

## **Penerapan Data Mining Untuk Memprediksi Jumlah Data Pasien Di Puskesmas Haekesak Menggunakan Metode ARIMA**

### **Implementation Of Data Mining To Predict The Number Of Patient Data In Haekesak Puskesmas Using ARIMA Method**

**Ermelinda Novita De Jesus<sup>1</sup>, Anief Fauzan Rozi<sup>2</sup>**

<sup>12</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta  
e-mail: <sup>1</sup> novitacaeiro97@gmail.com, <sup>2</sup> anief@mercubuana-yogya.ac.id

#### **ABSTRAK**

Puskesmas Haekesak adalah salah satu puskesmas yang berada di Kabupaten Belu Provinsi Nusa Tenggara Timur dan melayani 40.753 jiwa penduduk. Seiring dengan meningkatnya jumlah pasien di Puskesmas Haekesak maka data yang didapatkan dari rekam medis tersebut diharapkan dapat mengetahui penyebaran penyakit yang ada di Kabupaten Belu NTT. Tujuan penelitian ini adalah untuk mengetahui model persamaan metode Arima, mengetahui hasil analisis memprediksi jumlah pasien di puskesmas haekesak menggunakan tools Minitab dan mengetahui hasil memprediksi jumlah pasien yang dirawat di puskesmas haekesak periode 1 Januari hingga 10 Februari serta metode yang digunakan dalam penelitian adalah metode ARIMA. Hasil dari prediksi total pasien yang akan datang pada Puskesmas Haekesak dengan menggunakan metode ARIMA, analisa ini berakhir dengan jumlah pasien yang akan datang pada hari ke 26-hari ke 35 dimana total pasien tidak naik atau turun secara signifikan sehingga pihak puskesmas tidak perlu menambah tenaga medis ataupun penambahan stock obat yang berlebihan.

**Kata Kunci:** Forecasting, ARIMA, Penyakit

#### **ABSTRACT**

Puskesmas Haekesak is one of the puskesmas located in Belu Regency, East Nusa Tenggara Province and serves 40,753 people. Along with the increasing number of patients at the Haekesak Health Center, the data obtained from the medical records is expected to be able to determine the spread of diseases in Belu District, NTT. The purpose of this study was to determine the Arima method of equality model, to know the results of the analysis to predict the number of patients in the haekesak health center using Minitab tools and to know the results of predicting the number of patients treated at the Hasekesak health center from 1 January to 10 February and the method used in this study was the ARIMA method. The results of the prediction of the total number of patients who will come to the Haekesak Health Center using the ARIMA method, this analysis ends with the number of patients who will come on the 26th-35th day where the total number of patients does not increase or decrease significantly so that the puskesmas does not need to increase their mental health. or the addition of excessive drug stock.

**Keyword:** Forecasting, ARIMA, Disease

#### **1. PENDAHULUAN**

Puskesmas haekesak adalah salah satu Puskesmas di Kabupaten Belu Provinsi Nusa Tenggara Timur yang terletak di ibu kota Kabupaten Belu, melayani 40.753 jiwa penduduk, mayoritas bekerja sebagai wiraswata dan pegawai negeri sipil dengan berbagai latar belakang pendidikan. Dengan unit pelaksana teknis Dinas Kesehatan puskesmas haekesak yang bertanggung jawab menyelenggarakan

pembangunan kesehatan di 12 Kelurahan yang menjadi wilayah kerja dari Puskesmas Haekesak. Seiring dengan peningkatan yang signifikan jumlah pasien di Puskemas Haekesak, data yang didapatkan dari rekam medis dianggap dapat mewakili populasi data pasien di Kecamatan Raihat. Dari data tersebut, di harapkan pihak Puskemas dapat mengetahui penyebaran penyakit. Dengan demikian dalam memprediksi data penyakit dengan melihat jumlah penyakit

