

PERAMALAN GENRE FILM TERPOPULER BERDASARKAN DATASET MYMOVIE MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

Asrul¹, Wina Witanti², Fajri Rakhmat Umbara³

^{1,2,3}Informatika, Sains dan Informatika, Universitas Jenderal Achmad Yani

Email: 1asrul18@if.unjani.ac.id

ABSTRACT

At this time the film industry is experiencing very rapid progress, this is because extraordinary technological developments have had a major influence on the film industry. Successful films tend to have a large audience. To find out why the audience likes a film, there are several variables that must be considered, one of which is the genre of the film. This research was conducted to predict what film genres the audience is most interested in. To predict the genre of this film using the autoregressive integrated moving average (arima) method. The autoregressive integrated moving average (arima) method or commonly known as the Box-Jenkins method is a method used to make precise and accurate short-term forecasts, compared to long-term forecasts which usually tend to be flat (flat/constant). From this research a prediction of the popularity or number of viewers of each film genre will be generated which can be used as a reference to find out what genre of film the audience is interested in. So that film production companies can adjust film releases according to their interests. audience, in order to gain greater profits.

Keywords: *film industry, autoregressive integrated moving average (arima), forecasting.*

ABSTRAK

Pada saat ini industri perfilman mengalami kemajuan yang sangat pesat, itu dikarenakan perkembangan teknologi yang sangat luar biasa, Sehingga memberikan pengaruh yang besar terhadap industri perfilman. Film yang berhasil, cenderung memiliki jumlah penonton yang banyak. Untuk mengetahui penyebab penonton menyukai film ada beberapa variabel yang harus diperhatikan salah satunya yaitu genre film. Penelitian ini dilakukan untuk meramal genre film apa yang paling diminati oleh penonton. Untuk melakukan peramalan genre film ini menggunakan metode autoregressive integrated moving average (arima). Metode autoregressive integrated moving average (arima) atau yang biasa disebut metode Box-jenkins merupakan metode yang digunakan untuk melakukan peramalan jangka pendek yang tepat dan akurat, dibandingkan untuk peramalan jangka Panjang yang biasanya cenderung flat (mendatar/konstan). Dari penelitian ini akan menghasilkan peramalan jumlah popularitas atau jumlah penonton dari masing-masing genre film yang dapat menjadi acuan untuk mengetahui genre film apa yang diminati oleh penonton. Sehingga perusahaan produksi film bisa menyesuaikan perilis film sesuai dengan minat penonton, agar mendapatkan keuntungan yang lebih besar.

Kata Kunci: *industry film, autoregressive integrated moving average (arima), peramalan.*

Riwayat Artikel :

Tanggal diterima : 11-11-2023

Tanggal revisi : 13-11-2023

Tanggal terbit : 14-11-2023

DOI :

<https://doi.org/10.31949/infotech.v9i2.7358>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 1013 By Author



1. PENDAHULUAN

1.1. Latar Belakang

Industri perfilman adalah salah satu jenis industri yang sedang naik daun dan setiap tahun semakin meningkat baik dari munculnya industri film baru maupun dari industri film. Film adalah sarana yang dipergunakan untuk menyebarkan suatu hiburan yang telah menjadi kebiasaan terdahulu, dan menyajikan berbagai macam genre yaitu seperti horror, adventure, music, drama, comedy, dan kategori lainnya kepada masyarakat umum (Yosafat & Kurnia, 2019).

Film adalah sebuah rekaman gambar bergerak yang menceritakan sebuah cerita atau kejadian yang dapat ditonton pada bioskop atau televisi. Kualitas atau kesuksesan dari film dapat dinilai dengan menggunakan sistem angka yang disebut rating. Rating biasanya diasosiasikan dengan rentang angka 0 sampai 5 atau 0 sampai 10 (Prasetyo et al., 2022). Pada penelitian terdahulu yang berjudul "Rating prediction via generative convolutional neural networks based regression" yang mana pada penelitian ini menyebutkan rating adalah kriteria penting untuk mengevaluasi kualitas dari sebuah film dan merupakan indikator penting apakah penonton menyukai film tersebut (Ning et al., 2020). Ada beberapa variabel yang menjadi penilaian dari penonton dalam memberikan rating pada sebuah film yaitu informasi plot, genre dan pemeran. Untuk mengetahui pengaruh dari variabel tersebut khususnya genre film, peramalan (forecasting) dapat digunakan untuk menentukan strategi dalam pengambilan keputusan. Peramalan membentuk dasar perencanaan dan memungkinkan suatu perusahaan produksi film untuk dapat merespon lebih cepat dan akurat terhadap adanya perubahan pasar. Teknik peramalan yang saat ini berkembang adalah dengan menggunakan time series analysis (analisis deret waktu), Autoregressive integrated moving average (ARIMA) merupakan salah satu teknik analisis deret waktu yang banyak digunakan untuk peramalan data periode yang akan datang. Model ARIMA secara penuh mengabaikan independen variabel dalam membuat peramalan, ARIMA menggunakan nilai masa lalu dan masa sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat (Agribisnis et al., 2019).

Sebuah perusahaan produksi film bisa melakukan penyesuaian dalam merilis filmnya, jika mengetahui lebih dulu kemungkinan keberhasilan terkait film tersebut. Tujuannya untuk mendapatkan keuntungan maksimal setelah film dirilis (Winanda & Zega, 2019).

