

Penerapan Analisa Time Series Terhadap Nilai Matematika di SMAS Alfa Centauri Bandung.

Imam Nulhakim, Utriweni Mukhaiyar

Institut Teknologi Bandung, Jl.Tamansari no 64, Bandung; imamnul@gmail.com

Jl.Tamansari no 64, Bandung; utriweni@math.itb.ac.id

Abstrak

Time series merupakan serangkaian pengamatan berdasarkan pada urutan waktu dengan mempelajari pola gerakan nilai-nilai variable pada satu interval waktu. Berdasarkan pengamatan tersebut dapat dilihat suatu model yang dapat digunakan untuk memprediksi kejadian pada periode berikutnya. Hal yang akan dikaji dalam penelitian ini adalah memprediksi secara kuantitatif terjadinya perubahan dan perkembangan kemampuan penguasaan materi pada pelajaran matematika yang ada di SMAS Alfa Centauri Bandung yang dilihat dari pencapaian nilai Try Out Ujian Nasional siswa. Metode yang digunakan adalah Autoregressive Integrated Moving Average (ARIMA). Adapun data yang digunakan dalam penelitian ini adalah nilai Try Out siswa dari tahun 2010 sampai dengan ujian nasional pada tahun 2014. Didapatkan Model forecast hasil nilai ujian ke- t adalah ARI (3,1) dengan hasil ujian siswa dipengaruhi oleh tiga nilai ujian terakhir sebelum ujian yang akan dihadapi (waktu ke $t - 1$, $t - 2$ dan $t - 3$).

Kata Kunci : ARIMA, Stasioner, Forecast

1. PENDAHULUAN

Time series merupakan analisis yang berhubungan erat dengan prediksi di masa yang akan datang. Kondisi data yang ada sesuai dengan urutan waktu atau memiliki periode tertentu dengan pengamatan berdasarkan pada urutan waktu yang mempelajari pola gerakan nilai-nilai variable pada satu interval waktu. Secara umum, semua aktifitas yang dilakukan sering mengalami ketidak pastian dalam hal pengambilan keputusan sehingga diperlukan suatu peramalan untuk memprediksi kejadian di masa yang akan datang. Berdasarkan pengamatan tersebut dapat dilihat suatu model yang dapat digunakan untuk memprediksi kejadian pada periode berikutnya.

Time series banyak digunakan dalam berbagai disiplin ilmu seperti halnya ekonomi, pertanian, teknologi komunikasi. Akan tetapi, time series masih jarang digunakan dalam dunia pendidikan. Adapun beberapa peneliti di dunia pendidikan yang menggunakan salah satu diantaranya adalah Waren (2009, 293-296) dengan penelitiannya yang berjudul "Assessing the Effects of Multiple Educational Interventions in a Small-Enrollment Course". Penelitian Waren bertujuan untuk mengetahui bagaimana mengidentifikasi intervensi (kuliah, labs, pekerjaan rumah, tugas, dll) pada performa siswa dengan menggunakan data kuantitatif.

Dalam dunia pendidikan di Indonesia, Tolak ukur untuk untuk mengetahui mutu pendidikan adalah dengan ujian nasional. Ujian nasional merupakan sistem

evaluasi standar pendidikan dasar dan menengah secara nasional dan persamaan mutu tingkat pendidikan antar daerah yang dilakukan oleh pusat penilaian pendidikan. Merujuk pada hasil penelitian dari Tjala (2010) Ujian Nasional yang dilaksanakan dalam beberapa tahun terakhir bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi. Hasilnya digunakan sebagai:

1. Pemetaan mutu satuan dan/atau program pendidikan;
2. Seleksi masuk jenjang pendidikan berikutnya;
3. Penentuan kelulusan peserta didik dari program dan/atau satuan pendidikan;
4. Pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upaya peningkatan mutu pendidikan.

Berdasarkan hal tersebut, dengan menggunakan analisa time series, akan diprediksi hasil ujian nasional yang merupakan salah satu alat pengendali mutu pendidikan secara nasional berdasarkan hasil pencapaian dari nilai try out dan nilai rata-rata dari ulangan harian siswa. Adapun subjek yang dikaji dalam makalah ini siswa dari SMAS Alfa Centaury yang berada di Kota Bandung.

