



Klasifikasi Minat Masyarakat Dalam Berlangganan Netflix di Masa Pandemi Covid-19 dengan Algoritma Naïve Bayes (Studi Kasus: Kecamatan Karawang Timur)

Andreas Anggono¹, Agung Susilo Yuda Irawan², Purwantoro³

^{1,2,3}Universitas Singaperbangsa Karawang

Abstract

Received: 22 Februari 2023

Revised: 28 Februari 2023

Accepted: 4 April 2023

The first time the Covid-19 (Coronavirus Disease 2019) outbreak appeared was at the end of 2019 and the first time a Covid-19 case was found in Indonesia was in March 2020. Due to its fast spread, the World Health Organization (WHO) finally declared this outbreak a pandemic global. In order to prevent the spread of this epidemic, various countries implemented policies such as implementing social restrictions, lockdowns, and various other policies. With the ongoing Covid-19 outbreak, this has had quite a big impact on various sectors of life which of course has hampered the Indonesian economy. Netflix is a Video on Demand application that is quite popular and has been established since 1997. Therefore, this research was compiled with the aim of obtaining results of the classification of the people's interest in subscribing to Netflix during the Covid-19 pandemic and to find out how many people use it, interested and not interested in subscribing to Netflix by applying one of the data mining techniques, namely classification with the naïve Bayes algorithm. The data in this study were obtained from the results of filling out a questionnaire in the form of a google form which was distributed on several social media such as Instagram, Twitter and Telegram and was filled out by the people of East Karawang District. The total sample data used is 399 records obtained from the calculation of the slovin formula. Then the data will be processed using the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology which consists of six stages. The 399 records will be divided into two, namely training data and testing data. The process of dividing training data and testing data will be grouped into three ratios, including 90:10, 80:20, and 70:30 which will then be applied to RapidMiner tools to determine accuracy, precision, recall, and AUC values. The results obtained from applying data to the RapidMiner tools show that the 90:10 ratio has better results compared to other ratios, where the accuracy value is 92.50%, the precision value is 92.11%, the recall value is 100%, and AUC of 0.857.

Keywords: Naïve Bayes Algorithm, Data Mining, Classification, Netflix

(*) Corresponding Author: andreasanggono5440@gmail.com, agung@unsika.ac.id,
purwantoro.masbro@staff.unsika.ac.id

How to Cite: Anggono, A., Irawan, A. S., & Purwantoro, P. (2023). Klasifikasi Minat Masyarakat Dalam Berlangganan Netflix di Masa Pandemi Covid-19 dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Wahana Pendidikan*, 9(8), 31-44. <https://doi.org/10.5281/zenodo.7865546>

PENDAHULUAN

Penyakit Covid-19 (*Coronavirus Disease 2019*) pertama kali muncul pada akhir tahun 2019 dan pertama kali terdeteksi kasus Covid-19 di Indonesia pada Maret 2020. Karena kecepatan penyebarannya yang cepat, *World Health Organization* (WHO) akhirnya menyatakan wabah itu sebagai pandemi global. Untuk mencegah penyebaran wabah ini, banyak negara yang menerapkan kebijakan seperti menerapkan pembatasan sosial, *lockdown* dan masih banyak lagi lainnya. Dengan adanya wabah Covid-19 yang masih berlangsung, hal ini



memberikan dampak yang cukup besar di berbagai bidang kehidupan yang tentunya menghambat perekonomian Indonesia (Hasna, 2021).

Saat ini, banyak bisnis yang berhenti beroperasi akibat dampak pandemi Covid-19. Namun, ada satu bisnis yang saat ini sedang berkembang yaitu menjual layanan Netflix. Netflix termasuk dalam kategori *Video on Demand* yaitu sebuah fasilitas berupa sistem televisi interaktif yang dapat digunakan pengguna untuk memilih video atau film yang ingin ditonton dengan sebuah perangkat, perangkat tersebut tentunya harus terkoneksi dengan internet. Aplikasi Netflix sangat digemari oleh semua lapisan masyarakat dan alasan mengapa pengguna layanan Netflix ini tetap berlangganan setiap bulan adalah untuk mengisi waktu luang, hiburan, hobi atau tidak bisa menonton film di bioskop karena tidak bekerja karena pandemi covid-19 (Malia, Zakiyah, & R, 2021). Oleh karena itu, dapat disimpulkan bahwa keberadaan Netflix yang membuat penggunanya tidak bosan meski di rumah adalah karena kebijakan yang ditujukan untuk memutus mata rantai penyebaran Covid-19.