Berdasarkan uraian tersebut, penelitian akan berfokus kepada peramalan genre film yang banyak diminati oleh penonton dengan menggunakan metode *autoregressive integrated moving average (arima)*, untuk mengetahui jumlah popularitas dari masing-masing genre film tersebut.

1.2. Tinjauan Pustaka

a. Data Mining

Data mining adalah proses menemukan pola yang menarik dengan jumlah data yang besar, data tersebut dapat disimpan di database, data warehouse ataupun penyimpanan sejenis lainnya (Meilani et al., 2014).

Data mining juga dikenal sebagai knowledge discovery in database (KDD), yang berfungsi untuk menjelaskan proses pencarian informasi pada proses data mining (Arianto, 2019).

Tahap-tahapan di dalam data mining adalah sebagai berikut: (1)Data cleaning, (1)Data integration, (3)Data selection, (4)Data transformation, (5)Data mining, (6)Pattern evaluation, (7)Knowledge representation.

b. Peramalan (Forecasting)

Peramalan yaitu kegiatan yang dilakukan untuk memprediksi kejadian di masa mendatang dengan menggunakan data historis dan memproyeksikannya ke masa yang akan datang dengan model pendekatan sistematis. Dalam melakukan proses peramalan (forecasting), apapun bentuk dan jenis peramalan (forecasting) yang akan dilakukan, terdapat lima langkah proses peramalan (forecasting) yang bisa dilakukan, yaitu : (1)Formulasi masalah dan pengumpulan data, (1)Manipulasi dan pembersihan data, (3)Pembentukan dan evaluasi model, (4)Implementasi model (peramalan aktual/sebenarnya), (5)Evaluasi peramalan (forecasting). Hal yang perlu diperhatikan dalam melakukan peramalan (forecasting) yaitu pada error (galat). Untuk memperoleh hasil yang mendekati data asli, maka harus berusaha membuat error nya dengan nilai sekecil mungkin (Matematika & Ruspriyanty, 2018).

Penelitian yang menjadi acuan penulisan adalah penelitian yang dilakukan oleh Widhy Wahyani dan Achmad Syaichu (2015) yang berjudul "Penerapan metode peramalan sebagai alat bantu untuk menentukan perencanaan produksi di pt. Skk". Pada penelitian tersebut membahas tentang perbandingan hasil peramalan dan ukuran akurasi metode peramalan exponential smoothing dan regresi linier, yang menghasilkan nilai dari Metode Regresi Linier lebih kecil dibandingkan dengan metode Exponential Smoothing yang memiliki nilai ukuran akurasi kesalahan yang lebih kecil baik secara manual maupun dengan menggunakan software QM for Windows (Perencanaan et al., n.d.).

c. Film

Film merupakan media komunikasi yang mempunyai sifat audio visual untuk menyampaikan suatu pesan kepada orang-orang yang sedang menyaksikan suatu film. Film juga dianggap sebagai media komunikasi massa, karena sifatnya yang audio visual, film mampu bercerita banyak dalam waktu yang singkat. Ketika menonton film, penonton seakan-akan dapat menembus ruang dan waktu yang dapat menceritakan kehidupan dan

bahkan dapat mempengaruhi khalayak ramai (Asri, 2020).

Ada berbagai macam situs web yang menyediakan film-film, salah satunya yaitu situs web yang menyediakan informasi mengenai film dari seluruh dunia adalah Internet Movie Database (IMDb), situs ini selain menampilkan rating film juga memberikan informasi orang-orang yang terlibat di dalamnya mulai dari aktor, sutradara, penulis sampai penata rias dan soundtrack (Mayangky et al., 2019).

d. Autoregressive Integrated Moving Average (ARIMA)

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik, biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang. Model Autoregresif Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent) (As'ad et al., 2017).

Kelompok model yang termasuk dalam metode Autoregressive Integrated Moving Average (ARIMA) yaitu (Salwa et al., 2018):

a) Autoregressive (AR)

Model Autoregressive (AR) diperkenalkan pertama kali oleh Yule pada tahun 1916 dan kemudian dikembangkan oleh Walker pada tahun 1931. Asumsi yang dimiliki oleh model ini adalah data periode sekarang dipengaruhi oleh data pada periode sebelumnya. Disebut model Autoregressive dikarenakan pada model ini diregresikan terhadap nilai-nilai sebelumnya dari variabel itu sendiri. Model Autoregressive dengan ordo p disingkat menjadi AR(p) atau ARIMA (p,0,0) sebagai berikut:

$$Z_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - a_t \quad (1)$$

Dimana Z_t adalah deret waktu stasioner, μ' adalah konstanta, Z_{t-p} adalah variabel bebas, ϕ_p adalah koefisien parameter autoregressive ke-p, dan a_t adalah sisaan pada saat ke-t.

b) Moving Average (MA)

Model Moving Average (MA) pertama kali diperkenalkan oleh Slutsky pada tahun 1973, dengan orde q ditulis MA (q) atau ARIMA (0,0,q) dan dikembangkan oleh Wadsworth pada tahun 1989. Model MA (q) dapat dituliskan sebagai berikut:

$$Z_t = \mu' + a_t - \phi_1 a_{t-1} - \dots - \phi_q a_{t-q} \quad (2)$$

Dimana Z_t adalah deret waktu stasioner, μ' adalah konstanta, a_{t-1} adalah variabel bebas, ϕ_q adalah

koefisien parameter moving average ke-q, dan a_t adalah sisaan pada saat ke-t.

