
Implementation of KNN Method to Determine Prospective Student Interests
Penerapan Metode KNN untuk Menentukan Minat Calon Mahasiswa

Tiara Riyanto¹, Gomal Juni Yanris², Mila Nirmala Sari Hasibuan³

Manajemen Informatika, Fakultas Sains dan Teknologi, Universitas Labuhanbatu^{1,2,3}

Email : tiarariyanto124@gmail.com¹, gomaljunianris@gmail.com²,
milanirmalasari7@gmail.com³

Corresponding Author : tiarariyanto124@gmail.com

Abstract

This study focuses on the implementation of data mining to determine the interests of prospective male and female students in the Informatics Management Department using the K-Nearest Neighbors (KNN) method. The analysis process is carried out through the Knowledge Discovery in Databases (KDD) stages, which include data selection, pre-processing, transformation, data mining, and pattern evaluation. The KDD stage ensures that the data used has been prepared and processed properly to produce an accurate and relevant model. The KNN method is used to classify sample data consisting of 82 prospective male and female students. The results of this study indicate that 63 out of 82 prospective students are interested in the Informatics Management Department, while 19 other prospective students are not interested. This classification process shows that the KNN method is able to identify the interests of prospective students with a high level of accuracy, providing useful information for universities in understanding the preferences of their prospective students. Evaluation of the research results using two evaluation tools, namely Test and Score and Confusion Matrix, showed perfect results with an accuracy of 100%. Both of these evaluation tools are consistent in assessing the performance of the KNN model, confirming that this model works very well in classifying prospective student interests. In conclusion, the KNN method is proven to be effective and reliable in determining prospective students' interest in the Informatics Management Department, providing a strong foundation for similar applications in the future.

Keywords: *Data Mining, K-Nearest Neighbor (KNN) method, Knowledge Discovery in Databases (KDD), Confusion Matrix, Box Plots.*

I. Pendahuluan

Kuliah merupakan tahap pendidikan lanjutan yang memberikan kesempatan kepada individu untuk mendalami bidang ilmu tertentu secara lebih mendetail dan terstruktur. Di

perguruan tinggi, mahasiswa tidak hanya belajar melalui perkuliahan di kelas, tetapi juga melalui penelitian, praktek laboratorium, magang, dan berbagai kegiatan ekstrakurikuler yang

mendukung perkembangan akademik dan profesional mereka. Kurikulum yang lebih kompleks dan mendalam dibandingkan sekolah menengah menuntut mahasiswa untuk lebih mandiri dan proaktif dalam mengelola waktu serta sumber daya belajar mereka. Selain aspek akademik, kuliah juga merupakan masa penting dalam perkembangan pribadi dan sosial mahasiswa. Lingkungan kampus yang beragam memungkinkan mereka untuk bertemu dengan orang-orang dari berbagai latar belakang budaya dan sosial, memperluas wawasan, dan membangun jaringan yang berharga untuk masa depan. Berpartisipasi dalam organisasi mahasiswa, klub, dan kegiatan sosial lainnya juga membantu mahasiswa mengembangkan keterampilan kepemimpinan, komunikasi, dan kerjasama tim. Semua pengalaman ini, baik di dalam maupun di luar kelas, membentuk mahasiswa menjadi individu yang lebih siap menghadapi tantangan dunia kerja dan kehidupan setelah lulus.

Calon mahasiswa yang akan memasuki dunia perkuliahan menghadapi berbagai tantangan dan peluang baru dalam perjalanan pendidikan mereka. Mereka harus beradaptasi dengan lingkungan akademik yang lebih mandiri dibandingkan sekolah menengah, di mana mereka dituntut untuk mengembangkan kemampuan berpikir kritis, manajemen waktu, dan keterampilan penelitian. Di sisi lain, mereka juga berkesempatan untuk memperluas wawasan dan jaringan sosial melalui berbagai kegiatan ekstrakurikuler, seminar, dan proyek kolaboratif yang diselenggarakan oleh kampus. Selain itu, calon mahasiswa