Didirikan pada tahun 1997, Netflix kini tersedia di 190 negara termasuk Indonesia dan per 18 Oktober 2018 bahasa Indonesia pertama kali muncul di Netflix (Susilawati, Restu, Alamanda, & Wahid, 2021). Netflix adalah layanan *online* yang memerlukan langganan agar anggota dapat menonton acara TV dan film di perangkat mereka tanpa iklan, misalnya melalui *smartphone*, smart TV, laptop, PC, dan laptop yang terhubung ke internet (Sudarsono, Leo, Santoso, & Hendrawan, 2021).

Mengutip BBC (2021), Netflix kini memiliki lebih dari 200 juta pelanggan berbayar, berdasarkan angka tersebut terlihat bahwa pelanggan berbayar Netflix meningkat 30% dibandingkan tahun 2019. Di tahun 2020, terdapat 37 juta pelanggan baru layanan Netflix. termasuk 8,5 juta pelanggan baru dalam tiga bulan terakhir tahun ini. *Streaming* akan terus berlanjut karena perkembangan yang semakin maju didukung oleh pengguna internet setiap tahunnya, seperti pada akhir tahun 2020 pengguna internet di Indonesia akan mencapai 173 juta orang atau setengah dari populasi pada tahun 2020 dan awal tahun 2021 pengguna internet akan terus meningkat untuk 210 juta jiwa (Kompas, 2021).

Dengan aplikasi Netflix, para pelaku industri film memanfaatkan sepenuhnya pertumbuhan teknologi ini. Menggunakan aplikasi platform *streaming* seperti aplikasi Netflix sangat relevan di masa pandemi seperti ini. Selain dapat mengikuti protokol pemerintah, juga dapat menghemat biaya perjalanan, makan, dan tiket bioskop. Namun, untuk dapat mengakses aplikasi Netflix, Anda hanya memerlukan internet dan berlangganan aplikasi Netflix. Platform *streaming* menghadirkan kenyamanan dan kemudahan layanan (Hasna, 2021).

Untuk mengetahui tingkat minat masyarakat terhadap langganan Netflix di masa pandemi Covid-19, dilakukan penelitian dengan menggunakan teknik *data mining*. Penelitian ini menggunakan data yang diperoleh dari hasil pengisian kuesioner berupa *Google Form* oleh masyarakat Kecamatan Karawang Timur dengan total *record* data sebanyak 399 data. Menurut Badan Pusat Statistik (BPS) Kabupaten Karawang, bahwa jumlah penduduk Kecamatan Karawang Timur pada tahun 2020 sebanyak 131.446 jiwa (BPS, 2023). Data tersebut akan diolah menggunakan salah satu teknik *data mining* yaitu klasifikasi menggunakan

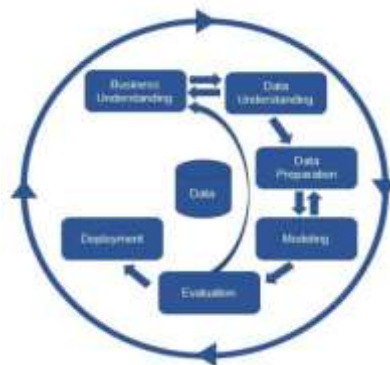
algoritma *naive bayes*. Klasifikasi adalah teknik yang digunakan untuk mengelompokkan data menurut data latih dan data uji serta membuat perbandingan antara data lama dan baru (Wulandari, Jusia, & Jasmir, 2020).

Dalam penelitian yang dilakukan oleh Dahlia, Wuryani, Hadianti, Gata, & Selawati (2021) untuk tujuan klasifikasi data Covid-19. Data yang digunakan berasal dari kaggle bernama *Science for COVID-19 (DS4C)*, yaitu dataset informasi pasien yang berisi informasi tentang pasien positif Covid-19 di Korea Selatan. Total data yang digunakan adalah 2420 *record* dan meliputi 5 atribut diantaranya jenis kelamin, usia, kota, kasus, dan negara bagian. Berdasarkan hasil penerapan teknik klasifikasi data mining dengan algoritma *naive bayes*, C4.5, dan K-NN pada alat RapidMiner, didapatkan hasil akurasi sebesar 80,79% dan AUC sebesar 0,881 dengan algoritma *naive bayes*. Sedangkan hasil akurasi sebesar 56,20% dan AUC adalah 0,743 dengan algoritma C4.5. Dan juga hasil akurasi sebesar 60,74% dan AUC sebesar 0,567 dengan algoritma K-NN. Oleh karena itu, dapat disimpulkan bahwa algoritma dengan akurasi tertinggi adalah algoritma *naive bayes*.

