

Peramalan Jumlah Kasus Penyakit Menular Influenza dan Demam Berdarah di Kota Surabaya Berdasarkan Data Google Trends dan Faktor Iklim Menggunakan Variasi Metode Autoregressive Integrated Moving Average (ARIMA)

Laras Aristiani¹, Wiwik Anggraeni, S.Si, M.Kom²

^{1,2}Sistem Informasi, Fakultas Teknologi Informasi (FTIf), Institut Teknologi Sepuluh Nopember (ITS) Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: ¹larasarist@gmail.com, ²wiwik@its-sby.edu

Abstrak— Penanganan efektif terhadap pengendalian penyakit menular memerlukan pendeteksian dini persebaran penyakit melalui surveilans yang tepat waktu dan akurat beserta respon yang cepat dari hasil deteksi dini. Jumlah kejadian penyakit menular dipengaruhi beberapa faktor eksternal. Angka kejadian demam berdarah dan Influenza berkorelasi dengan faktor iklim curah hujan, suhu dan kelembaban udara pada suatu daerah. Selain faktor iklim terdapat variabel eksternal lain yang bisa digunakan untuk pemodelan peramalan kasus demam berdarah dan influenza, yakni jejak data di internet, salah satu sumbernya adalah Google Trends. Penelitian sebelumnya menunjukkan bahwa data hasil pencarian query tertentu berkaitan dengan penyakit menular dalam Bahasa Korea pada Google Trend berkorelasi dengan data jumlah kasus penyakit di Korea Selatan.

Tujuan dari penelitian Tugas Akhir ini adalah untuk menyusun model peramalan yang melibatkan variabel eksternal faktor iklim dan Google Trends. Tugas Akhir ini juga menganalisis kelayakan Google Trends dengan lokasi Indonesia untuk dipakai sebagai data alternatif untuk sistem surveilans dan meramalkan jumlah penyakit demam berdarah dan influenza. Data yang digunakan dalam penelitian ini adalah (1) data jumlah kasus demam berdarah dan influenza yang diambil dari RSUD Dr. Soetomo (2) data faktor iklim yang meliputi suhu, kelembaban, dan curah hujan yang diambil dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Juanda Surabaya dan Perak I (3) Data hasil pencarian Google Trends dengan query yang sudah ditentukan sebelumnya. Seluruh data ini tersedia untuk periode Desember 2010 hingga Agustus 2015. Teknik interpolasi dan ekstrapolasi digunakan untuk menangani kekosongan data. Pemodelan ramalan dilakukan dengan metode ARIMA untuk model univariat dan ARIMAX untuk model multivariat.

Hasil dari penelitian ini menunjukkan bahwa penambahan variabel eksternal pada model ARIMAX mampu meningkatkan performa model 7,56% lebih baik dibandingkan model ARIMA untuk kasus demam berdarah. Kasus influenza tidak cocok dimodelkan dengan ARIMA, dengan jangkauan MAPE testset berada pada 72% - 102%. Hal ini dikarenakan pola data influenza penelitian ini naik turun secara drastis, tidak cocok dimodelkan dengan ARIMA. Walaupun model peramalan influenza masih belum layak, model ARIMAX dengan variabel iklim dan Google Trends bisa memperbaiki performa hingga 19,38%.

Penelitian ini menunjukkan bahwa variabel eksternal iklim dan Google Trends mampu meningkatkan performa model. Model dengan variabel eksternal iklim menunjukkan performa yang lebih baik dibandingkan dengan model dengan variabel eksternal Google Trends. Google Trend dengan lokasi Indonesia masih belum layak dipakai sebagai data alternatif untuk sistem surveilans penyakit DBD dan Influenza karena error yang masih besar.

Kata Kunci— Peramalan, Univariat, Multivariat, ARIMA, ARIMAX, Faktor Iklim, Google Trends, Influenza, Demam berdarah

I. PENDAHULUAN

Indonesia merupakan peringkat kedua di dunia yang memiliki kasus DBD terbanyak [3]. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia angka kesakitan atau *incidence rate* (IR) DBD cenderung meningkat sejak tahun 1968, dengan data terakhir pada tahun 2013 nilai IR adalah 41,25 per 100.000 penduduk [4]. Influenza merupakan penyakit yang tersebar di seluruh dunia dan bersirkulasi sepanjang tahun. Virus influenza musiman dan pandemi yang sangat mudah menular antar manusia. Di Indonesia, kasus infeksi influenza musiman berlangsung terus sepanjang tahun yang mengikuti pola sirkulasi virus influenza musiman [2].

Angka kejadian DBD dan Influenza dipengaruhi oleh faktor iklim curah hujan, suhu dan kelembaban udara pada suatu daerah [6, 7]. Fenomena penyakit menular perlu perhatian khusus, sehingga pengendalian penyakit menular merupakan salah satu dari rencana pembangunan jangka menengah bidang kesehatan oleh Kementerian Kesehatan RI [8]. Penanganan efektif terhadap pengendalian penyakit menular memerlukan pendeteksian dini persebaran penyakit melalui surveilans yang tepat waktu dan akurat beserta respon yang cepat dari hasil deteksi dini [9]. Pendeteksian dini dapat diperoleh dengan menggunakan teknik peramalan.

Akhir-akhir ini jejak data di internet dianggap sebagai sumber data potensial untuk monitoring penyakit menular. Generous et.al (2014) menggunakan data log akses artikel di Wikipedia untuk memantau dan memprediksi penyakit menular di Haiti, Brazil, Thailand, Uganda, Cina, Jepang, Polandia, Amerika Serikat dan Norwegia [12, 13]. Pada tahun 2008 Google meluncurkan Google Flu Trends dan Google Dengue Trends yang menggunakan data hasil pencarian di Google sebagai alternatif sistem surveilans dan peramalan data penyakit influenza dan DBD [14, 15, 16]. Walaupun saat ini layanan Google Flu Trends dan Google Dengue Trends sudah dibatasi hanya untuk penelitian, data hasil pencarian di Google bisa dengan mudah diakses di <http://www.google.com/trends>.

Sungjin Cho et.al (2013) menganalisis korelasi antara data surveilans nasional dan Google Trend di Korea Selatan. Hasil dari penelitian ini menunjukkan data hasil pencarian *query* tertentu dalam Bahasa Korea pada Google Trend berkorelasi dengan data surveilans nasional di Korea Selatan. Walaupun belum ada penelitian serupa yang dilakukan dengan menggunakan *query* Bahasa Indonesia, hasil penelitian ini menunjukkan hasil di Google Trend layak untuk dilakukan analisis lebih lanjut.

