

Penerapan Model Forecasting Arima Dan Artificial Neural Network Di Kabupaten Sragen

Hendra Setiawan^{1*}, Ema Utami¹, Hanif Al Fatta¹

¹Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Condong Catur, Sleman,
Yogyakarta, Indonesia

* Corresponding Author. E-mail : hendra.gooaccount@gmail.com. Telp : +6287835115884

Abstrak

Jumlah penderita DBD di kabupaten Sragen beberapa tahun terakhir terus meningkat yang kemudian diperparah dengan perubahan musim. Kurangnya kesadaran masyarakat pada permasalahan DBD ini disebabkan ketidakpedulian dan ketidaktahuan masyarakat pada bahaya DBD dan penyebarannya. Berdasarkan penelitian yang telah ada, pola persebaran DBD juga dipengaruhi oleh factor lingkungan seperti pola pergerakan penduduk yang dipengaruhi oleh seberapa baik sistem dan sarana transportasi yang ada. Berdasarkan fakta ini, penulis berusaha menerapkan beberapa metode peramalan jumlah penderita DBD untuk melihat adakah pengaruh yang nyata antara seberapa baik system transportasi dengan jumlah penderita DBD dan meningkatkan tingkat kesadaran masyarakat dan mempermudah pihak-pihak yang membutuhkan data penyebaran penderita DBD di Kabupaten Sragen pada periode-periode yang dibutuhkan. Model-model yang digunakan pada penelitian adalah model ARIMA, Backpropagation, dan Estimasi Titik. Proses prediksi dilakukan dengan membagi proses peramalan menjadi dua berdasarkan dataset yang digunakan yaitu dataset mulai tahun 2013 sampai tahun 2106 untuk data yang masih lengkap dan dataset mulai tahun 2010 sampai tahun 2016 untuk dataset yang tidak lengkap karena data yang hilang atau data yang rusak. Data-data yang digunakan sebagai dataset adalah data-data yang diambil dari data jumlah penderita di kabupaten Sragen. Proses peramalan pada penelitian ini dilakukan dengan melakukan proses peramalan dengan metode ARIMA dan LSTM pada dua dataset yang telah dipersiapkan.

Kata Kunci: Peramalan, DBD, Transportasi, ARIMA, LSTM, Estimasi Titik

Abstract

The number of DHF patients in Sragen regency in the last few years continued to increase which was then compounded by seasonal changes. The lack of public awareness of the problem of DHF was caused by ignorance of the public on the dangers of DHF and its spread. Based on existing research, DHF distribution patterns were also influenced by environmental factors such as population movement patterns that were influenced by how well the existing transportation systems and facilities. Based on this fact, the author tries to apply several methods of forecasting the number of DHF sufferers to see whether there is a real influence between how good the transportation system is with the number

of DHF sufferers and increase the level of public awareness and facilitate parties who need data on the spread of DHF sufferers in Sragen Regency in the period of required period. The models used in this study are ARIMA, Backpropagation, and Point Estimation models. The prediction process was done by dividing the forecasting process into two based on the dataset used, namely the dataset from 2013 to 2106 for complete data and the dataset from 2010 to 2016 for incomplete datasets due to missing or damaged data. The data used as a dataset were data taken from data on the number of patients in Sragen regency. The forecasting process in this study was carried out by forecasting using the ARIMA and LSTM methods on two prepared datasets.

Keywords: forecasting, DHF, Transportation, ARIMA, LSTM, Point Estimation

PENDAHULUAN

Penyakit Demam Berdarah *Dengue* adalah salah satu penyakit dengan tingkat morbiditas dan mortalitas tinggi yang disebabkan oleh virus *Dengue* yang ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan *Aedes Albopictus*. (Kusuma dan Sukendra, 2016). Pola sebaran kasus DBD juga sangat dipengaruhi oleh pola pergerakan dan kepadatan penduduk, iklim dan ketinggian tempat. (Muliansyah dan Baskoro, 2016). Pada penelitian ini iklim dan ketinggian tempat dapat diabaikan karena ketinggian tempat dan iklim di kabupaten Sragen relative sama. Lain halnya dengan pergerakan dan kepadatan penduduk dengan indikasi jumlah sarana dan prasarana transportasi. Penelitian ini salah satunya dilakukan untuk melihat apakah ada korelasi antara jumlah sarana transportasi dengan jumlah penderita DBD di Kabupaten Sragen. Jumlah kasus Demam Berdarah *Dengue* (DBD) di Indonesia tergolong tinggi. Hingga tahun 2013, jumlah penderita DBD yang ditemukan sebanyak 112.511 kasus dengan jumlah kematian 871 orang (*Incidence Rate*/Angka kesakitan = 45,85 per 100.000 penduduk dan *CFR*/angka kematian = 0,77%). Sementara pada Tahun 2014 di provinsi Jawa Tengah, ditemukan beberapa fakta bahwa jumlah kasus DBD sampai bulan September ditemukan 7.928 kasus dan angka kematian sebesar 128 orang dengan *Incidence Rate* (IR) sebesar 23,82 dan *Case Fatality Rate* (CFR) sebesar 1,61%. (Sari dan Cahyati, 2016).

Berdasarkan data yang diperoleh dari Dinas Kesehatan Kabupaten (DKK) di kabupaten Sragen, jumlah kasus Demam Berdarah *Dengue* di kabupaten Sragen ditemukan relatif tinggi dari tahun ke tahun. Pada tahun 2010 jumlah penderita DBD tercatat sebanyak 664 kasus, tahun 2011 hingga akhir bulan Agustus sebanyak 155 kasus, tahun 2012 hingga akhir bulan Juni sebanyak 130 kasus, tahun 2013 sebanyak 389 kasus, tahun 2014 sebanyak 575 kasus, tahun 2015 sebanyak 527 kasus dan tahun 2016 hingga akhir bulan Juli sebanyak 575 kasus. Oleh karena jumlah penderita penyakit demam berdarah *dengue* yang relatif tinggi dari waktu ke waktu, maka perlu dilakukan peramalan atau *forecasting* data penderita demam berdarah *dengue* (DBD) di Kabupaten Sragen ini agar pemerintah dapat mengambil tindakan antisipatif dengan tepat.

