TUGAS AKHIR – KS 141501

PERAMALAN JUMLAH PENDERITA PENYAKIT MENULAR INFLUENZA DAN DEMAM BERDARAH DI KOTA SURABAYA BERDASARKAN GOOGLE TRENDS DAN FAKTOR IKLIM MENGGUNAKAN VARIASI METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

LARAS ARISTIANI
NRP 5211 100 025

Dosen Pembimbing :
Wiwik Anggraeni, S.Si, M.Kom.

JURUSAN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2016

TUGAS AKHIR - KS 141501

Forecasting of Influenza and Dengue Fever Cases in Surabaya with Google Trends and Climate Data using variation of Autoregressive Integrated Moving Average (ARIMA) Method

LARAS ARISTIANI
NRP 5212 100 022

SUPERVISOR :
Wiwik Anggraeni, S.Si, M.Kom.

DEPARTMENT OF INFORMATION SYSTEMS
Faculty of Information Technology
Institute of Technology Sepuluh Nopember
Surabaya 2016

LEMBAR PENGESAHAN

PERAMALAN JUMLAH KASUS PENYAKIT MENULAR INFLUENZA DAN DEMAM BERDARAH DI KOTA SURABAYA BERDASARKAN *GOOGLE TRENDS* DAN FAKTOR IKLIM MENGGUNAKAN VARIASI METODE *AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)*

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

LARAS ARISTIANI
NRP. 5212 100 022

Surabaya, 2016
KETUA JURUSAN SISTEM INFORMASI

Dr. Ir. Aris Tjahyanto M.Kom.
NIP. 19650310 199102 1 001

LEMBAR PERSETUJUAN

PERAMALAN JUMLAH KASUS PENYAKIT MENULAR INFLUENZA DAN DEMAM BERDARAH DI KOTA SURABAYA BERDASARKAN *GOOGLE TRENDS* DAN FAKTOR IKLIM MENGGUNAKAN VARIASI METODE *AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)*

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Oleh :

LARAS ARISTIANI
NRP. 5212 100 022

Disetujui Tim Penguji:

Tanggal Ujian : Januari 2016

Periode Wisuda : September 2016

Wiwik Anggraeni S.Si, M.Kom. **(Pembimbing)**

Bekti Cahyo Hidayanto, S.Si.,M.Kom **(Penguji 1)**

Amalia Utamima, S.Kom, MBA **(Penguji 2)**

LEMBAR PERNYATAAN
PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS

Sebagai mahasiswa Institut Teknologi Sepuluh Nopember Surabaya, yang bertanda tangan di bawah ini saya :

Nama : LARAS ARISTIANI

Nrp. : 5212100022

Jurusan / Fak. : SISTEM INFORMASI / FTIf

Alamat kontak : TAMAN PONDOK JATI CO 19, TAMAN. SIDOARJO

a. Email : larasarist@gmail.com

b. Telp/HP : 085693129802

Menyatakan bahwa semua data yang saya *upload* di Digital Library ITS merupakan hasil final (revisi terakhir) dari karya ilmiah saya yang sudah disahkan oleh dosen penguji. Apabila dikemudian hari ditemukan ada ketidaksesuaian dengan kenyataan, maka saya bersedia menerima sanksi.

Demi perkembangan ilmu pengetahuan, saya menyetujui untuk memberikan **Hak Bebas Royalti Non-Ekslusif *(Non-Exclusive Royalti-Free Right)*** kepada Institut Teknologi Sepuluh Nopember Surabaya atas karya ilmiah saya yang berjudul :

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

Dengan Hak Bebas Royalti Non-Ekslusif ini, Institut Teknologi Sepuluh Nopember Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data *(database)*, mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta. Saya bersedia menanggung secara pribadi, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya Ilmiah saya ini tanpa melibatkan pihak Institut Teknologi Sepuluh Nopember Surabaya.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Surabaya

Pada tanggal :

Yang menyatakan,

Dosen Pembimbing 1

WIWIK ANGGRAENI, S.Si, M.Kom

NIP. 197601232001122002

LARAS ARISTIANI

Nrp. 5212100022

PERAMALAN JUMLAH KASUS PENYAKIT MENULAR INFLUENZA DAN DEMAM BERDARAH DI KOTA SURABAYA BERDASARKAN GOOGLE TRENDS DAN FAKTOR IKLIM MENGGUNAKAN VARIASI METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

Nama Mahasiswa : Laras Aristiani
NRP : 5212 100 022
Jurusan : Sistem Informasi FTIf-ITS
Dosen Pembimbing : Wiwik Anggraeni S.Si, M.Kom.

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 surveillans dan meramalkan jumlah penyakit demam berdarah dan influenza.

Data yang digunakan dalam penelitian adalah (1) data jumlah kasus demam berdarah dan influenza yang diambil dari RSUD Dr. Soetomo (2) data faktor iklim yang meliputisuhu, kelembaban, dan curah hujan yang diambil dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasion 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 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 naik turun secara drastis, tidak cocok dimodelkan dengan ARIMA.

Hasil lain yang ditemukan adalah 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***

Forecasting of Influenza and Dengue Fever Cases in Surabaya with Google Trends and Climate Data using variation of Autoregressive Integrated Moving Average (ARIMA) Method

Name	**: Laras Aristiani**
NRP	**: 5212 100 022**
Department	**: Information Systems FTIf-ITS**
Supervisor 1	**: Wiwik Anggraeni S.Si, M.Kom.**

ABSTRACT

Early detection is required in order to manage communicable disease such as dengue fever and influenza. The number of fever and influenza cases are affected by several external factors. There is a correlation between climate factors (such as humidity, rainfall, and temperature) and number of cases in certain area. Other than climate factors, there is a potential external variable that is widely available. Google Trends records the search queries that have been searched in Google. Previous search has shown that certain query search onn Google Trends are highly correlated with number of communicable disease cases in South Korea.

The aim of this final project research us to generate a forecasting model with climate factors and Google Trends as external variables. This research analyzes the feasibility of Google Trends in Indonesia to be used as surveillance system alternatives, and to forecast number of influenza and dengue fever cases.

This research uses several data (1) Number of dengue fever and influenza cases obtained from local general hospital of Dr. Soetomo (2) rainfall, humidity and tempature obtained from Indonesian Agency for Meteorological, Climatological and Geophysics (3) Google Trends search index of certain queries. All of the data spans from December 2010 – August 2015. Interpolation and extrapolation techniques are used to handle the missing data. ARIMA and ARIMAX methods are used to model the forecast.

The result of this research shows that the existence of external variables in dengue fever ARIMAX can improve model performance by 7,56%. It performs better than dengue fever ARIMA. It turns out that ARIMA is not suitable for influenza cases, with the range value of MAPE testset 72% - 102%. It is due to fluctuative influenza data pattern in this research. Although it is not suitable to forecast influenza cases with ARIMA, ARIMAX model with climate factors and Google Trends can improve the performance by 19,38%.

This research shows that climate factors and Google Trends are able to improve model performance. Models with climate factors perform better than Google Trends. Google Trends with Indonesian geographic location is not feasible yet to be used as surveillance system alternative data of dengue fever and influenza, due to high error variance.

Keywords: ***Forecasting, Univariate, Multivariate, ARIMA, ARIMAX, Climate Factors, Google Trends, Influenza, Dengue Fever***

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadirat Tuhan Yang Maha Esa atas segala berkat dan rahmat-Nya lah penulis dapat menyelesaikan buku tugas akhir dengan judul "***PERAMALAN JUMLAH KASUS PENYAKIT MENULAR INFLUENZA DAN DEMAM BERDARAH DI KOTA SURABAYA BERDASARKAN GOOGLE TRENDS DAN FAKTOR IKLIM MENGGUNAKAN VARIASI METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)*"** yang merupakan salah satu syarat kelulusan pada Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Secara khusus penulis akan menyampaikan ucapan terima kasih yang sedalam-dalamnya kepada:

1) Allah SWT yang telah memberikan kesehatan dan kesempatan untuk bisa menyelesaikan tugas akhir ini.
2) Kedua orang tua dan keluarga penulis yang senantiasa memberi dukungan selama masa kuliah.
3) Ibu Wiwik Anggraeni selaku dosen pembimbing yang telah membantu dan membimbing penulis dalam tugas akhir ini.
4) Bapak Holil Noor Ali selaku dosen wali yang memberikan nasihat dan dukungan selama masa kuliah penulis
5) Dosen-dosen yang memberikan bimbingan dan masukan untuk kompetisi yang diikuti oleh penulis; Ibu Hanim Maria Astuti dan Ibu Irmasari Hafidz. Terima kasih atas inspirasi dan semangatnya untuk terus berani mencoba kesempatan baru.
6) Sahabat dari Sistem Informasi angkatan 2012 : Putlen, Aulia, Amel, Anggun, Ika, Asti, Fadhil, Ulvi, Eka, Kiki Rafika, yang telah memberikan warna unik selama masa perkuliahan penulis.
7) Rekan lomba tim Adroit Fortunatus : Tito dan Ayik yang telah berjuang bersama dalam mengikuti

kompetisi nasional pertama penulis, yang mengajarkan untuk tidak menyerah sebelum mencoba

8) Rekan laboratorium RDIB : Rifqi, Ilham, Tiar, Ari, Gifari, Adi, Indi, Yukita, Yogi, Mba Fifi, dan seluruh member laboratorium RDIB.
9) Seluruh teman-teman Sistem Informasi Angkatan 2012 (SOLA12IS).
10) Sahabat penulis pada masa SMA : Hana, Danastri, Tissa, Solo, Almer, Faza dan Muhaji. Terima kasih atas silaturahmi yang masih berlanjut sampai saat ini.
11) Rekan-rekan AIESEC Surabaya : Candra, Neni, Della, Farah, Herlin, Jacqueline, Gashif, dan Rafika yang telah berjuang bersama selama satu tahun mengemban tanggung jawab di organisasi.
12) Bapak ibu dosen Jurusan Sistem Informasi yang telah berbagi ilmu dan pengalaman selama masa studi S1 ini, serta selalu menginspirasi.
13) Seluruh karyawan Jurusan Sistem InformasiTerima kasih atas senyum dan pelayanan yang baik sehingga penulis tidak bosan untuk hadir di kampus tercinta ini.

