

PENCARIAN METODE ELIMINASI PENCILAN TERBAIK UNTUK MEMPERBAIKI KINERJA *FUZZY TIME SERIES* PADA PERAMALAN HARGA SAHAM BURSA EFEK INDONESIA

Akhmad Tajuddin Tholaby MS ¹⁾ Arif Djunaidy ²⁾

Fakultas Teknologi Informasi – Institut Teknologi Sepuluh November
Keputih – Sukolilo Surabaya ¹⁾²⁾
Email : udik_gileee1@yahoo.com ¹⁾

ABSTRAK

Metode peramalan Fuzzy Time Series (FTS) merupakan salah satu metode peramalan yang paling banyak digunakan terutama untuk mengolah data dengan tingkat variasi tinggi dan tidak linier terhadap waktu. Namun, peluang untuk mengembangkan dan memperbaiki kinerja peramalan dengan menggunakan metode FTS masih terbuka lebar. Kinerja suatu metode peramalan dapat dilihat dari distribusi *error*-nya. FTS tidak melihat dan tidak memilah jenis data *input* seperti apa yang akan diprosesnya menjadi suatu nilai peramalan, apakah data *input*-nya memiliki variasi kecil atau besar, atau memiliki data pencilan (*outlier*), semuanya di proses untuk menjadi nilai peramalan. Tentu saja jika data *input* memiliki pencilan, maka pencilan itu dapat merusak distribusi *error* sehingga menjadikan kinerja peramalannya menjadi tidak bagus.

Salah satu upaya untuk meningkatkan kinerja peramalan FTS adalah dengan melakukan filterisasi data-data pencilan. Dalam penelitian ini akan dilakukan identifikasi pencilan dengan metode Cook's Distance, leverage value, DfFiTs dan Boxplot untuk meningkatkan kinerja peramalan FTS yang diaplikasikan pada peramalan. Hasil penelitian akan dilakukan observasi perhitungan untuk menentukan metode identifikasi pencilan yang paling optimal meningkatkan kinerja peramalan FTS, yaitu yang paling kecil distribusi *error*-nya. Hasil penelitian menunjukkan bahwa metode yang terbaik untuk memperbaiki kinerja FTS adalah Metode Cook's Distance dengan nilai MSE dan MAPE yang paling kecil.

Kata kunci: perbaikan kinerja peramalan, identifikasi pencilan, FTS

ABSTRACT

Fuzzy time series (FTS) is one of forecasting methods that is mainly used to forecast data with high variation and relatively time invariant. FTS proceeds all input data, irrespective of data variance as well as outliers that may exist in input data. Distribution error of data can affect the FTS performance where the larger the error the worse the performance that may be obtained. Among some efforts that can be achieved to enhance the FTS performance is to filter the outliers prior to executing the forecast process. In this research, four methods for filtering the outliers are studied and compared; i.e. Cook's distance, leverage value, DFiTs, and boxplot. Analysis and comparison among these outlier filtering methods are performed in terms of the forecasting error that may be produced before and after employing each filter. In this regard, forecasting error is computed using mean square error (MSE) and mean absolute percentage error (MAPE). Experimental results showed that among four methods employed for filtering outliers, Cook's distance produces better forecasting errors among the other methods.

Keywords: forecasting performance enhancement, outlier, FTS

1. Pendahuluan

Setiap perusahaan yang menghasilkan produk baik dalam bentuk barang maupun jasa pasti menginginkan perusahaannya dapat berjalan dengan baik dan terus mengalami kemajuan. Untuk mencapai kondisi tersebut, maka semua perusahaan tanpa terkecuali secara berkelanjutan harus melakukan tiga siklus utama manajemen, yaitu perencanaan, pelaksanaan dan evaluasi yang nantinya hasil evaluasi yang dihasilkan akan menjadi *input* dasar perencanaan tahap berikutnya.

Salah satu wujud evaluasi tersebut adalah melakukan perhitungan peramalan permintaan akan produk yang dihasilkan oleh perusahaan. Ada banyak sekali metode peramalan yang dapat diaplikasikan. Salah satunya adalah metode peramalan Fuzzy Time Series (FTS). Metode peramalan FTS merupakan salah satu metode peramalan yang paling banyak digunakan terutama untuk mengolah data dengan tingkat variasi tinggi dan tidak linier terhadap waktu. Namun, peluang untuk mengembangkan dan memperbaiki kinerja peramalan dengan menggunakan metode FTS masih terbuka lebar. Kinerja suatu metode peramalan dapat dilihat dari distribusi *error*-nya. FTS tidak melihat dan tidak memilah jenis data *input* seperti apa yang akan diprosesnya menjadi suatu nilai peramalan, apakah data *input*-nya memiliki variasi kecil atau besar, atau memiliki data pencilan, semuanya diproses untuk menjadi nilai peramalan. Tentu saja jika data *input* memiliki pencilan, maka pencilan itu bisa merusak distribusi *error* sehingga menjadikan kinerja peramalannya menjadi tidak bagus.