Salah satu metode yang dapat digunakan untuk melakukan proses peramalan atau *forecasting* data adalah metode *time series* atau metode runtun waktu. Metode peramalan *time series* dikelompokkan menjadi dua metode yaitu metode statistic seperti pemulusan eksponensial dan *Auto Regressive Integrated Moving Average* (ARIMA) dan metode kecerdasan buatan seperti *Artificial Neural Network* (ANN) dan Algoritma Genetika. Beberapa contoh metode-metode dasar pada peramalan data *time series* ini adalah: (Octavia, Yulia dan Lydia, 2015). Metode ARIMA sering digunakan sebagai metode standar untuk melakukan forecasting data runtun waktu linear. (Janah, Sulandari dan Wiyono, 2014). Akan tetapi data-data pada kehidupan nyata tidak hanya terdiri dari data dengan hubungan linear tetapi juga nonlinear sehingga metode ARIMA tidak baik diteapkan pada data dengan pola campuran ini. (Wiyanti dan Pulungan, 2012). Untuk mengatasi permasalahan ini, maka digunakan juga metode *Artificial Neural Network* (ANN) dengan teknik *Backpropagation* yang dapat digunakan untuk melakukan proses peramalan data hubungan nonlinear. (Janah, Sulandari dan Wiyono, 2014).

Walaupun metode yang akan digunakan untuk melakukan forecasting telah ada, namun sering kali data set yang digunakan hilang sebagian atau rusak seperti data jumlah penderita demam berdarah *dengue* (DBD) yang hilang mulai dari bulan September 2011 hingga bulan Desember 2011 dan mulai bulan Juli 2012 hingga bulan Desember 2012. Kurangnya data ini tentu mengganggu tingkat akurasi hasil peramalan data. Untuk menangani masalah ini, maka diperlukan proses estimasi data. Estimasi data adalah proses perkiraan atau taksiran data lampau yang menghasilkan data keluaran atau output berupa angka atau numerik, bukan kategori, yang didasarkan pada data-data yang telah ada. Terdapat beberapa metode proses estimasi ini, yaitu *point estimation* atau estimasi titik dan *confidence interval estimations* atau estimasi interval kepercayaan. (Larose, 2005). Estimasi Titik dilakukan untuk mendapatkan data pengganti dari data yang hilang sedangkan estimasi interval digunakan untuk menentukan seberapa valid data yang diestimasi karena data yang hilang tidak bisa diganti secara sembarangan sehingga menghasilkan *noise* data.

Berdasarkan uraian di atas, penulis tertarik untuk mengkaji dan menganalisis ulang penggunaan kombinasi metode *Auto Regressive Integrated Moving Average* (ARIMA) dan *Artificial Neural Network* (ANN) dengan Estimasi Titik dan Interval untuk meningkatkan akurasi prediksi penderita demam berdarah *dengue* (DBD) di kabupaten Sragen.

KAJIAN LITERATUR

Terdapat beberapa penelitian sebelumnya yang penulis gunakan sebagai referensi. Penelitian penelitian ini adalah sebagai berikut. A Comparison of Nineteen Various Electricity Consumption Forecasting Approaches and Practicing to Five Different Households in Turkey yang ditulis oleh Benli T.O pada *arXiv preprint arXiv:1607.05660* di tahun 2016. Tujuan penelitian ini adalah untuk menentukan model terbaik dari 19 model data *forecasting* dengan mengambil data konsumsi listrik di Turki pada bulan Juni 2016 sebagai output target dan menyusun model lain berdasarkan model yang telah ada untuk mendapatkan model *forecasting* yang terbaik. Penelitian ini berhasil melakukan perbandingan

semua model dan berhasil membuat model baru yang diklaim memiliki tingkat akurasi paling baik, namun hanya berlaku untuk data bulan Juni 2016. Perbedaan antara penelitian ini dengan penelitian yang akan dilakukan penulis adalah objek dan tujuan penelitian, penelitian ini menggunakan data konsumsi listrik di Turki sebagai objek penelitiannya dan metode yang digunakan adalah komparasi sembilan belas model *forecasting* untuk menyusun model baru sedangkan penelitian yang penulis teliti menggabungkan model *forecasting* dengan estimasi data untuk mendapatkan data *forecasting* dengan tingkat akurasi yang diharapkan lebih baik.

Penelitian kedua adalah penelitian yang berjudul *A framework for predictive analysis of stock market indices—a study of the Indian auto sector* yang ditulis oleh Jaydip Sen pada *Calcutta Business School (CBS) Journal of Management Practices 2.2: 1-19* di tahun 2016. Penelitian ini bertujuan untuk menyusun lima metode *forecasting* baru yang dapat mengolah data dengan tiga spesifik atribut yaitu musim, tren dan komponen acak dan mengelompokkan data prediksi berdasarkan tren, musim dan komponen acaknya untuk mempermudah proses pemahaman data apakah data modal yang diprediksi ini baik untuk pembelian jangka pendek atau jangka panjang. Penelitian ini berhasil merumuskan lima pendekatan yang dapat digunakan untuk melakukan proses *forecasting* data pada kondisi yang berbeda-beda tergantung komponen acak yang ditambahkan pada data set. Perbedaan antara penelitian ini dengan penelitian yang akan dilakukan adalah pada objek dan tujuan penelitian, penelitian ini bertujuan merumuskan sebuah pendekatan yang menggunakan lima model *forecasting* data dengan memperhatikan tiga unsur utama yaitu tren, musim dan unsur acak untuk melakukan prediksi data pasar modal sedangkan penelitian yang penulis teliti bertujuan untuk meningkatkan akurasi prediksi data penderita demam berdarah menggunakan kombinasi *Auto Regressive Integrated Moving Average (ARIMA)* dan *Artificial Neural Network (ANN)* dan Estimasi Titik dan Interval.