Terima kasih atas segala bantuan, dukungan, serta doanya. Semoga Tuhan senantiasa memberkati dan membalas kebaikan-kebaikan yang telah diberikan kepada penulis.

Penulis pun menyadari bahwa Tugas Akhir ini masih belum sempurna dengan segala kekurangan di dalamnya. Oleh karena itu penulis memohon maaf atas segala kekurangan yang ada di dalam Tugas Akhir ini dan bersedia menerima kritik dan saran. Semoga Tugas Akhir ini dapat bermanfaat bagi seluruh pembaca.

Surabaya, Juni 2016

DAFTAR ISI

ABSTRAK ii
ABSTRACT iv
KATA PENGANTAR vi
DAFTAR ISI viii
DAFTAR GAMBAR xi
DAFTAR TABEL xiii
BAB I PENDAHULUAN 1
1.1. Latar Belakang Masalah 1
1.2. Rumusan Masalah 5
1.3. Batasan Tugas Akhir 5
1.4. Tujuan Tugas Akhir 5
1.4. Manfaat Penelitian 6
1.5. Relevansi 6
BAB II TINJAUAN PUSTAKA 7
2.1. Studi Sebelumnya 7
2.1. Dasar Teori 11
2.2.1. Influenza 11
2.2.2. Demam Berdarah Dengue 11
2.2.3. Faktor Iklim 12
2.2.4. Google Trends 12
2.2.5. Teknik Peramalan 13
2.2.6. *Autoregressive Integrated Moving Average* (ARIMA) 14
2.2.7. ARIMA dengan variabel eksternal (ARIMAX) 16
2.2.8. *Autocorrelation Function* (ACF) & *Partial Autocorrelation Function* (PACF) 16
2.2.9. *Cross Correlation Function* (CCF) 19
2.2.10. Uji Korelasi Spearman 20
2.2.11. Uji Stasioneritas, ADF (*Augmented Dickey Fuller*) 21
2.2.12. Uji Ljung Box 21
2.2.13. MAPE 22
2.2.14. SAS Software 22

BAB III METODOLOGI PENELITIAN 23
3.1. Studi Literatur 24
3.2. *Pre processing data* 24
3.3. Pembuatan Model ARIMA Univariat 24
3.4. Pembuatan Model ARIMA Multivariat 25
3.5. Evaluasi model berdasarkan MAPE 29
3.6. Melakukan peramalan terhadap data testing 29
3.7. Validasi hasil peramalan 29
3.8. Analisis Hasil 29
3.9. Penyusunan laporan Tugas Akhir 29
BAB IV PERANCANGAN 31
4.1. Pengumpulan dan deskripsi data 31
4.2. *Pre-processing Data* 35
4.2.1. Perubahan Satuan Waktu Menjadi Per Bulan . 35
4.2.2. Penanganan *missing value* 35
4.2.3. Pembagian Trainset dan Testset 37
BAB V PEMODELAN 39
5.1. Model Univariat 39
5.1.1. Model ARIMA Univariat untuk Kasus Demam Berdarah 39
5.1.2. Model ARIMA Univariat untuk Kasus Influenza 48
5.1.3. Evaluasi dan Pemilihan Model Univariat 53
5.2. Model Multivariat 57
5.2.1. Model ARIMA Multivariat (ARIMAX) untuk Kasus Demam Berdarah 57
5.2.2. Model ARIMA Multivariat (ARIMAX) untuk Kasus Influenza 76
5.2.3. Evaluasi Model Multivariat 92
5.3. Rangkuman dan persamaan hasil pemodelan 97
5.3.1. Kasus demam berdarah 97
5.3.2. Kasus influenza 101
BAB VI HASIL DAN PEMBAHASAN 107
6.1. Hasil peramalan 107
6.2. Analisis hasil dan diskusi 111
BAB VII KESIMPULAN DAN SARAN 119
7.1 Kesimpulan 119

7.2 Saran 120
DAFTAR PUSTAKA 121
LAMPIRAN 126
BIODATA PENULIS 137

DAFTAR GAMBAR

Gambar 2.1 - Pencarian di Google Trends 13
Gambar 2.2 - Nilai ACF AR(1) 17
Gambar 2.3 - Nilai PACF AR(1) 17
Gambar 2.4 - Nilai ACF AR(2) 18
Gambar 2.5 - Nilai PACF AR(2) 18
Gambar 2.6 - Nilai PACF MA(1) 18
Gambar 2.7 - Nilai ACF MA(2) 19
Gambar 2.8 - Nilai PACF MA(2) 19
Gambar 2.9 - Nilai ACF MA(2) 19
Gambar 3.1 - Alur pengerjaan tugas akhir 23
Gambar 3.2 - Pembuatan Model ARIMA Univariat 26
Gambar 3.3 - Pembuatan Model Multivariat 27
Gambar 3.4 - Tahapan pemodelan ARIMAX 28
Gambar 4.1 - Plot VDBD 32
Gambar 4.2 - Plot HMD 33
Gambar 4.3 - Plot VFLU 33
Gambar 4.4 - Plot SUHU 33
Gambar 4.5 - Plot CH 33
Gambar 4.6 - Plot GTDBD 34
Gambar 4.7 - Plot GTFLU 34
Gambar 4.8 - Contoh Missing Value pada Google Trends 36
Gambar 5.1 - Plot VDBD 40
Gambar 5.2 - Plot VDBD setelah didiferensiasi 1 kali 40
Gambar 5.3 - Plot VDBD setelah didiferensiasi 2 kali 41
Gambar 5.4 - Plot VDBDLOG2 42
Gambar 5.5 - Plot VDBDLOG2 setelah didiferensiasi 1 kali 42
Gambar 5.6 - PACF VDBDLOG2 d=0 44
Gambar 5.7 - PACF VDBDLOG2 d=1 44
Gambar 5.8 - ACF VDBDLOG2 d=0 45
Gambar 5.9 - ACF VDBDLOG2 d=1 45
Gambar 5.10 - Plot VFLU 48
Gambar 5.11 - Plot VFLU setelah didiferensiasi 1 kali 49
Gambar 5.12 - PACF VFLU d=1 50
Gambar 6.1 - Hasil peramalan VDBD trainset 108
Gambar 6.2 - Hasil peramalan VFLU trainset 110

Gambar 6.3 - Confidence interval 95% model ARIMA kasus demam berdarah 114
Gambar 6.4 - Kasus influenza di RSUD Dr. Soetomo 115
Gambar 6.5 - Kasus influenza di Puskemas Kalirungkut Surabaya 115
Gambar 6.6 - Related search suggestion pada Google Trends 117

DAFTAR TABEL

Tabel 2.1 - Penelitian Sebelumnya 7
Tabel 2.2 - Interpretasi Nilai Koefisien Spearman Correlation 20
Tabel 2.3 - Penafsiran nilai MAPE 22
Tabel 4.1 - Data mentah yang digunakan 31
Tabel 4.2 - Keterangan *search query* dengan *missing value* 36
Tabel 4.3 - MAPE hasil ekstrapolasi variabel Google Trends 37
Tabel 5.1 - Hasil Uji Stasioneritas VDBD 40
Tabel 5.2 - Hasil Uji Stasioneritas VDBDLOG2 42
Tabel 5.3 - Hasil Uji Stasioneritas VDBDLOG2 setelah didiferensiasi 1 kali 43
Tabel 5.4 - Identifikasi nilai parameter p,d, dan q ARIMA VDBDLOG2 46
Tabel 5.5 - Identifikasi Model ARIMA(p,d,q) VDBDLOG2 46
Tabel 5.6 - Estimasi parameter VDBDLOG2 47
Tabel 5.7 - Uji Stasioneritas VFLU 48
Tabel 5.8 - Uji stasioneritas VFLU d=1 49
Tabel 5.9 - Identifikasi nilai parameter p,d, dan q ARIMA VFLU 51
Tabel 5.10 - Identifikasi model ARIMA(p,d,q) VFLU 51
Tabel 5.11 - Estimasi parameter VFLU 52
Tabel 5.12 - Uji kelayakan model ARIMA demam berdarah 54
Tabel 5.13 - Uji kelayakan model ARIMA kasus influenza 54
Tabel 5.14 - Hasil peramalan ARIMA(p,d,q) univariat untuk kasus demam berdarah 55
Tabel 5.15 - Hasil peramalan ARIMA(p,d,q) univariat untuk kasus influenza 56
Tabel 5.16 - Kandidat variabel eksternal ARIMAX demam berdarah 58
Tabel 5.17 - uji korelasi kandidat variabel eksternal ARIMAX demam berdarah 60
Tabel 5.18 - Variabel dengan korelasi signifikan terhadap VDBDLOG2 61
Tabel 5.19 - Variabel eksternal iklim demam berdarah 63
Tabel 5.20 - Contoh kombinasi variabel iklim 64