Salah satu upaya untuk meningkatkan kinerja peramalan FTS adalah dengan melakukan filterisasi data-data pencilan. Pencilan adalah data yang tidak menggambarkan karakteristik data secara keseluruhan, atau dengan kata lain pencilan

adalah data yang nilainya jauh melenceng dari karakteristik data secara keseluruhan.

Salah satu penelitian yang melakukan hal tersebut adalah S. Sureh dan K. Senthamarai Kannan yang melakukan filterisasi dan mengidentifikasi pencilan dengan menggunakan metode Cook's Distance. Namun ternyata pencilan yang dihasilkan melenceng dari definisi pencilan itu sendiri. Sehingga sangat perlu dilakukan observasi dan penghitungan lebih dalam untuk memperbaiki kondisi tersebut.

Di dalam ilmu statistik, disebutkan bahwa ada beberapa metode identifikasi pencilan, salah satunya adalah metode Boxplot (R. Dawson, 2011). Metode ini dapat mengidentifikasi adanya pencilan yang dapat merusak distribusi *error* pada kinerja peramalan FTS. Metode ini menggunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, Interquartile Range) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q3 - Q1$. Data-data pencilan dapat ditentukan yaitu nilai yang kurang dari $1.5 \cdot IQR$ terhadap kuartil 1 dan nilai yang lebih dari $1.5 \cdot IQR$ terhadap kuartil 3. Salah satu penelitian yang menggunakan metode ini adalah R. Dawson (2011) dan Soemartini (2007) yang sama-sama melakukan filterisasi data pencilan pada data untuk mengetahui dampak keberadaan dan ketiadaan pencilan dalam analisis data (analisis regresi).

Selain metode Boxplot, dalam ilmu statistik ada juga metode identifikasi pencilan lainnya yaitu metode Cook's Distance, Leverage Value dan DfFITS. Adapun penjelasan dari masing-masing metode adalah sebagai berikut:

1. Metode DfFITS atau *Standardized DfFIT* menampilkan nilai perubahan dalam harga yang diprediksi bilamana *case* tertentu dikeluarkan, yang sudah distandarkan, di mana data dari hasil perhitungan DfFITS yang lebih besar dari $2 \cdot \sqrt{p/n}$ akan dihitung sebagai pencilan. Salah satu penelitian yang menggunakan metode ini adalah Mintu Kr.Das dan Bipin Gogoi (2015) yang

melakukan filterisasi pengukuran influential observations (pengamatan berpengaruh/pencilan) pada data yang diperkenalkan oleh Belsley, Kuh dan Welsch (1980), untuk menfilterisasi pencilan hasil dari observasi ke- i dari suatu peramalan.

2. Metode cook's Distance menampilkan nilai jarak cook. Nilai jarak cook yang lebih besar dari tabel distribusi $F(0.5; p, n-p)$ maka data tersebut adalah data pencilan. Salah satu penelitian yang menggunakan metode ini adalah Mintu Kr.Das dan Bipin Gogoi (2015) yang melakukan filterisasi pengukuran informasi *studentized residuals*, nilai varian untuk mendeteksi adanya pencilan pada data berdasarkan besar kecilnya jarak cook.
3. Metode Leverage Values menampilkan nilai leverage (pengaruh) terpusat, dimana data hasil perhitungan dari Leverage Values yang lebih besar dari $(2p-1)/n$ akan dihitung sebagai pencilan. Salah satu penelitian yang menggunakan metode ini adalah Soemartini (2007) yang melakukan Diagnostic Plot pada pencilan regresi dan Leverage Point. Untuk mengidentifikasi adanya pencilan berdasarkan ukuran leverage pada data.

Ket: n = Jumlah observasi (sampel); p = jumlah parameter.

Data yang akan disaring menggunakan tiap metode-metode eliminasi pencilan di atas nantinya akan dihitung pada peramalan FTS setelah itu akan dihitung distribusi *error-nya*.

Dari beberapa pembahasan di atas dapat ditarik permasalahan bagaimana cara mencari metode eliminasi pencilan terbaik untuk memperbaiki kinerja peramalan FTS. Dalam penelitian ini akan dilakukan perbaikan kinerja FTS melalui optimasi metode identifikasi pencilan pada peramalan permintaan. Optimasi metode identifikasi pencilan yang dimaksud adalah melakukan

perhitungan identifikasi pencilan dengan metode Boxplot, Cook's distance, Leverage value dan DfFITS, yang kemudian akan diobservasi apakah nilai pencilan yang dihasilkan sama atau tidak. Jika tidak sama, maka akan diobservasi yang mana diantara lima metode tersebut yang menghasilkan nilai peramalan yang terbaik.

