

NON LINEAR AUTOREGRESSIVE MOVING AVERAGE UNTUK PREDIKSI KUNJUNGAN WISATAWAN

Ahmad Ashril Rizal¹, Abdurrahim², Andi Sofyan Anas³

^{1,2,3} Program Studi Teknik Infomatika, STMIK Bumigora Mataram

¹ashril.rizal@stmikbumigora.ac.id, ²aim.math13@gmail.com, ³andi.sofyan@stmikbumigora.ac.id

Abstract

Prediction is one of the elements for decision support in the future. Because of the unavailability of natural resources such as oil and gas, forest products or large-scale manufacturing industries in Lombok, tourism has become a leading sector in economics development. Prediction of tourist arrivals needs to be done to support policies related to tourism development. There are two basic methods of prediction: arima and neural network. Arima is good for prediction with stationary dataset, while neural networks are either used for prediction with stationary or non-stationary data. Previous research related to the prediction of tourist arrivals using Recurrent Neural Network approach with Extended Kalman Filter. This research tries to predict time series data from tourist arrivals by Nonlinear Autoregressive Moving Average (NARMA) approach. Predicted results based on Mean Square Error (MSE), the best prediction result is given on ARMA model (5,1,0) with MSE 0.178.

Keyword: Prediction, Time Series, Nonlinear, Autoregressive, Moving Average, Narma

I. Pendahuluan

Dengan tidak tersedianya sumber daya alam seperti migas, hasil hutan ataupun industri manufaktur yang berskala besar di provinsi Nusa Tenggara Barat, khususnya di pulau Lombok, maka pariwisata telah menjadi sektor andalan dalam pembangunan. Kontribusi sektor pariwisata menunjukkan *trend* yang semakin meningkat dari tahun ke tahun. Penukaran valuta asing dan pengeluaran wisatawan di daerah pariwisata adalah salah satu pemberi sumbangsih yang baik terhadap pendapatan masyarakat.

Pengaruh jumlah Wisatawan mancanegara dan lama tinggal secara parsial berpengaruh positif terhadap penerimaan Produk Domestik Regional Bruto. Selain itu industri pariwisata juga menyumbang peran terhadap pendapatan masyarakat. Jika ditinjau dari kesempatan kerja maka hampir seluruh kesempatan kerja yang ada di pulau Lombok merupakan kontribusi dari sektor pariwisata. Hal ini

bisa saja terjadi karena kesempatan kerja adalah implikasi dari pengeluaran Wisatawan dan investasi di sektor pariwisata. Implikasi lainnya dari pengeluaran Wisatawan terhadap perekonomian terdistribusi ke berbagai sektor, bukan saja hotel dan restoran. Distribusi tersebut juga terserap ke sektor pertanian, sektor industri dan kerajinan, sektor pengangkutan dan komunikasi, sektor jasa dan sebagainya. Hal ini sejalan dengan data mengenai distribusi pengeluaran Wisatawan. Data menunjukkan bahwa selama berkunjung, pengeluaran Wisatawan yang terserap ke dalam perekonomian rakyat cukup tinggi. Dari beberapa penelitian terkait dapat digeneralisasi bahwa peran Wisatawan memiliki pengaruh positif terhadap perekonomian di Indonesia.

Penelitian sebelumnya terkait prediksi kunjungan Wisatawan dilakukan [1] dengan menerapkan metode *Recurrent Neural Network*. Pada penelitian ini dilakukan prediksi kunjungan wisatawan dengan

pendekatan NARMA. Data yang digunakan sebagai data *training* adalah data kunjungan wisatawan mancanegara yang terekam dalam data wisatawan pada Dinas Pariwisata NTB. Selain sebagai kajian pustaka dalam analisa metode terbaik dalam melakukan prediksi kunjungan wisatawan, penelitian ini juga diharapkan dapat membantu pelaku wisata dalam mempersiapkan diri menyambut kedatangan wisatawan.