c) Autoregressive Moving Average (ARMA)

Model Autoregressive Moving Average (ARMA) merupakan model gabungan dari Autoregressive (AR) dan Moving Average (MA). Dan model ini memiliki asumsi bahwa data periode sekarang dipengaruhi oleh data periode sebelumnya dan nilai sisaan dari periode sebelumnya. Model ARMA dapat dituliskan sebagai berikut:

$$Z_t = \mu' + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \phi_1 - a_{t-1} - \dots - \phi_q a_{t-q} \quad (3)$$

Dimana Z_t adalah deret waktu stasioner, μ' adalah konstanta, Z_{t-p} adalah variabel bebas, dan ϕ_p adalah koefisien parameter autoregressive ke-p.

d) Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA merupakan model yang dituliskan dengan notasi (p,d,q), dimana (d) menyatakan Autoregressive (AR), lalu (d) menyatakan proses differencing dan (q) menyatakan Moving Average (MA). Berikut merupakan bentuk umum dari model ARIMA:

$$\phi_p(B)(1-B)^d Z_t = \phi_0 + \phi_q(B)a_t \quad (4)$$

Dimana Z_t adalah data observasi pada waktu ke-t, ϕ_p adalah koefisien parameter model Autoregressive ke-p, ϕ_0 adalah nilai konstanta, ϕ_q adalah koefisien parameter model Moving Average ke-q, a_t adalah nilai residual (error) pada waktu ke-t, dan d adalah banyaknya differencing yang dilakukan untuk menstasionerkan data terhadap mean.

e. Time Series

Metode deret waktu berkala (time series) merupakan metode yang cocok digunakan untuk meramalkan suatu permintaan pasar terhadap barang karena dapat menghitung dalam jenis data nominal. Metode ini termasuk kedalam metode peramalan jenis kuantitatif. Analisis deret waktu pada dasarnya digunakan untuk melakukan analisis data yang mempertimbangkan pengaruh waktu. Data dikumpulkan secara periodik berdasarkan urutan waktu, bisa dalam jam, hari, minggu, bulan, kuartal dan tahun. Analisis deret waktu dapat dilakukan untuk membantu dalam menyusun perencanaan ke depan (Maulana, 2018).

Dalam melakukan perhitungan terhadap metode kuantitatif dibutuhkan langkah-langkah peramalan, yaitu (Doktor et al., n.d.) : (1)Definisikan tujuan peramalan, (1)Pembuatan diagram pencar (scatter diagram), (3)Pemilihan minimal dua metode peramalan yang dianggap sesuai, (4)Perhitungan terhadap parameter-parameter fungsi peramalan, (5)Perhitungan kesalahan setiap metode peramalan, (6)Pemilihan metode yang terbaik dengan cara melihat hasil perhitungan metode dengan kesalahan terkecil, dan (7)Melakukan verifikasi peramalan.

f. Identifikasi Model

Identifikasi model digunakan untuk mengetahui model apa yang terbentuk, selain itu juga berfungsi untuk menentukan model ARIMA (p,d,q) dan menentukan data yang digunakan stasioner atau tidak.

a) Kestasioneran Data

Kestasioneran data time series merupakan suatu syarat yang harus diperhatikan dalam pembentukan model ARIMA. Stasioner data time series adalah suatu keadaan dimana proses pembangkitan yang mendasari suatu deret berkala didasarkan pada nilai rata-rata (mean) dan nilai varians yang saling konstan. Secara umum, uji untuk mengetahui kestasioneran data time series dapat dikategorikan menjadi tiga, yakni melalui grafik, korelogram, dan uji unit root.

b) Differencing Data

Data runtun waktu yang tidak stasioner dapat distasionerkan dengan melakukan differencing derajat d, sedangkan untuk mendapatkan kestasioneran dibuat deret baru yang terdiri dari differencing antara periode yang beruntun dengan menggunakan persamaan (5).

$$\nabla Z_t = Z_t - Z_{t-1} \tag{5}$$

Deret baru akan mempunyai n-1 buah nilai. Apabila differencing pertama tidak menunjukkan stasioner tercapai maka dapat dilakukan differencing kedua dengan menggunakan persamaan (6).

$$\nabla^1 Z_t = \nabla Z_t - \nabla Z_{t-1} = Z_t - 1Z_{t-1} + Z_{t-1} \tag{6}$$

dinyatakan sebagai deret differencing orde kedua. Deret ini akan mempunyai n-1 buah nilai.

c) ACF dan PACF

Autocorrelation Function (ACF) merupakan suatu hubungan linier antara pengamatan Z_t dengan pengamatan Z_{t+k} pada data yang dipisahkan oleh waktu ke- k. Berikut merupakan fungsi autokorelasi untuk sampel data yang diambil.

$$\rho_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \tag{7}$$

Dimana Z_t adalah nilai actual pada waktu ke-t dan ρ_k adalah nilai estimasi fungsi autokorelasi lag ke-k.