2. Fuzzy Time Series

Fuzzy Time Series adalah sebuah konsep baru yang diusulkan oleh Song dan Chissom berdasarkan teori *Fuzzy Set* dan konsep variabel linguistik dan aplikasinya oleh Zadeh. *Fuzzy Time Series* digunakan untuk menyelesaikan masalah peramalan yang mana data historis adalah nilai-nilai linguistik. Misalnya, dalam masalah peramalan, data historis tidak dalam bentuk angka real, namun berupa data linguistik. Dalam hal ini, tidak ada model *time series* konvensional yang dapat diterapkan, akan tetapi model *Fuzzy Time Series* dapat diterapkan dengan lebih tepat.

Berdasarkan teori himpunan *Fuzzy*, logika *Fuzzy* dan penalaran perkiraan, Song dan Chissom mengajukan definisi *Fuzzy Time Series* dan garis besar pemodelan dengan cara persamaan relasional *Fuzzy* dan penalaran perkiraan. Kemudian oleh Chen (pada tahun 1996) diperkenalkan sebuah metode peramalan *Fuzzy Time Series* menggunakan operasi *Arithmetic*. Huarng (pada tahun 2001), menyajikan model *Heuristic* untuk peramalan *Time Series* menggunakan *Heuristic Increasing and Decreasing Relations* untuk memperbaiki peramalan *Enroll-ments* dan *Exchange* di Taiwan. Kemudian oleh Singh tahun 2007, diajukan algoritma komputasi sederhana, sehingga dapat mengurangi waktu untuk menghasilkan persamaan relasi dengan menggunakan operasi komposisi *max-min* yang kompleks dan mengurangi waktu untuk proses *defuzzifikasi* pada metode Song dan Chissom. Metode Singh dapat menyelesaikan masalah dalam mencari prosedur *defuzzifikasi* yang cocok untuk

menghasilkan nilai *output crisp* dengan akurasi yang lebih baik.

Perbedaan utama antara FTS dan *Traditional Time Series* adalah bahwa nilai dari FTS direpresentasikan dengan set *Fuzzy*. Misalkan U merupakan kelompok *universal* dengan $U = \{u_1, u_2, \dots, u_n\}$. maka set *Fuzzy* dinyatakan dalam U mengikuti representasi berikut ini:

$$A = f_A(u)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n$$

(2.1) Dengan f_A menyatakan fungsi keanggotaan dari set *Fuzzy* A $f_A : U \rightarrow [0,1]$ dan $f_A(u_i)$ menyatakan derajat keanggotaan u_i yang termasuk set *Fuzzy* A dan $f_A(u_i) \in [0,1]$

A u_i and $1 \leq i \leq n$.

Definisi 1. Asumsikan $Y(t) \subset R$ (*real line*), $t = \dots, 0, 1, 2, \dots, n$ menjadi bentuk universal yang dinyatakan dengan set *Fuzzy* $f_i(t)$. $F(t)$ yang mengandung $f_i(t)$, $i = 1, 2, \dots, n$ didefinisikan sebagai FTS pada $Y(t)$. Dalam hal ini, $F(t)$ dapat dipahami sebagai *linguistic variable*, dengan $f_i(t)$ $i = 1, 2, \dots, n$ adalah nilai-nilai *linguistic* yang memungkinkan dari $F(t)$ (Şah dan Degtiarev (2005); Chen dan Chung (2006)).

Definition 2. Jika $F(t)$ dipengaruhi oleh $F(t-1)$ dan dinyatakan dengan $F(t-1) \rightarrow F(t)$ maka hubungan ini dapat direpresentasikan dengan $F(t) = F(t-1) \circ R(t, t-1)$ dengan simbol “ \circ ” menyatakan simbol operator; $R(t, t-1)$ adalah hubungan *Fuzzy* antara $F(t)$ dan $F(t-1)$ dan disebut sebagai *first-order model* dari $F(t)$, (Chen dan Hsu (2004); Nasser et al. (2008)).

Definition 3. Dengan menyatakan $F(t-1)$ dengan A_i dan $F(t)$ dengan A_j maka hubungan antara $F(t-1)$ dan $F(t)$ dapat dinyatakan dengan *logical relationship* $A_i \rightarrow A_j$ (Şah dan Degtiarev (2005); Lotfi dan Firozja (2007)).

Definition 4. Misalkan $R(t; t-1)$ sebagai *first-order* model dari $F(t)$. Jika untuk nilai t sembarang, $R(t; t-1) = R(t-1; t-2)$, maka

$F(t)$ disebut *time-invariant* FTS. (Chen and Chung, 2006).