2. METODE PENELITIAN

Data yang digunakan adalah nilai rata-rata matematika siswa yang dijadikan sampel dimulai dari tahun ajaran 2010/2011 sampai 2014/2015. Menurut Dettling [1] Suatu data time series $\{Z_t\}$ disebut mengikuti model *auto regressive integrated moving average* jika *difference* ke- d $W_t = (1 - B)^d Z_t$ merupakan proses stasioner ARMA. Jika W_t adalah ARMA (p,q) maka Z_t adalah ARIMA (p,d,q). Secara umum model ARIMA dinyatakan sebagai

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t$$

dengan

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

adalah koefien komponen MA non musiman dengan order q

a_t : residual white noise, $a_t \sim iid N(0, \sigma_a^2)$

B : operator mundur

$(1 - B)^d$: pembedaan tak musiman dengan order pembedaan tak musiman d
Order pembedaan yang bernilai bulat tak negatif dapat memberikan indikasi terhadap kestasioneran suatu model ARIMA.

Adapun untuk memilih model terbaik pada analisa time series, kriteria model didasarkan pada RMSE (Root Mean Square Error), MAPE (Mean Squared Deviation) yang terkecil. Pada metode ARIMA tahapan-tahapan yang digunakan adalah tahapan Box-Jenkins, langkah-langkahnya sebagai berikut.

- a. Identifikasi model sementara, pada tahap ini dilakukan identifikasi stasioneritas data, baik dalam mean atau varians. Pemeriksaan stasioneritas dilakukan dengan memeriksa time series plot pada keseluruhan data.
- b. Estimasi Parameter Model. Secara umum ada tiga metode dalam model ARIMA, yaitu metode moment, least Square atau Maximum likelihood.

- c. Cek Diagnosa. Cek diagnosa kesesuaian model yang digunakan yaitu uji asumsi apakah telah random (white noise) serta berdistribusi normal

Dasar utama dari analisa *time series* adalah stasioner (1999). Sebuah time series $\{r_t\}$ disebut sebagai *strictly stasionary* jika gabungan distribusi dari $(r_{t_1}, \dots, r_{t_k})$ identik dengan $(r_{t_1+t}, \dots, r_{t_k+t})$ untuk setiap t , yang mana k merupakan sebuah bilangan bulat sembarang dan (t_1, \dots, t_k) merupakan kumpulan dari bilangan bulat positif k . Dengan kata lain, *strictly stasionary* mengharuskan gabungan dari distribusi $(r_{t_1}, \dots, r_{t_k})$ merupakan invariant dari sebuah deret waktu. Dengan mempertimbangkan lemahnya kestasioneran dari suatu deret r_t (Tsay, 2005). Ketika kelinieran bergantung antara r_t dan nilai sebelumnya dari r_{t-i} . Konsep korelasi diperumum dengan auto korelasi. Hubungan dengan koefisien antara r dan r_{t-l} dikatakan lag- l auto korelasi dari r_t dan umumnya dinotasikan sebagai ρ_t . Untuk lebih khususnya definisikan

$$\rho_t = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}}$$

Jika \bar{r} merupakan sebuah rata-rata sampel. Maka untuk lag- l autokorelasi

$$\hat{\rho}_l = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}$$

Tes individual ACF (Auto Correlation Function). Diberikan sebuah bilangan buat l , hasil kedepannya dapat digunakan dengan menguji $H_0: \rho_t = 0$ dibandingkan dengan $H_1: \rho_t \neq 0$. Dengan uji stasistik

$$t - rasio = \frac{\hat{\rho}_l}{\sqrt{(1 + 2 \sum_{i=1}^{l-1} \hat{\rho}_i^2) / T}}$$

Selain dengan ACF, perlu dilakukan dengan PACF (Partial Auto Correlation Function). PACF dari stasioner time series merupakan sebuah fungsi dari ACF dan sangat berguna untuk menentukan order dari ρ dari sebuah AR model.