Ada beberapa penelitian yang sudah melakukan peramalan jumlah penderita penyakit menular dengan menggunakan data Google Trend. Dugas et.al (2013) melakukan peramalan penyakit influenza dengan data Google Flu Trend dan data iklim di Amerika Serikat [17]. Araz et.al

(2014) menggunakan data Google Flu Trend untuk meramalkan jumlah pasien *influenza-like-illness* yang datang ke ruang darurat rumah sakit [17]. Kedua penelitian menggunakan metode *Autoregressive Moving Average* (ARMA) dengan variasi yang berbeda. Hasil dari keduanya menunjukkan data Google Trend layak digunakan untuk meningkatkan performa peramalan.

Pada penelitian sebelumnya peramalan penyakit menular dilakukan dengan mengintegrasikan data historis pada periode sebelumnya dengan variabel eksternal seperti data absensi siswa sekolah [18], data penjualan obat di apotek [19] dan data iklim (suhu, kelembaban, curah hujan, dan tekanan udara) [20]. Soebiyanto et.al (2010) menganalisis pengaruh faktor iklim terhadap epidemiologi penyakit influenza di daerah bersuhu hangat dengan menggunakan beberapa model *Autoregressive Integrated Moving Average* (ARIMA) dan menggunakan model SARIMA (*Seasonal Autoregressive Integrated Moving Average*) saat ada unsur seasonalitas pada data. Hasil penelitian menunjukkan model peramalan ARIMA & SARIMA multivariat yang dilengkapi dengan data iklim memiliki performa yang lebih baik dibandingkan dengan model univariat yang bergantung pada data periode sebelumnya [20].

Melihat dari penelitian yang sudah dilakukan, Google Trend dan data iklim merupakan variabel yang potensial sebagai atribut alternatif selain data surveilans untuk melakukan peramalan penyakit menular. Data klimatologi memiliki keuntungan lebih karena data yang dihasilkan spesifik untuk suatu daerah tertentu, mudah diperoleh dan tersedia secara *real-time*. Begitu juga dengan Google Trend yang bisa diambil sampai ke level kota. Data penyakit menular yang cenderung memiliki seasonalitas, variasi metode ARIMA (ARIMA dan ARIMAX) bisa digunakan sebagai metode peramalan yang sesuai untuk meramalkan jumlah penyakit menular di Kota Surabaya. Penelitian Tugas Akhir ini akan membuat variasi model peramalan metode ARIMA univariat dan multivariate dengan inputan variabel eksternal berupa faktor iklim dan data Google Trend.

Tujuan dari penelitian ini adalah untuk mencari model ARIMA univariat dan multivariat dengan performa terbaik untuk meramalkan jumlah penyakit demam berdarah dan influenza. Penelitian ini juga menganalisis kelayakan Google Trends dengan lokasi geografis Indonesia untuk dipakai sebagai data alternatif sistem surveilans dan meramalkan jumlah penyakit demam berdarah dan influenza. Kelayakan dinilai dari hasil analisis MAPE dari model yang menggunakan data Google Trend serta *error variance* dari data kasus penyakit dan data hasil dari model Google Trend.

II. TINJAUAN PUSTAKA

A. Faktor Iklim

Soebiyanto et.al (2010) melakukan peramalan penyakit influenza dengan menggunakan faktor iklim sebagai inputan variabel eksternal selain data historis penyakit tersebut. Hasilnya menunjukkan bahwa peramalan yang dilakukan dengan model dengan faktor iklim memiliki performa yang lebih baik [20]. Pada penelitian ini faktor iklim yang dilibatkan sebagai inputan variabel eksternal meliputi :

- Data curah hujan
- Data suhu udara
- Data kelembaban relatif

B. Google Trends

Google Trends menganalisis persentase hasil pencarian di Google untuk mencari berapa banyak pencarian yang telah dilakukan pada beberapa periode tertentu. Google Trends menampilkan data yang sudah disesuaikan berupa skala dari 0 sampai 100. Hal ini bertujuan untuk mempermudah perbandingan antar hasil pencarian. Penyesuaian dilakukan dengan membagi seluruh total pencarian pada lokasi tertentu dan waktu pencarian [24]. Pada penelitian ini data Google Trend digunakan sebagai variabel inputan eksternal.

C. Teknik Peramalan

Peramalan adalah aktivitas untuk memprediksi masa depan dengan cara seakurat mungkin. Prediksi dilakukan berdasarkan data historis dan pengetahuan mengenai kejadian masa depan yang mungkin bisa mempengaruhi peramalan [25]. Secara umum terdapat dua metode peramalan yakni kualitatif dan kuantitatif. Pada penelitian ini metode yang digunakan adalah metode kuantitatif univariat dan multivariate.

D. ARIMA

Model ARIMA, atau juga dikenal dengan model Box-Jenkins adalah gabungan proses *autoregressive* (AR) dan *Moving Average* (MA), dengan *integration* dalam konteks ini adalah derajat *differencing* untuk mencapai pola data yang stasioner, karena asumsi awal metode ARIMA data awal harus stasioner. Secara umum persamaan ARIMA memiliki parameter p, d dan q. Dengan p adalah parameter order untuk AR, d adalah derajat *differencing*, dan q adalah parameter order untuk MA. Penggabungan model AR, *differencing* dan MA akan menghasilkan persamaan

$$(1 - \phi_1 B - \dots - \phi_p B^p) (1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) e_t$$

\uparrow
AR(p)

\uparrow
d differences

\uparrow
MA(q)

Persamaan II.1 - ARIMA

Untuk melakukan identifikasi model ARIMA bisa dilakukan dengan melakukan analisis terhadap ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) dari hasil plot korelogram data. Tahapan peramalan dengan menggunakan model ARIMA adalah sebagai berikut :

- i. Identifikasi
 - a. Uji stasioneritas
 - b. Analisis model berdasarkan hasil ACF dan PACF untuk mengestimasi parameter p dan q
- ii. Estimasi parameter dari model
- iii. Pengecekan diagnostik untuk menentukan model yang paling baik
- iv. Melakukan peramalan dengan menggunakan model yang paling baik.

E. ARIMAX

Model ARIMAX, atau dikenal dengan ARIMA dengan *multiple regressor* merupakan pengembangan dari model ARIMA dasar dengan variabel eksternal lainnya. Model ARIMA yang ditambahkan beberapa variabel yang memiliki pengaruh yang signifikan bisa menambah performa model peramalan [29].

F. Uji Korelasi Spearman

Uji korelasi Spearman digunakan untuk mengukur kekuatan asosiasi antara 2 jenis variabel. Nilai rho (ρ) merupakan nilai koefisien Spearman yang menunjukkan kekuatan korelasi antar variabel. Tabel 1 menunjukkan petunjuk interpretasi nilai dari rho [31].