Definition 5. Hubungan logis *Fuzzy*, yang memiliki sisi ruas kiri yang sama dapat dikelompokkan bersama kedalam kelompok relasi logis *Fuzzy*. Contoh, untuk sisi ruas kiri yang identik A_i pengelompokan yang sama dapat diuraikan sebagai berikut (Şah dan Degtiarev, 2005):

3. Ukuran Kesalahan Peramalan

Karena permintaan dipengaruhi oleh banyak faktor, dan nilai di masa mendatang tidak dapat diketahui secara pasti sehingga tidak masuk akal jika ingin mendapatkan peramalan yang tepat setiap waktu. Perhitungan rata-rata kesalahan yang dibuat oleh model peramalan setiap waktu merupakan ukuran seberapa tepat peramalan. Dua pengukuran kesalahan yang sering dipakai adalah *Mean Absolute Percentage Error (MAPE)* dan *Mean Squared Error (MSE)*. Hal yang pertama digunakan karena sangat berguna untuk menentukan *tracking signal*. *MAD* merupakan nilai rata-rata *error* dalam peramalan menggunakan nilai *absolute*, Metode Peramalan, melalui tahapan berikut: (M. Şah and Y.K. Degtiarev, 2005)

Jika X_t merupakan data aktual untuk periode t dan F_t merupakan ramalan (*fitted value*) untuk periode yang sama, maka kesalahan (e_t) didefinisikan sebagai berikut:

$$e_t = D_{(t)} - F_{(t)}$$

Pertimbangan diterimanya sebuah metode peramalan adalah melalui kriteria berikut ini:

- a. Persentase kesalahan (*PE*)

$$PE = ((D_{(t)} - F_{(t)}) / D_{(t)}) \times 100\% \quad (2.3)$$
- b. *Mean Squared Error (MSE)*

$$MSE = ((D_1 - F_1)^2 + (D_2 - F_2)^2 + (D_3 - F_3)^2 + \dots + (D_m - F_m)^2) / m$$
- c. *Mean Absolute Percentage Error (MAPE)*

$$MAPE = MAPE = (PE_1 + PE_2 + PE_3 + \dots + PE_m) / m$$

$$PE_m = |e_m| / D_m \times 100\%$$

Di mana:

$D_{(t)}$ = Harga aktual pada periode t

$F_{(t)}$ = Harga peramalan pada periode t

m = Jumlah periode peramalan

4. Pencilan

Pencilan adalah suatu data yang menyimpang dari sekumpulan data yang lain (Ferguson, 1961). Pencilan adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data (Barnett, 1981). Jika terdapat masalah yang berkaitan dengan pencilan, maka diperlukan alat diagnosis yang dapat mengidentifikasi masalah pencilan, salah satunya dengan menyisihkan pencilan dari kelompok data kemudian menganalisis data tanpa pencilan (Weissberg, 1985). Terdapat beberapa metode untuk menentukan batasan pencilan dalam sebuah analisis, yaitu:

1. Boxplot

Metode ini merupakan yang paling umum yakni dengan mempergunakan nilai kuartil dan jangkauan. Kuartil 1, 2, dan 3 akan membagi sebuah urutan data menjadi empat bagian. Jangkauan (IQR, *Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q3 - Q1$. Data-data pencilan dapat ditentukan yaitu nilai yang kurang dari $1.5 \cdot IQR$ terhadap kuartil 1 dan nilai yang lebih dari $1.5 \cdot IQR$ terhadap kuartil 3 (R. Dawson, 2011).

2. Leverage Values, DfFITS, Cook's Distance

- Leverage Values*, menampilkan nilai *leverage* (pengaruh) terpusat (P.J. Rousseeuw, 1991).
- DfFITS* atau *Standardized DfFIT* yang menampilkan nilai perubahan dalam harga yang diprediksi bilamana *case* tertentu dikeluarkan, yang sudah distandarkan (Mintu Kr.Das dan Bipin Gogoi, 2015).

- Cook's Distance*, menampilkan nilai jarak Cook (Mintu Kr.Das dan Bipin Gogoi, 2015).
- DfBETA(s)*; menampilkan nilai perubahan koefisien regresi sebagai hasil perubahan yang disebabkan oleh pengeluaran *case* tertentu. Digunakan untuk mendeteksi pencilan pada variabel bebas (Dan Liao & Richard Valliant, 2012).
- Leverage point* memberikan informasi tentang apakah sebuah pengamatan dari data variabel **independen** (*predictor*) dikelompokkan sebagai data yang luar biasa (*ekstrim/outlier*) jika dibandingkan dengan data lainnya.
- Leverage* adalah sebuah ukuran jarak antara sebuah data dalam variabel **independen** dengan nilai rata-rata untuk semua pengamatannya.
- Sebuah data disebut sebagai *high leverage point* (HLP) jika menunjukkan nilai pengamatan yang sangat jauh dari rata-rata untuk semua pengamatan.
- Suatu data yang mengandung HLP dapat memberikan pengaruh yang besar terhadap garis regresi atau model regresi.