perlu mempersiapkan diri secara mental dan emosional untuk menghadapi perubahan gaya hidup dan tanggung jawab yang lebih besar. Kemandirian dalam mengelola keuangan, menyesuaikan diri dengan jadwal perkuliahan yang padat, dan menjaga keseimbangan antara akademik dan kehidupan pribadi menjadi aspek penting dalam kesuksesan mereka di perguruan tinggi. Dengan persiapan yang matang dan dukungan yang memadai, calon mahasiswa dapat memanfaatkan masa perkuliahan sebagai waktu yang produktif dan berharga dalam membentuk masa depan mereka.

Setiap calon mahasiswa sering kali menghadapi dilema dalam memilih jurusan atau program studi yang tepat, yang sering kali menjadi sumber kebingungan dan kecemasan. Kebingungan ini dapat disebabkan oleh berbagai faktor, seperti kurangnya pemahaman mengenai minat dan bakat pribadi, tekanan dari orang tua atau lingkungan sekitar, serta ketidakpastian mengenai prospek karir di masa depan. Sebagai hasilnya, ada kalanya calon mahasiswa memilih jurusan yang kurang sesuai dengan minat atau keahlian mereka, yang dapat berdampak negatif terhadap motivasi dan performa akademik mereka di kemudian hari. Salah memilih jurusan dapat membawa konsekuensi serius, seperti rasa ketidakpuasan, penurunan prestasi akademik, hingga keinginan untuk berpindah jurusan atau bahkan berhenti kuliah. Oleh karena itu, penting bagi calon mahasiswa untuk melakukan refleksi diri yang mendalam, berkonsultasi dengan bimbingan karir, dan mencari informasi yang lengkap tentang berbagai jurusan serta prospek

kerjanya. Dengan demikian, mereka dapat membuat keputusan yang lebih bijak dan terinformasi, yang akan membantu mereka menjalani pendidikan tinggi dengan lebih percaya diri dan antusias.

Dari persoalan di atas, penulis ingin melakukan penelitian kepada para calon mahasiswa dan mahasiswi untuk memahami minat mereka dalam memilih jurusan Manajemen Informatika di Universitas Labuhanbatu. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi keputusan mereka, seperti pemahaman tentang kurikulum, prospek karir, minat pribadi dalam bidang teknologi informasi, serta dukungan dan dorongan dari lingkungan sekitar. Dengan menggali informasi ini, penulis berharap dapat memberikan wawasan yang lebih mendalam mengenai preferensi dan kebutuhan calon mahasiswa, sehingga pihak universitas dapat merancang strategi yang lebih efektif dalam menarik minat dan meningkatkan pendaftaran di jurusan tersebut. Penelitian ini juga akan mengungkapkan sejauh mana calon mahasiswa memiliki kesadaran dan pemahaman yang tepat tentang jurusan Manajemen Informatika, serta apakah mereka merasa bahwa jurusan ini dapat membantu mereka mencapai tujuan karir jangka panjang. Dengan menganalisis data yang diperoleh, penulis dapat memberikan rekomendasi yang berguna bagi Universitas Labuhanbatu dalam mengembangkan program promosi yang lebih tepat sasaran dan menyediakan informasi yang lebih komprehensif dan akurat tentang jurusan ini. Hasil penelitian ini diharapkan dapat membantu calon

mahasiswa membuat keputusan yang lebih terinformasi dan meningkatkan kepuasan mereka dalam memilih jurusan yang sesuai dengan minat dan bakat mereka.

II. Landasan Teori K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah sebuah metode machine learning yang digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini sederhana, mudah diimplementasikan, dan efektif pada data yang lebih besar. KNN juga kuat terhadap data training yang noise, dan dapat menghasilkan data yang lebih akurat.