Penelitian ketiga adalah penelitian yang ditulis oleh Tamal Datta Chaudhuri dan Indranil Ghosh pada *arXiv preprint arXiv:1607.02093* di tahun 2016 dengan judul *Artificial Neural Network and Time series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework*. Tujuan dari penelitian ini adalah untuk mengusulkan pendekatan prediksi data pertukaran mata uang Rupee dengan Dolar Amerika menggunakan kombinasi metode *Multilayer Feed Forward Neural Network (MLFFNN)* dan *Nonlinear Autoregressive, Exogenous Input (NARX) Neural Network* dengan model *Generalized Autoregressive Conditional Heteroskedastic (GARCH)* dan *Exponential Generalized Autoregressive Conditional Heteroskedastic (EGARCH)*. Hasil dari penelitian ini adalah mendapatkan model terbaik dari keempat model yang digunakan, yaitu model MLFFNN dan NARX. Perbedaan utama penelitian ini dengan penelitian yang akan dilakukan adalah metode digunakan pada proses *forecasting* di penelitian ini adalah kombinasi metode *Multilayer Feed Forward Neural Network (MLFFNN)* dan *Nonlinear Autoregressive, Exogenous Input (NARX) Neural Network* dengan model *Generalized Autoregressive Conditional Heteroskedastic (GARCH)* dan *Exponential Generalized Autoregressive Conditional Heteroskedastic (EGARCH)*

untuk melakukan prediksi pertukaran Rupee dengan Dolar Amerika sedangkan metode yang diusulkan pada penelitian yang penulis teliti adalah metode kombinasi *Auto Regressive Integrated Moving Average* (ARIMA) Dan *Artificial Neural Network* (ANN) dengan Estimasi Titik dan Interval untuk meningkatkan akurasi data prediksi penderita penyakit demam berdarah.

Penelitian selanjutnya adalah penelitian yang berjudul *A Time-varying Parameter Based Seasonally-adjusted Bayesian State-space Model for Forecasting* yang ditulis pada tahun 2015 oleh Arnab Hazra pada *arXiv preprint arXiv:1512.02149*. Penelitian ini bertujuan untuk mendapatkan model *forecasting* data dengan mengoptimasi teknik *Seasonally-adjusted Bayesian State-space* menggunakan parameter variasi waktu. Penelitian ini berhasil menyusun model *forecasting* dengan menghilangkan komponen musim dan tren dan bebas dari ketergantungan waktu. Perbedaan penelitian ini dengan penelitian yang akan dilakukan penulis adalah pada metode *forecasting* yang digunakan yaitu penelitian menggunakan metode *Seasonally-adjusted Bayesian State-space* yang dioptimasi dengan penambahan parameter variasi waktu sedangkan metode yang diusulkan oleh peneliti adalah model kombinasi *Auto Regressive Integrated Moving Average* (ARIMA) dan *Artificial Neural Network* (ANN) Dengan Estimasi Titik dan Interval.

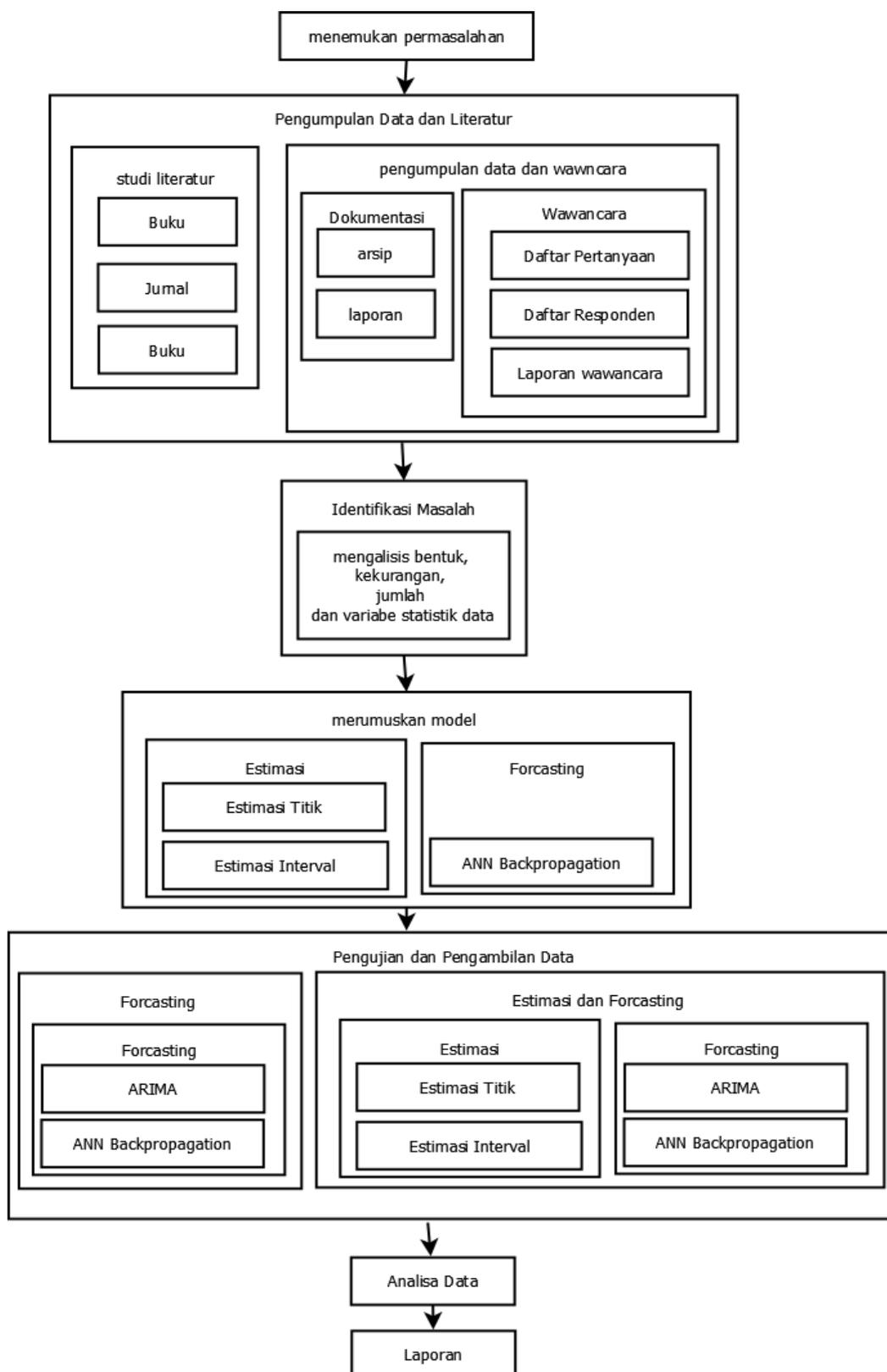
Penelitian kelima adalah penelitian yang dilakukan oleh Lutz F. Gruber dan Mike West pada tahun 2014. Penelitian yang berjudul *Bayesian Forecasting And Scalable Multivariate Volatility Analysis Using Simultaneous Graphical Dynamic Models* yang dipublikasikan pada *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. Tujuan penelitian ini adalah untuk menganalisis model *Forecasting* Bayesian dan volatilitas multivariat terskala menggunakan pendekatan *Simultaneous Graphical Dynamic Models* (SGDLMs). Penelitian ini berhasil membuktikan bahwa pendekatan *Simultaneous Graphical Dynamic Models* (SGDLMs) dapat dan baik digunakan untuk melakukan proses analisis dan *forecasting* volatilitas dan co-volatilitas data menggunakan model *forecasting* Bayesian. Perbedaan penelitian ini dengan penelitian yang dilakukan penulis adalah penelitian ini bertujuan untuk menganalisis model *Forecasting* Bayesian dan volatilitas multivariat terskala menggunakan pendekatan *Simultaneous Graphical Dynamic Models* (SGDLMs) sedangkan penelitian yang diusulkan bertujuan untuk meningkatkan akurasi prediksi data penderita DBD dengan menggabungkan algoritma *Forecasting* ARIMA dan ANN *Backpropagation* dengan Estimasi Interval dan Titik.