Rumus-rumus yang digunakan:

- H_i (**leverage**) diformulasikan sebagai:

$$b. H_i = 1/n + (x_i - \bar{x})^2 / ((n - 1)S_x)$$

$$c. R_{student_i} = e_i / (S_i(1 - H_i)^{1/2}),$$

$$\text{Di mana } S_i = (S_e^2 (n - p - R_{standarti}^2)) / (n - p - 1)$$

$$d. R_{standarti_i} = e_i / (MSE (1 - H_i)^{1/2}),$$

$$e. \text{Cook's Distance}_i = (R_{standarti}^2 / 2) \times (H_i / (1 - H_i))$$

$$f. DFITs_i = R_{student_i} (H_i / (1 - H_i))^{1/2}$$

Kriteria pengambilan keputusan adanya pencilan atau tidak :

Jika hasil perhitungan Leverage Value $> (2p - 1) / n \Rightarrow$ dianggap pencilan

Jika hasil perhitungan DfiTs $> 2 (p / n)^{1/2} \Rightarrow$ dianggap pencilan

Jika hasil perhitungan Cook's Distance $>$ nilai tabel $F(0.5; p, n - p) \Rightarrow$ dianggap pencilan

Jika hasil perhitungan DfETA $> 2 / n^{1/2} \Rightarrow$ dianggap pencilan

Ket. : n = Jumlah observasi (sampel); p = Jumlah parameter

5. Metodologi

Penelitian ini akan dibagi menjadi 2 uji coba, yaitu:

- Uji coba 1: Data awal yang digunakan untuk peramalan terdiri dari awal tahun 2011 sampai akhir 2013 untuk meramalkan periode tahun 2014. Dengan jumlah data sebanyak 737 data.
- Uji coba 2: Data awal yang digunakan untuk peramalan terdiri dari awal tahun 2011 sampai akhir 2014 untuk meramalkan periode tahun 2015. Dengan jumlah data sebanyak 978 data.

Pada semua uji coba, nilai kinerja FTS yang menggunakan filterisasi pencilan akan dibandingkan dengan kinerja FTS yang tanpa menggunakan filterisasi pencilan.

Tabel 1. Pengurutan data

No	Date	Close	Pengurutan A – Z
1	04/10/2011	667,95	569,457
2	05/10/2011	673,05	574,088
3	26/09/2011	677,72	574,153
4	22/09/2011	666,08	578,207
5	03/10/2011	642,54	584,218
6	24/01/2011	611,75	585,22
7	21/01/2011	605,37	590,236
8	10/02/2011	626,49	590,609

9	23/09/2011	629,01	592,718
10	11/02/2011	629,07	595,119
11	31/01/2011	622,29	597,854
12	09/02/2011	624,32	599,641
13	14/02/2011	621,74	600,168
14	07/10/2011	605,72	600,231

Filterisasi Boxplot

Untuk melakukan filterisasi Boxplot, langkah pertama adalah mengurutkan data secara alfabetis dari terkecil sampai terbesar. Perhatikan Tabel 1.

Setelah data diurutkan, langkah selanjutnya adalah mencari Q1 dan Q3. Dimana Q1 dan Q3 adalah nilai kuartil 1 dan kuartil 3 dari seluruh data. Adapun nilai Q1 dan Q3 pada uji coba 1 adalah 673,575 dan 746,8. Sedangkan nilai Q1 dan Q3 pada uji Coba 2 adalah 682,045 dan 814,445.

Setelah Q1 dan Q3 sudah diketahui, langkah selanjutnya adalah menentukan nilai IQR ($IQR = Q3 - Q1$).

Nilai data akan bernilai pencilan jika $> Q3 + 1,5 * IQR$ dan $< Q1 - 1,5 * IQR$

Filterisasi Leverage Value

Metode ini menampilkan nilai *leverage* (pengaruh) terpusat, dengan rumus:

$$H_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)S_x^2}$$

Hasil perhitungan leverage value akan bernilai pencilan jika nilai data hasil perhitungan $> (2p - 1) / n$

Di mana:

n = jumlah data; X_i = data ke- i ; \bar{X} = rata-rata data; S_x = standar deviasi X ;

p = jumlah parameter yang digunakan

Filterisasi pencilan DFITS

Metode ini merupakan metode filterisasi pencilan yang menampilkan nilai perubahan dalam harga yang diprediksi bilamana *case* tertentu dikeluarkan dan sudah distandarkan.