Dalam penelitian ini, digunakan beberapa referensi sumber pustaka yang berasal dari penelitian yang sudah dilakukan sebelumnya. Dalam melakukan prediksi terdapat dua pendekatan yang digunakan. Pendekatan pertama disebut dengan *time-series* yaitu model yang tidak memperlihatkan kecenderungan dari data masa lalu yang tersedia, sedangkan pendekatan yang kedua adalah pendekatan yang memperlihatkan hubungan sebab akibat (*cause-effects method*) atau pendekatan yang menjelaskan terjadinya suatu keadaan (*eksplanatory method*) oleh sebab-sebab tertentu. Permasalahan yang muncul kemudian adalah bagaimana melakukan prediksi. Pada awalnya untuk melakukan prediksi digunakan metode peramalan seperti *Autoregressive Integrated Moving Average Model* (Arima). Metode ini memiliki keterbatasan pada pengabaian kemungkinan hubungan non linear serta stationeritas data dan homokedastitas residual [5]. Kini metode peramalan data dengan *time-series* telah berkembang dengan pendekatan *Neural Network*. Model Arima dan *Neural Network* memiliki perbedaan yaitu Arima baik digunakan untuk meramal data *time series* yang linear sementara *Neural Network* baik digunakan untuk data linear maupun non linear [6].

II. METODE PREDIKSI

Analisis *time series* (runtun waktu) adalah suatu metode kuantitatif untuk melihat pola data pada masa lalu yang telah dikumpulkan secara runtun[3]. Analisis runtun waktu juga merupakan salah satu metode peramalan yang menjelaskan bahwa deretan observasi pada suatu variabel

dipandang sebagai realisasi dari variabel acak yang berdistribusi normal. Gerakan musiman adalah gerakan rangkaian waktu sepanjang tahun pada bulan-bulan yang sama selalu menunjukkan pola yang identik. Dalam hal ini, kunjungan wisatawan adalah data runtun waktu yang tidak berbentuk musiman.

Terdapat beberapa perbedaan antara ARIMA (*Autoregressive Integrated Moving Average*) dengan NARMA. Data kunjungan wisatawan data *stationer*. Namun dengan banyaknya faktor yang mempengaruhi kunjungan wisatawan seperti masa libur, kurs rupiah, keamanan, *event-event* tertentu dan lain-lain maka dapat dimungkinkan terdapat data yang tidak *stationer*. Metode ARMA merupakan analisa model runtun waktu untuk data *stationer*. Sedangkan ARIMA merupakan analisa model runtun waktu untuk data *non-stationer* [2].

N : Nonlinear

AR : *Autoregression* adalah sebuah model menggunakan hubungan ketergantungan antara pengamatan dan beberapa pengamatan tertinggal (*lag observastion*)

MA : *Moving Average* adalah sebuah model yang menggunakan ketergantungan antara pengamatan dan kesalahan residual dari model *moving average* yang diterapkan pada *lag observation*.

Proses ARMA (p,q) adalah suatu model campuran antara *autoregressive* orde p dan *moving average* orde q . *Autoregressive* (AR) merupakan suatu observasi pada waktu t dinyatakan sebagai fungsi linear terhadap p waktu sebelumnya ditambah dengan sebuah *residual* acak a_t yang *white noise* atau independen dan berdistribusi normal dengan rata-rata 0 dan varian konstan σ_a^2 , ditulis $a_t \sim N(0, \sigma_a^2)$. Bentuk umum model *autoregressive* orde p atau lebih ringkas ditulis model AR(p). *Moving average* (MA) digunakan untuk menjelaskan suatu fenomena bahwa suatu observasi pada waktu t dinyatakan sebagai kombinasi linear dari sejumlah *random*

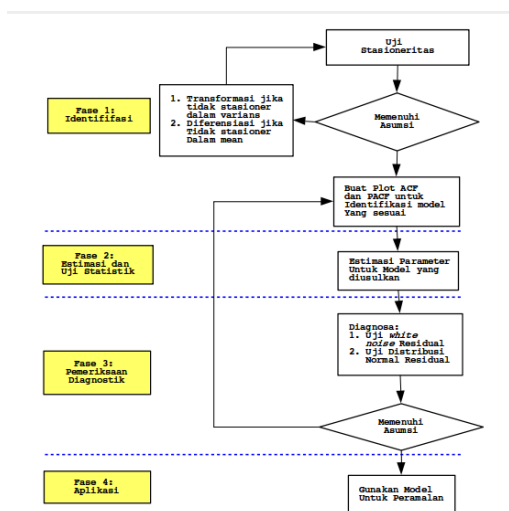
error. Bentuk umum model *moving average* orde q atau lebih ringkas ditulis model $MA(q)$ [4].