Partial Autocorrelation Function (PACF) digunakan untuk mengukur tingkat keeratan hubungan antara pasangan data Z_t dan Z_{t+k} setelah pengaruh variabel $Z_{t+1}, Z_{t+1} \dots Z_{t+k-1}$ dihilangkan. Perhitungan nilai PACF sampel lag ke-k dimulai dengan menghitung $\phi_{1,1} = \rho_k$, sedangkan untuk menghitung $\phi_{k,k}$ dilakukan dengan menggunakan persamaan 8.

$$\phi_{k+1,k+1} = \frac{\rho_{k+1} - \sum_{j=1}^k \phi_{kj} \rho_{k+1-j}}{1 - \sum_{j=1}^k \phi_{kj} \rho_j} \tag{8}$$

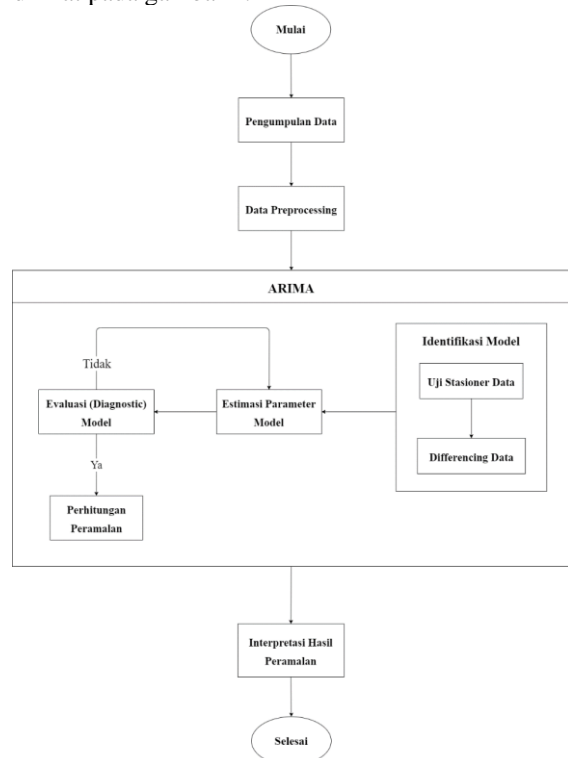
g. Supervised Learning

Supervised learning adalah sebuah pemodelan algoritma yang dapat memetakan input ke output yang diinginkan. Proses supervised learning memerlukan data training yang digunakan dalam melakukan prediksi ataupun mengklasifikasikan data (Somvanshi, 2016).

Dalam algoritma supervised learning, terdapat dua variabel, yaitu variabel input atau yang biasa disebut variabel X dan variabel output yang biasa disebut variabel Y. Tujuan algoritma supervised learning adalah untuk mempelajari fungsi pemetaan dari variabel X ke variabel Y. Rumus umum pemetaan variabel X dan Y adalah $Y = f(X)$. Tujuan akhir dari algoritma supervised learning adalah untuk memperkirakan fungsi pemetaan (f) agar dapat memprediksi variabel Y ketika memiliki data input (variabel X) yang baru.

1.3. Metodologi Penelitian

Berikut adalah alur metode penelitian yang dilakukan untuk melakukan peramalan genre film terpopuler berdasarkan dataset mymovie dengan menggunakan metode Autoregressive Integrated Moving Average (ARIMA) (Hartati, 2017), dapat dilihat pada gambar 1.



Gambar 1. Alur Metode Penelitian

a. Pengumpulan Data

Tahapan pertama yang dilakukan yaitu pengumpulan data dan informasi mengenai database movie, dimana data yang didapatkan berasal dari dataset publik situs Kaggle. Data tersebut adalah data movie TMDB yang memiliki 9 atribut yaitu tanggal rilis, judul film, sinopsis, popularitas, total voting, rata-rata voting, Bahasa film, kategori film dan poster url, dengan jumlah record 7298 data.

b. Data Preprocessing

Preprocessing data adalah membuang data yang lebih dari satu, menghapus data yang tidak diperlukan atau tidak konsisten, dan memperbaiki kesalahan pada data atau salah dalam penulisannya, karena tidak semua data atau atribut data yang ada didalam data digunakan dalam proses data mining. Proses ini dilakukan agar data yang akan digunakan sesuai dengan kebutuhan. Pada tahap preprocessing ini melakukan beberapa tahap yaitu Data Cleaning dan Data Transformation. Tahap Data cleaning penulis akan melakukan pemilihan data yang tidak lengkap, data noisy, data yang tidak konsisten, dan data yang duplikat. Selanjutnya melakukan Data transformation yaitu mengubah bentuk data yang sudah ada menjadi data yang lebih efektif untuk metode penelitian.

a) Data Cleaning

Tahapan berikutnya yaitu data cleaning, pada tahapan ini dilakukan proses menyeleksi data kosong dan atribut yang tidak akan digunakan pada penelitian (Sulastrri & Gufroni, 2017). Dari data yang didapatkan, hanya digunakan 6 atribut yaitu tanggal rilis, judul film, popularitas, total voting, rata-rata voting dan kategori film, dengan jumlah record 7023 data.

b) Data Transformation

Tahapan berikutnya yaitu data transformation, pada tahapan ini dilakukan proses mengubah bentuk data yang sudah ada menjadi data yang lebih efektif untuk metode penelitian (Rerung, 2018). Pada penelitian ini, atribut yang dilakukan transformation yaitu kategori film, dimana jumlah popularitas dari masing-masing kategori film diurutkan menjadi dalam bentuk per-bulan dan per-tahun.