Penelitian lainnya dilakukan oleh Firdaus, Walid, & Anwari (2022) dengan menerapkan algoritma *naive bayes* untuk mengklasifikasikan kasus pasien Covid-19 di RSUD Kabupaten Pamekasan ke dalam *cluster* kasus tinggi atau rendah. Data yang digunakan adalah data pasien Covid-19 di RSUD Pamekasan periode Januari sampai dengan Desember 2021. Dengan total data training sebanyak 23 *record*, diperoleh hasil kasus Covid-19 di Pamekasan tergolong tinggi dengan nilai akurasi (0,9688), presisi (0,97), recall (1,00), dan skor f-1 (0,98).

METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Metodologi ini terdiri dari enam langkah meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran (Fitriani, Nama, & Mardiana, 2022) seperti pada **Gambar 1**. berikut ini.



Gambar 1. Tahapan CRISP-DM
Sumber : (Rifai, Jatnika, & Valentino, 2019)

Pemahaman Bisnis (Business Understanding)

Pada tahap ini akan dilakukan pemahaman bisnis mengenai kebutuhan dan tujuan untuk mencapai hasil yang diinginkan. Setelah pemahaman bisnis sudah

diketahui, maka akan didapatkan suatu informasi mengenai *data mining* untuk mengklasifikasi tingkat minat masyarakat dalam berlangganan Netflix di masa pandemi Covid-19.

Pemahaman Data (Data Understanding)

Pada tahap ini, data akan ditentukan dan dikumpulkan untuk diolah pada proses klasifikasi *data mining*. Data yang dibutuhkan merupakan data yang diperoleh dari hasil pengisian kuisioner oleh masyarakat Kecamatan Karawang Timur mengenai minat masyarakat dalam berlangganan Netflix di masa pandemi Covid-19. Dalam menentukan jumlah sampel yang mewakili jumlah penduduk Kecamatan Karawang Timur, maka akan dilakukan perhitungan menggunakan rumus slovin. Adapun atribut-atribut berdasarkan data yang diperoleh dari hasil pengisian kuisioner akan digunakan pada tahap selanjutnya.

Persiapan Data (Data Preparation)

Tahap ini merupakan tahap mempersiapkan data yang kemudian akan digunakan pada tahap pemodelan. Data yang diperoleh akan dipersiapkan dengan cara menentukan atribut-atribut apa saja yang akan diolah pada tahap selanjutnya serta menentukan rasio pembagian *data training* dan *data testing* yang akan digunakan.

Pemodelan (Modelling)

Pada tahap ini akan dilakukan penerapan algoritma yang sebelumnya telah dipilih. Penelitian ini menggunakan algoritma *naïve bayes* yang akan diterapkan pada data yang telah disiapkan sebelumnya yaitu data mengenai minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19. Klasifikasi menggunakan algoritma *naïve bayes* ini akan diterapkan pada *tools* RapidMiner.

Evaluasi (Evaluation)

Pada tahap ini akan digunakan untuk mengukur tingkat keakuratan terhadap pemodelan pada tahap sebelumnya, yaitu klasifikasi menggunakan algoritma *naïve bayes*. Metode evaluasi yang digunakan yaitu *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan AUC.

Penyebaran (Deployment)

Pada tahap ini akan digunakan sebagai tahap untuk membuat laporan dalam bentuk tugas akhir atau skripsi yang memuat informasi dan pengetahuan yang didapatkan dari hasil klasifikasi data mengenai minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19.

HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan dengan menerapkan algoritma *naïve bayes* memperoleh hasil klasifikasi tingkat minat masyarakat dalam berlangganan Netflix di masa pandemi Covid-19 dan mengetahui seberapa banyak masyarakat yang berminat dan tidak berminat untuk berlangganan Netflix. Penelitian ini menggunakan metodologi CRISP-DM untuk mengelola data yang di peroleh dari

hasil penyebaran kuisisioner dalam bentuk *google form* kepada masyarakat khususnya daerah Kecamatan Karawang Timur pada tahun 2022 dengan total data yang di peroleh sebanyak 399 *record*. Kemudian seluruh data tersebut akan diproses berdasarkan enam tahapan pada metodologi CRISP-DM menggunakan *tools* RapidMiner.