Masing-masing komponen ini (p dan q) secara eksplisit ditentukan dalam model sebagai parameter. Sebuah notasi standar digunakan untuk ARMA (p, q) di mana parameternya diganti dengan nilai integer untuk segera menunjukkan model ARMA spesifik yang digunakan. Parameter model ARIMA didefinisikan sebagai berikut

p : Jumlah observasi lag yang termasuk dalam model atau disebut juga *lag order*.

q : Ukuran *window* dari *moving average* atau disebut juga urutan *moving average*.

Langkah prediksi dengan NARMA selengkapnya ditunjukkan dalam Gambar 1.



Gambar 1. Rancangan NARMA

Setelah data yang diperlukan diperoleh, kemudian berdasarkan teori-teori yang ada dan dengan menggunakan metode analisis *time series*, kemudian digunakan untuk proses prediksi kunjungan wisatawan.

Langkah dalam proses peramalan dengan menggunakan metode NARMA adalah sebagai berikut:

1. Kestasioneran Data

Hal yang pertama dilakukan dalam peramalan analisis runtun waktu adalah menghasilkan data yang stasioner artinya data mempunyai rata-rata dan varians

yang sama sepanjang waktu. Apabila data yang menjadi *input* dari model NARMA tidak stasioner, perlu dimodifikasi untuk menghasilkan data yang stasioner. Metode yang digunakan adalah metode *differencing*.

2. Identifikasi Model

Tahap kedua dalam melakukan peramalan dengan metode ini adalah menentukan model analisis runtun waktu berdasarkan ACF dan PACF. Salah satu fungsi ACF digunakan untuk menentukan kestasioneran data runtun waktu, jika dari ACF data asli ternyata data belum stasioner, maka dilakukan modifikasi data, yaitu dengan cara *differencing* pertama. ACF dan PACF digunakan untuk mengidentifikasi model analisis runtun waktu. Berdasarkan nilai ACF dan PACF, apabila nilai ACF turun lambat maka yang berperan dalam penentuan model adalah PACF. Artinya jika PACF terputus pada *lag* pertama berarti modelnya $AR(1)$, jika terputus pada *lag* ke-2 berarti modelnya $AR(2)$. Sebaliknya pada saat PACF turun lambat, maka jika ACF terputus pada *lag* ke-1 maka modelnya adalah $MA(1)$ dan jika terputus pada *lag* ke-2 berarti modelnya $MA(2)$.

3. Estimasi Parameter

Tahapan selanjutnya setelah diketahui model yang tepat dari data tersebut yaitu mencari nilai estimasi parameter model.

4. Verifikasi

Langkah selanjutnya adalah verifikasi yaitu memeriksa apakah model yang diperoleh cocok dengan data observasi. Pengujian kelayakan model dapat dilakukan dengan beberapa cara :

a. *Overfitting*

Overfitting dilakukan apabila disangka bahwa mungkin diperlukan model yang lebih luas. Namun, dalam hal ini perlu diperhatikan bahwa

dalam metode NARMA berlaku prinsip parsimoni artinya model yang dipilih adalah model yang paling sederhana yaitu yang jenjangnya paling rendah dan komponennya paling sedikit.

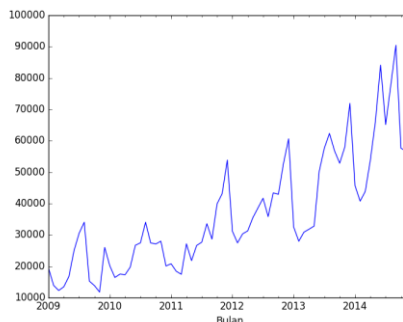
- b. Membandingkan nilai *mean square error*
Mean square error (MSE) prediksi merupakan suatu kriteria untuk memilih model terbaik. Semakin kecil MSE yang dihasilkan suatu model, maka model semakin baik.
- c. Menguji residual
 Secara sistematis residual (*error*) dapat dihitung dengan cara mengurangi data asli dengan data hasil ramalan. Residual model harus berdistribusi normal, identik, dan independen. Dapat diamati dengan melihat independensi antar *lag* dengan membuat plot ACF dan PACF residual. Bila tidak ada 1 *lag* pun yang keluar dari batas signifikansi maka dapat disimpulkan residual model bersifat independen. Dan model dianggap memadai untuk digunakan sebagai model peramalan.

- 5. Peramalan
 Setelah diperoleh model yang cocok (siapa digunakan untuk peramalan), maka model tersebut siap digunakan untuk prediksi.

