

Klasifikasi Diagnosis HCV di Mesir Berbasis Biomarker Menggunakan Regresi Logistik Multinomial

Azmy Minanda Hermiadi *, Fauziah Roshafara

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

dindaazmy95@gmail.com, fauziahroshafara@unisba.ac.id

Abstract. Hepatitis C Virus (HCV) infection is a major public health concern that can progress to liver fibrosis and cirrhosis if not properly managed. This study aims to identify the effects of clinical biomarkers and demographic factors on HCV diagnosis categories and to evaluate the performance of a statistical classification model. A quantitative approach was employed using secondary data from the UCI Machine Learning Repository, consisting of 615 patients as research subjects. The research procedures included data *Preprocessing*, variable standardization, and multinomial logistic regression modeling. The results show distinct variations in biomarker distributions across diagnosis categories. Albumin and cholesterol levels tended to be higher in the blood donor group, while bilirubin and alkaline phosphatase levels were elevated among patients with fibrosis and cirrhosis. Age also played an important role in differentiating diagnoses. The model demonstrated satisfactory classification performance, although the suspect blood donor group exhibited higher rates of misclassification due to intermediate clinical characteristics. Overall, this study highlights the potential use of basic clinical laboratory data to support risk assessment and provides insights into the progression of HCV-related liver disease.

Keywords: *Hepatitis C Virus, Clinical Biomarkers, Multinomial Logistic Regression.*

Abstrak. Infeksi Hepatitis C Virus (HCV) merupakan salah satu permasalahan kesehatan yang dapat berkembang menjadi fibrosis maupun sirosis hati apabila tidak ditangani secara tepat. Penelitian ini dilakukan untuk mengidentifikasi pengaruh biomarker klinis dan faktor demografis terhadap kategori diagnosis HCV, serta menilai kemampuan model statistik dalam mengklasifikasikan kondisi klinis pasien. Penelitian menggunakan pendekatan kuantitatif dengan data sekunder yang diperoleh dari UCI Machine Learning Repository, terdiri dari 615 pasien sebagai objek penelitian. Tahapan penelitian meliputi *Preprocessing* data, standarisasi variabel numerik, dan pemodelan menggunakan regresi logistik multinomial. Hasil penelitian menunjukkan adanya perbedaan distribusi biomarker klinis antar kategori diagnosis, di mana Albumin dan Kolesterol cenderung lebih tinggi pada kelompok donor darah, sedangkan Bilirubin dan Alkaline Phosphatase meningkat pada kelompok fibrosis dan sirosis. Faktor usia juga berperan dalam membedakan diagnosis. Model dapat memprediksi kategori diagnosis dengan performa yang memadai, meskipun kelas *suspect blood donor* menunjukkan misklasifikasi yang lebih tinggi. Secara keseluruhan, penelitian ini mendukung pemanfaatan data klinis sederhana dalam membantu proses analisis risiko penyakit hati dan memberikan *insight* mengenai progresivitas HCV.

Kata Kunci: *Hepatitis C Virus, Biomarker Klinis, Regresi Logistik Multinomial.*

A. Pendahuluan

Hepatitis C Virus (HCV) merupakan salah satu penyebab utama penyakit hati kronis di dunia dan menjadi masalah kesehatan masyarakat global. Menurut (Waked et al., 2020), terdapat sekitar 58 juta kasus infeksi HCV kronis dan lebih dari 290.000 kematian setiap tahunnya akibat komplikasi penyakit hati, termasuk sirosis dan karsinoma hepatoseluler. Salah satu negara dengan prevalensi HCV tertinggi adalah Mesir. Studi epidemiologi menunjukkan bahwa prevalensi HCV pada populasi dewasa di Mesir pernah mencapai 14,7%, jauh lebih tinggi dibandingkan rata-rata global (Guerra & Garenne, 2012). Fenomena tingginya angka infeksi ini banyak dikaitkan dengan praktik pengobatan schistosomiasis pada masa 1950–1980, yang menggunakan jarum suntik tidak steril secara massal sehingga memicu peningkatan transmisi HCV secara luas (Strickland, 2006) dan (Lehman & Wilson, 2009). Kendati demikian, pemerintah Mesir telah melakukan kampanye skrining dan eliminasi HCV secara nasional sejak 2018 dan berhasil menyaring lebih dari 49 juta penduduk dalam kurun waktu kurang dari dua tahun (Hassanin et al., 2021).

Selain sebagai masalah epidemiologis, HCV juga merupakan isu klinis yang berkaitan dengan fungsi hati dan metabolisme tubuh. Teori klinis menunjukkan bahwa beberapa parameter laboratorium seperti Albumin Blood Test (ALB), Alkaline Phosphatase (ALP), Bilirubin (BIL), dan Kolesterol (CHOL) dapat memberikan indikasi penting terhadap kondisi hepatologis pasien. Albumin mencerminkan kemampuan hati dalam mensintesis protein, sedangkan ALP berhubungan dengan kelainan saluran empedu dan proses kolestasis. Bilirubin meningkat pada gangguan hepatoseluler dan hemolisis, sementara kolesterol dapat menurun pada sirosis kronis karena gangguan sintesis lipid di hati (Wang et al., 2024). Faktor usia dan jenis kelamin juga dilaporkan memiliki hubungan dengan progresivitas infeksi serta respons pengobatan (Kandeel et al., 2017).