yang diderita pasien dapat membantu pihak puskesmas haekesak menentukan perkiraan jumlah obat atau vaksin yang harus disediakan untuk pengobatan penyakit pada masa yang akan datang dengan menggunakan data yang ada seperti jumlah pasien dan penyakit yang diderita maka perlu adanya peramalan atau prediksi untuk mengetahui jumlah penderita penyakit setiap bulannya agar semua kebutuhan pengobatan penderita dapat terpenuhi. Diperlukan suatu metode untuk melakukan pemetaan terhadap prediksi jumlah penyakit pada pasien berdasarkan penyebaran penyakit yang sering menjadi salah satu faktor penting yang mempengaruhi terjadinya penyakit serta perjalanan suatu penyakit.

## 2. TINJAUAN PUSTAKA

### 2.1 Data Mining

Data adalah catatan atas kumpulan fakta. Data merupakan bentuk jamak dari datum, berasal dari bahasa Latin yang berarti "sesuatu yang diberikan". (Fince,2017)

### 2.2 Peramalan (Forecasting)

Prediksi adalah suatu usaha untuk meramalkan keadaan di masa mendatang. melalui pengujian keadaan di masa lalu. Dalam kehidupan sosial segala sesuatu itu serba tidak pasti, sukar untuk diperkirakan secara tepat. (Linda, 2018)

### 2.3 Auto Regressive Integrated Moving Average (ARIMA)

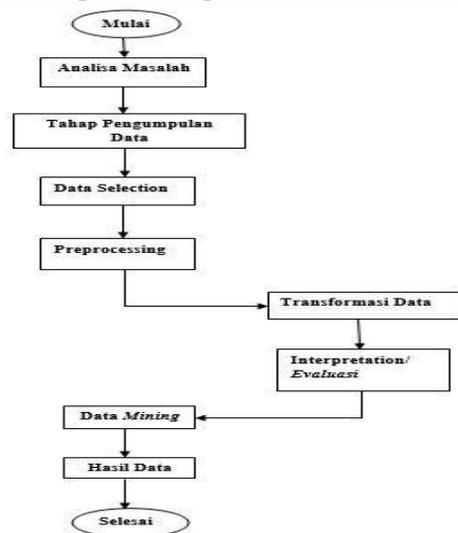
ARIMA merupakan salah satu metode peramalan yang telah dikenalkan oleh G.E.P. Box dan G.M. Jenkins. Adabeberapa model yang telah dihasilkan dengan menggunakan metodeBox-Jenkins yaitu model moving average (MA), autoregressive (AR), satu kelas model yang berguna untuk time series yang merupakan kombinasi proses MA dan AR yaitu ARMA. Model-model ini adalah model dari metode Box-Jenkins yang linier dan stasioner (stationary). Sedangkan model untuk data tidak stasioner yaitu model ARIMA. Klasifikasi model ARIMA terbagi kedalam empat kelompok, yaitu: model Autoregressive (AR), Moving Average (MA) dan model campuran Autoregressive Integrated Moving Average (ARIMA) yang memiliki karakteristik dari dua model pertama serta Seasonal Autoregressive Integrated Moving Average(SARIMA) yang merupakan turunan

dari ARIMA untuk mendapatkan prediksi data yang seasonal. (Syarfi Ahmad, & Mustaki 2017)

## 3. METODE PENELITIAN

Penelitian dilakukan dengan melalui beberapa tahapan dimulai dari analisa masalah, pengumpulan data, data selection, pre-processing data, transformasi data, data mining, interpretation, hingga menghasilkan suatu data analisis. Tahapan jalannya penelitian dapat dilihat pada gambar dibawah ini.

Secara garis besar tahap-tahap dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

## 4. HASIL DAN PEMBAHASAN

Pada subbab ini akan membahas penelitian yang dilakukan. Ada beberapa hal yang dihasilkan pada penelitian ini antara lain: pemrosesan raw data menjadi dataset, pencarian model terbaik agar dapat digunakan sebagai data forecasting, lalu melakukan *forecasting*.