c. Identifikasi Model

Tahapan berikutnya yaitu indentifikasi model, pada tahapan ini dilakukan proses pemeriksaan tren dan pola data. Setelah pemeriksaan pola data, selanjutnya dilakukan uji stasioner data, jika data belum stasioner maka akan dilakukan proses differencing yang juga menentukan nilai (d) pada model, jika data tidak mengalami proses differencing maka nilai (d) adalah 0, tetapi jika data stasioner setelah differencing ke-1 maka nilai (d) adalah 1, dan seterusnya. Untuk menentukan nilai ordo (p,q) dapat dilakukan dengan mengamati plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF) (Nurfadila & Ilham Aksan, 2020).

d. Estimasi Parameter Model

Tahapan berikutnya yaitu estimasi parameter model, pada tahapan ini dilakukan untuk menentukan apakah estimasi model ARIMA (p,d,q) sementara yang dipilih signifikan (John & Latupeirissa, 2021). Untuk mengetahui apakah model tersebut signifikan adalah dengan melihat nilai signifikansi kurang dari alpha (α), dengan nilai α adalah 0,05.

e. Evaluasi (Diagnostic) Model

Tahapan berikutnya yaitu uji diagnostic model, pada tahapan ini dilakukan untuk mengetahui apakah model yang dipilih sudah cukup baik digunakan dalam peramalan. Adapaun evaluasi yang dilakukan yaitu (Simbolon, 2022):

a) Uji White Noise

Uji white noise pada residual menggunakan uji Ljung-Box dengan hipotesis yang diuji adalah residual sudah white noise. Residual sudah white noise jika nilai p-value lebih besar dari alpha (α) dengan nilai (α) adalah 0,05.

b) Uji Normalitas Residual

Uji normalitas digunakan uji statistik Shapiro-Wilks. Hipotesis yang diuji adalah residual atau error berdistribusi normal. Residual dikatakan berdistribusi normal jika p-value lebih besar dari (α) dengan (α) adalah 0,05.

f. Perhitungan Peramalan

Tahapan berikutnya yaitu perhitungan peramalan, setelah tahapan uji diagnostic dilakukan dan model ARIMA terbaik didapatkan, maka akan dilakukan perhitungan peramalan untuk mengetahui jumlah popularitas dari masing-masing kategori film.

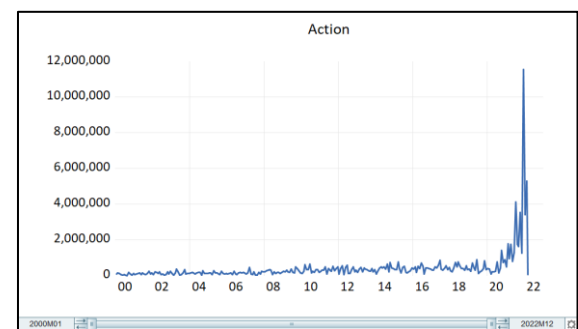
g. Interpretasi Hasil Peramalan

Tahapan berikutnya yaitu interpretasi hasil peramalan, tahapan ini dilakukan untuk mengetahui hasil akhir dari peramalan genre film. Dimana pada tahap ini akan menghasilkan suatu pengetahuan (knowledge), hasil dari interpretasi ini akan menjadi kesimpulan dari penelitian yang telah dilakukan.

2. PEMBAHASAN

2.1. Uji Stationer Data

Pada tahapan ini dilakukan proses uji stasioner data, dimana data yang sudah ditransformasikan dimasukan ke dalam tools Eviews. Pada percobaan analisis peramalan ini data yang digunakan yaitu kategori Action. Setelah data dimasukan, maka untuk mengetahui data tersebut stasioner atau tidak adalah dengan melihat grafik dari data tersebut, terlihat bahwa adanya indikasi data tidak stasioner pada rata-rata. Hal itu terlihat pada grafiknya yang meningkat drastis pada tahun 2021. Untuk mengatasi masalah data tidak stasioner, maka dilakukan proses differencing.



Gambar 2. Uji Stasioner Data

2.2. Differencing Data

Pada tahapan ini dilakukan untuk menentukan model sementara (p,d,q), dimana dilakukan proses differencing untuk mengetahui nilai (d). Proses yang dilakukan untuk mengetahui nilai (d) yaitu dengan melakukan uji ADF (Augmented Dickey-Fuller).

Augmented Dickey-Fuller Unit Root Test on D(ACTION,2)		
Null Hypothesis: D(ACTION,2) has a unit root		
Exogenous: Constant		
Lag Length: 10 (Automatic - based on SIC, maxlag=15)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.102855	0.0000
Test critical values:		
1% level	-3.455990	
5% level	-2.872720	
10% level	-2.572802	

Gambar 3. Uji Augmented Dickey-Fuller

Berdasarkan output uji ADF 2nd difference p-value = 0.0000 < alpha = 0.05 yang artinya data tidak mempunyai unit root (data stasioner). Karena data sudah stasioner pada difference 2nd, maka nilai d = 2. Selanjutnya menentukan nilai p dan q dengan cara melihat plot ACF (Autocorrelation Function) dan PACF (Partial Autocorrelation Function). Untuk melihat plot ACF dan PACF yaitu dengan menggunakan correlogram. Karena data stasioner pada differencing 2 kali, maka akan menggunakan correlogram 2nd difference.