Pemahaman Bisnis (Business Understanding)

Pada tahap ini akan dilakukan pendefinisian masalah dan tujuan yang akan dicapai. Permasalahan yang ada yaitu belum adanya informasi mengenai minat berlangganan Netflix di masa pandemi Covid-19. Oleh karena itu penulis berkeinginan untuk mengolah data minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19 guna memperoleh hasil klasifikasi dan berapa banyak masyarakat Kecamatan Karawang Timur yang berminat berlangganan Netflix di masa pandemi Covid-19.

Pemahaman Data (Data Understanding)

Pada tahap ini akan dilakukan penentuan atribut yang akan menyelesaikan permasalahan yang ada pada tahap pemahaman bisnis. Data pada penelitian ini diperoleh dengan cara menyebarkan kuisisioner berupa *google form* di beberapamedia sosial seperti Instagram, Twitter, dan Telegram. Adapun ketentuan responden, diantaranya responden merupakan warga yang berdomisili Kecamatan Karawang Timur, usia responden antara 18-64 tahun, dan responden telah memiliki penghasilan. Dari data tersebut sebanyak tujuh atribut merupakan atribut penentu minat berlangganan Netflix di masa pandemi Covid-19 dan sebanyak satu atribut sebagai label dapat dilihat pada **Tabel 1.** berikut ini.

Tabel 1. Variabel, Keterangan, Kategori

Variabel	Keterangan	Kategori
X1	Usia	1 : 18 – 25 Tahun 2 : 26 – 35 Tahun 3 : 36 – 45 Tahun 4 : 46 – 55 Tahun 5 : 55 – 64 Tahun
X2	Jenis Kelamin	1 : Pria 2 : Wanita
X3	Apakah Anda suka menonton film di masa pandemi Covid-19?	1 : Ya 2 : Tidak
X4	Apa genre film yang paling sering Anda tonton?	1 : Horror/Thriller 2 : Komedi 3 : Anime/Kartun 4 : Laga/Action 5 : Drama/Drama Korea 6 : Dokumenter 7 : Fiksi Ilmiah/Fantasi
X5	Berapakah penghasilan Anda perbulan?	1 : < Rp 1.500.000 2 : Rp 1.500.000 – Rp 2.500.000 3 : Rp 2.500.000 – Rp 3.500.000 4 : > Rp 3.500.000
X6	Paket berlangganan Netflix yang akan Anda pilih?	1 : Paket Ponsel (Rp 54.000) 2 : Paket Dasar (Rp 120.000) 3 : Paket Standart (Rp 153.000) 4 : Paket Premium (Rp 186.000)

X7	Apakah Anda mengunduh aplikasi Netflix di masa pandemi Covid-19?	1 : Ya 2 : Tidak
Y(status)	Apakah Anda berminat berlangganan Netflix di masa pandemi Covid-19?	1 : Ya 2 : Tidak

Perhitungan dengan rumus slovin untuk menentukan jumlah sampel yang mewakili jumlah penduduk Kecamatan Karawang Timur.

$$n = \frac{N}{1 + N(e^2)}$$

$$n = \frac{131446}{1 + 131446(0,05^2)}$$

$$n = \frac{131446}{1 + 131446(0,0025)}$$

$$n = \frac{131446}{1 + 328,615}$$

$$n = \frac{131446}{329,615}$$

$$n = 398,7864$$

Maka jumlah sampel yang diperlukan yaitu sebanyak 399 *record*.