III. Metode Penelitian

Metode K-Nearest Neighbor (KNN) digunakan dalam penelitian ini untuk menganalisis minat calon mahasiswa terhadap jurusan Manajemen Informatika di Universitas Labuhanbatu. Dalam konteks ini, Setelah pengumpulan data, langkah berikutnya adalah preproses data untuk membersihkan data yang tidak relevan atau yang hilang, serta melakukan transformasi data agar siap untuk analisis lebih lanjut. Setelah data dipreproses, proses utama menggunakan metode KNN akan dilakukan. KNN akan mengklasifikasikan minat calon mahasiswa berdasarkan kesamaan mereka dengan calon mahasiswa lain yang memiliki minat serupa terhadap jurusan Manajemen Informatika. Evaluasi dan interpretasi hasil dari proses ini akan memberikan wawasan yang lebih dalam tentang preferensi calon mahasiswa, yang dapat digunakan

untuk mendukung kebijakan penerimaan mahasiswa dan pengembangan program akademik yang lebih sesuai dengan kebutuhan mereka.

IV. Hasil dan Pembahasan Selection

Untuk tahap selection merupakan tahapan yang dilakukan untuk mengumpulkan data yang akan digunakan pada penelitian ini. Dengan adanya data, maka penelitian dapat dilakukan. Untuk data set yang akan digunakan pada Penelitian ini terdapat 2 data set yang akan digunakan. Untuk data set pertama yaitu data training dan data set yang kedua adalah data testing.

Data Training

Data training merupakan data yang digunakan untuk melatih data testing yang akan diklasifikasikan pada data mining dengan menggunakan metode KNN. Untuk data training yang akan digunakan sebanyak 16 data training.

Tabel 1. Data Training

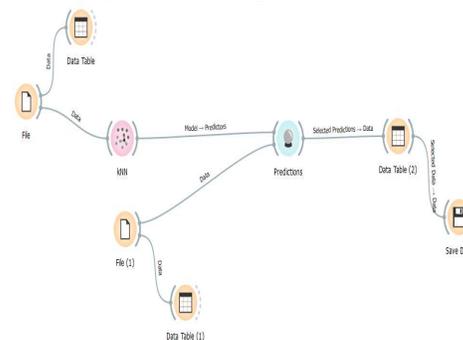
Nama	Akses Menuju Lokasi	Pelayanan Biro	Kebersihan Kampus	Keamanan	Kategori
Andi Pratama	Mudah Dijangkau	Baik	Bersih	Aman	Minat
Arif Wibowo	Sulit Dijangkau	Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Ayu Kartika	Mudah Dijangkau	Baik	Tidak Bersih	Aman	Minat
Budi Santoso	Sulit Dijangkau	Kurang Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Dewi Lestari	Mudah Dijangkau	Baik	Bersih	Aman	Minat
Dina Puspita	Sulit Dijangkau	Kurang Baik	Tidak Bersih	Aman	Tidak Minat
Fahri Aulia	Mudah Dijangkau	Kurang Baik	Bersih	Aman	Minat
Hendra Wijaya	Sulit Dijangkau	Kurang Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Indra Kusuma	Sulit Dijangkau	Baik	Bersih	Aman	Minat
Lina Marlina	Mudah Dijangkau	Kurang Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Lisa Amalia	Sulit Dijangkau	Kurang Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Rahmat Hidayat	Sulit Dijangkau	Kurang Baik	Tidak Bersih	Kurang Aman	Tidak Minat
Ratna Sari	Mudah Dijangkau	Baik	Bersih	Kurang Aman	Minat
Rizki Ramadhan	Mudah Dijangkau	Baik	Bersih	Aman	Minat
Siti Nurhaliza	Mudah Dijangkau	Baik	Bersih	Aman	Minat
Yusuf	Sulit Dijangkau	Kurang	Bersih	Kurang	Tidak

Data Testing

Untuk data testing merupakan data sampel penelitian yang digunakan untuk diklasifikasikan pada data mining dengan menggunakan metode KNN. Untuk data testing ini juga merupakan data sampel Penelitian. Untuk data yang digunakan yaitu sebanyak 82 data sampel.