### 4.1 Transformasi Data

Data Transformation adalah tahap mengubah data menjadi bentuk yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Dalam penelitian ini akan dilakukan tahapan transform mengubah data dari vertikal menjadi horizontal dan mengubah atribut tanggal menjadi harian dan penyakit dikelompokkan menjadi satu seperti pada gambar 2 dibawah ini.

Nama penyakit	TBC	Paru-paru	Sakit kepala	Cluster	Infeksi saluran kemih	Maag	Radang lambung	Asma	Stroke	Demam		
Hari1		1		0		1	2		3	2	0	4
Hari2		1		0		1	1		1	3	0	3
Hari3		0		0		0	4		1	0	1	3
Hari4		0		0		0	2		1	2	1	2
Hari5		1		0		1	3		1	0	0	2
Hari6		0		0		0	4		0	0	1	6
Hari7		1		2		1	1		0	0	0	5
Hari8		1		0		0	2		0	0	0	4
Hari9		0		0		0	3		2	1	0	1
Hari10		1		0		0	1		1	2	0	1
Hari11		0		0		0	0		4	0	0	3
Hari12		0		0		0	2		1	2	1	2
Hari13		1		0		1	2		1	0	0	1
Hari14		2		0		0	1		0	2	0	0
Hari15		0		0		0	2		0	0	0	4

Gambar 2. Hasil transformasi

#### 4.2 Data Reduction

Pada tahap ini akan dilakukan proses data reduction dilakukan untuk menyesuaikan dengan cara menyesuaikan data yang akan diproses, sehingga atribut harus disesuaikan. Dimana pada penelitian ini data yang akan digunakan sebagai bahan analisis ialah total data pasien pada Puskesmas Haekesak sehingga perlu dilakukan proses operasi sum untuk total pasien yang ada sehingga akan seperti table 1 dibawah ini

Tabel 1. Hasil reduction

Hari	Total Pasien
Hari1	29
Hari2	24
Hari3	32
Hari4	32
Hari5	27
Hari6	34
Hari7	29
Hari8	34
Hari9	24
Hari10	24
Hari11	36
Hari12	21
Hari13	20
Hari14	21
Hari15	17

#### 4.3 Proses Data Mining

Pada subbab ini akan membahas tentang proses data mining yang dilakukan terhadap data hasil-hasil yang didapat dari data pasien Puskesmas Haekesak.

##### Stationer dalam varian

Pada hasil uji coba dengan box chox maka didapatkan hasil jika rounded value bernilai -1,00 sehingga belum stationer maka akan dilakukan proses transformasi data agar varian dapat stationer. Maka langkah selanjutnya yaitu kita akan melakukan pengubahan  $\lambda = -1$  seperti pada tabel 2 dan gambar 3 di bawah ini.

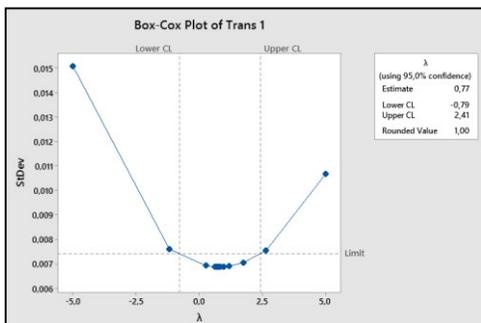
Tabel 2. Transform model  $\lambda = -1$  manual

Total Pasien	Hasil Transform data $\lambda=-1$
29	$\frac{1}{29} = 0,0344828$
24	$\frac{1}{24} = 0,0416667$
32	$\frac{1}{32} = 0,0312500$
32	$\frac{1}{32} = 0,0312500$
27	$\frac{1}{27} = 0,0370370$
34	$\frac{1}{34} = 0,0294118$
29	$\frac{1}{29} = 0,0344828$
34	$\frac{1}{34} = 0,0294118$
24	$\frac{1}{24} = 0,0416667$
24	$\frac{1}{24} = 0,0416667$