Pada sub Bab ini akan dijelaskan tahap-tahap perhitungan metode filterisasi pencilan DFITs untuk memperbaiki kinerja peramalan FTS pada BEI. Adapun rumus DFITs yang digunakan untuk menghitung filterisasi pencilan adalah sebagai berikut:

$$R_{student_i} = \frac{e_i}{S_i \sqrt{1-H_i}}$$

Di mana $S_i = \frac{S_e^2(n-p-R_{Standard_i}^2)}{(n-p-1)}$;

$$R_{standard_i} = \frac{e_i}{MSE \sqrt{1-H_i}}$$

$$DFIT_{S_i} = R_{student_i} \sqrt{\frac{H_i}{1-H_i}}$$

Hasil perhitungan DFITs akan bernilai pencilan jika nilai hasil perhitungan $> 2 * \sqrt{p/n}$

Di mana: H_i = nilai leverage value data ke- i ;
 p = jumlah parameter; n = jumlah data; e = $Y - Y_{taksiran}$; Se = standart deviasi e

Filterisasi Cook’s Distance

Metode *cook’s distance* adalah metode filterisasi pencilan yang menampilkan nilai jarak cook. Nilai jarak *cook* yang lebih besar dari tabel distribusi $F(0.5; p, n-p)$ maka data tersebut adalah data pencilan. Pada sub bab ini akan dijelaskan tahap-tahap perhitungan metode filterisasi pencilan *Cook’s Distance* untuk memperbaiki kinerja peramalan FTS pada BEI. Adapun rumus *Cook’s Distance* yang digunakan untuk menghitung pencilan adalah sebagai berikut:

$$R_{student_i} = \frac{e_i}{S_i \sqrt{1-H_i}}$$

Di mana $S_i = \frac{S_e^2(n-p-R_{Standard_i}^2)}{(n-p-1)}$;

$$R_{standard_i} = \frac{e_i}{MSE \sqrt{1-H_i}}$$

$$Cook’s\ distance_i = \left[\frac{R_{Standard_i}^2}{2} \right]_x \left[\frac{H_i}{1-H_i} \right]$$

Hasil perhitungan Cook’s Distance akan bernilai pencilan jika nilai hasil perhitungan $> F(0.5; p, n-p)$ (dapat dilihat pada tabel F)

Di mana:
 H_i = nilai leverage value data ke- i ;
 p = jumlah parameter;
 n = jumlah data;
 e = $Y - Y_{taksiran}$;
 Se = standart deviasi e

Setelah dilakukan filterisasi, langkah selanjutnya dilakukan perhitungan peramalan FTS. Dan setelah dilakukan peramalan FTS, langkah selanjutnya adalah menghitung nilai MSE dan MAPE pada semua uji coba. Untuk mengukur tingkat akurasi kinerja peramalan, maka dihitung nilai *Mean Square Error* (MSE) dan *Mean Absolute Percentage Error* (MAPE).

6. Hasil dan Analisa

Setelah melakukan perhitungan perhitungan filterisasi seperti yang sudah dibahas diatas, maka didapatkan nilai *error* (MSE dan MAPE) sebagai berikut:

Tabel 2. Perbandingan hasil nilai *error* dari kinerja FTS dengan dan tanpa filterisasi pencilan periode tahun 2011-2013

No	Date	Close	Pengurutan A – Z
1	04/10/2011	667,95	569,457
2	05/10/2011	673,05	574,088
3	26/09/2011	677,72	574,153
4	22/09/2011	666,08	578,207
5	03/10/2011	642,54	584,218
6	24/01/2011	611,75	585,22
7	21/01/2011	605,37	590,236
8	10/02/2011	626,49	590,609
9	23/09/2011	629,01	592,718
10	11/02/2011	629,07	595,119
11	31/01/2011	622,29	597,854
12	09/02/2011	624,32	599,641
13	14/02/2011	621,74	600,168

14 | 07/10/2011 | 605,72 | 600,231

Tabel 3. Perbandingan hasil nilai *error* dari kinerja FTS dengan dan tanpa filterisasi pencilan periode tahun 2011-2014

No	Date	Close	Pengurutan A – Z
1	04/10/2011	667,95	569,457
2	05/10/2011	673,05	574,088
3	26/09/2011	677,72	574,153
4	22/09/2011	666,08	578,207
5	03/10/2011	642,54	584,218
6	24/01/2011	611,75	585,22
7	21/01/2011	605,37	590,236
8	10/02/2011	626,49	590,609
9	23/09/2011	629,01	592,718
10	11/02/2011	629,07	595,119
11	31/01/2011	622,29	597,854
12	09/02/2011	624,32	599,641
13	14/02/2011	621,74	600,168
14	07/10/2011	605,72	600,231

