**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 Jesus1, Anief Fauzan Rozi, S.Kom., M.Eng2**

Program Studi Sistem Informasi, Teknologi Informasi, Universitas Mercu Buana Yogyakarta

Jl. Jembatan Merah 84 C Gejayan Yogyakarta 55283

Email: 1novitacaeiro97@gmail.com,2 **anief@mercubuana-yogya.ac.id**

**ABSTRAK**

Tujuan penelitian ini adalah untuk mengetahui model persamaan metode *Arima*, mengetahui hasil analisi memprediksi jumlah pasien di puskemas haekesak menggunkan tools *Minitab* dan mengetahui hasil memprediksi jumlah pasien yang dirawat di puskemas hasekesak 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 mendis ataupun penambahan stock obat yang berlebihan.

**Kata Kunci:***Forecasting, ARIMA,* Penyaki*t*

**ABSTRACK**

 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 the 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 day 26- day 35 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.

 ***Keywords:*** *Forecasting, ARIMA,* Disease

**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 puskemas 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 memprediski data penyakit dengan melihat jumlah penyakit yang diderita pasien dapat membantu pihak puskemas 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 Memprediksi jumlah penyakit pada pasien berdasarkan penyebaran penyakit yang sering menjadi salah satu faktor penting yang mempengaruhi terjadinya penyakit serta perjalanan suatu penyakit.

**LANDASAN TEORI**

**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)

### **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)

### **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 statsioner 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 pertam aserta Seasonal Autoregressive Integrated Moving Average(SARIMA) yang merupakan turunan dari ARIMA untuk medapatkan prediksi data yang seasonal. (Syarfi Ahmad, & Mustaki 2017)

**METODOLOGI 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.



**ANALISA 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*

#### **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 dibawah ini.



**Gambar Hasil Transformasi**

**Data Reduction**

Pada tahap ini akan dilakukan proses data reduction dilakukan untuk menyesuaikan denngan cara menyesuaikan data yang akan diproses, sehingga atribut harus diesuaikan. 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 tabel dibawah ini

**Tabel 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 |

**Proses Data Mining**

1. **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 $λ$ **=** -1 seperti pada tabel dan gambar di bawah ini

**Tabel Transform model** $λ$ **= -1 manual**

| **Total Pasien** | **Hasil Transform data** $λ$**=-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 | $$\frac{1}{24}=0,0416667$$ |



**Gambar Transform model** $λ$ **= -1 dengan minitab**

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



**Gambar Box Cox transform model λ = 1**

**2. 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.

**Tabel Nilai ACF Non stasioner**

|  |  |
| --- | --- |
| **Lag** | **Nilai ACF Non-Stasioner** |
| 1 | 0,547367 |
| 2 | 0,543085 |
| **Lag** | **Nilai ACF Non-Stasioner** |
| 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 ∝ 5 % sehingga diperlukan proses *differencing.* namun sebelum melakukan proses *differencing.* peneliti akan menyajikan cara mendapatkan nilai acf secara manual Tabel dibawah ini menunjukan 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 dibawah ini.

**Tabel 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 :

$$r\_{1}\_{=\frac{\sum\_{t=1}^{25-l}(y\_{t}-\hat{y})(1-\hat{y)}}{\sum\_{t=1}^{n}(y\_{t=1}-\hat{y})^{2}}}$$

=(0,03448-0,04453608)\*( 0,04167- 0,04453608)+ ( 0,04167- 0,04453608)\*( 0,03125-0,04453608)+...............( 0,0416667-0,04453608)\*( 0,0526316-0,04453608)

 (0,03448-0,04453608)2+(0,04167-0,04453608)2+(0,03125-0,04453608)2................ (0,0526316-0,04453608)2  *(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 niali 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.

**Tabel 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 |
| 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 |
| Data SQRT (Transformasi) | Data Hasil Differencing Minitab | Data Hasil Differencing Manual |
| 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 siginifikansi dari nilai autokorelasi pada *lag* 1 sampai dengan lag 6 di batas signifikansi ∝ = 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 Nilai acf proses stasioner**

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

**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 teoristis, 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.

**Tabel 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 |  |
| 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.

**KESIMPULAN DAN SARAN**

## **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 mendis ataupun penambahan stock obat yang berlebihan.

**Saran**

Jika ada peneliti lain yang akan menggunakan hasil analisa dari penelitian ini disarankan untuk menggunakan aplikasi dan metode lain sebagai pembanding.

# **REFERENSI**

1. Azhari, P. &. (2017). PrediksiKerawanan Wilayah Terhadap Tindak Pencurian Sepeda Motor MenggunakanMetode(S)ARIMA Dan CART. *JCCS* , 1.
2. 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.
3. samsul. (2019). Peramalan data penduduk miskin provinsi nusa tenggara barat (NTB) model auto regressive integrated moving average(ARIMA) . *urnal Pemikiran dan Penelitian Pendidikan Matematika* , 1.
4. Sendy, A. &. (2016). Peramalan Jumlah Produksi Tanaman Kelapa Sawit Dengan menggunakan Metode Arima. *URNAL MANAJEMEN* , 1.
5. Syarfi, A. &. (2017). enerapan Metode ARIMA untuk Peramalan PengunjungPerpustakaan UIN Suska Riau. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)* , 3.