	Total Pasien	Trans 1
1	29	0,0344828
2	24	0,0416667
3	32	0,0312500
4	32	0,0312500
5	27	0,0370370
6	34	0,0294118
7	29	0,0344828
8	34	0,0294118
9	24	0,0416667
10	24	0,0416667

Gambar 3. Transform model  $\lambda = -1$  manual dengan minitab

Maka selanjutnya dilakukan lagi uji coba dengan box cox menggunakan nilai  $\lambda = -1$  dengan menggunakan *software* minitab seperti pada gambar 4 dibawah ini.



Gambar 4. Box cox transform model  $\lambda = 1$

### Stationer dalam mean

Tujuan dari uji autokorelasi yang dapat dilihat pada gambar dibawah ini adalah untuk melihat apakah terjadi korelasi antara suatu periode (t) dengan periode sebelumnya (t-1) yang digambarkan pada sumbu x yaitu Lag, sedangkan sumbu y merupakan nilai korelasi pada setiap Lag. Untuk tabel nilai ACF non stationer dapat dilihat pada tabel 3 di bawah ini.

Tabel 3. Nilai ACF non stationer

Lag	Nilai ACF Non-Stationer
1	0,547367
2	0,543085
3	0,553507
4	0,423021
5	0,310359
6	0,183362

Data time series yang diolah menggunakan software minitab 16.0 belum stationer dalam mean dan melewati batas signifikansi  $\alpha 5\%$  sehingga diperlukan proses differencing. namun sebelum melakukan proses differencing, peneliti akan menyajikan cara mendapatkan nilai acf secara manual Tabel dibawah ini menunjukkan data pengamatan untuk mencari nilai acf secara manual pada lag 1, pada proses ini akan dilakukan pencarian dengan menggunakan data model yang telah di transform pada tabel 4 dibawah ini.

Tabel 4. Pengamatan data ACF

Jumlah Pengamatan	Total Pasien Transform
1	0,0344828
2	0,0416667
3	0,0312500
4	0,0312500
5	0,0370370
6	0,0294118
7	0,0344828
8	0,0294118
9	0,0416667
10	0,0416667
11	0,0277778
12	0,0476190
13	0,0500000
14	0,0476190
15	0,0588235
16	0,0526316
17	0,0526316
18	0,0555556
19	0,0588235
20	0,0454545
21	0,0476190
22	0,0555556
23	0,0666667
24	0,0416667
25	0,0526316
Jumlah	1,113402

Mean = Jumlah data pengamatan/ jumlah data =  $5,237938/25 = 0,04453608$

Mencari nilai ACF pada lag 1 :

$$r_1 = \frac{\sum_{t=1}^{25} (y_t - \bar{y})(1 - \bar{y})}{\sum_{t=1}^{25} (y_t - \bar{y})^2}$$

$$= (0,03448 - 0,04453608) * (0,04167 - 0,04453608) + (0,04167 - 0,04453608) * (0,03125 - 0,04453608) + \dots + (0,0416667 - 0,04453608) * ($$

$$0,0526316-0,04453608) \\ = (0,03448-0,04453608)^2 + (0,04167- \\ 0,04453608)^2 + \dots + (0,03125- \\ 0,04453608)^2 \dots \dots \dots (0,0526316- \\ 0,04453608)^2 \text{ (sampai baris hari ke 25)} \\ = 0,547367$$

Untuk melakukan pencarian pacf akan dilakukan dibantu dengan software minitab untuk mendapatkan hasil peramalan yang akurat. maksud dari peneliti menyajikan nilai acf secara manual, untuk mengetahui perhitungan nilai acf yang dan mengerti arti dari nilai yang ada pada sumbu Y dan sumbu X, oleh karena itu peneliti menggunakan *software* minitab sebagai alat bantu pengolahan data untuk peramalan. Hasil differencing secara manual dapat dilihat pada tabel 5 di bawah ini.