Dalam penelitian ini digunakan data sekunder pasien HCV di Mesir yang terdiri dari 615 baris dan diperoleh dari UCI Machine Learning Repository. Data mencakup satu variabel dependen berupa Diagnosis (Blood Donor, Suspect Blood Donor, Hepatitis, Fibrosis, Cirrhosis), serta beberapa variabel independen berupa Age, Sex, dan indikator laboratorium ALB, ALP, BIL, dan CHOL. Variabel tersebut merupakan biomarker klinis yang umum digunakan dalam pemeriksaan hepatologi dan memiliki potensi keterkaitan dengan status infeksi dan keparahan penyakit. Pada penelitian terdahulu, biomarker serupa digunakan untuk mendeteksi derajat fibrosis dan sirosis, serta untuk memprediksi keparahan hepatopati (Seto et al., 2017).

Urgensi penelitian ini didasarkan pada perlunya pemahaman mengenai karakteristik klinis pasien HCV di negara dengan beban epidemi yang tinggi seperti Mesir. Oleh karena itu penelitian ini diarahkan untuk mengidentifikasi hubungan antara variabel laboratorium dengan status diagnosis pasien, sekaligus memberikan gambaran mengenai profil klinis berdasarkan usia dan jenis kelamin.

Penelitian ini bertujuan untuk mengetahui bagaimana variabel klinis seperti ALB, ALP, BIL, CHOL, Age, dan Sex berhubungan dengan status diagnosis HCV, serta memberikan gambaran mengenai pola pengelompokan klinis yang dapat terjadi di antara kelompok pasien. Dengan demikian dapat diketahui apakah terdapat kecenderungan tertentu pada pasien dengan kategori diagnosis tertentu, misalnya apakah kadar bilirubin cenderung lebih tinggi pada pasien sirosis dibandingkan pada kelompok donor darah. Selain itu penelitian ini diharapkan dapat menambah pengetahuan mengenai karakteristik spesifik populasi Mesir dalam konteks HCV, mengingat struktur epidemiologinya berbeda dengan negara lain.

Manfaat penelitian dapat dilihat dari perspektif klinis maupun akademik. Dari perspektif klinis, hasil penelitian ini dapat mendukung upaya deteksi dini dan penilaian risiko pasien HCV berdasarkan biomarker terukur, sedangkan dari sisi akademis penelitian ini dapat memperkaya literatur mengenai hubungan biomarker hepatologis dengan status diagnosis HCV. Penelitian ini juga memiliki unsur kebaruan karena memanfaatkan dataset real-world dalam jumlah cukup besar dari populasi.

Hepatitis C merupakan infeksi yang disebabkan oleh Hepatitis C Virus (HCV), sebuah virus RNA yang termasuk dalam famili *Flaviviridae* dan bersifat hepatotropik. Infeksi HCV dapat berlangsung dalam jangka panjang dan menimbulkan penyakit hati kronis yang progresif. Sebagian besar pasien pada fase awal tidak menunjukkan gejala klinis sehingga HCV sering disebut sebagai *silent epidemic* (Hassanin et al., 2021). Infeksi kronis HCV dapat berkembang menjadi fibrosis, sirosis, dan karsinoma hepatoseluler (HCC), dengan risiko progresivitas yang meningkat apabila tidak dilakukan penanganan yang tepat (Wang et al., 2024). Dalam konteks klinis, karakteristik penyakit hati dievaluasi melalui pemeriksaan enzim dan biomarker hematologis karena indikator tersebut dapat mencerminkan derajat inflamasi dan kerusakan hepatoseluler.

Studi epidemiologi menunjukkan bahwa Mesir memiliki karakteristik epidemi HCV yang unik. Seperti dijelaskan oleh (Strickland, 2006) dan (Lehman & Wilson, 2009), tingginya prevalensi infeksi HCV di Mesir berkaitan dengan program pengobatan schistosomiasis yang menggunakan jarum suntik tidak steril pada pertengahan abad ke-20. Selain praktik medis tersebut, penelitian juga mengidentifikasi transfusi darah tanpa skrining, penggunaan alat medis invasif, dan prosedur keperawatan yang tidak mengikuti standar sterilisasi modern sebagai faktor risiko penyebaran HCV (El-Zanaty & Way, 2015). Studi terbaru juga menyoroti bahwa kelompok usia lanjut memiliki prevalensi lebih tinggi karena paparan historis terhadap praktik tersebut, sehingga struktur usia menjadi salah satu variabel risiko penting dalam penelitian HCV di Mesir (Kandeel et al., 2017).