Correlogram of D(ACTION,2)							
Date: 01/07/23 Time: 01:33							
Sample (adjusted): 2000M03 2022M03							
Included observations: 265 after adjustments							
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob		
		1	-0.792	-0.792	168.01	0.000	
		2	0.457	-0.454	224.31	0.000	
		3	-0.134	0.111	229.19	0.000	
		4	-0.106	-0.057	232.22	0.000	
		5	0.191	-0.130	242.16	0.000	
		6	-0.132	0.109	246.92	0.000	
		7	0.014	0.000	246.97	0.000	
		8	0.070	-0.092	248.31	0.000	
		9	-0.104	-0.026	251.31	0.000	
		10	0.113	0.103	254.84	0.000	
		11	-0.110	-0.068	258.23	0.000	
		12	0.095	-0.030	260.76	0.000	
		13	-0.078	0.017	262.45	0.000	
		14	0.056	-0.005	263.33	0.000	
		15	-0.019	0.006	263.43	0.000	
		16	-0.021	-0.003	263.56	0.000	
		17	0.039	-0.021	263.99	0.000	
		18	-0.024	0.030	264.16	0.000	
		19	-0.005	0.010	264.17	0.000	
		20	0.024	-0.043	264.34	0.000	
		21	-0.024	0.023	264.50	0.000	
		22	0.007	0.011	264.52	0.000	
		23	0.013	-0.022	264.57	0.000	
		24	-0.039	-0.059	265.02	0.000	
		25	0.053	0.018	265.85	0.000	
		26	-0.038	0.044	266.27	0.000	
		27	0.009	-0.025	266.29	0.000	
		28	0.025	0.024	266.48	0.000	
		29	-0.054	0.003	267.36	0.000	
		30	0.060	-0.026	268.46	0.000	

Gambar 4. Uji Correlogram

Berdasarkan output dari Gambar 3.6 yaitu correlogram 2nd difference, pada plot ACF terlihat bahwa pada lag 1 dan 2 signifikan dan cutoff pada lag 3 dan seterusnya. Sedangkan pada plot PACF juga terlihat bahwa pada lag 1 dan 2 signifikan dan cutoff pada lag 3 dan seterusnya. Maka dapat diduga sementara bahwa nilai p = 2 dan nilai q = 2. Berdasarkan hasil dari estimasi model, terdapat 8

kombinasi model yang dapat digunakan yaitu (0,2,1), (0,2,2), (1,2,0), (1,2,1), (1,2,2), (2,2,0), (2,2,1) dan (2,2,2).

2.3. Estimasi Parameter Model

Karena sudah terdapat 8 kandidat model yang akan digunakan, maka selanjutnya yaitu melakukan estimasi parameter model. Berdasarkan hasil estimasi masing-masing parameter model ARIMA, maka dapat ditentukan model terbaik. Pemilihan model terbaik didasarkan pada goodness of fit dari masing-masing model yaitu nilai R-squared yang lebih besar serta nilai AIC (Akaike Info Criterion) dan SC (Schwarz Criterion) yang terkecil (Teapon, 2015). Adapun rangkuman nilai masing-masing model ARIMA dapat dilihat pada gambar 5.

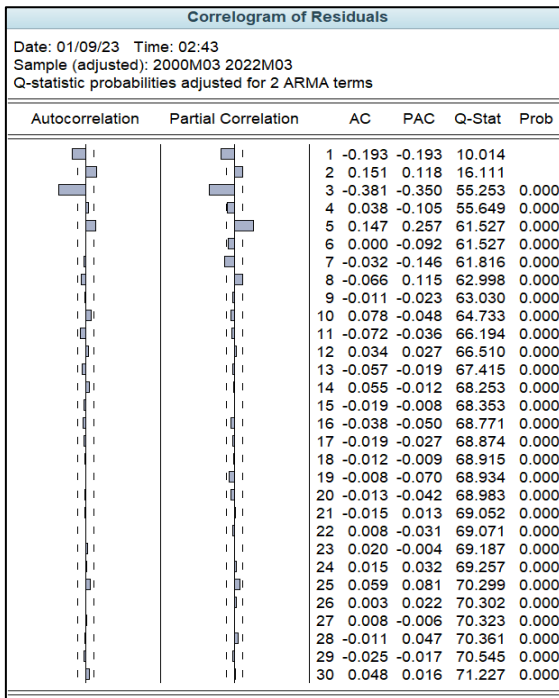
Model Arima	Parameter	Prob.	Keterangan	R-squared	SC	AIC
Arima (0,2,1)	MA(1)	0.9838	Tidak Signifikan	0.681147	30.45971	30.41918
Arima (0,2,2)	MA(1)	0.0000	Signifikan	0.839552	29.80894	29.75490
	MA(2)	0.0000	Signifikan			
Arima (1,2,0)	AR(1)	0.0000	Signifikan	0.672419	30.47037	30.42985
Arima (1,2,1)	AR(1)	0.0000	Signifikan	0.822285	29.90279	29.84876
	MA(1)	0.9824	Tidak Signifikan			
Arima (1,2,2)	AR(1)	0.0060	Signifikan	0.835867	29.85302	29.78547
	MA(1)	0.3405	Tidak Signifikan			
	MA(2)	0.5863	Tidak Signifikan			
Arima (2,2,0)	AR(1)	0.0000	Signifikan	0.737118	30.27275	30.21872
	AR(2)	0.0000	Signifikan			
Arima (2,2,1)	AR(1)	0.0000	Signifikan	0.822720	29.92175	29.85421
	AR(2)	0.0696	Tidak Signifikan			
	MA(1)	0.9832	Tidak Signifikan			
Arima (2,2,2)	AR(1)	0.0000	Signifikan	0.822873	29.94252	29.86147
	AR(2)	0.0129	Signifikan			
	MA(1)	0.9891	Tidak Signifikan			
	MA(2)	0.9911	Tidak Signifikan			

Gambar 5. Estimasi Parameter Model

Berdasarkan hasil nilai parameter model ARIMA pada gambar 5, menunjukkan bahwa model ARIMA (0,2,2) adalah model yang terbaik, karena memiliki nilai R-squared yang besar serta nilai AIC (Akaike Info Criterion) dan SC (Schwarz Criterion) yang terkecil dibandingkan dengan model ARIMA yang lainnya.