Penelitian ini dilakukan pada kunjungan Wisatawan di pulau Lombok di mana data akan diperoleh Dinas Pariwisata Provinsi Nusa Tenggara Barat Data yang telah diperoleh dari Dinas Pariwisata Nusa Tenggara Barata akan dianalisa dengan mengukur *Mean Square Error* dan akurasi hasil prediksi dari kedua pendekatan yang digunakan.

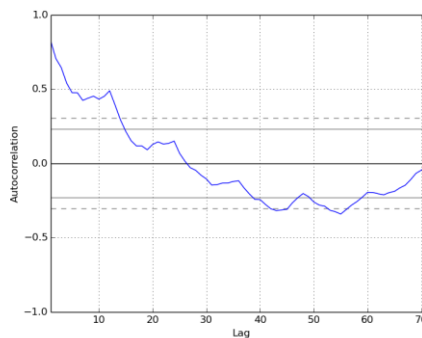
III. HASIL DAN PEMBAHASAN

Data kunjungan wisatawan dimuat dalam sebuah grafik. Grafik kunjungan wisatawan dari Januari 2009 hingga Desember 2014 ditunjukkan pada Gambar 2.



Gambar 2. Grafik Kunjungan Wisatawan

Data kunjungan wisatawan dari Gambar 2 menunjukkan bahwa deret waktu dari kunjungan wisatawan tidak stasioner dan akan memerlukan perbedaan untuk membuatnya stasioner, setidaknya urutan perbedaan. Plot autokorelasi dari rangkaian waktu dapat dilihat pada Gambar 3. Autokorelasi ini dengan dengan *built-in* pada python-pandas.



Gambar 3. Plot Autokorelasi

Dari hasil ujicoba yang dilakukan, hasil dengan beberapa model *arma* (p,q) memberikan hasil prediksi yang berbeda-beda. Hasil prediksi dengan NARMA (5,1) memberikan hasil terbaik dari beberapa uji coba yang dilakukan. Berikut hasil lengkap NARMA dengan model NARMA (5,1).

```

NARMA
Model Results
=====
    
```

```

Dep. Variable:          D.#Arrivals
No. Observations:      71
Model:                  ARIMA(5, 1, 0)
Log Likelihood          -747.454
Method:                 css-mle
S.D. of innovations    8991.343
Date:                   Thu, 19 Oct 2017
AIC                    1508.909
Time:                   11:23:30
BIC                    1524.748
Sample:                 02-01-2009
HQIC                   1515.207
                       - 12-01-2014
    
```

```

=====
=====
=====
    
```

z	P> z	coef	std err
[95.0% Conf. Int.]			
const		753.3909	490.630
1.536	0.130	-208.225	1715.007
ar.L1.D.#Arrivals		-0.2363	0.115
-2.062	0.043	-0.461	-
0.012			
ar.L2.D.#Arrivals		-0.3316	0.115
-2.894	0.005	-0.556	-
0.107			
ar.L3.D.#Arrivals		-0.1753	0.125
-1.399	0.167	-0.421	-
0.070			
ar.L4.D.#Arrivals		-0.2169	0.123
-1.766	0.082	-0.458	-
0.024			
ar.L5.D.#Arrivals		-0.2695	0.123
-2.199	0.031	-0.510	-
0.029			

```

Roots
=====
=====
    
```

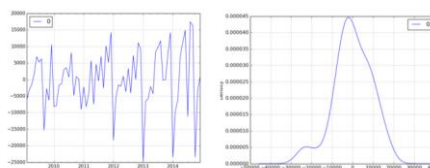
	Real	Imaginary	Modulus
AR.1	0.8126	-	-
0.9487j	1.2491	-	-
0.1373			
AR.2	0.8126	-	-
+0.9487j	1.2491	-	-
0.1373			
AR.3	-0.4832	-	-
1.1796j	1.2748	-	-
0.3119			
AR.4	-0.4832	-	-
+1.1796j	1.2748	-	-
0.3119			
AR.5	-1.4637	-	-
0.0000j	1.4637	-	-
0.5000			

```

-----
count      0
mean     71.000000
std     -57.633200
min     9064.902912
25%    -24706.911510
50%    -5148.864385
75%    -178.409423
max     6919.105287
max     17439.492162
    
```