Selain faktor medis, aspek demografis seperti jenis kelamin juga ditemukan memiliki hubungan dengan infeksi HCV. Pada beberapa penelitian, prevalensi HCV ditemukan lebih tinggi pada laki-laki dibandingkan perempuan, terutama pada kelompok usia dewasa dan lanjut usia, yang dikaitkan dengan perbedaan paparan risiko dan faktor hormonal (Guidance, 2020). Kombinasi faktor usia dan sex ini sering muncul dalam penelitian-penelitian berbasis populasi, termasuk penelitian di Mesir dan Asia Selatan.

Penelitian mengenai HCV tidak hanya berfokus pada faktor paparan, tetapi juga pada karakteristik biomarker laboratorium. Parameter Albumin (ALB) merupakan indikator fungsi hati yang menggambarkan kemampuan organ dalam mensintesis protein. Kadar albumin yang rendah merupakan salah satu indikator sirosis dan kerusakan hati jangka panjang. Alkaline Phosphatase (ALP) merupakan enzim yang meningkat pada kondisi kolestasis dan gangguan saluran empedu. Sementara itu Bilirubin (BIL) merupakan produk degradasi hemoglobin yang meningkat pada kerusakan hepatoseluler atau hemolisis. Kolesterol (CHOL) berperan dalam metabolisme lipid dan sering turun pada pasien sirosis akibat gangguan sintesis lipid di hati (Guidance, 2020). Dalam penelitian hepatologi, biomarker tersebut sering digunakan untuk mendeteksi progresivitas penyakit dan menilai respons pengobatan.

Studi-studi terdahulu di berbagai negara juga melaporkan temuan serupa. Di Pakistan dan India, penelitian (Kanaani et al., 2018) dan (Mukhopadhyaya, 2008) mengidentifikasi penggunaan jarum suntik tidak steril, transfusi darah, dan prosedur medis invasif sebagai faktor risiko dominan, sedangkan di Eropa Timur dan Rusia penyalahgunaan narkoba suntik menjadi faktor utama penularan HCV (Nelson et al., 2010). Meski demikian, struktur epidemiologi HCV di Mesir berbeda karena dominasi kasus berasal dari paparan historis dan bukan penyalahgunaan narkoba, sehingga konteks Mesir menjadi menarik sebagai model epidemiologi infeksi massal berbasis praktik medis.

Dalam literatur juga ditemukan bahwa respons pengobatan terhadap HCV dipengaruhi oleh faktor biologis dan klinis. Penggunaan obat antiviral generasi baru seperti *direct-acting antivirals (DAA)* telah menunjukkan tingkat kesembuhan klinis yang tinggi, namun beberapa penelitian menemukan variasi respons berdasarkan faktor usia, jenis kelamin, serta kondisi metabolik pasien (Altinel et al., 2016) dan (Wei et al., 2018). Studi pada pasien Mesir memperlihatkan bahwa keberhasilan terapi juga dipengaruhi oleh derajat fibrosis awal sebelum pengobatan (Hassanin et al., 2021).

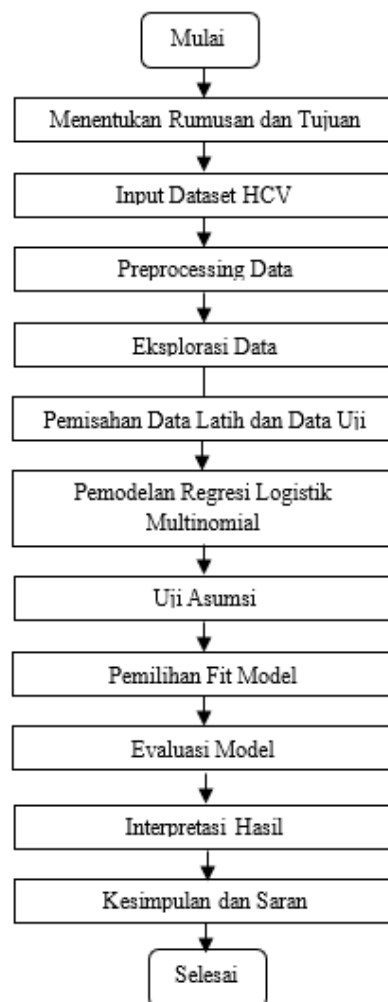
Penelitian terdahulu yang relevan dengan topik ini telah mengevaluasi faktor risiko HCV melalui pendekatan klinis dan epidemiologis, namun penelitian yang menggunakan pendekatan analisis biomarker laboratorium sebagai variabel utama masih terbatas. Dataset yang digunakan dalam penelitian ini berasal dari UCI Machine Learning Repository dan memiliki keunikan karena mencakup biomarker klinis dalam jumlah data yang relatif besar ($n = 615$), serta memuat variasi diagnosis mulai dari Blood Donor hingga Cirrhosis. Hal ini membuka peluang untuk melakukan klasifikasi klinis berdasarkan kombinasi variabel demografis dan klinis yang dapat dianalisis lebih lanjut menggunakan pendekatan statistik maupun machine learning.