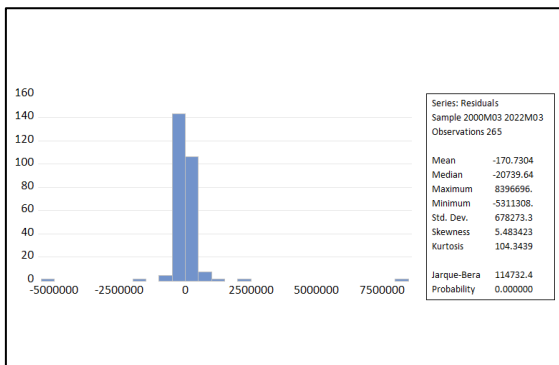
2.4. Evaluasi (Diagnostic) Model

Setelah model terbaik didapatkan yaitu model ARIMA (0,2,2), selanjutnya pemeriksaan diagnostic model yang akan dibagi menjadi dua bagian yaitu uji signifikansi parameter dan uji asumsi residual. Uji signifikansi parameter model ARIMA (0,2,2) sudah dilakukan pada gambar 5 yang hasilnya parameter model ARIMA (0,2,2) sudah signifikan. Untuk melakukan uji asumsi residual dilakukan dengan menggunakan Correlogram Of Residuals dan Histogram Normality. Adapun hasil dari correlogram of residuals dapat dilihat pada gambar 6.



Gambar 6. Correlogram of Residuals

Berdasarkan hasil dari correlogram of residuals terlihat bahwa dari lag 1 sampai 30 tidak ada lag yang signifikan. Hal tersebut menandakan bahwa residual sudah white noise dan model ARIMA (0,2,2) dapat dikatakan cocok.



Gambar 7. Histogram Normality

Berdasarkan hasil dari histogram normality terlihat bahwa residual tidak normal yang ditunjukkan dari nilai p-value < dari 0.05.

2.5. Perhitungan Peramalan

Setelah menemukan model yang sesuai yaitu ARIMA (0,2,2), selanjutnya melakukan perhitungan peramalan dengan menggunakan persamaan 4. Adapun hasil peramalan dari kategori Action dapat dilihat pada Tabel 1.

Tabel 1. Perhitungan Kategori Action

Periode	Waktu	Peramalan
1	2022-04	1.144.075
2	2022-05	2.179.638
3	2022-06	2.767.638

4	2022-07	1.108.309
5	2022-08	838.703

Untuk kategori film yang lainnya juga melakukan tahapan yang sama, yaitu melakukan uji stasioner data, differencing data, estimasi parameter model, evaluasi (diagnostic) model dan perhitungan peramalan dengan menggunakan model yang sesuai.

3. KESIMPULAN

Dalam penelitian Peramalan Genre Film Terpopuler Berdasarkan Dataset Mymovie menggunakan Metode Autoregressive Integrated Moving Average yang telah dilakukan, dapat mengetahui jumlah popularitas (penonton) dari kategori film. Terdapat 19 kategori film yang akan dilakukan peramalan jumlah popularitas yaitu Action, Adventure, Animation, Comedy, Crime, Documentary, Drama, Family, Fantasy, History, Horror, Music, Mystery, Romance, Science Fiction, Thriller, TV Movie, War dan Western. Tahapan yang dilakukan pada proses peramalan yaitu melakukan uji stasioner data terhadap kategori film, jika data tersebut tidak stasioner maka lakukan proses differencing agar data menjadi stasioner. Proses differencing juga digunakan untuk mendapatkan nilai (d) pada model ARIMA (p,d,q), sedangkan nilai (p dan q) didapatkan dari mengamati plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF). Setelah model (p,d,q) sementara didapatkan, Lalu model ARIMA yang didapatkan akan dilakukan uji asumsi residuals dengan menggunakan Correlogram Of Residuals dan Histogram Normality. Jika pada uji Correlogram Of Residuals data sudah white noise dan pada uji Histogram Normality, Maka model tersebut cocok digunakan untuk perhitungan peramalan. Periode peramalan yang dilakukan pada penelitian ini hanya 5 periode yang akan datang, hal ini dikarenakan metode ARIMA lebih cocok digunakan untuk peramalan jangka pendek dibandingkan peramalan jangka panjang yang hasilnya relatif datar/konstan.

PUSTAKA

Agribisnis, J., Utara, S., Statistisi, F., Pusat, B., Utara, P. S., Kajian, A., Utara, S., Autoregressive, M., Moving, I., Utara, P. S., Utara, S., & Utara, S. (2019). Penerapan Autoregressive Integrated Moving Average (ARIMA) Pada Peramalan Produksi Kedelai di Sumatera Utara. 9(2).