Persiapan Data (Data Understanding)

Data awal yang diperoleh dari hasil pengisian kuisioner berupa *google form* oleh masyarakat Kecamatan Karawang Timur mengenai tingkat minat berlangganan Netflix di masa pandemi Covid-19 terdiri dari 9 atribut dan 399 *record*. Pada tahap ini akan dilakukan pemilihan atribut yang akan digunakan pada tahap selanjutnya, dimana atribut “Nama Lengkap” akan dihapus. Sehingga total atribut yang akan digunakan sebanyak 8 atribut. Dari 399 *record* tersebut akan dikelompokkan menjadi dua data yaitu *data training* dan *data testing* dengan melakukan tiga perbandingan rasio yaitu 90:10, 80:20, dan 70:30. Berdasarkan perbandingan 90:10 maka dari 399 *record* sebanyak 90% (359 *record*) akan digunakan sebagai *data training* dan 10% (40 *record*) akan digunakan sebagai *data testing*. Berdasarkan perbandingan 80:20 maka dari 399 *record* sebanyak 80% (319 *record*) akan digunakan sebagai *data training* dan 20% (80 *record*) akan digunakan sebagai *data testing*. Berdasarkan perbandingan 70:30 maka dari 399 *record* sebanyak 70% (279 *record*) akan digunakan sebagai *data training* dan 30% (120 *record*) akan digunakan sebagai *data testing*.

Tabel 2. Dataset

X1	X2	X3	X4	X5	X6	X7	Y
1	2	1	5	1	3	1	1
1	1	1	2	4	4	1	1
1	1	1	2	4	4	1	1
1	2	1	5	3	3	1	1
1	1	1	4	2	3	1	1
...

...
1	2	1	5	1	1	1	1
1	1	1	3	4	2	1	2
1	2	1	5	2	2	1	1

Pemodelan (Modelling)

Pada tahap ini akan dilakukan pengujian teknik *data mining* klasifikasi dengan algoritma *naïve bayes* pada *tools* RapidMiner untuk mengklasifikasikan minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19 berdasarkan tiga perbandingan *data training* dan *data testing* yang telah ditentukan. Hasil rasio 90:10, 80:20, dan 70:30 dari *tools* RapidMiner akan dibandingkan sehingga diperoleh hasil yang paling baik diantaranya ketiga perbandingan tersebut.

Pada **Gambar 2.** merupakan proses penyambungan seluruh operator untuk mendapatkan hasil dari klasifikasi minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19. Lakukan proses penyambungan tersebut ke seluruh rasio, yaitu pada rasio 90:10, 80:20, dan 70:30.



Gambar 2. Sambungkan seluruh operator

1. Hasil klasifikasi algoritma *naïve bayes* pada *tools* RapidMiner untuk rasio 90:10

Pada **Gambar 3.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *accuracy* sebesar 92,50% untuk rasio 90:10.

accuracy: 92.50%

	true Ya	true Tidak	class precision
pred. Ya	35	3	92.11%
pred. Tidak	0	2	100.00%
class recall	100.00%	40.00%	

Gambar 3. Accuracy rasio 90:10

Pada **Gambar 4.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *precision* Ya sebesar 92,11% untuk rasio 90:10.

precision: 100.00% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	35	3	92.11%
pred. Tidak	0	2	100.00%
class recall	100.00%	40.00%	

Gambar 4. *Precision* rasio 90:10

Pada **Gambar 5.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *recall* Ya sebesar 100% untuk rasio 90:10.

recall: 40.00% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	35	3	92.11%
pred. Tidak	0	2	100.00%
class recall	100.00%	40.00%	

Gambar 5. *Recall* rasio 90:10

Pada **Gambar 6.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil AUC sebesar 0,857 untuk rasio 90:10.



Gambar 6. AUC rasio 90:10

- Hasil klasifikasi algoritma *naïve bayes* pada *tools* RapidMiner untuk rasio 80:20

Pada **Gambar 7.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *accuracy* sebesar 86,25% untuk rasio 80:20.

accuracy: 86.25%

	true Ya	true Tidak	class precision
pred. Ya	87	11	85.90%
pred. Tidak	0	2	100.00%
class recall	100.00%	15.38%	

Gambar 7. *Accuracy* rasio 80:20

Pada **Gambar 8.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *precision* Ya sebesar 85,90% untuk rasio 80:20.

precision: 100.00% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	87	11	85.90%
pred. Tidak	0	2	100.00%
class recall	100.00%	15.38%	

Gambar 8. Precision rasio 80:20

Pada **Gambar 9.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *recall* Ya sebesar 100% untuk rasio 80:20.

recall: 15.38% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	87	11	85.90%
pred. Tidak	0	2	100.00%
class recall	100.00%	15.38%	

Gambar 9. Recall rasio 80:20

Pada **Gambar 10.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil AUC sebesar 0,660 untuk rasio 80:20.