Dengan demikian, tinjauan pustaka menunjukkan bahwa masih terdapat ruang penelitian untuk mengevaluasi hubungan biomarker klinis dengan diagnosis HCV secara lebih komprehensif, terutama pada populasi dengan struktur epidemi yang unik seperti Mesir. Penelitian ini diharapkan dapat berkontribusi dalam memperkaya literatur mengenai karakteristik klinis pasien HCV sekaligus mendukung upaya pengembangan strategi diagnosis, pengelompokan pasien, dan pemantauan penyakit secara lebih terukur.

B. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data sekunder. Data penelitian diperoleh dari UCI Machine Learning Repository berupa dataset klinis pasien Hepatitis C Virus (HCV) di Mesir sebanyak 615 observasi. Data mencakup variabel Diagnosis sebagai variabel dependen yang terdiri atas kategori Blood Donor, Suspect Blood Donor, Hepatitis, Fibrosis, dan Cirrhosis, serta variabel independen berupa Age, Sex, Albumin (ALB), Alkaline Phosphatase (ALP), Bilirubin (BIL), dan Cholesterol (CHOL). Unit analisis dalam penelitian ini adalah pasien, dan ruang lingkup penelitian terletak pada upaya mengidentifikasi pengaruh biomarker klinis dan demografis terhadap diagnosis HCV. Populasi penelitian adalah pasien terdiagnosis maupun tidak terdiagnosis HCV di Mesir pada tahun 2020, dengan teknik sampling berupa total sampling, karena seluruh observasi dalam dataset digunakan sebagai sampel. Waktu pelaksanaan penelitian dimulai pada bulan Oktober hingga Desember 2024 yang meliputi pengumpulan data, transformasi data, pemodelan, evaluasi, dan interpretasi hasil penelitian.

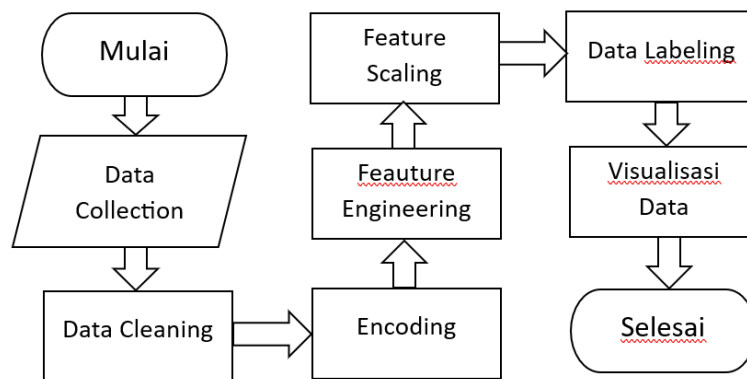
Tahapan penelitian dilakukan melalui beberapa fase utama mulai dari pengumpulan data, *Preprocessing*, pemodelan regresi logistik multinomial, evaluasi model hingga interpretasi hasil penelitian. Secara garis besar, tahapan metode penelitian ini dapat digambarkan melalui alur proses berikut:



Gambar 1. Tahapan Riset.

Setelah tahap awal selesai, dilakukan tahapan *Preprocessing* untuk menyiapkan dataset sebelum masuk ke dalam tahap pemodelan. *Preprocessing* meliputi pemeriksaan missing value, deteksi outlier, encoding variabel kategorikal, standarisasi variabel numerik, dan pemberian label diagnosis. Proses ini bertujuan untuk menjaga kualitas data, menghindari bias estimasi, dan memastikan seluruh variabel dapat diproses dalam model statistik.

Tahapan *Preprocessing* tersebut dapat digambarkan melalui alur proses sebagai berikut:



Gambar 2. Tahapan *Preprocessing*.

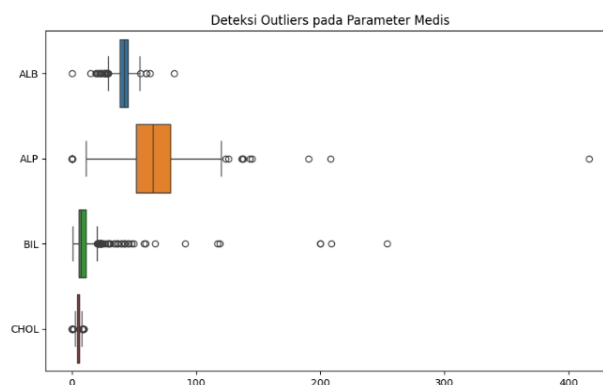
Setelah *Preprocessing* selesai, dataset dipisahkan menjadi data latih dan data uji dengan proporsi tertentu untuk mengevaluasi performa model. Metode regresi logistik multinomial digunakan dalam penelitian ini karena variabel dependen Diagnosis memiliki lebih dari dua kategori. Model ini mengestimasi peluang suatu diagnosis klinis berdasarkan variabel klinis dan demografis serta diestimasi melalui metode *Maximum Likelihood Estimation* (MLE). Secara matematis, peluang suatu diagnosis dapat dinyatakan sebagai:

$$P(Y = k|X) = \frac{e^{\beta_{0k} + \beta_{1k}X_1 + \dots + \beta_{pk}X_p}}{\sum_{j=1}^K e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{pj}X_p}} \quad \dots (1)$$