Data Mining

Pada tahapan ini penulis akan merancang suatu model yang dapat digunakan untuk melakukan klasifikasi pada data testing yang sudah diperoleh sebelumnya. Dengan adanya model yang dirancang, maka proses klasifikasi akan dengan mudah dapat dilakukan.



Gambar 1. Klasifikasi Data Menggunakan Metode KNN

Pada gambar diatas merupakan model yang digunakan untuk melakukan klasifikasi data pada data mining dengan menggunakan metode KNN. Untuk hasil klasifikasinya nantinya akan memberikan 2 kelompok klasifikasi yaitu kelas minat dan kelas tidak minat.

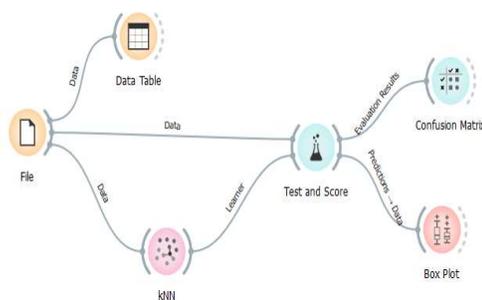
Hasil Klasifikasi

Untuk hasil klasifikasi merupakan hasil penelitian yang sudah dilakukan. Hasil klasifikasi yang diperoleh akan dipaparkan oleh penulis dalam bentuk tabel. Hal ini dikarenakan

untuk memudahkan untuk dapat memahami hasil dari klasifikasi. Pada penelitian hasil klasifikasi yang sudah dilakukan menggunakan metode KNN pada data mining. Dari hasil penelitian dengan menggunakan 82 data sampel, ditemukan bahwa sebanyak 63 dari data tersebut menunjukkan minat yang kuat terhadap Jurusan Manajemen Informatika, sedangkan sisanya, sebanyak 19 data sampel, menunjukkan ketidakminatan terhadap jurusan yang sama. Temuan ini memberikan gambaran yang jelas tentang preferensi calon mahasiswa terhadap program studi ini di Universitas Labuhanbatu. Analisis ini dapat digunakan untuk menyusun strategi penerimaan mahasiswa yang lebih efektif, serta untuk meningkatkan penawaran program akademik dan fasilitas yang sesuai dengan minat dan kebutuhan calon mahasiswa di masa mendatang.

Interpretation/Evaluation

Untuk evaluasi yang dilakukan pada tahapan ini untuk menentukan hasil kemampuan dari metode yang digunakan. Jadi untuk menentukan hasil kemampuan dari metode, penulis akan merancangn sebuah model yang akan digunakan untuk melakukan evaluasi.



Gambar 2. Model Evaluasi Menggunakan Metode KNN

Pada gambar diatas merupakan model evaluasi yang akan digunakan untuk melakukan evaluasi pada metode yang digunakan. Tujuan nya dilakukan evaluasi untuk mengetahui sebesar besar kemampuan metode yang digunakan pada penelitian ini ataupun cocok atau tidaknya metode KNN digunakan pada penelitian ini.

Hasil Test and Score

Tabel 1. Result of Test and Score

Model	AUC	CA	F1	Precision	Recall	MCC
K-Nearest Neighbor	1.000	1.000	1.000	1.000	1.000	1.000

Pada tabel diatas merupakan hasil dari Test and Score yang diberikan menggunakan metode KNN. Dari hasil penelitian yang diperoleh, semua metrik evaluasi seperti Area Under the Curve (AUC), Classification Accuracy (CA), F1 Score, Precision, Recall, dan Matthew's Correlation Coefficient (MCC) menunjukkan hasil sempurna dengan nilai 100%. Ini menunjukkan bahwa model yang digunakan dalam penelitian ini memiliki kinerja yang sangat baik dan mampu mengklasifikasikan data dengan akurasi dan konsistensi yang sangat tinggi pada semua metrik yang diuji.