Tabel 1 - Interpretasi Nilai Koefisien Spearman Correlation

Nilai rho	Interpretasi hasil
.90 s/d 1.00 -.90 s/d -1.00	Korelasi positif/negative yang sangat kuat
.70 s/d .90 -.70 s/d -.90	Korelasi positif/negative yang kuat
.50 s/d .70 -.50 s/d -.70	Korelasi positif/negative yang cukup/moderat
.30 s/d .50 -.30 s/d -.50	Korelasi positif/negative yang lemah
.00 s/d .30 .00 s/d -.30	Tidak ada korelasi, tidak perlu dihiraukan

Uji korelasi ini akan dilakukan untuk melihat korelasi data antara data hasil pencarian Google Trend terhadap data historis kejadi penyakit.

G. Uji Stasioneritas, ADF (Augmented Dickey Fuller Test)

Uji ADF (Augmented Dickey Fuller) adalah uji akar unit yang bertujuan untuk menganalisis kestasioneritasan data. Uji ini diperlukan karena metode ARIMA memerlukan inputan yang stasioner. Hipotesis null dari uji ini adalah terdapat keberadaan akar unit pada data *timeseries* [32].

- Jika nilai uji T lebih besar dari nilai kritis ADF maka terima H_0 dan disimpulkan bahwa data tersebut memiliki akar unit atau tidak stasioner
- Jika nilai uji T kurang dari nilai kritis ADF maka tolak H_0 dan disimpulkan bahwa data tersebut tidak memiliki akar unit atau stasioner.

H. Uji Ljung Box

Uji Ljung-Box digunakan untuk melakukan uji kelayakan terhadap model yang sudah diestimasi. Tujuan dari uji Ljung-Box adalah untuk mengetahui keberadaan autokorelasi pada residual model. Residual merupakan selisih dari data aktual dengan data hasil peramalan. Jika diketahui masih ada autokorelasi pada residual model, maka model tersebut masih bisa dikembangkan kembali. Dari uji Ljung-Box terhadap residual, hasil yang dicari adalah model yang residualnya tidak memiliki autokorelasi (*white noise*). Keberadaan *white noise* pada residual model menunjukkan bahwa model tersebut sudah cukup [25] [40]. Hipotesis awal (H_0) pada Ljung-Box adalah data tersebut *white noise*. Hipotesa alternatif (H_a) dari Ljung-Box adalah tidak *white noise*.

I. MAPE

Mean Absolute Percentage Error adalah satuan yang digunakan untuk mengetahui persentase error hasil peramalan. Satuan MAPE lebih sering digunakan untuk *reporting* karena konsep yang mudah dipahami. Semakin rendah nilai MAPE menunjukkan performa yang lebih baik dari suatu model. MAPE dirumuskan sebagai berikut

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{xt - ft}{xt} \right|}{n} \times 100\%$$

Penafsiran nilai MAPE dijelaskan pada Tabel 2

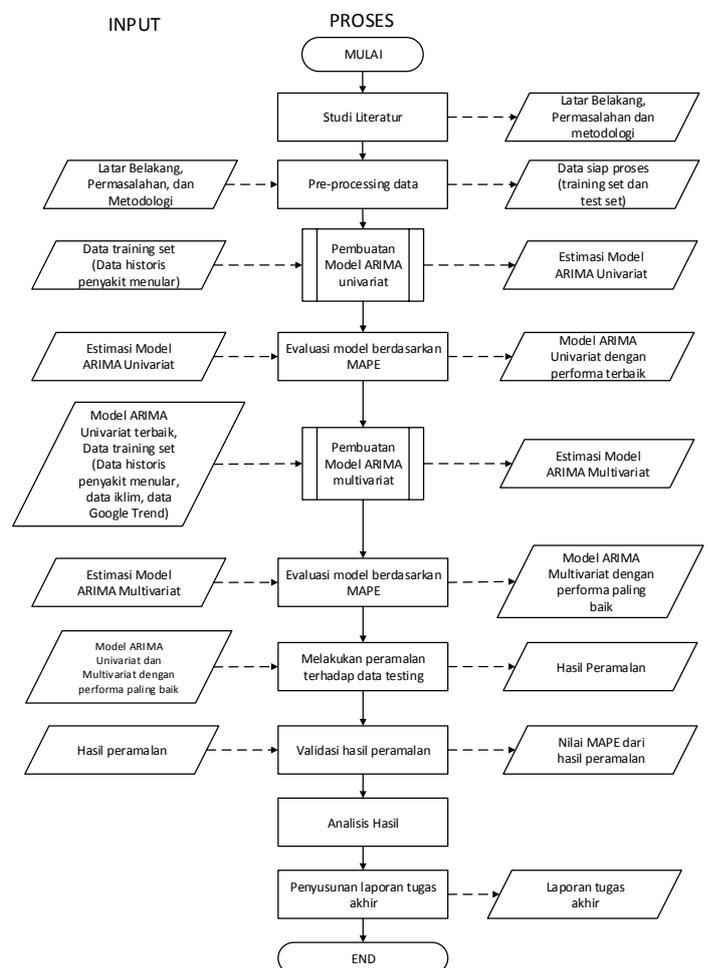
Tabel 2 - Penafsiran nilai MAPE

MAPE	Hasil Peramalan
<10%	Sangat Baik
10-20%	Baik
20-50%	Layak/ Cukup
>50%	Buruk

J. SAS Software

SAS merupakan aplikasi statistik yang digunakan untuk melakukan berbagai aktivitas analisis data. SAS menjadi pilihan utama aplikasi analisis statistik pada berbagai industri [35]. Salah satu komponen SAS adalah SAS/ETS (Econometric and Time Series Analysis) yang memberikan fasilitas analisis time series dan analisis finansial. Komponen SAS/ETS memiliki berbagai macam fungsi salah satunya adalah fungsi peramalan dengan metode ARIMA [36]. Untuk penelitian ini SAS/ETS digunakan untuk mengembangkan model ARIMA, dan peramalan. metode penelitian.

III. METODOLOGI



Gambar 1. Alur pengerjaan tugas akhir

A. Pre processing data

Kegiatan pre-processing data meliputi pengumpulan data dan persiapan data sebelum dianalisis. Pada penelitian Tugas Akhir ini data yang digunakan sebagai bahan penelitian:

- Data historis jumlah kasus penyakit demam berdarah dari RSUD dr. Soetomo Surabaya.