Dengan $k = 1, 2, \dots, K$ sebagai jumlah kategori diagnosis. Parameter β yang diperoleh kemudian diuji signifikansinya menggunakan uji Wald maupun likelihood ratio test untuk menilai pengaruh variabel terhadap diagnosis tertentu. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, confusion matrix, dan cross-validation untuk menguji reliabilitas dan stabilitas model. Interpretasi dilakukan berdasarkan koefisien regresi dan signifikansi statistik untuk mengetahui variabel yang mempengaruhi diagnosis, sehingga dapat diperoleh gambaran mengenai peran biomarker laboratorium dalam membedakan kelompok klinis seperti donor darah, hepatitis, fibrosis, hingga sirosis.

C. Hasil Penelitian dan Pembahasan

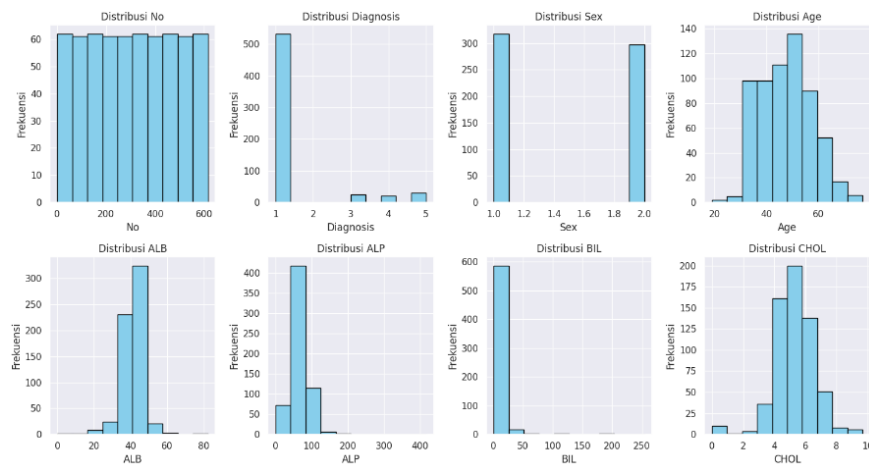
Tahap *Preprocessing* menunjukkan bahwa dataset tidak mengandung missing value yang signifikan, sehingga data dapat diproses secara langsung tanpa prosedur imputasi. Hasil pemeriksaan outlier dilakukan melalui visualisasi boxplot terhadap variabel klinis seperti ALB, ALP, BIL, dan CHOL. Hasil eksplorasi tersebut memperlihatkan adanya nilai-nilai ekstrem pada ALP dan BIL yang tetap dipertahankan karena dinilai sebagai representasi kondisi klinis pasien HCV.



Gambar 3. Boxplot Variabel Laboratorium ALB, ALP, BIL, dan CHOL

Berdasarkan Gambar 3 dapat dilihat bahwa variabel ALP dan BIL memiliki persebaran yang lebih lebar dibandingkan variabel lainnya, yang mengindikasikan variasi klinis yang lebih tinggi pada pasien yang mengalami fibrosis maupun sirosis. Sementara itu, nilai ALB dan CHOL cenderung stabil pada kelompok pasien donor darah. Hasil ini konsisten dengan literatur klinis yang menyatakan bahwa progresivitas HCV dapat mempengaruhi fungsi hati melalui akumulasi bilirubin serta peningkatan enzim fosfatase alkali sebagai indikator kerusakan hepatoseluler.

Visualisasi deskriptif lebih lanjut dilakukan dengan membandingkan distribusi biomarker klinis antar kategori diagnosis. Distribusi ini memberikan gambaran bagaimana kondisi klinis pasien mengalami perubahan dari donor darah hingga sirosis.