Penelitian terakhir adalah penelitian yang berjudul *Bayesian nonparametric forecasting of monotonic functional time series* yang ditulis pada tahun 2016 oleh Antonio Canale dan Matteo Ruggiero yang dipublikasikan pada *Electronic Journal of Statistics 10.2: 3265-3286*. Tujuan penelitian ini adalah untuk mengusulkan model nonparametrik Bayesian untuk memodelkan dan memprediksi sebuah kelas dari *time series* fungsional. Kesimpulan yang didapatkan pada penelitian ini adalah model Bayesian nonparametric yang diterapkan menghasilkan prediksi data yang baik pada level fungsional data dan level variabel data tunggal sehingga menjadikannya instrument atau tool yang dapat diandalkan oleh pelaku pasar untuk melakukan prediksi harga pasar. Perbedaan

penelitian ini dengan penelitian yang diusulkan terletak pada metode dan objek penelitiannya. Penelitian ini menggunakan metode prediksi Nonparametrik Bayesian dan objek penelitian yang diteliti adalah harga pasar dari komoditas ekonomi seperti gas alam di Eropa khususnya di Italia, sedangkan penelitian yang diusulkan menggunakan metode kombinasi *Auto Regressive Integrated Moving Average* (ARIMA) dan *Artificial Neural Network* (ANN) dengan Estimasi Titik dan Interval sedang objek penelitian yang diusulkan adalah data penderita DBD di kabupaten Sragen Indonesia dari tahun 2010 hingga 2015.

METODE PENELITIAN

Metode penelitian yang diterapkan pada penelitian ini adalah penelitian riset eksperimental. Metode riset eksperimental adalah sebuah metode dengan melakukan suatu tindakan terhadap seseorang/objek dan mengobservasi reaksinya dalam kondisi di mana performanya dapat diukur (Hasibuan, 2007). Penelitian ini dilakukan dengan membagi penelitian berdasarkan dataset penelitian. Pembagian dataset ini dilakukan karena terdapat data hilang atau data rusak pada data sumber penelitian. Dari dataset yang tidak lengkap ini kemudian dipotong dan diambil data runtun waktu yang masih utuh untuk dibuat dataset sebagai dataset pertama dan dataset kedua didapatkan dengan melengkapi atau mengganti data yang rusak dengan metode estimasi titik. Kedua dataset ini kemudian dijadikan sebagai dataset prediksi ARIMA dan Backpropagation. Hasil prediksi dari keempat percobaan ini kemudian akan dibandingkan tingkat keakurasiannya dengan melihat kedekatan data hasil prediksi dengan data sebenarnya untuk mendapatkan model prediksi terbaik. Proses pengumpulan data pada penelitian ini dilakukan dengan metode wawancara dan studi literatur. Metode wawancara ini adalah suatu metode pengumpulan data dengan melakukan mengajukan pertanyaan-pertanyaan atau melakukan tanya jawab dengan pihak terkait dan berkompeten yang dalam penelitian ini adalah Dinas Kesehatan Kabupaten (DKK) Sragen. Metode wawancara ini dilakukan untuk mengetahui gambaran jumlah penderita demam berdarah di kabupaten Sragen dari tahun 2010 hingga 2015. Studi literatur yang dilakukan oleh peneliti pada penelitian ini adalah yaitu studi kepustakaan dengan media seperti : jurnal, buku, prosiding dan e-book. Alur penelitian ini dijelaskan melalui Gambar 1 sebagai berikut:



Gambar 1. Alur Penelitian

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan membuat sistem prototype yang berfungsi untuk memudahkan perhitungan prediksi dengan model ARIMA dan Backpropagation. Sistem ini dibagi menjadi beberapa bagian seperti proses forecasting untuk data yang berkelanjutan atau tidak ada data hilang dan forecasting untuk dataset yang terdapat kerusakan atau kehilangan data. kedua proses ini kemudian akan penulis namakan dengan proses 1 dan proses 2.

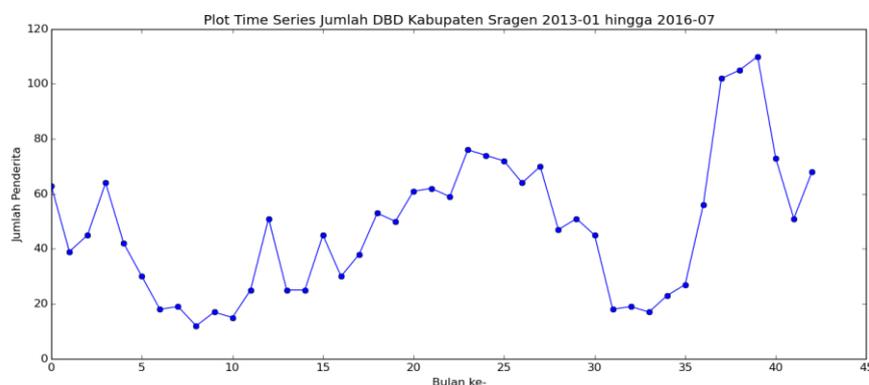
Model forecasting pertama pada proses 1 ini adalah model forecasting ARIMA. Tahapannya secara rinci adalah sebagai berikut:

1. Input dan Persiapan Data

Pada penelitian ini, data set yang digunakan untuk membangun *sistem forecasting* menggunakan format csv untuk memudahkan proses penelitian. Untuk memproses data mentah ini agar dapat dimasukkan ke dalam sistem, penulis memilih menggunakan *library pandas* dari python. *pandas* adalah *library* python *open source* untuk menganalisis data dan menyediakan struktur data yang mudah tapi berkinerja tinggi. Data atau file csv ini akan diproses menggunakan *pandas* sehingga menghasilkan dataset.