- Faktor iklim (data suhu, kelembaban relatif dan curah hujan), diperoleh dari website resmi BMKG (<http://dataonline.bmkg.go.id>).
- Data Google Trends

Aktivitas yang dilakukan pada pemrosesan data adalah menangani missing value dan agregasi data menjadi data per bulan

B. Pembuatan Model ARIMA Univariat

Pada pembuatan model ARIMA univariat data yang digunakan sebagai inputan adalah data historis penyakit menular. Langkah-langkah pembuatan model univariat meliputi uji stasioneritas, analisis lag ACF dan PACF untuk identifikasi model, estimasi nilai parameter dan uji kelayakan model.

C. Pembuatan Model ARIMA Multivariat

Penyusunan model ARIMA Multivariat memerlukan model ARIMA univariat dengan performa terbaik. Pada pembuatan model ARIMA multivariat data yang digunakan sebagai inputan adalah data historis penyakit menular, faktor iklim dan Google Trend. Variasi model yang akan dibentuk adalah sebagai berikut :

- Model multivariat dengan variabel iklim
- Model multivariat dengan variabel Google Trends
- Model multivariat dengan variabel iklim dan Google Trends

Variabel eksternal untuk pemodelan tersebut harus melalui proses transformasi yang sama dengan variabel dependen. Seluruh variabel eksternal melalui proses uji korelasi terhadap variable dependen, untuk memilih variabel eksternal yang digunakan untuk pemodelan ARIMAX. Tahapan pemodelan ARIMAX adalah identifikasi orde p , d dan q , identifikasi kandidat variable eksternal, estimasi parameter AR, MA dan variable eksternal, uji kelayakan model.

D. Evaluasi model berdasarkan MAPE

Uji kelayakan model dilakukan dengan menganalisis *white noise* dari residual dengan uji Ljung Box. Keberadaan *white noise* pada residual model menunjukkan bahwa model tersebut sudah layak [25]. Pemilihan model dilakukan berdasarkan nilai MAPE. Model dengan nilai MAPE yang paling kecil yang dipilih sebagai dasar model multivariat.

E. Melakukan peramalan terhadap data testing

Pada tahap ini dilakukan peramalan terhadap data testing. Peramalan dilakukan dengan menggunakan model dengan performa terbaik berdasarkan uji kelayakan model yang telah dilakukan.

F. Validasi hasil peramalan

Tahap ini mengukur tingkat error dari hasil peramalan terhadap data histori penyakit dengan menggunakan MAPE.

G. Analisis Hasil

Tahap ini menganalisis hasil peramalan dari berbagai model ARIMA yang sudah dibuat untuk meramalkan jumlah penyakit menular di Kota Surabaya. Analisis yang dilakukan berupa *diagnostic checking* melihat kecukupan model melalui pengecekan *white noise* pada residual, dan hasil MAPE dari model.

IV. PERANCANGAN

A. Deskripsi Data

Penelitian ini menggunakan beberapa variabel yang berbeda. Tabel 3 menjelaskan variabel yang digunakan untuk melakukan peramalan jumlah kasus penyakit demam berdarah dan influenza. Data yang digunakan untuk penelitian ini semuanya diambil dalam jangka waktu Januari 2010 – Agustus 2015.

Tabel 3 - Data Mentah yang digunakan

Nama Variabel	Keterangan	Satuan
VDBD	Jumlah Kasus Demam Berdarah per bulan	Jumlah Kasus per bulan
VFLU	Jumlah Kasus Influenza per bulan	Jumlah Kasus perbulan
GTDBD	Variabel Eksternal, Google Trend Data dengan search query : <ul style="list-style-type: none"> • demam berdarah • dbd • demam • dengue per bulan area Jawa Timur	Search Index per 2 minggu
GTFLU	Variabel Eksternal, Google Trend Data dengan search query : <ul style="list-style-type: none"> • Bersin • Influenza • Flu • Batuk • Pilek • Batuk pilek area Jawa Timur	Search Index per 2 minggu
SUHU	Variabel Eksternal, Suhu rata-rata	°C per hari
HMD	Variabel Eksternal, Kelembaban rata-rata	persen per hari
CH	Variabel Eksternal, Curah Hujan rata-rata	mm per hari

B. Pre-processing Data

Tahapan *pre-processing data* mengolah data-data mentah agak bias digunakan dalam pemodelan. Seluruh variabel eksternal diubah satuan waktunya menjadi bulanan. Data yang kosong atau *missing value* ditangani dengan teknik interpolasi dan ekstrapolasi. Penghitungan interpolasi data dilakukan dengan menggunakan Microsoft Excel digunakan untuk menangani *missing value* [35]. *Exponential smoothing* dipilih untuk teknik ekstrapolasi karena lebih mudah daripada metode Box-Jenkins, dan regresi. Metode *exponential smoothing* juga memiliki tingkat akurasi yang lebih baik jika dibandingkan dengan metode *moving average* [38]. Ekstrapolasi dengan *exponential smoothing* dilakukan dengan menggunakan tools SAS.

Data kemudian dibagi menjadi 2 yakni trainset dan testset. Perbandingan trainset dan testset adalah 70:30, sehingga pembagiannya adalah sebagai berikut:

- Trainset : Januari 2010 – November 2013
- Testset : Desember 2013 – Agustus 2015

V. HASIL PEMODELAN

Pada tahap ini ditunjukkan hasil model peramalan ARIMA dan ARIMAX untuk masing-masing kasus penyakit. Model-model ini sudah memenuhi syarat signifikansi parameter (pvalue < 0,05), dan memiliki residual yang white noise. Tabel 4 menjelaskan model ARIMA univariat untuk kasus demam berdarah. Pemodelan pada kasus demam berdarah, variabel dependen kasus demam berdarah melalui transformasi logaritmik sebanyak dua kali. Hasil MAPE peramalan dihitung dari nilai kasus yang sudah dikembalikan ke bentuk normal.

No	Model	AR		MA		MAPE
		Estimasi	Pr> t	Estimasi	Pr> t	
1	ARIMA(1,0,3)	0,56531	0,0034	-0,4436	0,0083	32,56%
2	ARIMA(1,0,1)	0,74963	<,0001	-0,4036	0,0117	34,19%
3	ARIMA(2,0,0)	-0,3792	0,0064	-	-	33,37%
4	ARIMA(2,0,1)	-0,7622	<,0001	0,53024	0,0328	32,34%
5	ARIMA(4,1,0)	-0,4357	0,0034	-	-	33,87%

Tabel 4 - Model ARIMA Univariat kasus Demam Berdarah

Model ARIMA (2,0,1) memiliki performa terbaik untuk peramalan kasus demam berdarah. Tabel 5 menjelaskan model ARIMA univariat untuk kasus influenza.

