1	13342	4,75	517809,72
60	1	13298	4,75	561820,83
61	1	13323	4,75	485123,85
62	1	13118	5	469541,7
63	1	13165	5,25	466501,59
64	1	13119	6,5	474245,9
65	1	13355	6,5	511294,54
66	1	13420	6,75	440659,81
67	1	13180	6,75	435295,81
68	1	13193	6,75	420213,6
69	1	13516	7	422149,44
70	1	13889	7,25	439871,75
71	1	13855	7,5	469534,21
72	1	13673	7,5	437756,2
73	1	13313	7,5	409713,13
74	1	13141	7,5	406499,02
75	1	12948	7,5	395686,64
76	1	13067	7,5	382004,92
77	1	12750	7,5	387889,28
78	1	12579	7,75	391255,5
79	1	12438	7,75	419261,84
80	1	12158	7,75	405694,05
81	1	12145	7,5	396112,97
82	1	11893	7,5	381637,54
83	1	11526	7,5	380473,75
84	1	11436	7,5	372341,57
85	1	11427	7,5	377437,65
86	1	11935	7,5	367651,74
87	1	12180	7,5	380070,16

88	1	12087	7,5	399606,17
89	1	11613	7,5	375784,44
90	1	11367	7,25	363797,12
91	1	11346	7,25	360078,55
92	1	10573	7	359417,43
93	1	10073	6,5	383931,57
94	1	9882	6	347146,05
95	1	9761	5,75	334033,38
96	1	9724	5,75	324333,2
97	1	9709	5,75	331168,76
98	1	9687	5,75	321483,32
99	1	9687	5,75	326828,94
100	1	9646	5,75	361966,71
101	1	9176	5,75	290860,56
102	1	9165	5,75	287046,49
103	1	9026	5,75	280102,64
104	1	9109	6	286241,88
105	1	9088	6	307759,79
106	1	9015	6	279066,19
107	1	8895	6,5	281340,94
108	1	8766	6,75	279223,58
109	1	8761	6,75	241617,68
110	1	8913	6,75	245326,65
111	1	9037	6,5	247480,97
112	1	9023	6,5	307759,79
113	1	8938	6,5	260226,78
114	1	8928	6,5	238500,01
115	1	8976	6,5	235709
116	1	9174	6,5	211390,29
117	1	9348	6,5	205083,05

118	1	9275	6,5	211708,21

Keterangan:

Y : Laju Inflasi

1. : Inflasi Naik

0. : Inflasi Turun

X1 : Nilai Tukar

Nilai Tukar uang domestik terhadap uang dolar Amerika

X2 : Suku Bunga

Suku Bunga yang ditetapkan oleh pemerintah

X3 : Jumlah Uang Beredar

Jumlah kartal yang beredar di Indonesia

Lampiran 4. Sintaks Pembersihan Data

```
> library(naniar)
> vis_miss(data)
> miss_var_summary(data)
# A tibble: 5 x 3
  variable                       n_miss pct_miss
  <chr>                           <int>    <num>
1 No                                  0        0
2 Inflasi (%)                         0        0
3 Nilai Kurs (NK) (Rp)                0        0
4 Tingkat Suku Bunga (TSB) (%)        0        0
5 Jumlah Uang Beredar (JUB) (Rp)      0        0

> dup=duplicated(data)
> print(dup)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE
  [7] FALSE FALSE FALSE FALSE FALSE FALSE
 [13] FALSE FALSE FALSE FALSE FALSE FALSE
 [19] FALSE FALSE FALSE FALSE FALSE FALSE
 [25] FALSE FALSE FALSE FALSE FALSE FALSE
 [31] FALSE FALSE FALSE FALSE FALSE FALSE
 [37] FALSE FALSE FALSE FALSE FALSE FALSE
 [43] FALSE FALSE FALSE FALSE FALSE FALSE
 [49] FALSE FALSE FALSE FALSE FALSE FALSE
 [55] FALSE FALSE FALSE FALSE FALSE FALSE
 [61] FALSE FALSE FALSE FALSE FALSE FALSE
 [67] FALSE FALSE FALSE FALSE FALSE FALSE
 [73] FALSE FALSE FALSE FALSE FALSE FALSE
 [79] FALSE FALSE FALSE FALSE FALSE FALSE
 [85] FALSE FALSE FALSE FALSE FALSE FALSE
 [91] FALSE FALSE FALSE FALSE FALSE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE FALSE
[103] FALSE FALSE FALSE FALSE FALSE FALSE
[109] FALSE FALSE FALSE FALSE FALSE FALSE
[115] FALSE FALSE FALSE FALSE
```