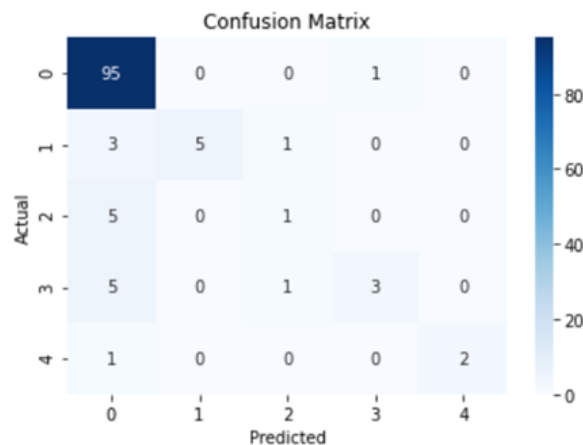


Gambar 4. Distribusi Biomarker Klinis Berdasarkan Diagnosis HCV

Gambar 2 menunjukkan bahwa kelompok Blood Donor memiliki nilai ALB dan CHOL yang relatif lebih tinggi, sedangkan kelompok Fibrosis dan Cirrhosis memiliki nilai BIL dan ALP yang lebih tinggi. Kelompok Hepatitis tampak berada pada posisi intermediate. Hal ini menguatkan temuan epidemiologi bahwa biomarker klinis dapat digunakan sebagai indikator progresivitas penyakit hati akibat infeksi HCV. Selain itu, variabel usia menunjukkan kecenderungan meningkat pada kelompok Fibrosis dan Cirrhosis, yang sejalan dengan literatur klinis yang menegaskan bahwa progresivitas HCV meningkat seiring bertambahnya usia.

Model regresi logistik multinomial diterapkan untuk menganalisis pengaruh variabel ALB, ALP, BIL, CHOL, Age, dan Sex terhadap kategori diagnosis. Hasil pemodelan menunjukkan bahwa beberapa variabel memberikan kontribusi signifikan dalam membedakan kategori diagnosis klinis. Misalnya, Albumin berperan dalam memisahkan donor darah dari fibrosis dan sirosis, sedangkan Bilirubin berperan dalam membedakan hepatitis dari fibrosis. Hasil ini mendukung teori bahwa bilirubin merupakan indikator penting pada penyakit hati kronis.

Evaluasi model dilakukan dengan menggunakan confusion matrix untuk menilai kinerja klasifikasi model. Hasil evaluasi memperlihatkan bahwa model memiliki performa yang baik dalam memprediksi kategori Blood Donor, meskipun terdapat misklasifikasi pada kategori Suspect Blood Donor yang berdekatan dengan kategori donor dan hepatitis secara klinis.



Gambar 5. Confusion Matrix Model Regresi Multinomial

Confusion matrix pada Gambar 3 memperlihatkan bahwa kelas Blood Donor merupakan kelas dengan tingkat prediksi tertinggi, sedangkan kelas Suspect Blood Donor merupakan kelas dengan tingkat misklasifikasi tertinggi. Hal ini dapat dimaklumi karena kelas tersebut merupakan transisi klinis antara normal dan hepatitis. Evaluasi tambahan dilakukan melalui cross-validation untuk menilai stabilitas model, dan hasilnya menunjukkan konsistensi performa yang baik antara data latih dan data uji.

Secara keseluruhan, hasil penelitian ini memperlihatkan bahwa biomarker klinis seperti ALB, ALP, BIL, dan CHOL serta faktor usia memainkan peran penting dalam membedakan kategori klinis penyakit HCV. Temuan ini konsisten dengan teori hepatologi dan literatur terdahulu yang menegaskan bahwa progresivitas fibrosis dan sirosis tidak hanya ditentukan oleh virus tetapi juga oleh kondisi fisiologis pasien.

Analisis dan Pembahasan

Hasil penelitian ini menunjukkan bahwa variasi biomarker klinis memiliki hubungan yang jelas dengan kategori diagnosis HCV. Nilai Albumin (ALB) yang lebih tinggi pada kelompok donor darah mengindikasikan fungsi hati yang masih optimal, sedangkan penurunan ALB pada kelompok fibrosis dan sirosis menunjukkan adanya gangguan sintesis protein yang merupakan salah satu tanda kerusakan hati kronis. Di sisi lain, peningkatan nilai Bilirubin (BIL) dan Alkaline Phosphatase (ALP) pada kelompok fibrosis dan sirosis mencerminkan proses patofisiologis yang melibatkan inflamasi dan gangguan metabolisme bilirubin. Temuan ini sejalan dengan literatur hepatologi yang menyatakan bahwa Bilirubin merupakan penanda klinis penting dalam fase lanjutan penyakit hati, sementara ALP berhubungan dengan proses kolestatik dan fibrosis hati (Guidance, 2020).

Analisis regresi logistik multinomial memberikan informasi yang lebih mendalam mengenai peran setiap variabel dalam membedakan kategori diagnosis. Parameter regresi yang bersifat signifikan menunjukkan bahwa variabel-variabel tersebut mampu memperkirakan probabilitas suatu individu masuk ke dalam kategori klinis tertentu. Misalnya, variabel Bilirubin menghasilkan kontribusi terbesar dalam memisahkan kelompok Hepatitis dan Fibrosis, sedangkan Albumin lebih kuat dalam memisahkan kelompok Donor Darah dan Sirosis. Faktor usia juga menunjukkan peran penting, di mana peningkatan usia berhubungan dengan peningkatan probabilitas masuk ke kelompok Fibrosis dan Sirosis. Hal ini sejalan dengan studi epidemiologis yang menyebutkan bahwa proses fibrosis hati meningkat seiring bertambahnya usia akibat akumulasi inflamasi dan penurunan kapasitas regeneratif hepatosit (Seto et al., 2017).