No	Model	AR		MA		MAPE
		Estimasi	Pr> t	Estimasi	Pr> t	
1	ARIMA(1,1,0)	-0,57483	<,0001	-	-	65,57%
2	ARIMA(5,1,0)	-0,41649	0,0165	-	-	54,78%
3	ARIMA(2,1,0)	-0,29509	0,0486	-	-	58,99%

Tabel 5 - Model ARIMA Univariat kasus influenza

Model ARIMA (5,1,0) memiliki performa terbaik untuk. Berikutnya adalah hasil pemodelan multivariat untuk kasus demam berdarah dan influenza. Tabel 6 merupakan hasil pemodelan ARIMAX kasus demam berdarah dengan performa terbaik.

Tabel 6 - ARIMAX kasus influenza

No	Model	Variabel Eksternal	MAPE	
			Trainset	Testset
1	ARIMAX(0,0,2)	CH_LOG2_8	24,78%	26%
		HMD_LOG2_1		
2	ARIMAX(2,0,1)	dengue_log2_2	29,11%	19%
		demamberdarah_log2_2		
3	ARIMAX(1,0,0)	HMD_LOG2_1	28,26%	30%
		dengue_log2_1		

Tabel 7 merupakan hasil pemodelan ARIMAX kasus influenza dengan performa terbaik.

Tabel 7 - ARIMAX kasus influenza

No	Model	Variabel Eksternal	MAPE	
			Trainset	Testset
1	ARIMAX(5,1,0)	HMD_9	42,20%	72,40%
2	ARIMAX(5,1,0)	batuk_1	46,24%	84,50%
3	ARIMAX(5,1,0)	HMD_9	35,40%	72%
		batuk_1		

Hasil akhir model-model tersebut berupa persamaan. Persamaan untuk model univariat dan multivariat kasus demam berdarah adalah sebagai berikut

1. Model univariat

$$y_t = 0,21923 + \varepsilon_t + 0,53024\varepsilon_{t-1} + 1,61305y_{t-1} - 0,76222y_{t-2}$$

$$y_t = \text{VDBDLOG2 pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

2. Model multivariat dengan variabel iklim

$$y_t = c + \varepsilon_t - 0,49893\varepsilon_{t-1} - 0,06692CH_{t-8} + 6,68784HMD_{t-1}$$

$$y_t = \text{VDBDLOG2 pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

$$HMD_t = \text{HMD_log2 pada periode } t$$

3. Model multivariat dengan variabel Google Trends

$$y_t = 0,120258 + \varepsilon_t + 1,18464y_{t-1} - 0,33857y_{t-2} + 0,40698x1_{t-1} + 0,1574x2_{t-2}$$

$$y_t = \text{VDBDLOG2 pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

$$x1_t = \text{search index query "dengue" yang sudah ditransformasi logaritmik (dengue_log2) pada periode } t.$$

$$x2_t = \text{search index query "demam berdarah" yang sudah ditransformasi logaritmik (demamberdarah_log2) pada periode } t.$$

4. Model multivariat dengan variabel Iklim dan Google Trends

$$y_t = -1,7539 + \varepsilon_t + 0,77804y_{t-1} + 6,04271HMD_{t-1} + 0,39142x1_{t-1}$$

$$y_t = \text{VDBDLOG2 pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

$$x1_t = \text{search index query "dengue" yang sudah ditransformasi logaritmik (dengue_log2) pada periode } t$$

$$HMD_t = \text{HMD_log2 pada periode } t$$

Persamaan untuk model ARIMA univariat dan multivariat kasus influenza adalah sebagai berikut :

1. Model univariat

$$y_t = c + \varepsilon_t - 0,76151y_{t-1} - 0,31349y_{t-2} - 0,12029y_{t-3} - 0,33234y_{t-4} - 0,41649y_{t-5} + y_{t-1}$$

$$y_t = \text{VFLU pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

2. Model multivariat dengan variabel iklim

$$y_t = c + \varepsilon_t - 1,03739y_{t-1} - 0,68632y_{t-2} - 0,47759y_{t-3} - 0,68504y_{t-4} - 0,60171y_{t-5} - 0,37865HMD_{t-9} + y_{t-1}$$

$$y_t = \text{VFLU pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

$$HMD_t = \text{HMD pada periode } t$$

3. Model multivariat dengan variabel Google Trends

$$y_t = c + \varepsilon_t - 0,90297y_{t-1} - 0,63271y_{t-2} - 0,57546y_{t-3} - 0,78566y_{t-4} - 0,62196y_{t-5} + 0,16187x3_{t-1} + y_{t-1}$$

$$y_t = \text{VFLU pada periode } t$$

$$\varepsilon_t = \text{random error pada periode } t$$

$$x3_t = \text{search index query "batuk" pada periode } t$$

4. Model multivariat dengan variabel iklim dan Google Trends

$$y_t = c + \varepsilon_t - 1,03457y_{t-1} - 0,73255y_{t-2} - 0,64346y_{t-3} - 0,90542y_{t-4} - 0,73025y_{t-5} - 0,23834HMD_{t-9} + 0,10049x3_{t-1} + y_{t-1}$$

y_t = VFLU pada periode t

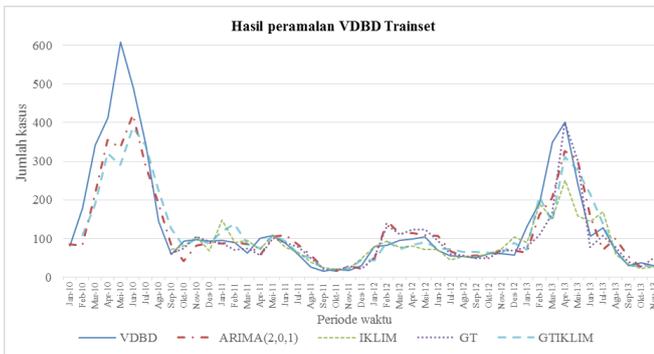
ε_t = random error pada periode t

$x3_t$ = search index query “batuk” yang sudah ditransformasi logaritmis (dengue_log2) pada periode t

HMD_t = HMD pada periode t

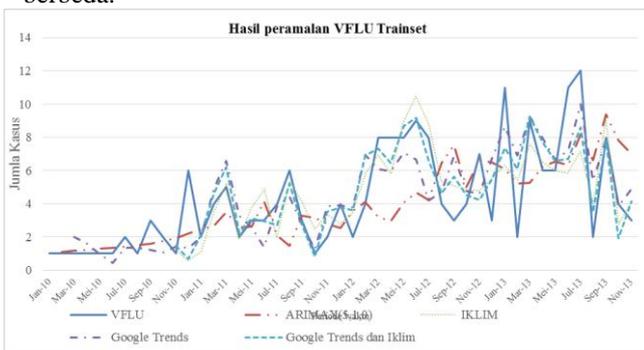
VI. ANALISIS HASIL DAN PEMBAHASAN

Terlihat pada Tabel 6 performa hasil peramalan kasus demam berdarah dengan menggunakan data trainset mengalami peningkatan seiring dengan penambahan variabel eksternal. Model ARIMAX memberikan performa yang lebih baik daripada model ARIMA. Dengan nilai MAPE testset berada di antara 20-50%, model peramalan untuk kasus demam berdarah sudah dikatakan layak, walaupun performanya masih belum baik. Gambar 2 merupakan plot perbandingan hasil peramalan menggunakan data trainset.