Untuk mendapatkan model yang sesuai, haruslah memenuhi berbagai macam uji pertama perlu diketahui White Noise. Sebuah time series r_t dikatakan white noise jika $\{r_t\}$ merupakan rangkaian yang independen dan berdistribusi identik variable acak dengan rata-rata terbatas dan bervariansi. Pada dasarnya, jika r_t merupakan distribusi normal dengan rata-rata nol dan variansi σ^2 , barisan tersebut dikatakan Gaussian white noise.

Selanjutnya dilakukan estimasi parameter. Untuk spesifik model AR(p) kondisi paling sedikit dari metode barisan yang mana dimulai dari barisan observasi ke $(p + 1)$, hal ini sering digunakan untuk estimasi parameter. Untuk lebih jelasnya, kondisional dari p pertama diperoleh

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p}, \quad t = p + 1, \dots, T$$

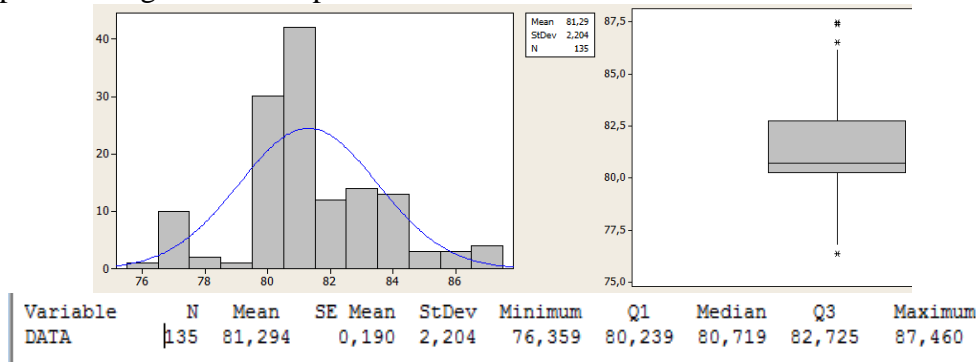
Dan terkait dengan residual

$$\hat{a}_t = r_t - \hat{r}_t$$

Adapun model yang cocok haruslah diperiksa secara hati-hati untuk memperoleh model yang memadai. Jika model yang memadai, maka residual haruslah bersifat white noise. ACF dan Ljung-Box statistic dari residual bisa di periksa sebagai kedekatan dari \hat{a} dari white noise.

3. HASIL PENELITIAN DAN PEMBAHASAN

Statistik deskriptif merupakan bagian dari statistik yang menitik beratkan pada pengumpulan, penyajian, pengolahan serta peringkasan data yang mana aktivitas ini tidak berlanjut pada penarikan kesimpulan. Statistik deskriptif disajikan untuk memberikan informasi mengenai karakteristik variable penelitian, antara lain *mean*, nilai maksimum, nilai minimum dan standar deviasi, dan untuk melihat pencilan digunakan box plot.

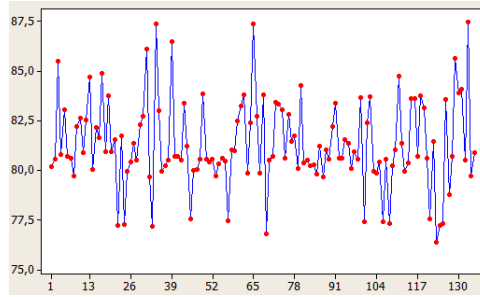


Gambar 1. Deskriptif data

Statistik deskriptif pada gambar 1 menunjukkan bahwa nilai matematika tersebut mempunyai nilai minimum yang lebih kecil dari *mean*, dan nilai maksimum yang lebih besar dari, serta standar deviasi yang lebih kecil dari mean. Hal tersebut menunjukkan bahwa variabel mengindikasikan hasil yang baik.

Dalam prosesnya terdapat asumsi yang harus dipenuhi sehingga analisa time series dapat dilaksanakan. Pertama, subjek yang diamati semenjak tahun 2010-2014 merupakan suatu subjek yang sama. Meskipun subjek siswa berbeda dari tiap angkatan, namun berada di suatu sekolah yang sama, dengan kurikulum serta aturan sekolah yang serupa dari tahun ke tahun menjadikan subjek yang diteliti dianggap satu kesatuan. Kedua, data yang digunakan dianggap sebagai suatu barisan waktu yang terjadi secara kontinu.