Tabel 5. Hasil differencing secara manual

Data SQRT (Transformasi)	Data Hasil Differencing Minitab	Data Hasil Differencing Manual
0,0344828	0	0,0344828-0,0344828 = 0
0,0416667	0,0071839	0,0344828-0,0416667 = 0,0071839
0,0312500	-0,0104167	0,0416667-0,0312500 = -0,0104167
0,0312500	0,0000000	0,0312500-0,0312500 = 0
0,0370370	0,0057870	0,0312500-0,0370370 = 0,0057870
0,0294118	-0,0076253	0,0370370-0,0294118 = -0,0076253
0,0344828	0,0050710	0,0294118-0,0344828 = 0,0050710
0,0294118	-0,0050710	0,0344828-0,0294118 = -0,0050710
0,0416667	0,0122549	0,0294118-0,0416667 = 0,0122549
0,0416667	0,0000000	0,0416667-0,0416667 = 0

Data SQRT (Transformasi)	Data Hasil Differencing Minitab	Data Hasil Differencing Manual
0,0277778	-0,0138889	0,0416667-0,0277778 = -0,0138889
0,0476190	0,0198413	0,0277778-0,0476190 = 0,0198413
0,0500000	0,0023810	0,0476190-0,0500000 = 0,0023810
0,0588235	0,0112045	0,0476190-0,0526316 = 0,0112045
0,0526316	-0,0061920	0,0588235-0,0526316 = -0,0061920
0,0526316	0,0000000	0,0526316-0,0526316 = 0
0,0555556	0,0029240	0,0526316-0,0555556 = 0,0029240
0,0588235	0,0032680	0,0555556-0,0588235 = 0,0032680
0,0454545	-0,0133690	0,0588235-0,0454545 = -0,0133690
0,0476190	0,0021645	0,0454545-0,0476190 = 0,0021645
0,0555556	0,0079365	0,0476190-0,0555556 = 0,0079365
0,0666667	0,0111111	0,0555556-0,0666667 = 0,0111111
0,0416667	-0,0250000	0,0666667-0,0416667 = -0,0250000
0,0526316	0,0109649	0,0416667-0,0526316 = 0,0109649

Setelah data memenuhi asumsi stationeritas dalam *mean* yang nilai- nilai fungsi autokorelasi batas signifikan dari nilai autokorelasi pada lag 1 sampai dengan lag 6 di batas signifikan  $\alpha$

= 5% sehingga data total pasien hari ke 1-hari ke 25 telah memenuhi asumsi stationeritas dalam *mean*. Langkah selanjutnya ada mengidentifikasi model dugaan arima sementara dengan melihat grafik acf yang telah stationer dan grafik pacf.

Tabel 6. Nilai ACF proses stasioner

Lag	Nilai ACF Proses Stationer
1	-0,464829
2	-0,160934
3	0,197578
4	0,026799
5	-0,014678
6	-0,094473

#### 4.4 Forecasting

Dalam tahap ini peneliti akan melakukan peramalan dengan menggunakan metode Arima, dengan menggunakan software minitab 16.0. dengan menggunakan data pada yang telah diolah peneliti membagi data menjadi 2 bagian yaitu : Data total pasien hari ke 1-hari ke 25 dengan jumlah 25 data digunakan untuk membentuk model dugaan Arima sementara. Secara teoritis, dalam analisis time series yang paling menentukan adalah kualitas atau keakuratan dari informasi atau data yang diperoleh serta waktu atau periode dari data tersebut dikumpulkan. Proyeksi total pasien bertujuan untuk mengetahui gambaran total pasien hari ke 26- hari ke 35. data proyeksi dapat dijadikan sebagai acuan atau pedoman oleh pihak puskesmas. Hasil forecasting dapat dilihat pada tabel 7 di bawah ini.