Setelah melihat nilai *error* (MSE dan MAPE) dan grafik perbandingan kinerja FTS dari hasil perhitungan yang sudah dilakukan pada seluruh uji coba, maka dapat ditarik hasil penelitian sebagai berikut:

- Dari segi analisa grafik: seluruh kinerja FTS yang menggunakan metode filterisasi pencilan pada uji coba 1 dan 2 memiliki alur grafik yang lebih halus jika dibandingkan dengan kinerja FTS yang tanpa menggunakan filterisasi pencilan.
- Metode Boxplot tidak cocok untuk data yang sangat banyak terutama mengenai harga saham, karena kemungkinan munculnya pencilan sangat kecil atau bahkan tidak muncul sama sekali. Boxplot cocok jika digunakan untuk jumlah data yang sedikit dan diaplikasikan untuk peramalan jangka pendek.
- Metode *Leverage Value* tidak cocok untuk data yang memiliki perubahan

yang sangat besar, seperti halnya pada Gambar 4.23 yang nilai varians-nya tinggi dibagian ahir peramalan.

- Metode filterisasi DFiTs bagus digunakan untuk data dengan jumlah yang banyak, variasi yang kecil dan menampilkan nilai perubahan dalam harga (untuk tiap naik-turun yang ekstrim) yang diprediksi bilamana *case* tertentu dikeluarkan dan sudah distandarkan.
- Metode filterisasi *Cook's Distance* bagus digunakan untuk menfilter data yang sangat banyak dan digunakan untuk peramalan dengan periode jangka panjang seperti halnya pada Gambar 4.25.
- Dari segi perbandingan MSE dan MAPE:
 - a. pada uji coba 1: nilai MSE dan MAPE paling kecil adalah DFiTs
 - b. pada uji coba 2: nilai MSE dan MAPE paling kecil adalah *Cook's Distance*.

Tabel posisi pergerakan harga saham ditinjau dari data awal X_1 sampai X_{terahir} :

Tabel 4. Posisi pergerakan harga saham pada uji coba 1

No	Nama Filterisasi	Uji Coba 1 (2011 - 2013)			
		X1	X737	X'737	X738
1	Boxplot	667,951	711,135	710,7463	723,516
2	Leverage Value	667,951	711,135	710,8092	723,516
3	DfiTs	667,951	711,135	715,2611	723,516
4	Cook's Distance	667,951	711,135	684,8269	723,516

Tabel 5 Posisi pergerakan harga saham pada uji coba 2

No.	Nama Filterisasi	Uji Coba 1 (2011 - 2014)			
		X1	X978	X''978	X979
1	Boxplot	667,95	898,581	8815377	903,130
2	Leverage Value	667,95	898,581	742,3040	903,130
3	DfiTs	667,95	898,581	883,2946	903,130
4	Cook's	667,95	898,581	837,7753	903,130

| Distance | | | |

Keterangan:

X1 = data ke-1;

X737 = data ke-737;

X978 = data ke-978;

X'737 = data peramalan ke-737;

X'978 = data peramalan ke-978

Hasil penelitian menunjukkan bahwa pergerakan nilai saham dari data awal (X1) sampai data ahir (X737) memiliki kecenderungan naik. Nilai peramalan pada periode berikutnya (X'737) ternyata juga mengalami kenaikan relatif terhadap X1. Nilai periode berikutnya yaitu X738 juga naik. Hal ini menunjukkan bahwa sebenarnya hasil peramalan yang muncul sangat sesuai dengan tren kecenderungan kenaikan harga saham tersebut. Sehingga walaupun ada nilai *error* peramalan yang relatif kecil terhadap nilai sebenarnya, maka itu menjadi tidak masalah selama dalam jalur tren yang benar. Hal yang sama juga terjadi pada uji coba 2.

7. Kesimpulan

Setelah melakukan penelitian yang sudah dibahas pada pembahasan sebelumnya, maka diperoleh kesimpulan sebagai berikut:

- Grafik hasil filterisasi pencilan memiliki fitur dengan tingkat variasi yang lebih halus dari pada tanpa filterisasi pencilan.
- Metode filterisasi Boxplot tidak cocok jika digunakan untuk mengolah data yang relatif sangat banyak dan cocoknya untuk mengolah data yang sedikit, dengan kata lain cocok digunakan untuk mendukung peramalan dalam jangka pendek atau sistem *trading*.
- Metode DFITs dan *Cook's Distance* cocok digunakan untuk mendukung sistem peramalan yang melibatkan jumlah data yang relatif sangat banyak atau untuk mendukung sistem peramalan jangka panjang atau sistem investasi, sehingga walaupun hasil peramalan yang dihasilkan tetap menghasilkan *error* yang

relatif kecil, tetap saja hasilnya baik, karena posisi nilai peramalan dan nilai yang muncul adalah sama-sama mengalami kenaikan dibandingkan dengan data awal X1.