Lampiran 5. Sintaks Regresi Logistik Biner

```
data=datas
head(data)
library(haven)
library(dplyr)
data%>%
  select(Y,X1,X2,X3)
library(dplyr)
data <- data %>%
  mutate_if(is.integer, as.factor) %>%
  mutate(Y = factor(Y, levels = c(0,1), labels = c("turun","naik")))
glimpse(data)
library(dplyr)
set.seed(123)
intrain <- sample(nrow(data), nrow(data)*0.8)
data_train <- data[intrain,]
data_test <- data[-intrain,]
data$Y %>%
  levels()
RLB<-glm(Y~X1+X2+X3, family = binomial(link = 'logit'),data = data_train)
RLB
library(pscl)
pR2(RLB)
qchisq(0.95,4)
summary(RLB)
library(car)
vif(RLB)
library(lmtest)
lrtest(RLB)
logLik(RLB)
library(generalhoslem)
modelRLB <- model.frame(RLB)
logitgof(modelRLB$Y, fitted(RLB))
library(ggplot2)
data_test$prob_data <- predict(RLB, type = "response", newdata = data_test)
ggplot(data_test, aes(x=prob_data))+
  geom_density(lwd=0.5) +
  labs(title = "Klasifikasi laju inflasi") +
  theme_minimal()
data_test$pred_data <- factor(ifelse(data_test$prob_data > 0.5, "naik","turun"))
data_test[1:10, c("pred_data","Y")]
library(caret)
confusionMatrix(data_test$Y, data_test$pred_data,positive="naik")
#accuracy <- (TP+TN)/(TP+FP+FN+TN)
accuracy <- (17+6)/(17+0+1+6)
accuracy
#error <- (FP+FN)/(TP+FP+FN+TN)
error <- (0+1)/(17+0+1+6)
error
#Re-call\sensitivity = TP/(TP+FN)
sensitivity <- 17/(17+1)
sensitivity
#specifity = TN/(TN+FP)
specificity <- 6/(6+0)
specificity
```

Lampiran 6. Sintaks Curva ROC Regresi Logistik Biner

```
#gambar kurva ROC
library(pROC)
# Menghitung nilai ROC
roc_data <- roc(data_test$Y, data_test$prob_data)
#Hitung AUC
auc_value <- auc(roc_data)
auc_value
# Menampilkan kurva ROC
plot(roc_data, main="Kurva ROC", col="#1c61b6", lwd=2)
# Menampilkan kurva ROC
abline(a=0, b=1, col = "red", lty = 2)
text(1.0, 0.9444, "(1.0, 0.9444)", col = "black", cex = 1,1)
```

Lampiran 7. Sintaks Algoritma C5.0

```
data=datas
head(data)
library(DMwR2)
data$Y <- as.factor(data$Y)
data$X1 <- as.factor(data$X1)
data$X2 <- as.factor(data$X2)
data$X3 <- as.factor(data$X3)
str(data)
table(data$Y)
prop.table(table(data$Y))
target_variable <- "Y"
predictor_variable <- setdiff(names(data), target_variable)
table(data$Y)
data_balanced <- SMOTE(as.formula(paste(target_variable,"~", paste(predictor_variable,
                       collapse = "+"))),data, perc.over = 110, perc.under = 20, k=3)
table(data_balanced$Y)
prop.table(table(data$Y))
#membuat data training dan testing
n <- round(nrow(data_balanced)*0.8);n
set.seed(123)
samp=sample(1:nrow(data_balanced),n)
data_train = data_balanced[samp,]
dim(data_train)
data_test = data_balanced[-samp,]
dim(data_test)
str(data_test$Y)
library(C50)
model <- C5.0(Y ~., data=data_train)
model
summary(model)
plot(model)
#melakukan prediksi pada data pengujian
predictions <- predict(model, data_test)
#menghitung matriks kebingungan
cofusion_matrix <- table(actual = data_test$Y, predicted = predictions)
print(cofusion_matrix)
class_predict <- predict(model, newdata = data_test, type = "class")
class_predict
#konversi variabel target ke format biner
binary_truth <- ifelse(data_test$Y == "naik", 1, 0)
binary_prediction <- ifelse(class_predict == "naik", 1, 0)
length(binary_prediction)
length(binary_truth)
#accuracy <- (TP+TN)/(TP+FP+FN+TN)
accuracy <- (18+4)/(18+2+0+4)
accuracy
#error <- (FP+FN)/(TP+FP+FN+TN)
error <- (2+0)/(18+2+0+4)
error
#Re-call\sensitivity = TP/(TP+FN)
sensitivity <- 18/(18+0)
sensitivity
#specifity = TN/(TN+FP)
specificity <- 4/(4+2)
specificity
```