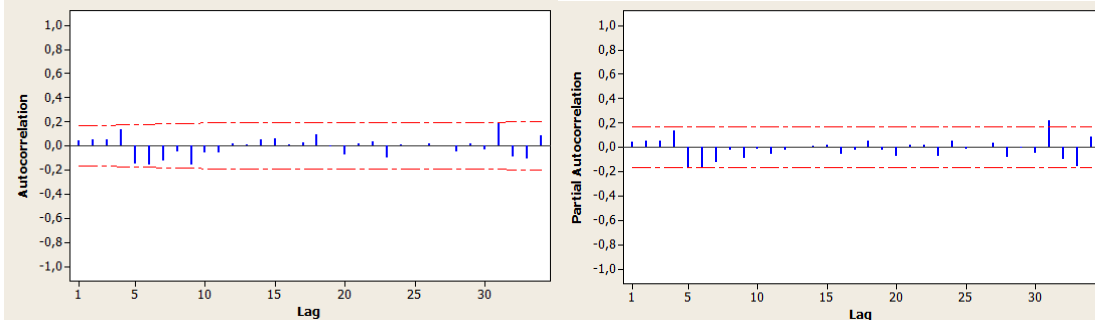
Penerapan model autoregresif mensyaratkan bahwa data yang digunakan adalah data yang stasioner. Data dinyatakan stasioner jika nilai rata-rata dan varian dari data tersebut tidak mengalami perubahan secara sistematis sepanjang waktu (Wei, 1994). Untuk melakukan pemodelan langkah awal yang dilakukan, yaitu memplot data untuk melihat kestasioneran data. Adapun hasil plotnya sebagai berikut.



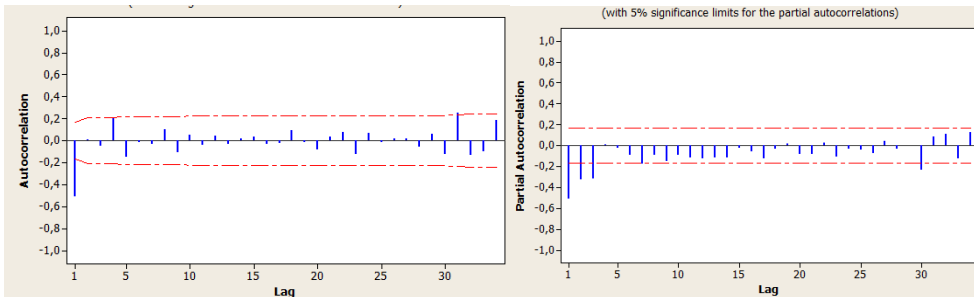
Gambar 2. Plot time series

Pada Gambar 2 terlihat variansi data yang cukup besar, sehingga dilakukan tindakan differensiasi untuk mendapatkan data yang stasioner hal ini dilakukan karena syarat terpenting yang harus dipenuhi agar data dapat dimodelkan oleh keluarga ARIMA adalah stasioner. Data yang diolah merupakan data yang stasioner baik dalam mean maupun varian. Pengujian dengan ACF dan PACF bersifat obyektif, maka diperlukan pengujian secara ilmiah untuk itu digunakan uji unit root test. Analisis selanjutnya adalah melihat grafik ACF dan PACF.

d=0



d=1



Gambar 3. Plot ACF dan PACF

Ada dua asumsi yang harus dipenuhi dalam menentukan kecukupan model, yaitu residual bersifat white noise dan berdistribusi normal (Roesadi, 2014). Pengujian

asumsi residual white noise dapat dilakukan dengan menggunakan uji *Ljung-Box* dengan $\alpha = 5\%$ sebagai berikut:

Tabel 1. Estimasi dan uji signifikansi

<i>Model</i>	<i>Parameter</i>	<i>P-Value</i>	<i>Keputusan</i>
ARI(3,1)	$\phi_1 = -0,65$	0,00	Signifikan
	$\phi_2 = -0,49$	0,00	Signifikan
	$\phi_3 = -0,27$	0,00	Signifikan
	$\mu = 0,04$	0,969	Tidak Signifikan
IMA (1,1)	$\theta_1 = 0,98$	0,00	Signifikan
	$\mu = 0,003$	0,529	Tidak Signifikan

Dari bagan Estimasi dan uji signifikansi, hanya terdapat 2 model yang signifikan Model ARI (3,1) dan IMA (1,1) dengan konstanta yang diabaikan karena tidak signifikan.