DAFTAR PUSTAKA

- [1] Arumugam, P., & Anithakumari, V. (2013). "Fuzzy Time Series Method for Forecasting Taiwan Export Data". International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue8- August 2013. India: Manonmaniam Sundaranar University.
- [2] Talal Jasim, H.; Ghafoor Jasim Salim, Abd.; & Ismail Ibraheem, K (1989). "A Novel Algorithm to Forecast Enrollment Based on Fuzzy Time Series". Applications and Applied Mathematics: An International Journal (AAM). ISSN: 1932-9466. Vol. 7, Issue 1 (June 2012), pp. 385 – 397. Iraq: Mosul University
- [3] Yolcu, Ufuk (2012). *The Forecasting of Istanbul Stock Market with a High Order Multivariate Fuzzy Time Series Forecasting Model*. TJFS: Turkish Journal of Fuzzy Systems (eISSN: 1309–1190). An Official Journal of Turkish Fuzzy Systems Association. Vol.3, No.2, pp. 118-135, 2012. Turkey: University of Giresun, Faculty of Arts and Science, Department of Statistics
- [4] H Amjad, Usman.; A. Jilani, Tahseen., & Yasmeen, Farah. (2012). *A Two Phase Algorithm for Fuzzy Time Series Forecasting using Genetic Algorithm and Particle Swarm Optimization Techniques*. International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, October 2012. University of Karachi
- [5] Alpaslan, Faruk.; Cagcag, Ozge.; Aladag, C.H.; Yolcu, U.; and Egrioglu, E.. (2012). "A Novel Seasonal Fuzzy Time Series Method" Hacettepe Journal

- of Mathematics and Statistics Volume 41 (3) (2012), 375 – 385. Turkey: Statistics Department, Ondokuz Mayıs University, Samsun.
- [6] Suresh, S., & Senthamarai Kannan, K. (2011). " *Identifying Outliers in Fuzzy Time Series*". Journal of Modern Applied Statistical Methods. India: Manonmaniam Sundaranar University, Tirunelveli
- [7] Pratama, Jaka.; Awallun, NK.; & Delyana, S (2012). " Analisis Residual dengan *Leverage Values*, *DfFITS*, dan *Cook's Distance*". Jakarta: Sekolah Tinggi Ilmu Statistik Jakarta hal: 3-10.
- [8] Şah, M. and Degtiarev, Y.K. (January, 2005). *Forecasting Enrollment Model Based on First-Order Fuzzy Time Series*, *World Academy of Science, Engineering and Technology*. Vol. 1, pp. 375-378, *The Board of Trustees The University of Alabama*.
- [9] Chen, S. M. and Hsu, C. C. (2004). *A New Method to Forecast Enrollments Using Fuzzy Time Series*, *International Journal of Applied Science and Engineering*, No. 2, Vol. 3, pp. 234-244.
- [10] Allahviranloo, T., Lotfi, F. M. and Firozja, A. (December 2007). *Fuzzy Efficiency Measure with Fuzzy Production Possibility Set*, *AAM*, Vol. 2, No. 2, pp. 152 – 166.
- [11] Chen, S. M. and Chung, N. Yi (2006). *Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms*, *Information and Management Sciences*, No.3, Vol.17, pp. 1-17.
- [12] Rousseeuw, P.J. (1991), " *Diagnostic Plot for Regression Outlier and Leverage Point*," *Statistical Software Newsletter*, 127-129
- [13] Imon, Rahmatullah. 2007, "Robust Regression," Postgraduate Lecture Series 3. Institute of Mathematical Sciences University of Malaya.
- [14] Atkinson A.C (1981). Two Graphical Display for Outlying and Influential Observation in Regression. *Technometrics*
- [15] Kr.Das, Mintu & Gogoi, Bipin (2015), " *Usage of Graphical Displays to Detect Outlying Observations in Linear Regression*". *Indian Journal of Applied Research*. Volume: 5, Issue: 5, May 2015, ISSN - 2249-555X
- [16] Dawson, R. (2011), " *How Significant Is A Boxplot Outlier?*". *Journal of Statistics Education*, Volume 19, Number 2(2011)
- [17] Dan Liao, Dan & Valliant, Richard (2012). " *Variance inflation factors in the analysis of complex survey data*". *Component of Statistics Canada Catalogue no. 12-001-X Business Survey Methods Division*
- [18] Soemartini (2007), "Pencilan (outlier)" Universitas Padjajaran, Jatinangor