Selain itu, hasil confusion matrix memperlihatkan bahwa kelas Blood Donor merupakan kelas yang paling mudah diprediksi oleh model. Hal ini dimungkinkan karena kelompok tersebut memiliki profil klinis yang relatif normal dan homogen. Sebaliknya, kelas Suspect Blood Donor menunjukkan tingkat misklasifikasi yang lebih tinggi. Secara klinis, kategori ini berada pada fase transisi antara keadaan normal dan fase hepatitis awal, sehingga nilai biomarker klinisnya tumpang tindih dengan dua kategori lainnya. Kondisi ini menyebabkan model menghadapi kesulitan dalam menentukan batas klasifikasi yang jelas. Hasil ini menegaskan bahwa ketepatan klasifikasi pada kondisi intermediate memerlukan variabel tambahan atau pendekatan model yang lebih kompleks, seperti model berbasis machine learning non-linear atau biomarker tambahan yang lebih sensitif.

Evaluasi melalui cross-validation menunjukkan performa yang cukup stabil pada model. Hal ini mengindikasikan bahwa model tidak mengalami overfitting dan mampu melakukan generalisasi terhadap data uji. Dari perspektif klinis, hal ini penting karena model klasifikasi biomarker hepatologis tidak hanya diuji berdasarkan akurasi statistik tetapi juga relevansi medis. Temuan ini membuka peluang penerapan model prediktif dalam praktik klinis untuk mendukung proses triase atau pengelompokan pasien berdasarkan risiko penyakit hati.

Secara keseluruhan, penelitian ini memberikan kontribusi teoritis dan praktis. Secara teoritis, temuan ini memperkuat literatur terkait hubungan biomarker klinis dengan progresivitas HCV. Secara praktis, hasil penelitian mendukung pemanfaatan data laboratorium sederhana sebagai dasar skrining awal terhadap pasien HCV, terutama pada populasi berisiko tinggi. Meskipun model menunjukkan performa yang memadai, penelitian ini juga mengindikasikan perlunya pendekatan multi-variabel dan model yang lebih kaya untuk meningkatkan sensitivitas klasifikasi pada kategori intermediate.

D. Kesimpulan

Penelitian ini menunjukkan bahwa biomarker klinis dan faktor demografis memiliki peran penting dalam membedakan kategori diagnosis pasien HCV. Hasil analisis deskriptif memperlihatkan bahwa nilai Albumin dan Kolesterol cenderung lebih tinggi pada kelompok donor darah, sedangkan peningkatan Bilirubin dan Alkaline Phosphatase ditemukan pada kelompok fibrosis dan sirosis yang mengindikasikan progresivitas penyakit hati. Variabel usia juga berperan dalam memisahkan diagnosis, di mana individu berusia lebih tua memiliki kecenderungan berada pada kategori fibrosis dan sirosis.

Model regresi logistik multinomial mampu memprediksi diagnosis HCV dengan performa yang memadai. Confusion matrix memperlihatkan bahwa kelas donor darah merupakan kelas yang paling mudah diprediksi, sedangkan kelas suspect blood donor memiliki tingkat misklasifikasi yang lebih tinggi karena kondisi klinisnya berada pada fase transisi. Evaluasi melalui cross-validation menunjukkan bahwa model memiliki tingkat konsistensi yang baik antara data latih dan data uji, sehingga dapat dikatakan model memiliki kemampuan generalisasi yang cukup stabil.

Secara keseluruhan, penelitian ini mendukung pemanfaatan data laboratorium klinis sederhana untuk mengidentifikasi dan mengklasifikasikan risiko penyakit hati pada pasien HCV. Temuan ini juga memperkuat literatur mengenai progresivitas HCV dan mengindikasikan bahwa pendekatan statistik dan data mining dapat digunakan sebagai alat bantu dalam analisis klinis dan epidemiologis.

Ucapan Terimakasih

Penulis menyampaikan ucapan terima kasih kepada seluruh pihak yang telah memberikan bimbingan dan arahan selama proses penelitian. Penulis juga berterima kasih kepada *UCI Machine Learning Repository* sebagai penyedia dataset sekunder yang digunakan dalam penelitian ini. Selain itu, apresiasi disampaikan kepada rekan dan pihak-pihak terkait yang turut memberikan masukan dalam proses analisis dan penyusunan artikel ilmiah ini.