Tabel 5.21 - Variabel eksternal Google Trends demam berdarah 65
Tabel 5.22 - Contoh kombinasi variabel iklim 65
Tabel 5.23 - Variabel eksternal Google Trends dan Iklim kasus demam berdarah 66
Tabel 5.24 - Kombinasi Google Trends dan Iklim kasus demam berdarah 67
Tabel 5.25 - Hasil Estimasi ARIMAX variabel iklim kasus demam berdarah 69
Tabel 5.26 - Hasil Estimasi ARIMAX variabel Google Trends kasus demam berdarah 71
Tabel 5.27 - Hasil estimasi parameter ARIMAX dengan Variabel Google Trends dan Iklim kasus demam berdarah....72
Tabel 5.28 - Uji white noise ARIMAX variabel eksternal Google Trends kasus demam berdarah.....73
Tabel 5.29 - Uji white noise ARIMAX variabel eksternal iklim kasus demam berdarah 74
Tabel 5.30 - Uji white noise ARIMAX variabel eksternal Google Trends dan iklim kasus demam berdarah.....75
Tabel 5.31 - Kandidat variabel eksternal ARIMAX influenza 76
Tabel 5.32 - uji korelasi kandidat variabel eksternal ARIMAX influenza 78
Tabel 5.33 - Variabel dengan korelasi signifikan terhadap VFLU.....79
Tabel 5.34 - Variabel eksternal iklim influenza 81
Tabel 5.35 - Contoh kombinasi variabel iklim kasus influenza 81
Tabel 5.36 - Variabel eksternal Google Trends influenza 82
Tabel 5.37 - Contoh kombinasi variabel Google Trends kasus influenza 83
Tabel 5.38 - Variabel eksternal Google Trends dan Iklim kasus influenza 84
Tabel 5.39 - Kombinasi Google Trends dan Iklim kasus influenza 84
Tabel 5.40 - Hasil Estimasi ARIMAX variabel iklim kasus influenza 86

Tabel 5.41 - Hasil Estimasi ARIMAX variabel Google Trends kasus influenza 87
Tabel 5.42 - Hasil estimasi parameter ARIMAX dengan Variabel Google Trends dan Iklim kasus influenza 88
Tabel 5.43 - Uji white noise ARIMAX variabel eksternal iklim kasus influenza 90
Tabel 5.44 - Uji white noise ARIMAX variabel eksternal Google Trends kasus influenza 91
Tabel 5.45 - Uji white noise ARIMAX variabel eksternal Google Trends dan iklim kasus influenza 92
Tabel 5.46 - Performa model ARIMAX variabel iklim kasus demam berdarah 93
Tabel 5.47 - Performa model ARIMAX variabel Google Trends kasus demam berdarah 94
Tabel 5.48 - Performa model ARIMAX variabel Google Trends dan iklim kasus demam berdarah 94
Tabel 5.49 - Performa model ARIMAX variabel iklim kasus influenza 95
Tabel 5.50 - Performa model ARIMAX variabel Google Trends kasus influenza 96
Tabel 5.51 - Performa model ARIMAX variabel Google Trends dan iklim kasus influenza 97
Tabel 5.52 - Rangkuman pemodelan kasus demam berdarah 97
Tabel 5.53 - Rangkuman pemodelan kasus influenza 101
Tabel 6.1 - Performa hasil peramalan kasus demam berdarah 107
Tabel 6.2 - Performa hasil peramalan kasus influenza 109
Tabel 6.3 - Model ARIMAX dengan variabel eksternal kekuatan korelasi sedang 111
Tabel 6.4 - Model dengan performa terbaik dan keterangan korelasi 112
Tabel 6.5 – Perbandingan nilai MAPE model univariat dengan rata-rata MAPE model ARIMAX performa terbaik 116

BAB I
PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang, rumusan masalah, batasan pengerjaan, tujuan, dan manfaat dari pengerjaan tugas akhir.

1.1. Latar Belakang Masalah

Surabaya merupakan kota terbesar kedua di Indonesia dan menjadi pusat perekonomian di Indonesia Timur dengan laju pertumbuhan penduduk tahunan sebesar 0,65 [1]. Berdasarkan hasil sensus penduduk tahun 2010, jumlah penduduk kota Surabaya sebesar 2.885.385. Dengan jumlah penduduk sebesar ini diperlukan fasilitas kesehatan yang memadai. Kota Surabaya memiliki 3 Rumah Sakit Umum Daerah, dengan salah satu yang terbesar adalah RSUD Dr.Soetomo yang setiap harinya melayani setidaknya 3000 pasien sehari. Untuk menangani pasien dalam jumlah besar pihak manajemen perlu melakukan perencanaan yang baik, salah satu caranya adalah dengan melakukan pendeteksian dini penyakit yang ada.

Penyakit menular disebabkan oleh mikroorganisme patogenik contohnya bakteri, virus, parasit dan jamur. Penyakit ini bisa ditularkan secara langsung atau tidak langsung dari penderita ke orang lain yang berada di sekitarnya melalui berbagai macam media, bisa melalui udara, vektor berupa nyamuk, dsb. Di Indonesia penyakit menular yang menjadi perhatian Kementrian Kesehatan Republik Indonesia diantaranya adalah Malaria, Influenza, Demam Berdarah Dengue (DBD), AIDS (HIV), TBC, disingkat dengan MADAT [2]. Penyakit-penyakit tersebut menjadi sorotan karena jumlah kasus dan dampak yang diberikan cukup tinggi.

Indonesia merupakan peringkat kedua di dunia yang memiliki kasus DBD terbanyak [3]. Berdasarkan data dari Kementrian 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 kasus penyakit DBD, dan Influenza setiap tahunnya bergerak secara musiman. Kecenderungan untuk meningkat pada saat musim hujan dan musim peralihan dari musim kemarau ke musim hujan, sehingga terdapat unsur seasonalitas pada angka kasus masing-masing penyakit. Angka kejadian DBD dan Influenza dipengaruhi oleh faktor iklim curah hujan, suhu dan kelembaban udara pada suatu daerah [5, 6]. Fenomena penyakit menular perlu perhatian khusus, sehingga pengendalian penyakit menular merupakan salah satu dari rencana pembangunan jangka menengah bidang kesehatan oleh Kementrian Kesehatan RI [7]. 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 [8]. Pendeteksian dini dapat diperoleh dengan menggunakan teknik peramalan.

Kegiatan surveilans kesehatan di Indonesia sudah diatur oleh Kementrian Kesehatan Republik Indonesia di Peraturan Menteri Kesehatan Republik Indonesia Nomor 45 Tahun 2014 tentang penyelenggaraan surveilans kesehatan. Salah satu contohnya adalah Grafik Surveilans Influenza Nasional Berbasis Virologi (ILI-SIBI) yang menampilkan grafik influenza positif pada beberapa minggu secara nasional [9]. Surveilans ini dilakukan berdasarkan data yang dikumpulkan oleh kementrian kesehatan dari puskesmas dan rumah sakit seluruh Indonesia.

Surveilans konvensional yang dilakukan di Indonesia bergantung pada laporan dari puskesmas dan rumah sakit untuk melacak aktivitas penyakit, yang berisiko keterlambatan pelaporan dan konfirmasi kasus. Hal ini bisa menunda pendeteksian dini dari persebaran penyakit, dan hasil peramalan yang dilakukan tidak dihasilkan sesuai dengan waktunya.

Sehingga diperlukan sumber data alternatif lainnya dan *real-time tools* untuk memantau persebaran penyakit [10].

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 [11, 12]. 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 [13, 14, 15]. 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.

Google Trends merupakan alternatif data yang bisa digunakan sebagai pendukung sistem surveilans dan prediksi jumlah penyakit menular karena data yang dihasilkan bersifat mudah diperoleh dan *real-time*. Sungjin Cho et.al (2013) menganalisis korelasi antara data surveilans nasional dan Google Trend di Korea Selatan. Hasil dari penelitian ini menunjukan 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 [16]. Araz et.al (2014) menggunakan data Google Flu Trend untuk meramalkan jumlah pasien *influenza-like-illness* yang datang ke ruang darurat rumah sakit [16]. Kedua penelitian menggunakan metode *Autoregressive Moving Average* (ARMA) dengan variasi yang berbeda. Hasil dari keduanya menunjukan 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 [17], data penjualan obat di apotek [18] dan data iklim (suhu, kelembaban, curah hujan, dan tekanan udara) [19]. 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 [19].

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 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 multivariat dengan inputan variabel eksternal berupa faktor iklim dan data Google Trend.

1.2. Rumusan Masalah

Rumusan masalah yang akan dititik beratkan dalam pengerjaan tugas akhir ini meliputi :

1. Bagaimana model ARIMA univariat yang sesuai untuk meramalkan DBD dan Influenza berdasarkan kejadian positif penyakit pada periode sebelumnya?
2. Bagaimana model ARIMA multivariat yang sesuai untuk meramalkan penyakit DBD dan Influenza berdasarkan kejadian positif penyakit, temperatur, kelembaban, dan curah hujan?
3. Bagaimana model ARIMA multivariat yang sesuai untuk meramalkan penyakit DBD dan Influenza berdasarkan data dari Google Trend?
4. Bagaimana model ARIMA multivariat yang sesuai untuk meramalkan penyakit DBD dan Influenza berdasarkan data Google Trend, temperatur, kelembaban, dan curah hujan?
5. Model manakah yang memiliki performa yang paling baik untuk meramalkan jumlah penyakit DBD dan Influenza?

1.3. Batasan Tugas Akhir

Batasan dalam pengerjaan tugas akhir ini adalah

1. Data jumlah kasus bulanan DBD dan Influenza yang digunakan diambil dari RSUD Dr. Soetomo periode Desember 2010 – Agustus 2015
2. Data harian temperatur, kelembaban, dan curah hujan diambil dari database online Badan Meteorologi Klimatologi dan Geofisika, diambil dari stasiun Juanda Surabaya dan Perak 1 Surabaya periode Desember 2010 – Agustus 2015.
3. Data mingguan Google Trend diambil dengan query yang sudah ditentukan sebelumnya, dengan jangka waktu Desember 2010-Agustus 2015. Data Google Trends diunduh pada tanggal 19-21 September 2016.

1.4. Tujuan Tugas Akhir

Tujuan dari tugas akhir ini adalah

1. Mencari model ARIMA univariat dan multivariat yang memiliki performa terbaik untuk meramalkan jumlah penyakit DBD, dan Influenza

2. Mengetahui kelayakan data Google Trend dengan lokasi Indonesia untuk dipakai sebagai data alternatif untuk sistem surveilans dan meramalkan jumlah penyakit DBD 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.