Tabel 2. Uji residual *White Noise*

Model	Lag	P-Value	Keputusan
ARI (3,1)	12	0,209	<i>White Noise</i>
	24	0,655	<i>White Noise</i>
	36	0,186	<i>White Noise</i>
	48	0,293	<i>White Noise</i>
IMA (1,1)	12	0,036	<i>Tidak white noise</i>
	24	0,303	<i>White Noise</i>
	36	0,299	<i>White Noise</i>
	48	0,488	<i>White Noise</i>

Dari dua model di atas model yang memenuhi uji white noise hanyalah ARI(3,1), selanjutnya kitamenguji residual (uji kenormalan menggunakan uji *Kolmogorov-Smirnov*)

Tabel 3. Uji Kenormalan Residual

Model	P-Value	Keputusan
AR (3,1)	0,150	Residual berdistribusi normal

Dari *Tabel Estimasi dan uji signifikansi, Tabel Uji residual White Noise dan Tabel Uji Kenormalan Residual*, jelas bahwanya terdapat satu model dan

merupakan model terbaik yaitu model ARI(3,1). Dengan model **ARI(3,1):** $W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} + e_t$.

Karena telah differensiasi satu kali menjadikan $Y_t = W_{(t-1)}$ dengan $Y_t = \phi_1(Y_{t-1} - W_{t-2}) + \phi_2(Y_{t-2} - W_{t-3}) + \phi_3(Y_{t-3} - W_{t-4}) + e_t$

Tabel 4. Hasil Prediksi

Periode	Forecast	Lower	Upper	Actual
Try Out Januari	81,072	75,2658	84,9747	79,221
Try Out Februari	82,377	77,9670	86,4083	80,205
Ujian Nasional	82,416	77,3843	86,6938	82,698

4. KESIMPULAN

ARI (3,1): $Y_t = -0,65(Y_{t-1} - W_{t-2}) - 0,46(Y_{t-2} - W_{t-3}) - 0,27(Y_{t-3} - W_{t-4}) + e_t$ sehingga model ARI(3,1) akan digunakan untuk prediksi Nilai Ujian Nasional (UN) mata pelajaran matematika. Artinya hasil Ujian Nasional mata pelajaran matematika, dipengaruhi oleh nilai try out tiga ulangan sebelumnya.

5. DAFTAR PUSTAKA

- Detting, Marcel. (2013). *Applied Time Series Anlysis*. Diakses dari https://stat.ethz.ch/education/semesters/ss2013/atsa/ATSA-Scriptum-SS2013_130415.pdf.
- Markidakis, Wheelwright. (1999). *Metode dan Aplikasi Peramalan*. Jakarta: Bina Rupa Aksara
- Rosadi, Dedi. (2012). *Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews*. Yogyakarta: Andi Yogyakarta
- Tjalla, Awaludin. (2010). *UN dan Peningkatan Mutu Pembelajaran di Sekolah*. Diakses dari pustaka.ut.ac.id/pdfartikel/TIG602.pdf
- Tsay, Ruey S. (2005). *Analysis Of Financial Time Series*. Diakses dari down.cenet.org.cn/upfile/28/2008111495713151.pdf
- Warren. Aaron R. (2009). *Time-Series Analysis: Assessing the Effects of Multiple Educational Interventions in a Small-Enrollment Course*: Physics Education Research Conference Volume 1179, Pages 293-296, July 29-30, 2009
- Wei,W.W.S. (1994). *Time Series Univariate and Multivariate Methods*. Addison Wesley Publishing company, inc.