Daftar Pustaka

- Altinel, K., Hashimoto, K., Wei, Y., Neuveut, C., Gupta, I., Suzuki, M., & Santos, D. (2016). Single-Nucleotide Resolution Mapping of Hepatitis B Virus Promoters in Infected Human Livers and Hepatocellular Carcinoma. *Journal of Virology*, Volume 90, 10811–10822. <https://doi.org/10.1128/JVI.01625-16.Editor>
- El-Zanaty, F., & Way, A. (2015). *Egypt Demographic and Health Survey 2014* (T. D. Program, Ed.). Ministry of Health and Population and ICF International.
- Guerra, J., & Garenne, M. (2012). HCV burden of infection in Egypt : results from a nationwide survey. *Journal of Viral Hepatitis*, 19, 560–567. <https://doi.org/10.1111/j.1365-2893.2011.01576.x>

- Guidance, P. (2020). Hepatitis C Guidance 2019 Update : American Association for the Study of Liver Diseases – Infectious Diseases Society of America Recommendations for Testing ., *Hepatology*, 71(2), 686–721. <https://doi.org/10.1002/hep.31060>
- Hassanin, A., Kamel, S., Waked, I., & Fort, M. (2021). Egypt ' s Ambitious Strategy to Eliminate Hepatitis C Virus : *Global Health: Science and Practice*, 9(1), 187–200.
- Kanaani, Z. Al, Mahmud, S., Silva, P., & Abu-raddad, L. J. (2018). The epidemiology of hepatitis C virus in Pakistan : systematic review and meta-analyses. *ROYAL SOCIETY OPEN SCIENCE*, 5: 180257, 1–21.
- Kandeel, A., Genedy, M., El-refai, S., Funk, A. L., Fontanet, A., & Talaat, M. (2017). The prevalence of hepatitis C virus infection in Egypt 2015 : implications for future policy on prevention and treatment. *Liver International*, (January 2016), 45–53. <https://doi.org/10.1111/liv.13186>
- Lehman, E. M., & Wilson, M. L. (2009). Epidemic hepatitis C virus infection in Egypt : estimates of past incidence and future morbidity and mortality. *Journal of Viral Hepatitis*, 16, 650–658. <https://doi.org/10.1111/j.1365-2893.2009.01115.x>
- Lestari, T. S., & Sirodj, D. A. N. (2022). Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest. *Jurnal Riset Statistika*, 1(2), 160–167. <https://doi.org/10.29313/jrs.v1i2.5jurnal>
- Mukhopadhyaya, A. (2008). Hepatitis C in India. *Department of Gastrointestinal Sciences*, 33(November), 465–473.
- Nelson, P. K., Mathers, B. M., Cowie, B., Hagan, H., Jarlais, D. Des, Horyniak, D., & Degenhardt, L. (2010). Global epidemiology of hepatitis B and hepatitis C in people who inject drugs : results of systematic reviews. *The Lancet*, 378(9791), 571–583. [https://doi.org/10.1016/S0140-6736\(11\)61097-0](https://doi.org/10.1016/S0140-6736(11)61097-0)
- Oktoriandi, D. (2022). Penerapan uji Q Cochran terhadap Atribut Produk Laptop Menggunakan Multiple Response Analysis (MRA). *Jurnal Riset Statistika*, 1(2), 127–134. <https://doi.org/10.29313/jrs.v1i2.521>
- Seto, W., Hui, R. W. H., Mak, L., Fung, J., Cheung, K., Liu, K. S. H., Wong, D. K., Lai, C., & Yuen, M. (2017). Association Between Hepatic Steatosis, Measured by Controlled Attenuation Parameter, and Fibrosis Burden in Chronic Hepatitis B. *Clinical Gastroenterology and Hepatology*. <https://doi.org/10.1016/j.cgh.2017.09.044>
- Strickland, G. T. (2006). Liver Disease in Egypt : Hepatitis C Superseded. *Perspective In Clinical Hepatology*, (May), 915–922. <https://doi.org/10.1002/hep.21173>
- Waked, I., Esmat, G., Elsharkawy, A., Serafy, M. El, Tabakh, H. El, Sc, M., Emad, E., Sc, M., Gemeah, H., Sc, M., Hashem, A., Sc, M., Abdalla, M., Akel, W. El, Doss, W., Zaid, H., & Sc, M. (2020). Spe ci a l R e p o r t Screening and Treatment Program to Eliminate Hepatitis C in Egypt. *The New Engl and Journal of Medicine Special, n engl j m*, 1166–1174.
- Wang, H., Wang, Y., Huang, Y., & Id, M. J. (2024). All-cause and cause-specific mortality risk among men and women with hepatitis C virus infection. *PLOS ONE*, 1–12. <https://doi.org/10.1371/journal.pone.0309819>

- Wei, B., Ji, F., Yeo, Y. H., Ogawa, E., Zou, B., Stave, C. D., Dang, S., Li, Z., Firusyo, N., Cheung, R. C., & Nguyen, M. H. (2018). Real-world effectiveness of sofosbuvir plus ribavirin for chronic hepatitis C genotype 2 in Asia: a systematic review and meta-analysis. *BMJ Open Gastroenterology*, 1–10. <https://doi.org/10.1136/bmjgast-2018-000207>
- Zain, M. N. (2022). Algoritma Artificial Neural Network dalam Klasifikasi Chest X-Rays Pasien COVID-19. *Jurnal Riset Statistika*, 137–144. <https://doi.org/10.29313/jrs.v2i2.1426>