1.4. Manfaat Penelitian

Tugas akhir ini diharapkan dapat memberikan manfaat adalah membantu dinas kesehatan kota Surabaya dalam pengambilan keputusan dalam menangani penyakit menular seperti DBD dan Influenza dan memberikan usulan alternatif sistem surveilans dengan sumber data Google Trend dan data iklim. Manfaat bagi akademik adalah sebagai penelitian potensi teknologi informasi di bidang kesehatan dengan memanfaatkan jejak yang ada di internet.

1.5. Relevansi

Relevansi penelitian tugas akhir ini adalah untuk mengantisipasi jumlah penyakit menular khususnya demam berdarah dan influenza pada perubahan musim. Antisipasi tersebut berguna untuk tindakan pencegahan penyebaran penyakit menular. Penelitian ini secara tidak langsung menganalisis *health seeking behavior* para pengguna internet, dengan menggunakan variabel eksternal jejak internet yakni Google Trends. Penelitian ini menganalisis kelayakan penggunaan Google Trends dengan lokasi geografis Jawa Timur, Indonesia untuk sistem surveillans secara *real time*.

BAB II
TINJAUAN PUSTAKA

Pada bab ini dijelaskan mengenai teori-teori terkait yang bersumber dari buku, jurnal, ataupun artikel yang berfungsi sebagai dasar dalam melakukan pengerjaan tugas akhir agar dapat memahami konsep atau teori penyelesaian permasalahan yang ada.

2.1. Studi Sebelumnya

Tabel 2.1 - Penelitian Sebelumnya

No	Judul	Sumber Data	Metode	Hasil Temuan
1.	*Influenza Forecasting with Google Flu Trends* Dugas et.al (2013) [8]	• Data jumlah pasien di UGD dengan keluhan seperti influenza tahun 2004-2011 • Data Google Flu Trends • Data Suhu lokal • Data kelembaban relatif • *Julian weeks*	• *Negative Binomial Generalized Autoregressive Moving Average* (GARMA) • *Negative Binomial Generalized Linear Models (GLM)*	Dengan menggunakan model peramalan GARMA(3,0) dengan distribusi *negative binomial* digabungkan dengan data Google Flu Trends menghasilkan prediksi yang paling akurat dibandingkan dengan model lainnya.

No	Judul	Sumber Data	Metode	Hasil Temuan
2.	*Modelling and Predicting Seasonal Influenza Transmission in Warm Regions Using Climatological Parameters* Soebiyanto et.al (2010) [19]	• Data kejadian positif influenza di Hongkong dan Maricopa County Januari 2005 – September 2008 • Data curah hujan, LST (*Land Surface Temperature*), kelembaban relatif	• Membandingkan performa masing-masing variasi model ARIMA yang dibentuk	Model ARIMA yang dilengkapi dengan variabel iklim (suhu, curah hujan dan kelembaban) memiliki performa prediksi yang lebih baik.
3.	*Using Google Flu Trends Data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska*	• Data Google Flu Trends • Data jumlah kunjungan UGD dengan keluhan seperti influenza di Omaha,	• Seasonal ARIMA untuk data jumlah kunjungan UGD • Model Holt-Winters untuk data jumlah kunjungan UGD	Model regresi melibatkan bisa menghasilkan prediksi dengan performa yang lebih baik dibandingkan dengan SARIMA dan Holt Winters dengan

No	Judul	Sumber Data	Metode	Hasil Temuan
	Araz et.al (2014) [16]	Nebraska tahun 2008-2012 • Data surveilans influenza departemen kesehatan Nebraska tahun 2008-2012	• Regresi Linear untuk data Google Flu Trends dan data dari surveilans	menggunakan data jumlah kunjungan UGD.
4.	*Using internet search queries for infectious disease surveillance: screening disease for suitability* Milinovich et.al (2014) [20]	• Data jumlah kejadian 64 jenis penyakit menular di Australia • Data 164 hasil pencarian Google Trends terkait penyakit menular di Australia	• Mengkorelasikan data kejadian penyakit menular dengan data hasil pencarian di Google Trends • Spearman's rank correlation • Time series cross correlations	17 jenis penyakit ternyata berkorelasi secara signifikan terhadap hasil pencarian di Google Trends
5.	*Correlation between National Influenza*	• Data surveilans ILI dan virologis dari	• Pearson Correlation	Data Google Trend dalam Bahasa Korea bisa digunakan untuk data

No	Judul	Sumber Data	Metode	Hasil Temuan
	Surveillance Data and Google Trends in South Korea Cho et.al (2013) [10]	Korea Centers for Disease Control and Prevention • Data hasil pencarian terkait dengan penyakit influenza di Google Trends	• Lag Correlation Analysis	pelengkap untuk surveilans influenza, namun belum cukup untuk dibuat sebagai model prediktif.

2.1. Dasar Teori

2.2.1. Influenza

Influenza merupakan infeksi virus yang mempengaruhi hidung, tenggorokan, bronkus dan paru-paru. Biasanya infeksi terjadi selama satu minggu. Gejala dari influenza adalah demam tinggi yang tiba-tiba, nyeri otot, pusing, batuk, radang tenggorokan, peradangan pada selaput lendir hidung. Virus influenza bisa ditularkan antar manusia dnegan mudah, melalui partikel kecil yang dihasilkan ketika penderita batuk atau bersin. Penyakit influenza akan menular secara cepat pada musim tertentu [21]. Menurut WHO, secara global epidemi tahunan influenza diperkirakan mencapai angka 3 sampai 5 juta kasus dan 250.000 sampai 500.000 kematian [22]. Di Indonesia, kasus infeksi influenza musiman berlangsung terus sepanjang tahun yang mengikuti pola sirkulasi virus influenza musiman [2]. Data kejadian influenza pada kota Surabaya menjadi variabel inputan pada penelitian ini.

2.2.2. Demam Berdarah Dengue

Demam berdarah dengue (DBD) adalah penyakit yang disebabkan oleh virus yang ditularkan ke manusia mlalui nyamuk *Aedes*. Dengue merupakan salah satu masalah kesehatan di dunia yang terus meningkat dalam 3 dekade terakhir. Diperkirakan 50 juta infeksi dengue terjadi di dunia seacara berkala. Indonesia merupakan peringkat kedua di dunia yang memiliki kasus DBD terbanyak [3]. Berdasarkan data dari Kementrian 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]. Data kejadian demam berdarah pada kota Surabaya menjadi variabel inputan pada penelitian ini.

2.2.3. Faktor Iklim

Soebiyanto et.al (2010) melakukan peramalan penyakit influenza dengan menggunakan faktor iklim sebagai inputan

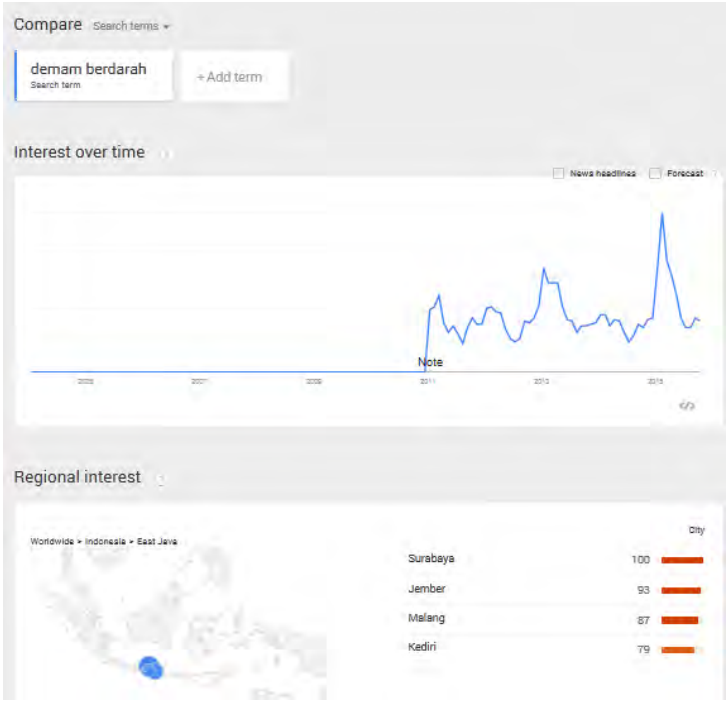
variabel eksternal selain data historis penyakit tersebut. Hasilnya menunjukkan bahwa peramalan yang dilakukan dengan model dengan faktor iklim memiliki performa yang lebih baik [19]. Pada penelitian ini faktor iklim yang dilibatkan sebagai inputan variabel eksternal meliputi :

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

2.2.4. Google Trends

Google Trends merupakan layanan dari Google Inc. yang menunjukan seberapa sering sebuah kata-kata tertentu dicari dengan menggunakan Google. 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 [23]. Kelebihan dari Google Trends adalah pembagian daerah sesuai dengan IP lokasi asal request berasal. Gambar 2.1 menunjukan contoh hasil pencarian yang dilakukan pada Google Trends untuk daerah Provinsi Jawa Timur.

Pada penelitian ini data Google Trend digunakan sebagai variabel inputan eksternal.



Gambar 2.1 - Pencarian di Google Trends

2.2.5. 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 [24]. Secara umum terdapat dua metode peramalan yakni kualitatif dan kuantitatif. Pada penelitian ini metode yang digunakan adalah metode kuantitatif univariat dan multivariat.

2.2.5.1. Peramalan Univariat

Metode peramalan *univariate* adalah prosedur untuk menghitung titik peramalan hanya berdasarkan nilai sekarang dan nilai masa lalu dari sebuah data, yang biasanya dilengkapi dengan fungsi waktu seperti tren linear [25]. Data historis dianalisa dengan tujuan untuk melihat pola data dengan asumsi

pola data tersebut akan berkelanjutan di masa depan. Model peramalan univariat akan sangat berguna apabila kondisi yang diharapkan akan tetap sama tanpa ada perubahan dari variabel eksternal lainnya [26]. Salah satu model yang bisa digunakan untuk melakukan metode peramalan univariat adalah ARIMA.