2. Pemeriksaan Kestasioneran Data

Kestasioneran data dapat diketahui melalui beberapa cara. Cara yang pertama adalah dengan melihat (secara visual) bentuk plot atau diagram garis yang dihasilkan dari data penelitian. Pada sistem yang dibangun, kode inti untuk membentuk plot data ini dimulai dari mengimport *library* atau *module* python yang berfungsi untuk memproses plot data. *Library* yang digunakan pada penelitian adalah *matplotlib*. *Library matplotlib* adalah *library* python yang berfungsi untuk memproses plot seperti menggambar, mengedit, menyimpan dan proses-proses lain. Setelah *library-library* ini diimport atau dipanggil pada kode sistem, *class*, *function* dan *variable* yang disediakan oleh *library-library* ini dapat digunakan pada sistem yang akan dibuat. Output data plot yang dihasilkan pada proses ini dapat dilihat pada gambar 2, yaitu:



Gambar 2. Plot data penelitian.

Setelah data penelitian berhasil diplot, maka langkah selanjutnya adalah menguji apakah data penelitian ini bersifat stasioner atau tidak. Dari

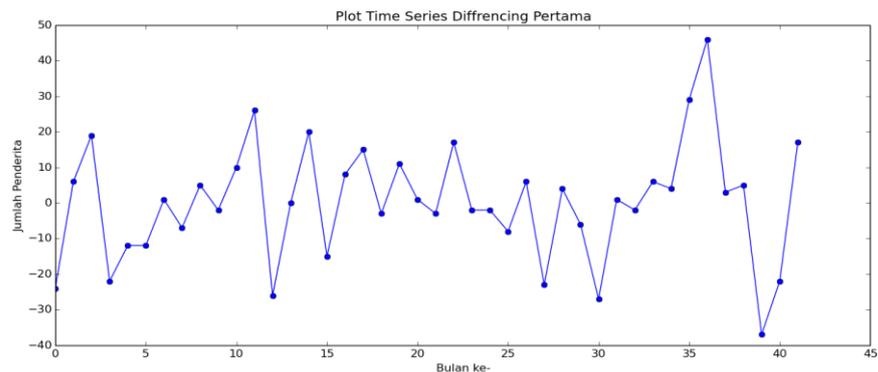
gambar 2 di atas dapat disimpulkan bahwa data penelitian ini bersifat tidak stasioner karena memiliki kecenderungan kenaikan (perubahan) data sepanjang waktu. Metode ini dilakukan secara manual dengan mengecek kondisi visual dari plot data. Metode kedua adalah dengan cara korelogram. Metode ketiga dan yang penulis gunakan adalah metode test ADF atau *Augmented Dickey-Fuller*. Test ADF adalah test yang memanfaatkan data statistik dari data yang akan ditest. Hasil output test ADF dapat dilihat pada tabel 1 di bawah ini:

Tabel 1. Hasil tes ADF proses 1.

```

=====CEK STASIONERITAS=====
(-2.4358918480143856,
 0.13186133409139636,
 3,
 32,
 {'1%': -3.653519805908203,
  '10%': -2.6175881640625001,
  '5%': -2.9572185644531253},
 18.150977799118841)
ADF Statistik: -2.435892
p-value: 0.131861
Critical Value:
 5%: -2.957
 1%: -3.654
10%: -2.618
Jenis Data: Data Non-Stasioner
    
```

Jika data penelitian sudah berbentuk stasioner, maka data penelitian ini bisa langsung diproses untuk tahap selanjutnya, yaitu tahap identifikasi model ARIMA. Namun karena data masih bersifat non-stasioner, maka data ini harus diubah dulu menjadi bentuk data stasioner dengan cara differencing. Hasil dari proses differencing ini dapat dilihat pada gambar 3 di bawah ini:

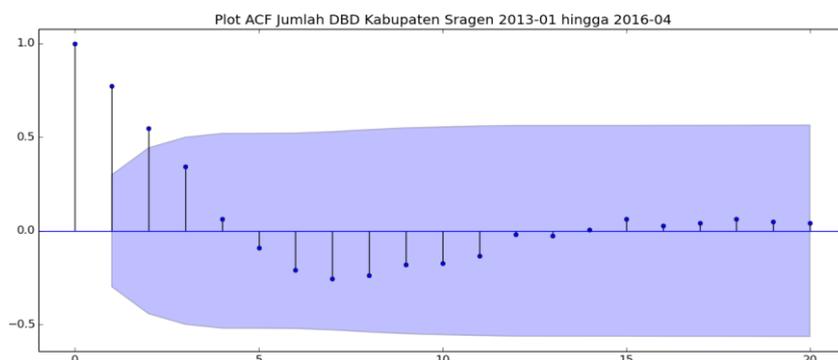


Gambar 3. plot differencing iterasi ke-1.

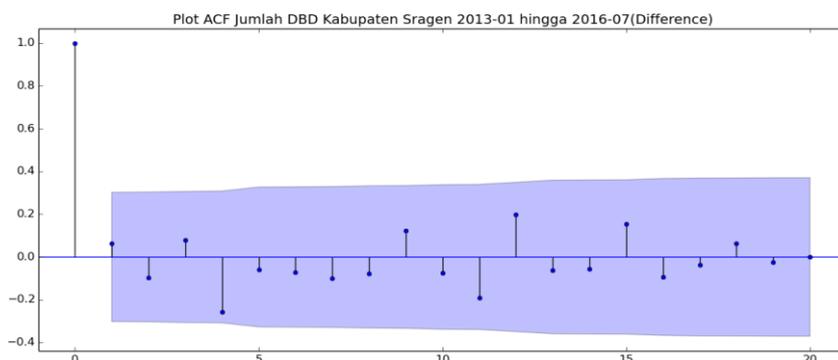
3. Identifikasi model dalam ARIMA dengan plot ACF dan PACF

Untuk membuat plot-plot ACF dan PACF, penulis terlebih dahulu mengimport module-module yang dibutuhkan system seperti library statsmodels. Setelah module-module yang dibutuhkan diimport, langkah

selanjutnya adalah memanggil library-library ini untuk membuat dan mengidentifikasi plot-plot ACF dan PACF dari data penelitian. Plot-plot ini dibuat berdasarkan dataset yang belum di-differencing dan sudah differencing sebagai perbandingan. Plot ACF dengan dataset sebelum dilakukan proses differencing yang dihasilkan pada proses ini ditampilkan pada gambar 4 di bawah ini:

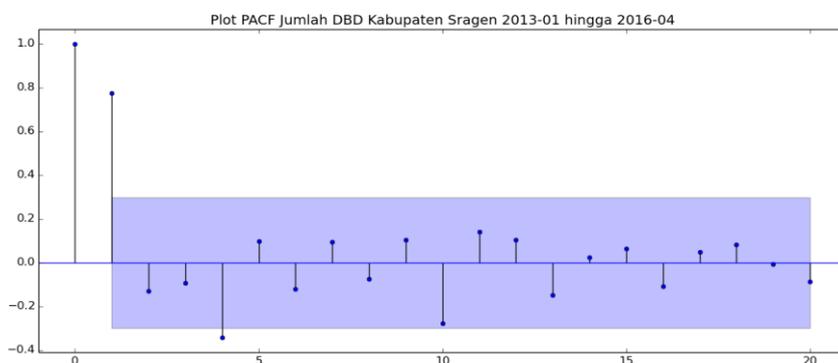


Gambar 4. Plot ACF data penelitian sebelum di-differencing. Kemudian plot dataset setelah data penelitian setelah di-differencing dapat dilihat pada gambar 5 di bawah ini:

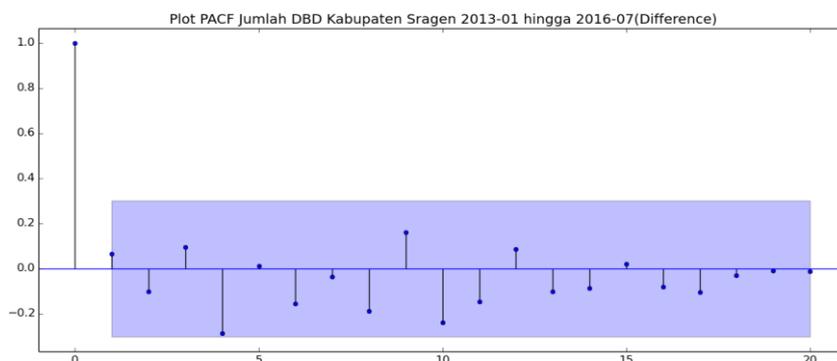


Gambar 5. Plot ACF data stasioner.

Setelah plot ACF didapatkan, langkah berikutnya adalah membuat plot data PACF. Plot PACF untuk data non-stasioner yang dihasilkan pada proses ini dapat dilihat pada gambar 6 di bawah ini:



Gambar 6. Plot PACF data penelitian dataset non-stasioner.
Sedangkan plot PACF untuk data stasioner ditampilkan pada gambar 7 di bawah ini:



Gambar 7. Plot PACF untuk data stasioner.

Berdasarkan keempat plot ACF dan PACF di atas, didapatkan lima kemungkinan model ARIMA, yaitu ARIMA (1, 0, 1), ARIMA (1, 0, 4), ARIMA (2, 0, 1), ARIMA (2, 0, 4), dan ARIMA (0, 1, 4).

4. Penentuan persamaan model ARIMA dan forecasting.

Setelah model-model ARIMA ditentukan, langkah selanjutnya adalah memilih model terbaik. Pada penelitian penulis, pemilihan model terbaik ini dilakukan secara visual dan perbandingan log likelihood terbesar. Berikut ini adalah daftar nilai Log Likelihood terbesar dari semua model ARIMA yang ada: ARIMA (1, 0, 1) dengan Log Likelihood = -141.948, ARIMA (1, 0, 4) dengan Log Likelihood = -137.399, ARIMA (2, 0, 1) dengan Log Likelihood = -141.947, ARIMA (2, 0, 4) dengan Log Likelihood = -137.398, dan ARIMA (0, 1, 4) dengan Log Likelihood = -135.657, maka dapat diambil kesimpulan, model ARIMA terbaik adalah model ARIMA(0, 1, 4).

Setelah proses 1 selesai, tahap berikutnya adalah melakukan proses 2. Pada bagian Proses 2 ini, forecasting ARIMA dilakukan dengan referensi data mulai dari data DBD tanggal 2010-01-31 hingga tanggal 2016-07-31. Selama selang waktu pada tahun 2010 hingga 2016, terdapat sepuluh bulan data DBD yang hilang, yaitu 2011-09, 2011-10, 2011-11, 2011-12 dan 2012-07, 2012-08, 2012-09, 2012-10, 2012-11, 2012-12. karena adanya data yang hilang proses forecasting dengan model ARIMA tidak dapat dilakukan. Untuk mengatasi masalah ini, perlu dilakukan estimasi titik untuk mengisi data yang kosong ini. Metode estimasi titik yang dipilih oleh penulis adalah metode trend moment. Persamaan-persamaan yang digunakan pada metode trend moment adalah sebagai berikut:

1. $y = a + bx$
2. $\sum y = n.a + \sum x.b$
3. $\sum xy = \sum x.a + \sum xx$

Dari data penelitian yang digunakan, didapatkan data-data sebagai berikut: $\sum y = 3007$, $\sum x = 3081$, $\sum xy = 130693$, $\sum xx = 161239$, $n = 78$. Berdasarkan data