Gambar 2 - Hasil Peramalan VDBD trainset

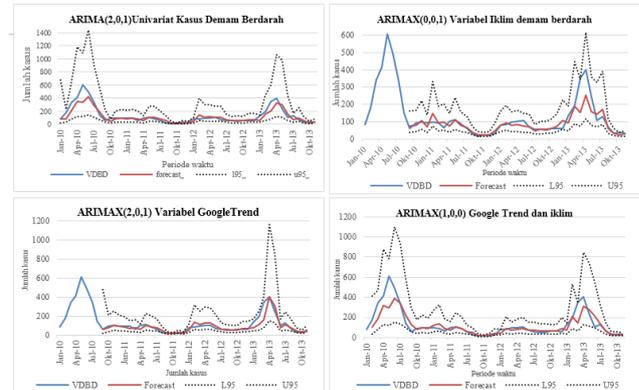
Tabel 7 merupakan performa hasil peramalan kasus influenza terhadap testset. Terlihat pada Tabel 7 performa hasil peramalan kasus influenza dengan menggunakan data trainset mengalami peningkatan seiring dengan penambahan variabel eksternal. Model ARIMAX memberikan performa yang lebih baik daripada model ARIMA. Dengan nilai MAPE testset berada di atas 50%, model peramalan untuk kasus influenza belum layak. Gambar 3 merupakan plot perbandingan hasil peramalan menggunakan data trainset. Hasil MAPE pada testset naik drastis dikarenakan pola data trainset dan testset variabel dependen VFLU yang sangat berbeda.



Gambar 3 - Hasil Peramalan VFLU trainset

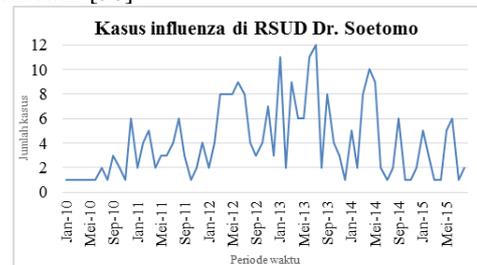
Kasus demam berdarah dan kasus influenza menunjukkan korelasi yang signifikan terhadap curah hujan, suhu udara, kelembaban, dan hasil pencarian di Google Trends. Kekuatan korelasi tersebut beragam, dari yang memiliki korelasi sedang, kuat dan sangat kuat. Namun, kekuatan korelasi variabel eksternal terhadap variabel dependen tersebut tidak berpengaruh dalam pembuatan model ARIMAX.

Hal ini ditunjukkan pada model ARIMAX dengan variabel eksternal Google Trends kasus demam berdarah dan model ARIMAX dengan variabel eksternal iklim kasus influenza. Model yang memenuhi syarat hasil estimasi parameter AR, MA, dan variabel eksternal signifikan adalah model dengan variabel eksternal berkorelasi signifikan sedang atau kuat. Pada penelitian ini terlihat bahwa model ARIMAX dengan performa terbaik kebanyakan menggunakan variabel eksternal dengan kekuatan korelasi kuat atau sangat kuat. Performa hasil peramalan kasus demam berdarah sudah layak, walaupun hasilnya masih belum baik. Model ARIMA univariat dan multivariat untuk kasus demam berdarah sudah layak karena residual dari seluruh modelnya sudah *white noise*, dan data aktual kasus demam berdarah berada di dalam *range 95% confidence interval* dari seluruh model. Hal ini ditampilkan pada Gambar 4.



Gambar 4 - Confidence Interval 95% model ARIMA(X) kasus demam berdarah

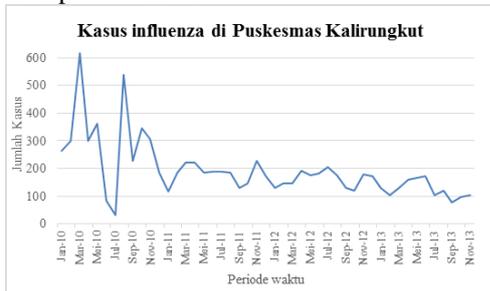
Model ARIMA untuk kasus influenza sudah layak, dilihat dari keberadaan *white noise* pada residual seluruh model, namun performa hasil peramalan model ARIMA untuk kasus influenza tidak layak. Hal ini disebabkan dari data kasus influenza yang digunakan tidak cocok untuk dimodelkan dengan ARIMA. Data kasus influenza yang diperoleh dari RSUD Dr. Soetomo sangat fluktuatif dengan naik dan turun yang tajam. Model ARIMA akan memiliki performa yang baik selama pergerakan datanya bertahap, dan tidak fluktuatif [38].



Gambar 5 - Kasus Influenza di RSUD Dr. Soetomo

Data kejadian penyakit influenza dari RSUD Dr. Soetomo pada penelitian ini memiliki kejanggalan. Rata-rata kasus influenza yang terjadi di RSUD Dr. Soetomo ada 5 kasus setiap bulannya, dengan jumlah kasus terbanyak dalam satu bulan adalah 12 kasus. Bahkan ada pada beberapa bulan hanya terjadi 1 kasus. Jumlah ini sangat berbeda jika

dibandingkan dengan kasus influenza pada puskesmas Kalirungkut Surabaya, dengan jumlah kasus influenza rata-rata tiap bulannya ada 193 kasus, dengan jumlah kasus terbanyak dalam satu bulan sebanyak 615 kasus. Grafik jumlah kasus influenza di Puskesmas Kalirungkut, Surabaya ditunjukkan pada Gambar 6.