2.2.5.2. Peramalan Multivariat

Peramalan multivariat atau disebut juga dengan model peramalan kausal melibatkan variabel lain yang berkaitan dengan variabel yang akan diramalkan. Peramalan multivariat dilakukan dengan asumsi ada variabel lain yang mempengaruhi hasil dari peramalan [26]. Salah satu model yang bisa digunakakn untuk melakukan metode multivariat adalah ARIMA dengan variabel eksternal, *Autoregressive Integrated Moving Average with Exogenous Variables* (ARIMAX).

2.2.6. *Autoregressive Integrated Moving Average* (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

$$\underset{\uparrow}{(1-\phi_1 B-\cdots-\phi_p B^p)}\quad \underset{\uparrow}{(1-B)^d y_t}\quad =\quad c+\underset{\uparrow}{(1+\theta_1 B+\cdots+\theta_q B^q)}e_t$$

$$\text{AR}(p) \qquad d \text{ differences} \qquad \text{MA}(q)$$

Persamaan 2.1 - 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.

Penelitian ini akan menggunakan variasi model ARIMA yakni ARIMA(p,d,q), dan ARIMAX(p,d,q).

2.2.6.1. Model *Autoregressive* (AR)

Pada model *autoregressive* (AR), peramalan dilakukan berdasarkan kombinasi linear dari nilai data pada periode sebelumnya [24]. Istilah *autoregeressive* mengindekasikan regresi terhadap dirinya sendiri. Sehingga model *autoregressive* dengan orde p bisa dituliskan dengan :

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

Persamaan 2.2 - Rumus Model AR

Dengan Y_t adalah variabel respon pada waktu t, Y_{t-k} adalah observasi pada waktu t-k, Φ_i merupakan nilai koefisien regresi yang akan diestimasi, dan e_t adalah error pada waktu t. Model ini disebut dengan AR(p) model. Untuk model dengan nilai p=1, AR(1) dan p=2, AR(2) berikut adalah batasan untuk nilai koefisien regresi yang diestimasi (Φ_i)

- Untuk model AR(1) : $-1 < \Phi_1 < 1$
- Untuk model AR(2) : $-1 < \Phi_2 < 1$, $-1 < \Phi_1 + \Phi_2 < 1$, $\Phi_2 - \Phi_1 < 1$

Untuk p>=3 batasan dari nilai koefisien regresi lebih kompleks, dan biasanya pada tools ARIMA (contoh : SPSS, Minitab, R) bisa dianalisis langsung.

2.2.6.2. Model *Moving Average* (MA)

Pada model *autoregressive* (AR), peramalan dilakukan berdasarkan kombinasi linear dari nilai eror pada observasi periode sebelumnya [24]. Untuk orde ke-q pada model MA dirumuskan dengan :

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q}$$

Persamaan 2.3 - Model MA

Dengan Y_t adalah variabel respon pada waktu t, μ adalah rata-rata konstan dari proses, ε_t adalah observasi eror data pada waktu t, ω_q merupakan nilai koefisien regresi yang akan diestimasi. Model ini disebut dengan MA(q) model. Untuk model dengan nilai q=1, MA(1) dan q=2, MA(2) berikut adalah batasan untuk nilai koefisien regresi yang diestimasi (ω_q)

- Untuk model MA(1) : $-1 < \omega_1 < 1$
- Untuk model MA(2) : $-1 < \omega_2 < 1$, $-1 < \omega_2 + \omega_1 < 1$, $\omega_1 - \omega_2 < 1$

2.2.7. ARIMA dengan variabel eksternal (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. Pada model ARIMA, estimasi parameter berdasarkan analisis ACF dan PACF, sedangkan pada model ARIMAX, estimasi parameter dilakukan berdasarkan CCF (*Cross Correlation Function*), ACF dan PACF [27].

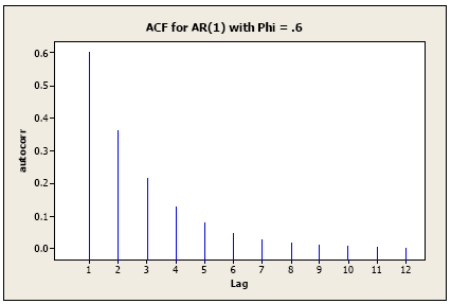
2.2.8. *Autocorrelation Function* (ACF) & *Partial Autocorrelation Function* (PACF)

Dalam menentukan parameter model AR dan MA dilakukan dengan melihat hasil korelogram ACF dan PACF. Contoh-contoh analisis yang dilakukan untuk menentukan parameter model ARIMA dengan melihat PACF dan ACF adalah sebagai berikut

2.2.8.1. Parameter AR(p)

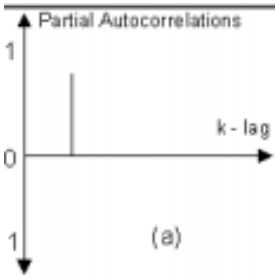
Analisis karakteristik plot ACF dan PACF dari AR(1) adalah sebagai berikut

- Hasil plot nilai ACF berupa eksponensial, seperti pada *Gambar 2.2*



Gambar 2.2 - Nilai ACF AR(1)

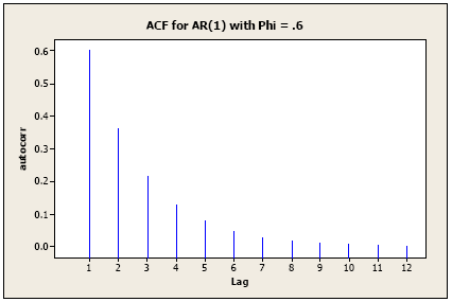
- Hasil plot nilai PACF memiliki 1 nilai yang berbeda secara signifikan



Gambar 2.3 - Nilai PACF AR(1)

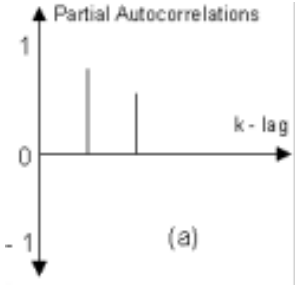
Analisis karakteristi plot ACF dan PACF dari AR(2) adalah sebagai berikut :

- Hasil plot nilai ACF berupa eksponensial, seperti pada Gambar 2.4



Gambar 2.4 - Nilai ACF AR(2)

- Hasil plot nilai PACF memiliki 2 nilai yang berbeda secara signifikan

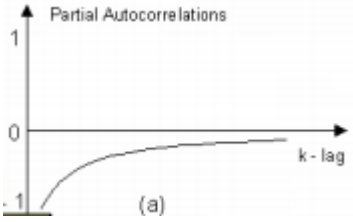


Gambar 2.5 - Nilai PACF AR(2)

2.2.8.2. Parameter MA(q)

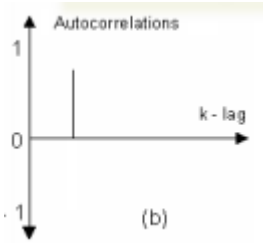
Analisis karakteristik plot ACF dan PACF dari MA(1) adalah sebagai berikut

- Hasil plot nilai PACF berupa eksponensial, ditunjukkan pada Gambar 6.



Gambar 2.6 - Nilai PACF MA(1)

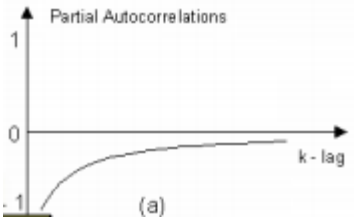
- Hasil plot nilai ACF memiliki 1 nilai yang berbeda secara signifikan, ditunjukkan pada Gambar 7



Gambar 2.7 - Nilai ACF MA(2)

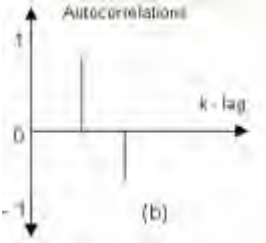
Analisis karakteristik plot ACF dan PACF dari MA(2) adalah sebagai berikut

- Hasil plot nilai PACF berupa eksponensial, seperti pada



Gambar 2.8 - Nilai PACF MA(2)

- Hasil plot nilai ACF memiliki 2 nilai yang berbeda secara signifikan



Gambar 2.9 - Nilai ACF MA(2)

2.2.9. *Cross Correlation Function* (CCF)

Analisis terhadap CCF dilakukan untuk mengetahui korelasi respon antar variabel pada sebuah model. Semakin besar nilai CCF menunjukkan semakin besar keterkaitan antar variabel. Pada penelitian ini analisis CCF digunakan untuk mencari *time lag* yang signifikan dengan P-value<0.05 [27]

antara data historis kejadian penyakit dengan data faktor iklim dan data Google Trend.

2.2.10. Uji Korelasi Spearman

Uji korelasi Spearman digunakan untuk mengukur kekuatan asosiasi antara 2 jenis variabel. Terdapat dua asumsi yang dijalankan untuk melakukan uji korelasi Spearman, yang pertama dua variabel yang akan diuji berupa ordinal, inteval atau rasio. Asumsi kedua adalah terdapat hubungan monoton antara dua variabel tersebut [28]. Secara rumus uji korelasi Spearman dirumuskan sebagai berikut.

i. Saat data memiliki peringkat yang berbeda

$$\rho = 1 - \frac{6 \sum {d_i}^2}{n(n^2 - 1)}$$

Persamaan 2.4 - Uji Korelasi Spearman

d adalah selisih pada peringkat dan n adalah jumlah kasus.

ii. Saat data memiliki peringkat yang sama

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Persamaan 2.5 - Uji Korelasi Spearman

Nilai rho (ρ) merupakan nilai koefisien Spearman yang menunjukkan kekuatan korelasi antar variabel. Tabel 2.2 menunjukkan petunjuk interpretasi nilai dari rho [29].