Gambar 10. AUC rasio 80:20

3. Hasil klasifikasi algoritma *naïve bayes* pada *tools* RapidMiner untuk rasio 70:30

Pada **Gambar 11.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *accuracy* sebesar 85,83% untuk rasio 70:30.

accuracy: 85.83%

	true Ya	true Tidak	class precision
pred. Ya	99	17	85.34%
pred. Tidak	0	4	100.00%
class recall	100.00%	19.05%	

Gambar 11. Accuracy rasio 70:30

Pada **Gambar 12.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *precision* Ya sebesar 85,34% untuk rasio 70:30.

precision: 100.00% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	99	17	85.34%
pred. Tidak	0	4	100.00%
class recall	100.00%	19.05%	

Gambar 12. Precision rasio 70:30

Pada **Gambar 13.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil *recall* Ya sebesar 100% untuk rasio 70:30.

recall: 19.05% (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	99	17	85.34%
pred. Tidak	0	4	100.00%
class recall	100.00%	19.05%	

Gambar 13. Recall

Pada **Gambar 14.** berdasarkan proses penyambungan seluruh operator yang di lakukan sebelumnya maka diperoleh hasil AUC sebesar 0,621 untuk rasio 70:30.



Gambar 14. AUC rasio 70:30

Evaluasi (Evaluation)

Pada tahap ini akan dilakukan evaluasi dari teknik *data mining* klasifikasi menggunakan algoritma *naïve bayes*. Hasil klasifikasi dari rasio 90:10, 80:20, dan 70:30 akan diterapkan ke dalam *confusion matrix* yang dapat dilihat pada **Tabel 15.** berikut ini.

1. Hasil *Confusion Matrix* Rasio 90:10

Tabel 3. Confusion Matrix rasio 90:10

Predicted	Class	
	Ya	Tidak
Ya	35	3
Tidak	0	2

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{35+2}{35+2+3+0} \\
 &= \frac{37}{40} \\
 &= 0,925 * 100\% \\
 &= 92,50\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP+FP} \\
 &= \frac{35}{35+3} \\
 &= \frac{35}{38} \\
 &= 0,921052 * 100\% \\
 &= 92,1052\%
 \end{aligned}$$

$$\begin{aligned}
 Recall &= \frac{TP}{TP+FN} \\
 &= \frac{35}{35+0} \\
 &= \frac{35}{35} \\
 &= 1 * 100\% \\
 &= 100\%
 \end{aligned}$$

2. Hasil *Confusion Matrix* Rasio 80:20

Tabel 4. Confusion Matrix rasio 80:20

Predicted	Class	
	Ya	Tidak
Ya	67	11
Tidak	0	2

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{67+2}{67+2+11+0} \\
 &= \frac{69}{80} \\
 &= 0,8625 * 100\%
 \end{aligned}$$

$$= 86,25\%$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{67}{67+11} \\ &= \frac{67}{78} \\ &= 0,858974359 * 100\% \\ &= 85,8974359\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{67}{67+0} \\ &= \frac{67}{67} \\ &= 1 * 100\% \\ &= 100\% \end{aligned}$$

3. Hasil *Confusion Matrix* Rasio 70:30

Tabel 5. Confusion Matrix rasio 70:30

Predicted	Class	
	Ya	Tidak
Ya	99	17
Tidak	0	4

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ &= \frac{99+4}{99+4+17+0} \\ &= \frac{103}{120} \\ &= 0,858333333 * 100\% \\ &= 85,8333333\% \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{99}{99+17} \\ &= \frac{99}{116} \end{aligned}$$

$$= 0,853448276 * 100\%$$

$$= 85,3448276\%$$

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{35}{35+0} \\ &= \frac{35}{35} \\ &= 1 * 100\% \\ &= 100\% \end{aligned}$$

Penyebaran (Deployment)

Pada tahap ini merupakan tahap pembuatan laporan hasil dari klasifikasi minat masyarakat Kecamatan Karawang Timur berlangganan Netflix di masa pandemi Covid-19 pada *tools* RapidMiner.