Gambar 6 - Kasus Influenza di Puskesmas Kalirungkut Surabaya

Hal ini dikarenakan RSUD Dr. Soetomo merupakan rumah sakit besar yang menjadi rujukan lanjutan bagi pasien-pasien yang tidak bisa ditangani di tempat lain. Penyakit influenza (influenza like illness) merupakan penyakit yang tidak terlalu serius bagi orang untuk sampai berobat ke RSUD Dr. Soetomo. Untuk penelitian ke depan mengenai peramalan penyakit influenza akan lebih baik jika mengambil data dari puskesmas. Dari hasil pemodelan ARIMAX, ditemukan bahwa model ARIMAX dengan variabel iklim memiliki performa rata-rata lebih baik dibandingkan dengan model ARIMAX dengan variabel Google Trends. Tabel 8 merupakan rata-rata MAPE model ARIMAX yang ditemukan pada penelitian ini.

Tabel 8 - Perbandingan nilai MAPE model univariat dengan rata-rata MAPE model ARIMAX performa terbaik

Model	MAPE kasus Demam berdarah	MAPE kasus Influenza
ARIMA	32,34%	54,78%
ARIMAX Iklim	29,30%	50%
ARIMAX Google Trends	30%	54,30%
ARIMAX Google Trends iklim	31%	47,40%

Variabel iklim merupakan hasil pengukuran cuaca dari BMKG, data yang diperoleh berdasarkan kejadian alam yang sebenarnya. Sedangkan variabel Google Trends dipengaruhi oleh sorotan media mengenai suatu penyakit tertentu, sehingga mengundang orang-orang untuk melakukan pencarian di Google mengenai penyakit tersebut [39]. Secara statistik, kedua variabel eksternal ini mampu meningkatkan performa model peramalan.

Pada penelitian-penelitian sebelumnya, hasil pencarian di Google Trends menunjukkan potensi untuk meramalkan kejadian kasus influenza [9] [17]. Data Google Trends dengan lokasi geografis Indonesia masih memiliki keterbatasan, yakni masih belum search index lokal di kota-kota tertentu masih belum tersedia, hanya ada untuk kota-kota besar seperti Surabaya, Jakarta dan Bandung. Tingkatan paling rendah yang bisa diambil hasil search index nya adalah provinsi. Sehingga Google Trend dengan lokasi Indonesia masih belum layak dipakai sebagai data alternatif untuk sistem surveilans penyakit DBD dan Influenza.

VII. KESIMPULAN DAN SARAN

Berdasarkan proses penelitian yang telah dilakukan, berikut ini merupakan kesimpulan yang dapat diambil :

1. Model ARIMA dan ARIMAX demam berdarah dikatakan sudah layak dengan nilai MAPE testset model berada pada jangkauan 19% - 32,34%.
2. Model peramalan demam berdarah dengan performa terbaik adalah model ARIMAX(0,0,1) dengan variabel iklim.
3. Model ARIMAX kasus demam berdarah dengan variabel iklim menunjukkan performa paling baik dalam meramalkan kasus demam berdarah. Model ARIMAX variabel iklim meningkatkan performa peramalan hingga 7,56% lebih baik daripada model ARIMA.
4. Model ARIMA dan ARIMAX influenza dikatakan belum layak dengan nilai MAPE testset model berada pada jangkauan 72% - 102%. Hal ini menunjukkan kasus influenza tidak cocok diramalkan dengan metode ARIMA.
5. Model peramalan influenza dengan performa terbaik adalah model ARIMAX(5,1,0) dengan variabel eksternal iklim dan Google Trends.
6. Walaupun model peramalan influenza masih belum layak, model ARIMAX dengan variabel iklim dan Google Trends bisa memperbaiki performa hingga 19,38%.
7. Variabel eksternal iklim dan Google Trends mampu meningkatkan performa model.
8. Model dengan variabel eksternal iklim menunjukkan performa yang lebih baik dibandingkan dengan model dengan variabel eksternal Google Trends.
9. Google Trend dengan lokasi Indonesia masih belum layak dipakai sebagai data alternatif untuk sistem surveilans penyakit DBD dan Influenza.
10. Pola data yang naik turun secara tajam tidak cocok dimodelkan dengan ARIMA.

Berikut ini merupakan saran untuk penelitian selanjutnya:

1. Pengambilan data untuk kasus influenza lebih baik jika diambil dari seluruh kecamatan di Surabaya
2. Ketersediaan data observasi yang lebih banyak diharapkan mampu menangkap faktor siklus lima tahunan pada kasus demam berdarah yang terjadi di Indonesia belum layak dipakai sebagai data alternatif untuk sistem surveilans penyakit demam berdarah dan Influenza. Namun masih bisa digunakan sebagai variabel eksternal lain untuk meningkatkan performa peramalan.
3. Model ARIMA tidak cocok untuk meramalkan data dengan fluktuasi yang sangat tinggi. Alternatif metode lain untuk meramalkan data dengan fluktuasi tinggi adalah artificial neural network.
4. Penentuan search query pada Google Trends sebaiknya tidak menggunakan hasil related search. Alternatif lainnya adalah menggunakan kuesioner yang menanyakan responden mengenai kata-kata yang terpikirkan oleh mereka saat ingin mencari informasi mengenai penyakit demam berdarah dan influenza.

DAFTAR PUSTAKA

- [1] D. R. Ostojic, R. K. Bose, H. Kramerberck, J. Lim and Y. Zhang, "Surabaya," in *Energizing Green Cities in Southeast Asia : Applying Sustainable Urban Energy and Emissions Planning*, World Bank, 2013, pp. 139-156.
- [2] Badan Penelitian dan Pengembangan Kesehatan, "Penyakit Menular Non-Neglected : kajian dan program penelitian," Kementrian Kesehatan Republik Indonesia, Jakarta, 2013.