Tabel 2.2 - 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

Nilai rho	Interpretasi hasil
.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. Uji korelasi Spearman akan digunakan untuk menentukan data hasil pencarian Google Trend yang akan digunakan. Uji korelasi ini digunakan untuk data Google Trend karena tidak sensitif terhadap adanya data yang *outlier.*

2.2.11. Uji Stasioneritas, ADF (*Augmented Dickey Fuller*)

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* [30].

- Jika nilai uji T lebih besar dari nilai kritis ADF maka terima H0 dan disimpulkan bahwa data tersebut memiliki akar unit atau tidak stasioner
- Jika nilai uji T kurang dari nilai kritis ADF maka tolak H0 dan disimpulkan bahwa data tersebut tidak memiliki akar unit atau stasioner.

2.2.12. 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 [24] [31].

Hipotesis awal (H0) pada Ljung-Box adalah data tersebut *white noise*. Hipotesa alternatif (Ha) dari Ljung-Box adalah tidak *white noise*.

2.2.13. 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

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

Penafsiran nilai MAPE dijelaskan pada Tabel 2.3

Tabel 2.3 - Penafsiran nilai MAPE

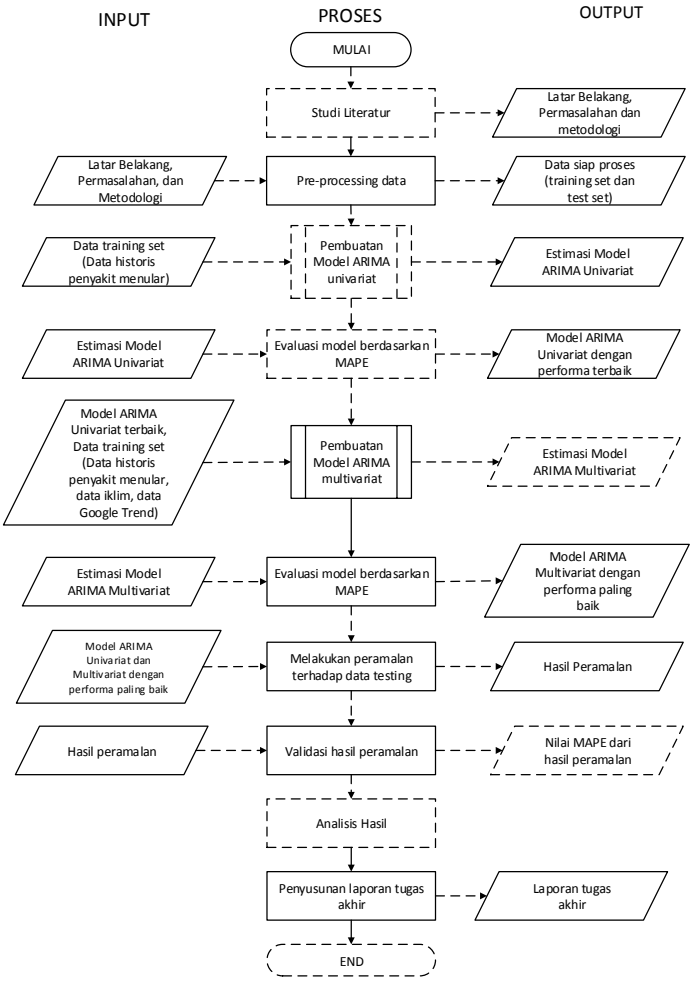
MAPE	Hasil Peramalan
<10%	Sangat Baik
10-20%	Baik
20-50%	Layak/ Cukup
>50%	Buruk

2.2.14. SAS Software

SAS merupakan aplikasi statistik yang digunakan untuk melakukan berbagai aktivitas analisis data. SAS menjadi pilihan utama aplikasi analisis statistuk pada berbagai industri [32]. 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 [33]. Untuk penelitian ini SAS/ETS digunakan untuk mengembangkan model ARIMA, dan peramalan.

BAB III
METODOLOGI PENELITIAN

Pada bagian ini akan dijelaskan mengenai proses pengerjaan tugas akhir. Proses pengerjaan tugas akhir dijelaskan pada Gambar 3.1



Gambar 3.1 - Alur pengerjaan tugas akhir

Penjelasan detil mengenai metodologi penelitian dijelaskan pada bagian berikut.

3.1. Studi Literatur

Penelitian Tugas Akhir diawali dengan melakukan kajian pustaka terkait permasalahan dan metode penelitian yang akan digunakan.

3.2. *Pre processing data*

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

- Data historis jumlah kasus penyakit DBD dan Influenza
- Faktor iklim (data suhu, kelembaban relatif dan curah hujan)
- Data Google Trends dari

Data historis jumlah kasus penyakit DBD dan Influenza diperoleh dari Dinas Kesehatan Kota Surabaya dan RSUD Dr. Soetomo Surabaya. Faktor iklim diperoleh dari website resmi BMKG (http://dataonline.bmkg.go.id/).

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

3.3. Pembuatan Model ARIMA Univariat

Pembuatan model ARIMA univariat digambarkan pada Gambar 3.2. Pada pembuatan model ARIMA univariat data yang digunakan sebagai inputan adalah data historis penyakit menular. Langkah pertama yang dilakukan adalah uji stasioneritas data dengan menggunakan uji ADF.

Jika data masih belum stasioner maka diperlukan diferensiasi dan transformasi data. Tahapan selanjutnya adalah menganalisis lag berdasarkan PACF dan ACF untuk mengidentifikasi nilai orde p, d, dan q untuk model ARIMA univariat. Order tersebut kemudian digunakan untuk mengestimasi nilai parameter AR, dan MA.

Hasil estimasi parameter selanjutnya diuji signifikansi nya (pvalue<0,05). Model dengan parameter AR dan MA tidak signifikan akan dibuang, atau diidentifikasi ulang. Model dengan parameter AR dan MA yang sudah signifikan selanjutnya melalui uji white noise pada residualnya. Model yang menunjukkan keberadaan white noise pada residualnya dinyatakan sebagai model yang sudah layak.

3.4. Pembuatan Model ARIMA Multivariat

Pembuatan model ARIMA multivariat digambarkan pada Gambar 3.3. 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 :

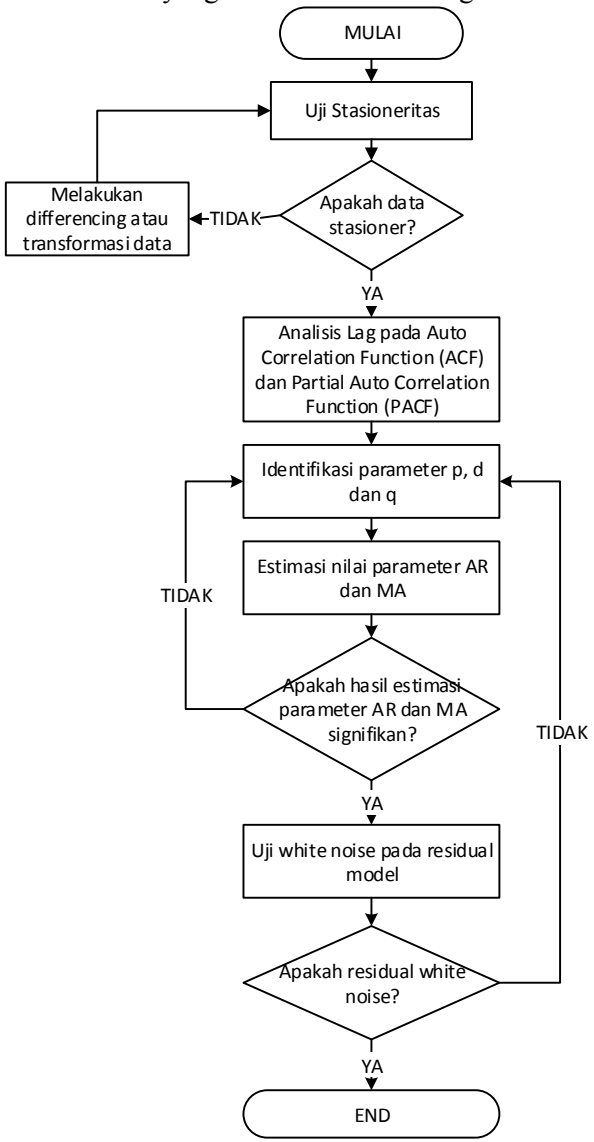
- Data historis penyakit menular dengan variabel eksternal data faktor iklim (suhu, curah hujan dan kelembaban relatif)
- Data historis penyakit menular dengan variabel eksternal data faktor iklim (suhu, curah hujan dan kelembaban relatif), dan data Google Trend
- Data Google Trend dengan variabel eksternal data faktor iklim (suhu, curah hujan dan kelembaban relatif)

Seluruh variabel eksternal tersebut harus melalui proses transformasi yang sama dengan variabel dependen. Masing-masing variabel eksternal yang ada akan dilakukan operasi lag 1-10. Hal ini dilakukan dengan tujuan menangani kemungkinan adanya waktu yang diperlukan untuk perubahan variabel eksternal yang berdampak terhadap variabel dependen.