Lampiran 8. Sintaks Curva ROC Algoritma C5.0

```
# Membuat kurva roc
library(pROC)
# Menghitung nilai ROC
roc_curve <- roc(binary_truth, binary_prediction)
#Hitung AUC
auc_value <- auc(roc_curve)
auc_value
# Menampilkan kurva ROC
plot(roc_curve, main="Kurva ROC", col="#1c61b6", lwd=2)
abline(a=0, b=1, col = "red", lty = 2)
text(0.6667, 1.0, "(0.6667, 1.0)", col = "black", cex = 1,1)
```

Lampiran 9. Output Regresi Logistik Biner

1. Uji Multikolinieritas

```
> vif(RLB)
       X1        X2        X3
1.717434 1.096447 1.692692
```

2. Uji Wald

```
> summary(RLB)

Call:
glm(formula = Y ~ X1 + X2 + X3, family = binomial(link = "logit"),
    data = data_train)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.512e+00  9.292e+00  -0.378 0.705448
X1          -1.379e-04  7.964e-04  -0.173 0.862536
X2           2.223e+00  6.286e-01   3.536 0.000407
X3          -5.492e-06  5.803e-06  -0.946 0.343914

(Intercept)
X1
X2          ***
X3
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 88.864  on 93  degrees of freedom
Residual deviance: 34.383  on 90  degrees of freedom
AIC: 42.383

Number of Fisher Scoring iterations: 8
```

3. Uji Rasio *Likelihood*

```
> lrtest(RLB)
Likelihood ratio test

Model 1: Y ~ X1 + X2 + X3
Model 2: Y ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   4 -17.191
2   1 -44.432 -3 54.481  8.858e-12 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. Uji Kesesuaian Model

```
> logitgof(modelRLB$Y, fitted(RLB))

        Hosmer and Lemeshow test (binary model)

data:  modelRLB$Y, fitted(RLB)
X-squared = 10.178, df = 8, p-value = 0.2528
```

5. *Confuxion Matrix*

```
> confusionMatrix(data_test$Y, data_test$pred_data,positive="naik")
Confusion Matrix and Statistics

          Reference
Prediction naik turun
     naik    17     1
     turun    0     6
```

6. Akurasi Klasifikasi

```
> #accuracy <- (TP+TN)/(TP+FP+FN+TN)
> accuracy <- (17+6)/(17+0+1+6)
> accuracy
[1] 0.9583333
> #error <- (FP+FN)/(TP+FP+FN+TN)
> error <- (0+1)/(17+0+1+6)
> error
[1] 0.04166667
> #Re-call\sensitivity = TP/(TP+FN)
> sensitivity <- 17/(17+1)
> sensitivity
[1] 0.9444444
> #specifity = TN/(TN+FP)
> specificity <- 6/(6+0)
> specificity
[1] 1
```

Lampiran 10. Output Algoritma C5.0

1. Pohon Klasifikasi

```
> library(C50)
> model <- C5.0(Y ~., data=data_train)
> model

Call:
C5.0.formula(formula = Y ~ ., data = data_train)

Classification Tree
Number of samples: 94
Number of predictors: 3

Tree size: 4

Non-standard options: attempt to
 group attributes
```

2. Tingkat Kepentingan Variabel

```
> summary(model)

Call:
C5.0.formula(formula = Y ~ ., data = data_train)


C5.0 [Release 2.07 GPL Edition]         Sun May 12 03:03:03 2024
-------------------------------

Class specified by attribute `outcome'

Read 94 cases (4 attributes) from undefined.data

Decision tree:

X2 > 4.5: naik (72/1)
X2 <= 4.5:
:...X3 <= 634149: naik (3)
    X3 > 634149:
    :...X3 <= 765015.1: turun (14)
        X3 > 765015.1: naik (5/2)

Evaluation on training data (94 cases):

          Decision Tree
        ----------------
        Size      Errors

           4    3( 3.2%)   <<


         (a)   (b)    <-classified as
        ----  ----
          14     3    (a): class turun
                77    (b): class naik


       Attribute usage:

       100.00% X2
        23.40% X3


Time: 0.0 secs
```

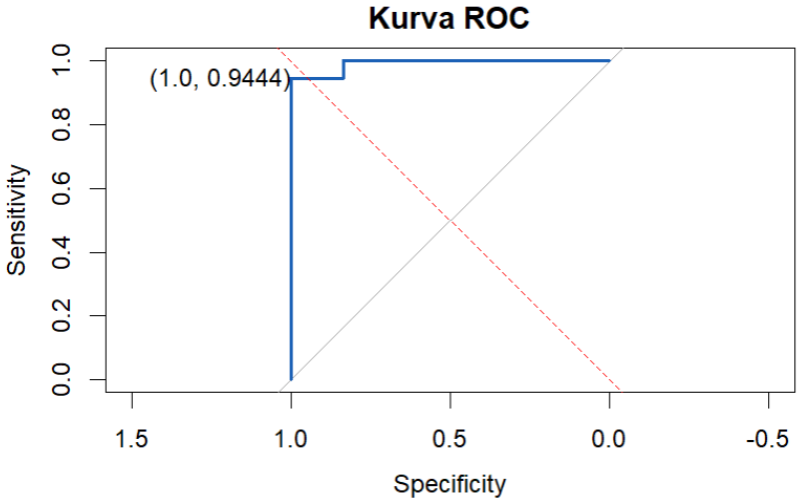
3. *Confussion Matrix*

```
> cofusion_matrix <- table(actual = data_test$Y, predicted = predictions)
> print(cofusion_matrix)
       predicted
actual  turun naik
  turun     4    2
  naik      0   18
```

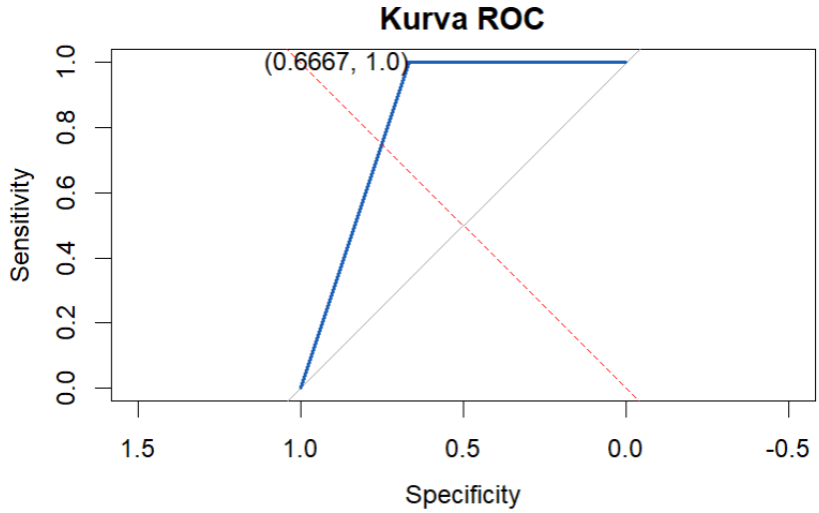
4. Tingkat Akurasi Ketepatan Klasifikasi

```
> #accuracy <- (TP+TN)/(TP+FP+FN+TN)
> accuracy <- (18+4)/(18+2+0+4)
> accuracy
[1] 0.9166667
> #error <- (FP+FN)/(TP+FP+FN+TN)
> error <- (2+0)/(18+2+0+4)
> error
[1] 0.08333333
> #Re-call\sensitivity = TP/(TP+FN)
> sensitivity <- 18/(18+0)
> sensitivity
[1] 1
> #specifity = TN/(TN+FP)
> specificity <- 4/(4+2)
> specificity
[1] 0.6666667
```

Lampiran 11. Output Curva ROC Regresi Logistik Biner



Lampiran 12. Output Curva ROC Algoritma C5.0



Lampiran 13. Riwayat Hidup

RIWAYAT HIDUP

A. Identitas Diri

1. Nama Lengkap : Muhamad Cahyo Nugroho
2. Tempat dan Tgl. Lahir : Semarang, 3 November 2001
3. Alamat Rumah : Jl. Karang Anyar Rt/04 Rw/12, Kel. Mukhtiharjo Kidul, Kec. Pedurungan , Semarang
4. No. HP : 0895421099000
5. Email : muhamadcahyo31@gmail.com

B. Riwayat Pendidikan

1. SD N 01 Sawah Besar Semarang
2. SMP N 6 Semarang
3. SMA N 10 Semarang

C. Prestasi

1. Prestasi Non Akademik
   a. Juara 2 Tunggal Putra Badminton event IKANMAS 2021
   b. Juara 3 Ganda Putra Badminton event PORMAWA UIN WS 2023