- [3] World Health Organization, "Prevention and control of dengue and dengue haemorrhagic fever," World Health Organization, India, 2003.
- [4] Pusat Data dan Informasi Kemenkes RI, "Situasi Demam Berdarah Dengue di Indonesia," Kementerian Kesehatan Republik Indonesia, Jakarta, 2014.
- [5] M. M. Sintorini, "Pengaruh Iklim terhadap Kasus Demam Berdarah Dengue," *Jurnal Kesehatan Masyarakat Nasional*, vol. 2, no. 1, pp. 11-18, 2007.
- [6] C. Sloan, M. L. Moore and T. Hartert, "Impact of pollution, climate, and sociodemographic factors on spatiotemporal dynamics of seasonal respiratory viruses," *Clin Transl Sci*, vol. 4, no. 1, pp. 48-54, 2011.
- [7] Kementerian Kesehatan, *Rencana Strategis Kementerian Kesehatan Tahun 2015-2019*, Jakarta: Kementerian Kesehatan Republik Indonesia, 2015.
- [8] A. F. Dugas, M. Jalalpour, Y. Gel, S. Levin, F. Torcaso, T. Igusa and R. E. Rothman, "Influenza Forecasting with Google Flu Trends," *PLOS ONE*, vol. 8, no. 2, 2013.
- [9] Badan Penelitian dan Pengembangan Kesehatan, "Grafik Surveilans Influenza Nasional Berbasis Virologi (ILI-SIBI)," [Online]. Available: <http://www.litbang.kemkes.go.id/grafik-surveilans-influenza-nasional-berbasis-virologi-ili-sibi/>. [Accessed 29 September 2015].
- [10] S. C. J. M. S. S.-Y. L. J. e. a. Cho S, "Correlation between National Influenza Surveillance Data and Google Trends in South Korea," *PLOS ONE*, vol. 8, no. 12, 2013.
- [11] N. Generous, G. Fairchild, A. Deshpande, S. Y. D. Valle and R. Priedhorsky, "Global Disease Monitoring and Forecasting with Wikipedia," *PLOS ONE*, vol. 10, no. 11, 2014.
- [12] K. S. Hickmann, G. Fairchild, R. Priedhorsky, N. Generous and J. M. H. et al, "Forecasting the 2013-2014 Influenza Season Using Wikipedia," *PLoS Computational Biology*, vol. 11, no. 5, 2015.
- [13] J. Ginsberg, M. H. Mohebbi, R. S. Patel and L. B. et al, "Detecting influenza epidemics using search engine query data," 19 February 2009. [Online]. Available: <http://static.googleusercontent.com/media/research.google.com/en/archive/papers/detecting-influenza-epidemics.pdf>. [Accessed 29 September 2015].
- [14] Google, "Google Dengue Trends," Google, 2015. [Online]. Available: <http://www.google.org/dengu Trends>. [Accessed 29 September 2015].
- [15] Google, "Google Flu Trends," Google, 2015. [Online]. Available: <http://www.google.org/flu Trends>. [Accessed 29 September 2015].
- [16] O. M. Araz, D. Bentley and R. L. Muellman, "Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska," *American Journal of Emergency Medicine*, vol. 32, 2014.
- [17] E. JR, H. AG, B. JS, B. DL and O. D. e. al, "Usefulness of school absenteeism data for predicting influenza outbreaks, United States," *Emerging infectious diseases*, vol. 18, no. 8, 2012.
- [18] P. A, "Comparison : Flu Prescription Sales Data from a Retail Pharmacy in the US with Google Flu Trends and US ILINet(CDC) Data as Flu Activity Indicator," *PLOS ONE*, vol. 7, 2012.
- [19] R. P. Soebiyanto, F. Adimi and R. K. Kiang, "Modelling and Predicting Seasonal Influenza Transmission in Warm Regions Using Climatological Parameters," *PLOS ONE*, vol. 5, no. 3, 2010.
- [20] G. J. Milinovich, S. M. R. Avril, A. C. A. Clements, J. S. Brownstein, S. Tong and W. Hu, "Using internet search queries for infectious disease surveillance : screening diseases for suitability," *BMC Infectious Disease*, vol. 14, 2014.
- [21] World Health Organization, "Influenza," WHO, [Online]. Available: <http://www.who.int/topics/influenza/en/>. [Accessed 4 10 2015].
- [22] World Health Organization, "Influenza (Seasonal) Fact Sheet," WHO, Maret 2014. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs211/en/>. [Accessed 4 Oktober 2015].
- [23] Google, "Google Trends," [Online]. Available: <http://www.google.com/trends>. [Accessed 5 Oktober 2015].
- [24] R. J. Hyndman and G. Athanasopoulos, "Forecasting Principles and Practice," OTexts, [Online]. Available: <https://www.otexts.org/book/fpp>. [Accessed 5 Oktober 2015].
- [25] C. Chatfield, "Basic of Time Series Analysis," in *Time-Series Forecasting*, Florida, CRC Press LLC, 2000, pp. 20-42.
- [26] Bowerman, O'Connel and Koehler, *Forecasting, Time Series and Regression*, Thomson Brooks/Cole, 2005.
- [27] S. Chadsuthi, C. Modchang, Y. Lenbury, S. Iamsirithaworn and W. Triampo, "Modelling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses," *Asian Pacific Journal of Tropical Medicine*, 2012.
- [28] Laerd Statistics, "Spearman's Rank-Order Correlation," [Online]. Available: <https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php>. [Accessed 5 Oktober 2015].
- [29] D. E. Hinkle, W. Wiersma and S. G. Jurs, *Applied Statistics for the Behavioral Sciences 5th Edition*, Boston: Houghton Mifflin, 2003.
- [30] "Dickey-Fuller Unit Root Test (Stasionarity Test)," [Online]. Available: <http://staffweb.hkbu.edu.hk/billhung/econ3600/application/app01/app01.html>. [Accessed 6 Oktober 2015].
- [31] SAS, *SAS/ETS 13.2 User's Guide*, North Carolina: SAS Institute Inc., 2014.
- [32] B. Plante, "R vs SAS, why is SAS preferred by private companies?," 3 September 2013. [Online]. Available: <http://stats.stackexchange.com/questions/33780/r-vs-sas-why-is-sas-preferred-by-private-companies>. [Accessed 7 Februari 2016].
- [33] SAS, "SAS/ETS Software," [Online]. Available: <http://support.sas.com/rnd/app/ets/>. [Accessed 2 Februari 2016].
- [34] Microsoft, "Method to calculate interpolation step value in Excel," 23 September 2015. [Online]. Available: <https://support.microsoft.com/en-us/kb/214096>. [Accessed 5 Desember 2015].
- [35] J. S. Armstrong, *Long-Range Forecasting From Crystal Ball to Computer*, Hoboken: Wiley-Interscience, 1985.
- [36] R. Nau, "Stationarity and differencing," Fuqua School of Business Duke University, 6 November 2015. [Online]. Available: <http://people.duke.edu/~rnau/411diff.htm#firstdiff>. [Accessed 31 Desember 2015].
- [37] D. J. Corliss, "Non-Temporal ARIMA Models in Statistical Research," in *SAS Global Forum 2013*, 2013.
- [38] G. Zhang, B. E. Patuwo and M. Y. Hu, "Forecasting with artificial neural networks : the state of art," *International Journal of Forecasting*, vol. 14, pp. 35-62, 1998.
- [39] A. R. Cook, M. I. Chen and R. T. P. Lin, "Internet Search Limitations and Pandemic Influenza, Singapore," October 2010. [Online]. Available: <http://dx.doi.org/10.3201/eid1610.100840>. [Accessed 13 May 2016].
- [40] D. E. Bowman, R. E. Ortega, M. L. Hamrick, J. R. Spiegel and T. R. Kohn, "Refining search queries by the suggestion of correlated terms from prior searches". Amerika Serikat Patent US6006225 A, 21 December 1999.