Hasil Confusion Matrix

		Predicted		Σ
		Interest	Not Interested	
Actual	Interest	63	0	63
	Not Interested	0	19	19
Σ		63	19	82

Gambar 3. Confusion Matrix Results in the Neural Network Method

Pada tabel di atas merupakan hasil evaluasi yang dilakukan pada metode KNN dengan menggunakan widget Confusion Matrix. Hasilnya

menunjukkan bahwa terdapat 63 data yang termasuk dalam kategori True Positif (TP), 19 data dalam kategori True Negatif (TN), 0 data dalam kategori False Positif (FP), dan 0 data dalam kategori False Negatif (FN). Meskipun hasil ini tampak sangat baik, untuk mengukur nilai akurasi secara tepat, data tersebut harus dihitung terlebih dahulu menggunakan rumus yang ada pada confusion matrix, yaitu sebagai berikut.

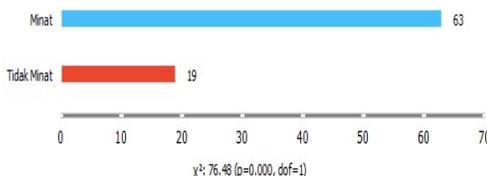
$$Accuracy = \frac{63+19}{63+19+0+0} + 100\% \quad \text{Then the Accuracy value} = 100\%$$

$$Presisi = \frac{63}{63+0} + 100\% \quad \text{Then the Precision value} = 100\%$$

$$Recall = \frac{63}{63+0} + 100\% \quad \text{Then the Recall value} = 100\%$$

Dari hasil perhitungan yang dilakukan, akurasi yang diperoleh dari widget confusion matrix untuk metode K-Nearest Neighbors (KNN) menunjukkan performa yang luar biasa, dengan tingkat akurasi mencapai 100%. Temuan ini mengindikasikan bahwa metode KNN sangat cocok diterapkan dalam penelitian ini untuk klasifikasi minat calon mahasiswa pada Jurusan Manajemen Informatika di Universitas Labuhanbatu. Akurasi sempurna tersebut mengungkapkan bahwa model KNN mampu mengklasifikasikan data dengan sangat akurat, sehingga dapat diandalkan untuk analisis serupa di masa depan.

Hasil Box Plot



Gambar 4. Hasil Box Plot

Untuk hasil Box Plot yang diberikan sesuai dengan hasil klasifikasi

yang diperoleh pada Penelitian ini. Hasil evaluasi menunjukkan bahwa metode K-Nearest Neighbors (KNN) sangat cocok digunakan untuk penelitian ini. Hal ini terlihat dari garis tebal berwarna biru yang menunjukkan 63 data calon mahasiswa dan mahasiswi yang berminat pada Jurusan Manajemen Informatika, sedangkan garis tebal berwarna merah mewakili calon mahasiswa dan mahasiswi yang tidak berminat pada jurusan tersebut. Temuan ini menguatkan keandalan metode KNN dalam mengklasifikasikan minat calon mahasiswa, sehingga dapat diandalkan untuk analisis serupa di masa mendatang.

V. Kesimpulan

Kesimpulannya, penelitian ini menunjukkan bahwa metode K-Nearest Neighbors (KNN) sangat efektif dalam mengklasifikasikan minat calon mahasiswa pada Jurusan Manajemen Informatika. Dengan hasil evaluasi yang sempurna dari widget Test and Score dan Confusion Matrix, yang keduanya memberikan akurasi 100%, metode KNN terbukti mampu mengidentifikasi minat calon mahasiswa dengan sangat tepat. Dari total data yang dianalisis, 63 calon mahasiswa menunjukkan minat pada jurusan tersebut, sementara 19 calon mahasiswa tidak berminat, menegaskan keandalan metode ini dalam proses klasifikasi. Keberhasilan metode KNN dalam penelitian ini tidak hanya menyoroti ketepatan dan stabilitas model, tetapi juga memberikan dasar yang kuat untuk penerapan KNN dalam analisis serupa di masa depan. Keselarasan hasil evaluasi antara widget Test and Score dan Confusion Matrix dengan rasio 1:1 memperkuat kesimpulan bahwa metode

KNN dapat diandalkan dalam mengklasifikasikan data dengan akurasi tinggi. Dengan demikian, metode KNN dapat dijadikan sebagai alat analisis yang andal dalam memahami minat dan preferensi calon mahasiswa di berbagai institusi pendidikan tinggi.