Tabel 7. Hasil forecasting pada minitab

Periode	Forecast	Lower	Upper	Actual
26	16,7335	7,7722	25,6948	17
27	20,5853	11,4298	29,7408	20
28	17,9826	8,5715	27,3937	20
29	16,8652	5,6340	28,0964	17
30	18,3906	6,8093	29,9719	19
31	16,9478	4,9584	28,9373	16
32	16,2891	3,3605	29,2176	17
33	16,7658	3,4222	30,1093	

Periode	Forecast	Lower	Upper	Actual
34	15,8810	2,0887	29,6733	
35	15,4013	0,9692	29,8333	

Pada tabel diatas didapatkan berupa hasil prediksi yang dilakukan oleh Tools minitab dengan model ARIMA (2,1,0) yang menghasilkan nilai lower itu dari nilai minimal forecast dan upper itu hasil kebalikan dari lower itu sendiri.

#### 5. KESIMPULAN

Hasil prediksi total pasien yang akan datang pada Puskesmas Haekesak dengan menggunakan metode ARIMA, analisa ini berakhir dengan jumlah pasien yang akan datang pada hari ke 26-hari ke 35 dimana total pasien tidak naik atau turun secara signifikan sehingga pihak puskesmas tidak perlu menambah tenaga medis ataupun penambahan stock obat yang berlebihan.

#### DAFTAR PUSTAKA

- A. Syarif, S. Ahmad, and Mustakim, "Penerapan Metode ARIMA untuk Peramalan Pengunjung Perpustakaan UIN Suska Riau," Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9, Pekanbaru, 2017, pp. 186-193.
- B. P. Muhammad, and W. Arief, "Aplikasi Metode ARIMA Box-Jenkins Untuk Meramalkan Kasus DBD Di Provinsi Jawa Timur," The Indonesian Journal Of Public Health, vol. 13, no. 2, Desember 2018: 181-194.
- Dewi, & R. (2015). Penerapan Data Mining Dalam Analisis Kejadian TanaH Longsor Di Indonesia Dengan Menggunakan Association Rule Algoritma Apriori. Prosiding Seminar Nasional Matematika dan Pendidikan MatematikaUMS , 1 dan 2.
- E. P. U. Pradipta, and S. N. Azhari. "Predikai Kerawanan Wilayah Terhadap Tindak Pencurian Sepeda Motor Menggunakan Metode (S) ARIMA dan CART," *IJCCS*, vol. 11, no. 2, pp. 119-130, July 2017. <https://doi.org/10.22146/ijccs.16643>
- H. Djoni, "Penerapan Model ARIMA Untuk Memprediksi Harga Saham PT. Telkom Tbk,

*Jurnal Ilmiah Sains*, vol. 11, no. 01, pp. 116–123, 2011.

L. Yonalthia, “Peramalan Penerimaan Pajak Hotel Dengan Metode Runtun Waktu-ARIMA,” M.T. tesis, Fakultas Teknologi Industri, ITS 10 November., Surabaya, 2017.

P. E. Sendy, R. U. Anis, and Y. Rizky, “Peramalan Jumlah Produksi Tanaman Kelapa Sawit Dengan menggunakan Metode ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE),” *JURNAL MANAJEMEN*, vol 8, (1), 2016, pp. 95-112.

R. P. A. Sri, and H. Putriaji, “Peramalan Inflasi di demak Menggunakan Metode ARIMA Berbantuan Software R dan MINITAB,” *Prosiding Seminar Nasional Matematika (PRISMA) 1*, Semarang, 2018, pp. 745-754.

Samsul. (2019). Peramalan data penduduk miskin provinsi nusa tenggara barat (NTB) model autoregressive integrated moving average (ARIMA). *urnal Pemikiran dan Penelitian Pendidikan Matematika*, 1.

W. Satrio, and S. Wellie, “Peramalan Produksi Teh Hijau Dengan Pendekatan Autoregressive Integrated Moving Average,” *Seminar Nasional Pendidikan, Sains dan Teknologi*, Semarang, 2017, pp. 273-282