KESIMPULAN

Dari hasil penelitian klasifikasi *data mining* menggunakan algoritma *naïve bayes*, maka dapat ditarik kesimpulan sebagai berikut :

1. Dengan mengimplementasikan algoritma *naïve bayes* untuk memperoleh hasil klasifikasi data tingkat minat masyarakat Kecamatan Karawang Timur dalam berlangganan Netflix di masa pandemi Covid-19 menggunakan metodologi CRISP-DM pada *tools* RapidMiner dapat berjalan dengan baik. Rasio 90:10 memiliki *accuracy* sebesar 92,50%, *precision* sebesar 92,11%, *recall* sebesar 100%. Rasio 80:20 memiliki *accuracy* sebesar 86,25%, *precision* sebesar 85,90%, *recall* sebesar 100%. Rasio 70:30 memiliki *accuracy* sebesar 85,83%, *precision* sebesar 85,34%, *recall* sebesar 100%. Sehingga dapat diketahui bahwa rasio 90:10 memiliki nilai *accuracy*, *precision*, dan *recall* yang paling baik.
2. Dengan menggunakan klasifikasi algoritma *naïve bayes* memiliki hasil AUC sebesar 0,857 dari rasio 90:10, hasil AUC sebesar 0,660 dari rasio 80:20, dan hasil AUC sebesar 0,621 dari rasio 70:30. Sehingga dapat diketahui bahwa rasio 90:10 memiliki nilai AUC yang paling baik.

REFERENCES

- BPS. (2023). Jumlah Penduduk Menurut Kecamatan dan Jenis Kelamin (Jiwa), 2018-2020. Retrieved from Badan Pusat Statistik (BPS) Kabupaten Karawang website: <https://karawangkab.bps.go.id/indicator/12/35/1/jumlah-penduduk-menurut-kecamatan-dan-jenis-kelamin.html>
- Dahlia, R., Wuryani, N., Hadianti, S., Gata, W., & Selawati, A. (2021). Penerapan Data Mining Terhadap Data Covid-19 Menggunakan Algoritma Klasifikasi. *Jurnal Informatika*, 21(1), 44–52. doi: <https://doi.org/10.30873/ji.v21i1.2868>
- Firdaus, A., Walid, M., & Anwari. (2022). Klasifikasi Kasus Covid-19

- Menggunakan Model Naive Bayes Classifier. *JATI: Jurnal Mahasiswa Teknik Informatika*, 6(2), 583–588. doi: <https://doi.org/10.36040/jati.v6i2.5333>
- Fitriani, M., Nama, G. F., & Mardiana. (2022). Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku Upt Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM. *JITET: Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1), 41–49. doi: 10.23960/jitet.v10i1.2263
- Hasna, S. K. (2021). “*Analisis Sentimen Data Ulasan menggunakan Algoritma Support Vector Machine (Studi Kasus: Aplikasi Iflix)*.” Skripsi. Yogyakarta: Universitas Islam Indonesia.
- Kompas. (2021). Pelanggan Melonjak di Tengah Pandemi, Netflix Raup Pendapatan Rp 350 Triliun. Retrieved October 10, 2022, from Kompas.com website: <https://money.kompas.com/read/2021/01/20/165327826/jumlah-pelanggan-melonjak-di-tengah-pandemi-netflix-raup-pendapatan-rp-350>
- Malia, Y. N., Zakiyah, & R, G. N. K. M. (2021). Analisis Minat Konsumen Berlangganan Netflix di Masa Pandemi Perspektif Ekonomi Islam. *UNISKA: Universitas Islam Kalimantan*, 1–6. Retrieved from <https://eprints.uniska-bjm.ac.id/7165/>
- Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *PETIR: Jurnal Pengkajian Dan Penerapan Teknik Informatika*, 12(2), 131–144. doi: 10.33322/petir.v12i2.471
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE: Journal of Business and Audit Information Systems*, 4(1), 13–21. doi: <http://dx.doi.org/10.30813/jbase.v4i1.2729>
- Susilawati, W., Restu, M. S., Alamanda, D. T., & Wahid, A. A. (2021). Uji Beda Layanan Netflix berdasarkan Jenis Paket Menggunakan Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2). *Jurnal Wacana Ekonomi*, 21(1), 25–34. doi: <http://dx.doi.org/10.52434/jwe.v21i1.1733>
- Wulandari, F., Jusia, P. A., & Jasmir. (2020). Klasifikasi Data Mining untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes pada Puskesmas Jambi Selatan. *Jurnal Ilmiah Mahasiswa Sistem Informasi*, 2(3), 214–227. Retrieved from <http://ejournal.stikom-db.ac.id/index.php/jimsi/article/view/895/661>