VI. Daftar Pustaka

- A. Ayu, D. Sulistyawati, and M. Sadikin, "SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan," vol. 10, pp. 516–526, 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- F. Rozi, M. Bagoes, and S. Junianto, "Penerapan Machine Learning Untuk Prediksi Harga Saham PT.Telekomunikasi Indonesia Tbk Menggunakan Algoritma K-Nearest Neighbors," *J. Inform. MULTI*, vol. 1, no. 1, pp. 18–24, 2023.
- H. A. Pratama, G. J. Yanris, M. Nirmala, and S. Hasibuan, "Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels," vol. 8, no. 3, pp. 1832–1851, 2023.
- H. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijiis.v4i1.73.
- I. P. Ninditama, I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, "Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang," vol. 15, no. 2, pp. 37–49, 2020.
- K. Ma, "Analisis Penerapan Algoritma ID3 dalam Mendiagnosis Kesuburan Pria," 2019.
- K. N. Di, K. P. P. Pratama, C. Dua, P. Studi, K. Akuntansi, and S. I. Cirebon, "JURNAL DATA SCIENCE & INFORMATIKA (JDSI) Klasifikasi Pemberian Sanksi Pajak Dengan Metode," vol. 1, no. 2, pp. 41–45, 2021.
- K. N. Neighbor, A. Pratama, F. Ali, I. Ade, and R. Rinaldi, "JURNAL DATA SCIENCE & INFORMATIKA (JDSI) Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma," vol. 1, no. 1, pp. 11–15, 2021.
- M. Ula, R. Zulhusna, R. P. Fhonna, and A. Pratama, "Penerapan Model Klasifikasi K-Nearest Neighbor Dalam Pencarian Kesesuaian Pekerjaan," 2022, doi: 10.47002/metik.v6i1.343.
- N. K. Sriwinarti and P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa (Analysis of K-Nearest Neighbors (K-NN) and Naive Bayes Methods in Predicting Student Graduation)," vol. 3, no. 2, pp. 106–112, 2021.
- P. Arsi, L. N. Hidayati, and A. Nurhakim, "Komparasi Model

- Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform Instagram,” vol. 6, pp. 459–466, 2022, doi: 10.30865/mib.v6i1.3509.
- R. Ali, M. M. Yusro, M. S. Hitam, and M. Ikhwanuddin, “Machine Learning With Multistage Classifiers For Identification Of Ectoparasite Infected Mud Crab Genus Scylla,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 2, pp. 406–413, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16724.
- R. G. de Luna, E. P. Dadios, A. A. Bandala, and R. R. P. Vicerra, “Tomato growth stage monitoring for smart farm using deep transfer learning with machine learning-based maturity grading,” *Agrivita*, vol. 42, no. 1, pp. 24–36, 2020, doi: 10.17503/agrivita.v42i1.2499.
- S. Diansyah, “Jurnal Sistim Informasi dan Teknologi Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN),” vol. 4, pp. 1–3, 2022, doi: 10.37034/jsisfotek.v4i1.114.
- W. Sudrajat, I. Cholid, and J. Petrus, “Wahyu Sudrajat et al, Penerapan Algoritma K-Means Untuk,” p. 27, 2022.
- Y. Indah Lestari and S. Defit, “Jurnal Informatika Ekonomi Bisnis Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C.45,” vol. 3, pp. 148–154, 2021, doi: 10.37034/infeb.v3i3.104.