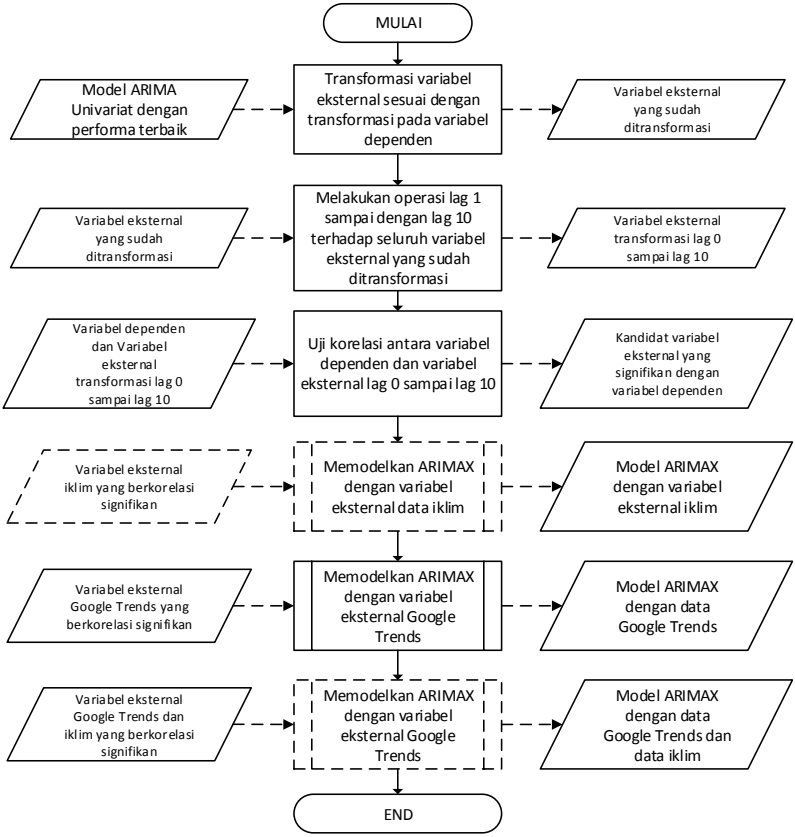
Sehingga ada 11 variasi variabel eksternal dengan lag 0 sampai dengan lag 10. Seluruh variabel eksternal diuji korelasinya terhadap variabel dependen. Variabel eksternal yang memiliki korelasi signifikan terhadap variabel dependen akan digunakan dalam menyusun model ARIMAX.

Nilai parameter p, d, dan q pada ARIMAX(p,d,q) ditentukan dari model univariat yang memiliki performa

terbaik. Dari model univariat tersebut akan ditambahkan variabel eksternal yang memiliki korelasi signifikan.



Gambar 3.2 - Pembuatan Model ARIMA Univariat

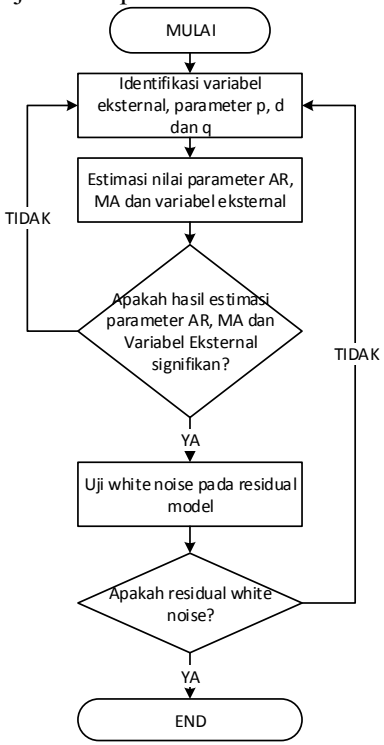


Gambar 3.3 - Pembuatan Model Multivariat

Terdapat tiga model multivariat untuk masing-masing kasus penyakit yang akan dihasilkan :

1. Model ARIMAX dengan variabel eksternal iklim
2. Model ARIMAX dengan variabel eksternal Google Trends
3. Model ARIMAX dengan variabel eksternal Google Trends dan iklim

Ketiga model tersebut melalui tahapan pemodelan yang sama, perbedaannya terletak di variabel eksternal yang digunakan. Gambar 3.4 menjelaskan pemodelan ARIMAX.



Gambar 3.4 - Tahapan pemodelan ARIMAX

Pemodelan diawali dengan mengidentifikasi orde p,d, dan q beserta kandidat variabel eksternal yang berkorelasi signifikan dengan variabel dependen. Kemudian model tersebut diestimasi nilai parameter AR, MA, dan variabel eksternalnya. Hasil estimasi parameter selanjutnya diuji signifikansi nya (pvalue<0,05). Model dengan parameter AR, MA dan variabel eksternal tidak signifikan akan dibuang, atau diidentifikasi ulang. Model dengan parameter AR, MA dan variabel eksternal yang sudah signifikan selanjutnya melalui uji white noise pada residualnya. Model yang menunjukkan keberadaan white noise pada residualnya dinyatakan sebagai model yang sudah layak.

3.5. Evaluasi model berdasarkan MAPE

Tahap berikutnya adalah menguji kelayakan model ARIMA univariat dan multivariat yang telah dihasilkan. Uji kelayakan model ini menggunakan MAPE (*Mean Absolute Percentage Error)*. Tahap ini bertujuan untuk memilih model yang memiliki performa terbaik.

3.6. 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.

3.7. Validasi hasil peramalan

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

3.8. 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.

3.9. Penyusunan laporan Tugas Akhir

Hasil akhir pada tahap ini adalah laporan Tugas Akhir yang memberikan penjelasan dokumentasi aktivitas yang dilakukan selama penelitian ini.

Halaman ini sengaja dikosongkan

BAB IV
PERANCANGAN

Pada bab ini akan dijelaskan bagaimana rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian akan dilakukan.

4.1. Pengumpulan dan deskripsi data

Penelitian ini menggunakan beberapa variabel yang berbeda. Tabel 4.1 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.

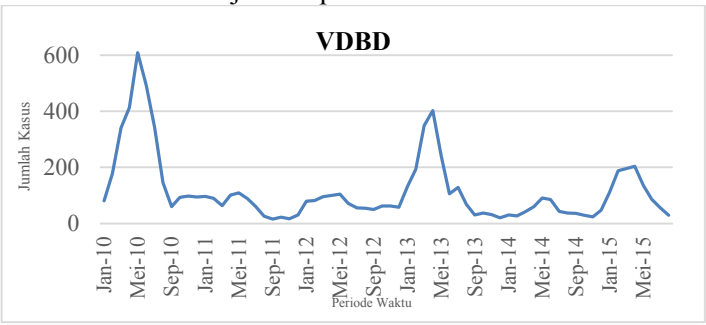
Tabel 4.1 - Data mentah yang digunakan

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 : • demam berdarah • dbd • demam • dengue per bulan area Jawa Timur	Search Index per 2 minggu
GTFLU	Variabel Eksternal, Google Trend Data dengan search query : • Bersin • Influenza	Search Index per 2 minggu

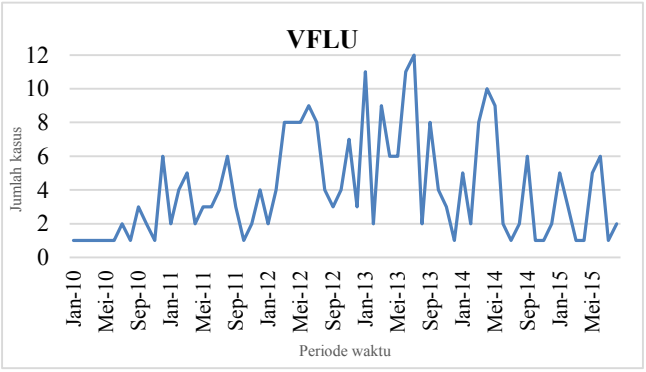
Nama Variabel	Keterangan	Satuan
	• Flu • Batuk • Pilek • Batuk pilek area Jawa Timur	
SUHU	Variabel Eksternal, Suhu rata-rata	^{O}C per hari
HMD	Variabel Eksternal, Kelembaban rata-rata	persen per hari
CH	Variabel Eksternal, Curah Hujan rata-rata	mm per hari

Untuk data Google Trends demam berdarah dan influenza *query* ditentukan berdasarkan hasil *related* search yang muncul setelah melakukan pencarian query "demam berdarah" dan "influenza" pada Google Trends. Data GTDBD dan GTFLU diunduh pada tanggal 19-21 September 2016.

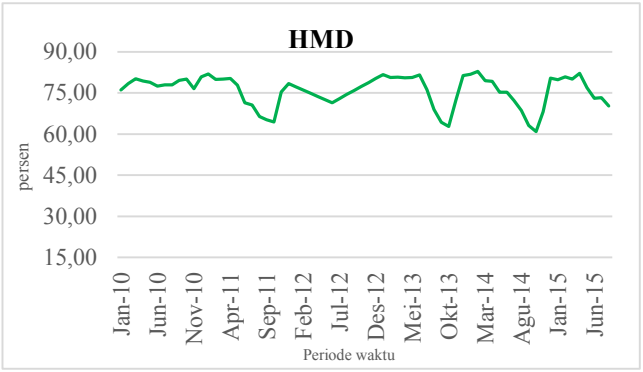
Plot data variabel VDBD dan VFLU ditunjukkan pada Gambar 4.1 dan Gambar 4.3. Plot data variabel eksternal iklim yakni kelembaban (HMD), curah hujan (CH), dan suhu udara (SUHU) ditunjukkan pada Gambar Gambar 4.2, Gambar 4.5, dan Gambar 4.4. Plot data variabel eksternal Google Trends untuk kasus demam berdarah ditunjukkan pada Gambar 4.6, sedangkan plot data variabel eksternal Google Trends untuk kasus influenza ditunjukkan pada Gambar 4.7.