Arianto, J. (2019). Penerapan Data Mining Untuk Pengelompokan Penduduk Kurang Mampu Desa Sambirejo Timur Dengan Algoritma K-Medoids (Studi Kasus Kantor Kepala Desa Sambirejo Timur). KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3(1), 569–573. <https://doi.org/10.30865/komik.v3i1.1660>

- As'ad, M., Wibowo, S. S., & Sophia, E. (2017). Peramalan Jumlah Mahasiswa Baru Dengan Model Autoregressive Integrated Moving Average (Arima). *JIMP - Jurnal Informatika Merdeka Pasuruan*, 2(3), 20–33. <https://doi.org/10.37438/jimp.v2i3.77>
- Asri, R. (2020). Membaca Film Sebagai Sebuah Teks: Analisis Isi Film “Nanti Kita Cerita Tentang Hari Ini (NKCTHI).” *Jurnal Al Azhar Indonesia Seri Ilmu Sosial*, 1(2), 74. <https://doi.org/10.36722/jaiss.v1i2.462>
- Doktor, P., Pertanian, I., Pertanian, F., Padjadjaran, U., & Jatinangor, K. U. (n.d.). (*the Importance of Time Series Data in Production Planning*). 2011, 582–589.
- Hartati, H. (2017). Penggunaan Metode Arima Dalam Meramal Pergerakan Inflasi. *Jurnal Matematika Sains Dan Teknologi*, 18(1), 1–10. <https://doi.org/10.33830/jmst.v18i1.163.2017>
- John, D. R., & Latupeirissa, S. J. (2021). Peramalan Harga Emas Di Indonesia Tahun 2014-2019 Dengan Metode Arima Box-Jenkins. *VARIANCE: Journal of Statistics and Its Applications*, 2(2), 53–62. <https://doi.org/10.30598/variancevol2iss2page53-62>
- Matematika, J. I., & Rusprianty, D. I. (2018). *MATH unesa*. 6(2), 75–80.
- Maulana, H. A. (2018). Pemodelan Deret Waktu Dan Peramalan Curah Hujan Pada Dua Belas Stasiun Di Bogor. *Jurnal Matematika Statistika Dan Komputasi*, 15(1), 50. <https://doi.org/10.20956/jmsk.v15i1.4424>
- Mayangky, N. A., Kholifah, D. N., Balla, I., & Thira, I. J. (2019). Pengaruh Rating Film Terhadap Jumlah Audience Yang Menonton Film. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), 113–120. <https://doi.org/10.31294/ijse.v5i2.6963>
- Meilani, B. D., Susanti, N., Informatika, J. T., Informasi, F. T., Teknologi, I., & Tama, A. (2014). Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode NAÏVE BAYES. *Jurnal Sistem Informasi Universitas Suryadarma*, 3(2), 182–189. <https://doi.org/10.35968/jsi.v3i2.66>
- Ning, X., Yac, L., Wang, X., Benatallah, B., Dong, M., & Zhang, S. (2020). Rating prediction via generative convolutional neural networks based regression. *Pattern Recognition Letters*, 132, 12–20. <https://doi.org/10.1016/j.patrec.2018.07.028>
- Nurfadila, K., & Ilham Aksan. (2020). Aplikasi Metode Arima Box-Jenkins Untuk Meramalkan Penggunaan Harian Data Seluler. *Journal of Mathematics: Theory and Application*, 2(1), 5–10. <https://doi.org/10.31605/jmta.v2i1.749>
- Perencanaan, M., Di, P., & Skk, P. T. (n.d.). *Kata kunci: perencanaan produksi, peramalan*,. 133–141.
- Prasetyo, V. R., Mercifia, M., Averina, A., Sunyoto, L., & Budiarto. (2022). Film Rating Prediction on Imdb Website Using Neural Network. *Jurnal Ilmiah NERO*, 7(1), 60293.
- Rerung, R. R. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*, 3(1), 89. <https://doi.org/10.31544/jtera.v3.i1.2018.89-98>
- Salwa, N., Tatsara, N., Amalia, R., & Zohra, A. F. (2018). Peramalan Harga Bitcoin Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average). *Journal of Data Analysis*, 1(1), 21–31. <https://doi.org/10.24815/jda.v1i1.11874>
- Simbolon, L. D. (2022). Penerapan Model Arima Dalam Memprediksi Harga Emas. *MES: Journal of Mathematics Education and Science*, 7(2), 30–41. <https://doi.org/10.30743/mes.v7i2.5139>
- Somvanshi, M. (2016). *Iccubea.2016.7860040*.
- Sulastri, H., & Gufroni, A. I. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 3(2), 299–305. <https://doi.org/10.25077/teknosi.v3i2.2017.299-305>
- Teapon, R. R. H. (2015). Pembentukan Model Arima Untuk Peramalan Inflasi Kelompok Bahan Makanan Di Kota Ternate. *Ekonomi*, 04(01), 13–30.
- Winanda, O. I., & Zega, S. A. (2019). *Prediksi Rating Film Animasi Berdasarkan Elemen Mise En Scene Menggunakan Neural Network*. <https://jurnal.polibatam.ac.id/index.php/JAMN/article/download/1427/862>
- Yosafat, A. R., & Kurnia, Y. (2019). Aplikasi Prediksi Rating Film dengan Perbandingan Metode Naïve Bayes dan KNN Berbasis Website Menggunakan Framework Codeigniter. *Jurnal ALGOR*, 1(1), 16–26.