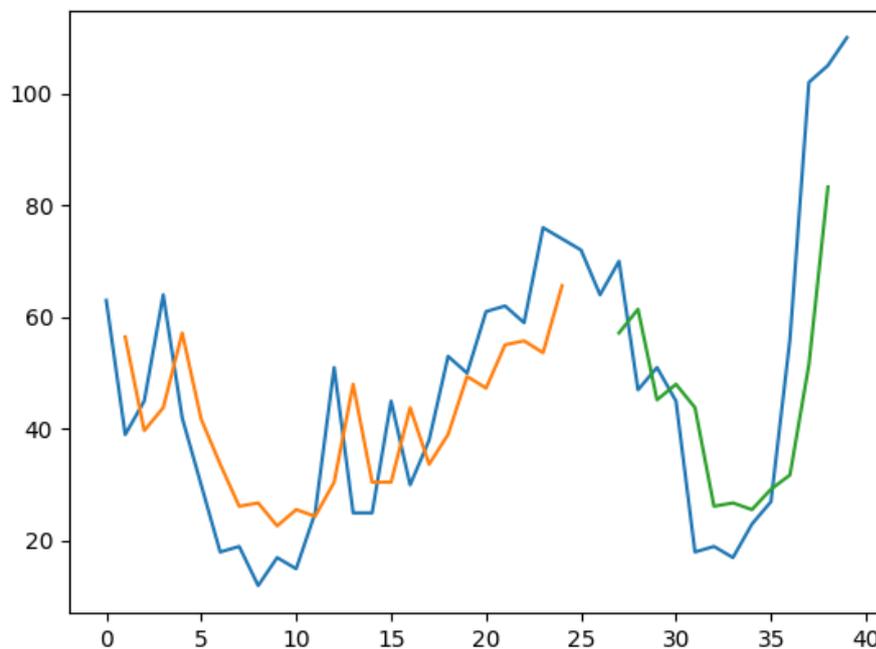
penelitian ini dan sampel data set penelitian DBD yang digunakan, persamaan $\sum y = n.a + \sum x.b$ akan diubah menjadi $3007 = 78*a + b*3081$, sedangkan persamaan $\sum xy = \sum x.a + \sum xx$ diubah menjadi $130693 = 3081.a + 161239.b$. Dari kedua persamaan ini didapatkan nilai $a = 26.646686652$ dan $b = 0.301382162$. kemudian dari persamaan $y = a + bx$, di dapatkan $y(2011-09) = 32.674329892$, $y(2011-10) = 32.975712054$, $y(2011-11) = 33.277094216$, $y(2011-12) = 33.578476378$ dan $y(2012-07) = 35.688151512$, $y(2012-08) = 35.989533674$, $y(2012-09) = 36.290915836$, $y(2012-10) = 36.592297998$, $y(2012-11) = 36.893680616$, $y(2012-11) = 37.195062322$. Setelah data ini dibulatkan, data ini dimasukkan pada data-data kosong sesuai tahun dan bulan masing-masing data, dan kemudian dilakukan proses forecasting ARIMA.

Setelah data kosong atau data rusak diganti, proses pengolahan data pada proses 1 dilakukan lagi pada proses 2, tetapi dengan menggunakan dataset yang telah dilengkapi. Hasil pengolahan data pada proses 2 ini berupa perbandingan log likelihood dari dua model ARIMA yang dihasilkan, yaitu ARIMA(1, 0, 1) dan ARIMA(2, 0, 1). Model terbaik dari model ARIMA(1, 0, 1) dan ARIMA(2, 0, 1) adalah model ARIMA(2, 0, 1) karena memiliki log likelihood sebesar -292.163, sedangkan model ARIMA(1, 0, 1) memiliki log likelihood = -292.429.

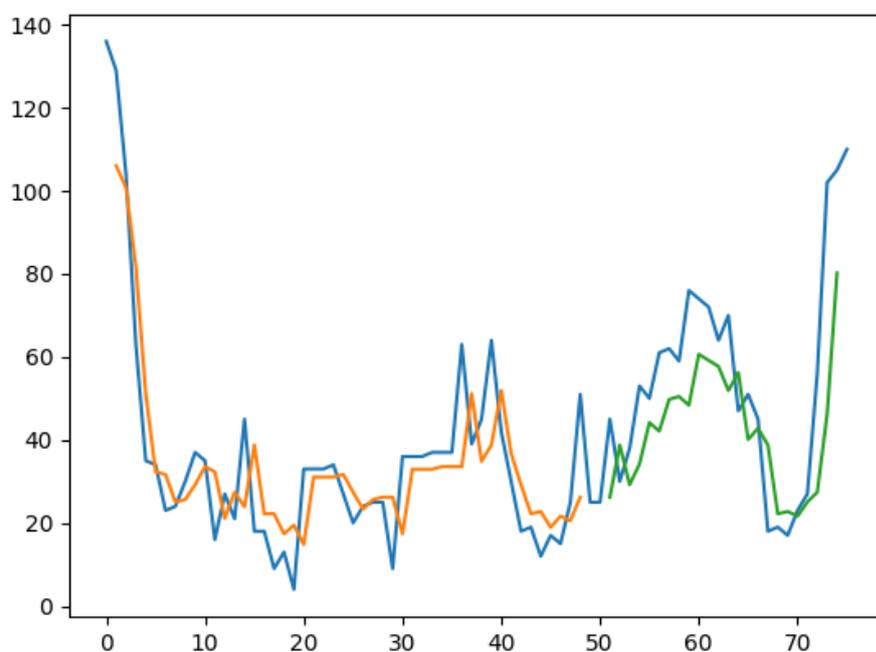
Sementara itu jika semua model ARIMA yang terbentuk dibandingkan, model terbaik adalah model ARIMA (0, 1, 4) dengan Log Likelihood = -135.657.

Setelah proses ARIMA pada proses 1 dan proses 2 dilakukan, proses selanjutnya adalah melakukan forecasting Backpropagation pada dataset 1 dan dataset 2. Model Backpropagation yang digunakan pada penelitian ini adalah model LSTM (*Long-Short Term Memory*). Tahap-tahap model LSTM yang dijabarkan adalah sebagai berikut:

1. Mengimport library-library yang dibutuhkan.
Library-library yang dibutuhkan pada proses LSTM yang digunakan pada penelitian ini di antaranya numpy, matplotlib.pyplot, pandas, math, keras.models, sklearn.preprocessing dan sklearn.metrics.
2. Memuat dataset dari file csv
3. Menormalisasi dataset
4. Membagi dataset menjadi data *train* dan data *test*
5. Mengubah data train dan test menjadi $X=t$ dan $Y=t+1$
6. Membuat model LSTM
7. Melakukan prediksi
8. Melakukan invert prediksi
9. Menghitung root mean square error
10. Membuat plot prediksi train dan test
11. Membuat plot perbandingan data prediksi dengan data sebenarnya
Output yang dihasilkan pada proses ini berupa plot 1 dan 2 untuk perbandingan data sebenarnya dan data hasil prediksi. Plot 1 dan plot 2 ini dapat dilihat pada gambar 8 dan gambar 9 di bawah ini:



Gambar 8. output proses LSTM 1

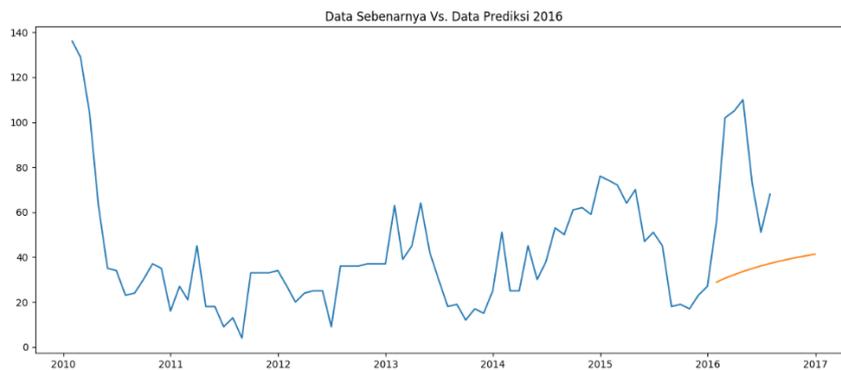


Gambar 9. Output proses LSTM 2

Dari kedua plot di atas, dapat dilihat bahwa plot output proses LSTM 2 lebih akurat dari output proses LSTM 1.

SIMPULAN

Jika dibandingkan output proses ARIMA dan proses LSTM yang dilakukan maka dapat disimpulkan bahwa model terbaik yang dijalankan pada dataset penelitian adalah model LSTM proses 2. Sebagai perbandingan berikut ini adalah gambar 10 yang merupakan plot prediksi dari model ARIMA (0, 1, 4):



Gambar 10. Plot prediksi ARIMA (0, 1, 4)

DAFTAR PUSTAKA

- Ariati, Jusniar, Athena Anwar. (2014). Model Prediksi Kejadian Demam Berdarah Dengue (DBD) Berdasarkan Faktor Iklim di Kota Bogor, Jawa Barat. *Buletin Penelitian Kesehatan (Bulletin of Health Reseach, p-ISSN: 0125. E-ISSN: 2338-3453)*. Badan Penelitian dan Pengembangan Kesehatan, Ministry of Health of Republic of Indonesia.
- Benli, T. O. (2016). A Comparison of Nineteen Various Electricity Consumption Forecasting Approaches and Practicing to Five Different Households in Turkey. *arXiv preprint arXiv:1607.05660*.
- Canale, Antonio, Matteo Ruggiero. (2016). Bayesian nonparametric forecasting of monotonic functional time series. *Electronic Journal of Statistics 10.2: 3265-3286*.
- Chaudhuri, Tamal Datta, Indranil Ghosh. (2016). Artificial Neural Network and Time series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework. *arXiv preprint arXiv:1607.02093*.
- Damayanti, Eka, Iah Solikhah, Nunung Nurjannah, dan Tita Azizah, (2013), *Metode Momen, Metode Likelihood, dan Estimator Bayes*. Cirebon: Fakultas Tarbiyah Institut Agama Islam (IAIN) Syekh Nurjati.
- Gutierrez, Nicolas, Manuela Wiesinger-Widi. (2016). AUGURY: A time-series based application for the analysis and forecasting of system and network performance metrics. *arXiv preprint arXiv:1607.08344*.
- Halim, Reinaldo Andhika, Alexander Setiawan, dan Djoni Haryadi Setiabudi. (2016). Aplikasi Sistem Forecasting CV. Karisma Motor Dengan Metode ARIMA. *Jurnal Infra 4.2: 105-109*.
- Hasibuan, Z. A. (2007). *Metodologi Penelitian Pada Bidang Ilmu Komputer dan Teknologi Informasi*. Depok: Universitas Indonesia.

- Hazra, Arnab. (2015). A Time-varying Parameter Based Seasonally-adjusted Bayesian State-space Model for Forecasting. *arXiv preprint arXiv:1512.02149*.
- Herjanto, Eddy. (2007). *Manajemen Operasi, Edisi Ketiga*. Daerah Khusus Ibukota Jakarta: Grasindo.
- Janah, Sufia Nur, Winita Sulandari, dan Santoso Budi Wiyono. (2014). Penerapan Model *Hybrid* ARIMA Backpropagation Untuk Peramalan Harga Gabah Indonesia. *MEDIA STATISTIKA* 7.2: 63-69.
- Kusuma, Agcrista Permata, dan Dyah Mahendrasari Sukendra. (2016). Analisis Spasial Kejadian Demam Berdarah Dengue Berdasarkan Kepadatan Penduduk. *Unnes Journal of Public Health* 5.1.
- Larose, Daniel T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: JOHN WILEY & SONS, INC.
- Octavia, Tanti, Yulia Yulia, dan Lydia Lydia. (2015). Peramalan Stok Barang untuk Membantu Pengambilan Keputusan Pembelian Barang pada Toko Bangunan XYZ dengan Metode Arima. *Seminar Nasional Informatika (SEMNASIF)*. Vol. 1. No. 1.
- Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- Puspitaningrum, Diyah. (2006). *Pengantar Jaringan Saraf Tiruan*. Yogyakarta: Andi.
- Rumagit, Silviani E., and S. N. Azhari. (2014). Prediksi Pemakaian Listrik Menggunakan Jaringan Syaraf Tiruan dan ARIMA Di Wilayah Sulluttenggo. *Berkala Ilmiah MIPA* 23.2.
- Sari, Lisa Anita, and Widya Hary Cahyati. (2016). Efektivitas Ekstrak Buah Belimbing Wuluh (*Averrhoa Bilimbi* L.) Dalam Bentuk Granul Terhadap Kematian Larva Nyamuk *Aedes Aegypti*. *VISIKES* 14.1.
- Sen, Jaydip, Tamal Datta Chaudhuri. (2016). A framework for predictive analysis of stock market indices—a study of the Indian auto sector. *Calcutta Business School (CBS) Journal of Management Practices* 2.2: 1-19.
- Sutojo, T., Edi Mulyanto, Vincent Suhartono. (2011). *Kecerdasan Buatan*. Yogyakarta: Andi.
- Syahputra, Irvandi. (2016). Aplikasi Algoritma Memetika untuk Peramalan Pergerakan Kurs Valuta Asing dengan Menggunakan Model ARIMA (Box-Jenkins). *Jurnal Sains dan Seni ITS* 4.2.
- Wiyanti, D. T., and R. Pulungan. (2012). Peramalan Deret Waktu Menggunakan Model Fungsi Basis Radial (RBF) Dan Auto Regressive Integrated Moving Average (ARIMA). *Jurnal MIPA* 35.2.