Dalam rangkaian waktu stasioner, sifat statistik dari waktu ke waktu harus konstan dan *autoconviance* harus independen. Biasanya, saat menjalankan regresi reguler, diasumsikan bahwa pengamatan bersifat independen satu sama lain. Dalam kasus *time series*, hasil pengamatan tergantung pada waktu. Untuk menggunakan teknik regresi pada variabel yang dependen terhadap waktu, data harus stasioner. Teknik yang berlaku untuk variabel acak independen juga berlaku untuk variabel acak stasioner. Ada dua cara untuk memeriksa kestasioneran data *time series*. Pertama adalah plot varians bergerak dan mengamati jika tetap konstan seiring berjalannya waktu. Namun metode tersebut tidak selalu dapat membuat kesimpulan secara visual. Cara kedua adalah dengan menggunakan tes *Dickey-Fuller* yang merupakan sebuah uji statistik dengan hipotesis yang menunjukkan bahwa data tidak stasioner.

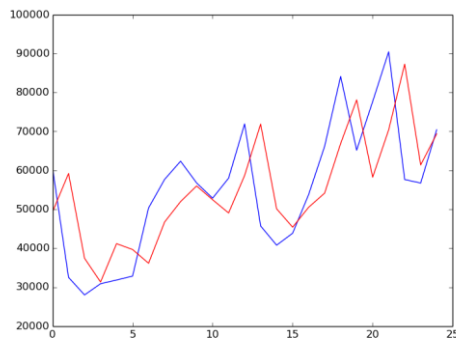
Setelah menguji kestasioneran data, dilakukan plot terhadap dari kesalahan residual yang akan menunjukkan bahwa mungkin masih ada beberapa informasi *trand* yang tidak ditangkap oleh model. Plot kerapatan dari nilai kesalahan residual, menunjukkan bahwa kesalahannya bersifat Gaussian, namun mungkin tidak berpusat pada nol. Gambar 4 menunjukkan plot dari fit residual dimana Gambar 4 bagian kiri adalah *line plot arma fit residual error* dan bagian kanan adalah plot *arma fit residual error density*.



Gambar 4. Fit Residual Error Density Plot

Model NARMA dapat digunakan untuk memperkirakan langkah waktu di masa depan. Fungsi *predict()* pada objek *ARIMAResults* digunakan untuk membuat prediksi. Fungsi ini menerima indeks dari deret waktu untuk membuat suatu prediksi. Indeks ini bersifat relatif terhadap awal dataset pelatihan yang digunakan untuk membuat prediksi. Grafik hasil prediksi

dengan $arma(5,1)$ ditunjukkan pada Gambar 2. Hasil prediksi tersebut memberikan nilai MSE 0.0178.



Gambar 5. Grafik Hasil Prediksi

IV. Kesimpulan

Berdasarkan analisa pada *training* dan *testing* dengan ARMA, maka dapat diambil kesimpulan bahwa metode *time series* dapat digunakan untuk memprediksi kunjungan wisatawan di pulau Lombok. Metode *time series* yang memiliki hasil paling mendekati dengan data nyata adalah ARMA (5,1) dengan nilai MSE yang diperoleh adalah 0.0178.

Saran pengembangan yang diberikan untuk penelitian lebih lanjut adalah melihat *sentiment analysis* dari faktor-faktor yang mempengaruhi kunjungan wisatawan. Dari data tersebut bisa dilakukan prediksi dengan analisa multivariate untuk melakukan prediksi kunjungan wisatawan beberapa waktu mendatang.

Referensi

- [1] A. A. Rizal and S. Hartati, "Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok," 2016 *International Conference on Informatics and Computing (ICIC)*, Mataram, 2016, pp. 180-185.
- [2] Brath, A., Castellarin, A., & Montanari, A. 1999. Detecting non stationarity in extreme rainfall data observed in Northern Italy. In *Proceedings of EGS-Plinius*

Conference on Mediterranean Storms, Maratea (pp. 219-231).

- [3] Mulyono, S. 2006. Statistik untuk Ekonomi dan Bisnis Edisi Ketiga. Jakarta: Lembaga Penerbit Fakultas Ekonomi UI.
- [4] Makridakis, S. dan Steven, W., 1993, Metode dan Aplikasi Peramalan, Jakarta : Penerbit, Erlangga.
- [5] Adnyani, L.P.W., 2012, General Regression Neural Network (GRNN) Pada Peramalan Data Time Series, Tesis, S2 Matematika UGM, Yogyakarta.
- [6] Munarsih, E., 2011, Penerapan Model Arima-Neural Network Hybrid Untuk Peramalan Data Time Series, Tesis, Program Studi S2 Matematika, Universitas Gadjah Mada, Yogyakarta.