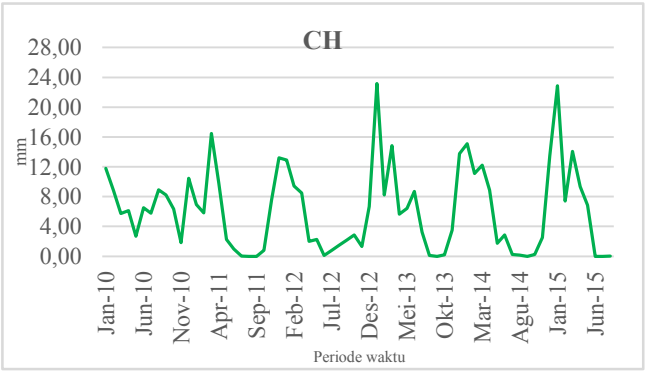
Gambar 4.1 - Plot VDBD



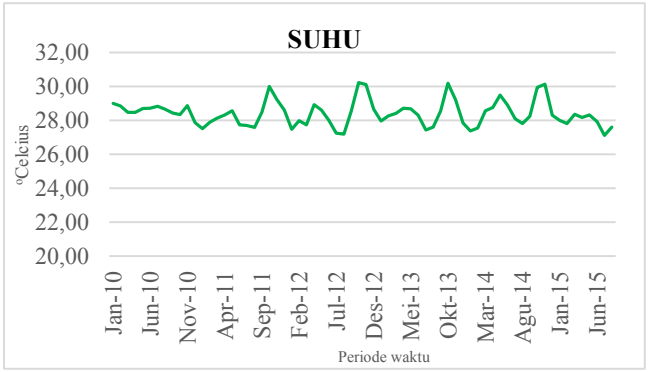
Gambar 4.3 - Plot VFLU



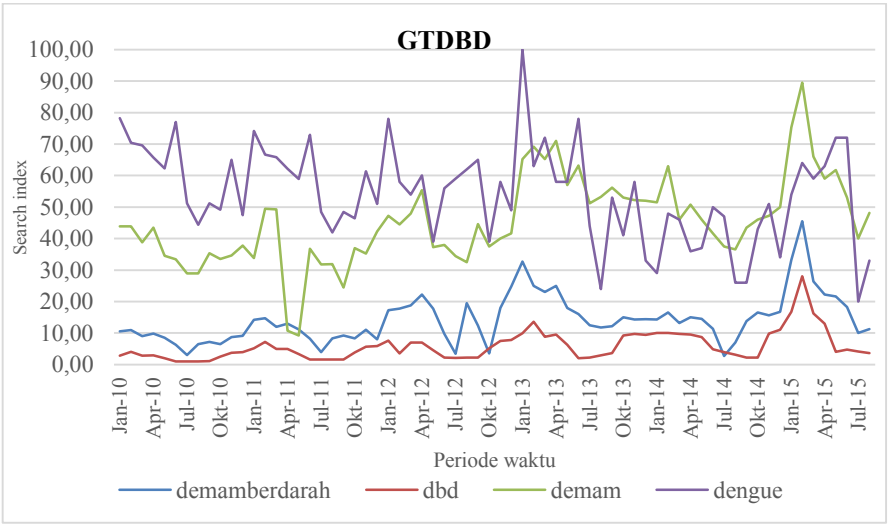
Gambar 4.2 - Plot HMD



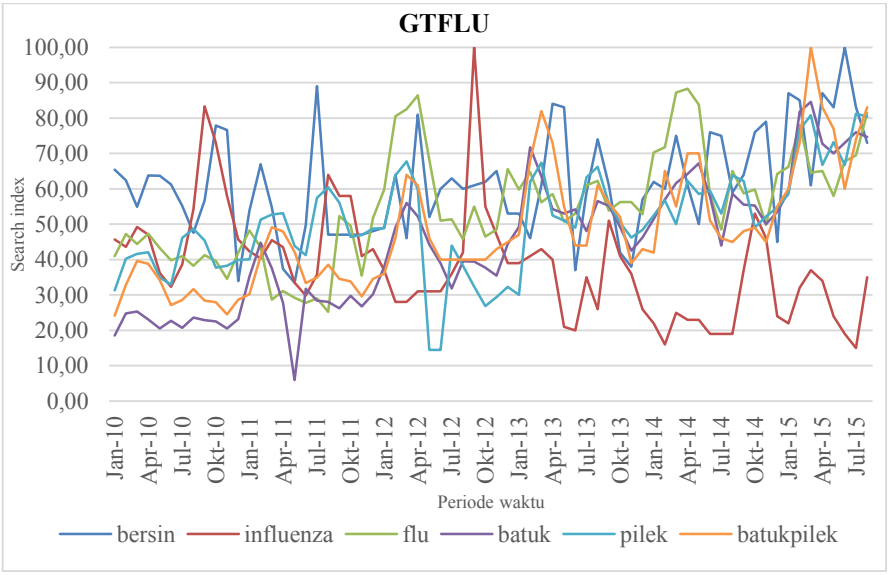
Gambar 4.5 - Plot CH



Gambar 4.4 - Plot SUHU



Gambar 4.6 - Plot GTDBD



Gambar 4.7 - Plot GTFLU

4.2. *Pre-processing Data*

4.2.1. Perubahan Satuan Waktu Menjadi Per Bulan

Dari data mentah yang diperoleh, semua data tersebut diubah menjadi satuan perbulan. Perubahan tersebut dilakukan dengan mencari nilai rata-rata jumlah data yang ada pada bulan tersebut. Gambar 4.1 sampai Gambar 4.7 merupakan plot variabel eksternal yang sudah diagregasi menjadi perbulan.

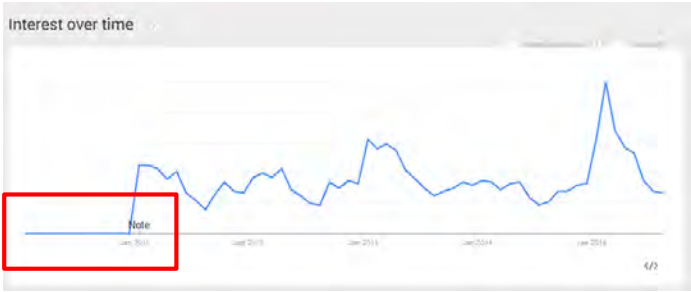
4.2.2. Penanganan *missing value*

4.2.2.1. *Missing Value* pada variabel iklim

Pada variabel HMD dan CH terdapat beberapa bulan data kelembaban dan curah hujan tidak tercatat oleh stasiun BMKG di Perak, sehingga mengakibatkan data yang kosong. Data yang kosong tersebut dicari dengan menghitung interpolasi linear antara data sebelum *missing value* dengan data setelah *missing value*. Penghitungan interpolasi data dilakukan dengan menggunakan Microsoft Excel [34].

4.2.2.2. *Missing Value* pada variabel Google Trends

Penelitian ini menggunakan hasil pencarian di Google Trends sebagai salah satu variabel eksternal. Namun pada hasil *search query* pada Google Trends area Jawa Timur terdapat beberapa data yang hilang. Gambar 4.8 merupakan contoh grafik hasil pencarian “demam berdarah” di Jawa Timur dari Januari 2010 – Agustus 2015.



Gambar 4.8 - Contoh Missing Value pada Google Trends

Pada grafik tersebut terlihat data dari Januari 2010 – Desember 2011 tidak tercatat oleh sistem Google Trends. Tabel 4.2 menunjukkan jumlah *missing value* untuk masing-masing *search query* yang digunakan untuk penelitian ini.

Tabel 4.2 - Keterangan *search query* dengan *missing value*

Search Query	**Jumlah titik data yang kosong**
"demam berdarah"	**23** (Jan-10 s/d Des-11)
"dbd"	**38** (Jan-10 s/d Feb-13)
"demam"	**12** (Jan-10 s/d Des-10)
"dengue"	**23** (Jan-10 s/d Nov-11)
"bersin"	**18** (Jan-10 s/d Juni-11)
"influenza"	**19** (Jan-10 s/d Jul-11)
"flu"	**12** (Jan-10 s/d Des-10) **5** (Mar-11 s/d Jul-11)
"batuk"	**12** (Jan-10 s/d Des-10)
"batuk pilek"	**24** (Jan-10 s/d Des-11)
"pilek"	**25** (Jan-10 s/d Jan-12) **5** (Agustus-12 s/d Des-12)

Data-data yang kosong ditangani dengan melakukan ekstrapolasi. Terdapat berbagai macam metode ekstrapolasi yang bisa dilakukan, salah satunya adalah *exponential smoothing* [35]. *Exponential smoothing* dipilih 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. Tabel 4.3 menjelaskan nilai MAPE dari hasil ekstrapolasi data.

Tabel 4.3 - MAPE hasil ekstrapolasi variabel Google Trends

Search Query	MAPE
“dbd”	28,20%
“demam”	9,90%
“demam berdarah”	36,50%
“dengue"	23,84%
“batuk”	16,70%
“batuk pilek”	11,08%
“bersin”	9,01%
“flu”	13,17%
“influenza”	21%
“pilek"	16,26%

4.2.3. Pembagian Trainset dan Testset

Data trainset digunakan untuk menentukan model ARIMA yang akan digunakan pada peramalan, sedangkan data testset digunakan untuk menguji model yang telah dibuat. Pada penelitian ini perbandingan trainset dan testset adalah 70:30, sehingga pembagiannya adalah sebagai berikut :

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

Halaman ini sengaja dikosongkan

BAB V
PEMODELAN

Pada bab ini akan dijelaskan proses pemodelan ARIMA univariat dan ARIMA multivariat untuk kasus demam berdarah dan influenza.

5.1. Model Univariat

Model ARIMA Univariat adalah model ARIMA yang digunakan untuk meramalkan jumlah kasus penyakit berdasarkan kasus-kasus yang terdahulu. Pemodelan ARIMA univariat meliputi uji stasioneritas, estimasi nilai parameter p dan q, diakhiri dengan evaluasi dan pemilihan model ARIMA univariat dengan performa terbaik.

5.1.1. Model ARIMA Univariat untuk Kasus Demam Berdarah

Pada bagian ini akan dijelaskan pemodelan model ARIMA univariat untuk kasus demam berdarah.

5.1.1.1. Uji Stasioneritas

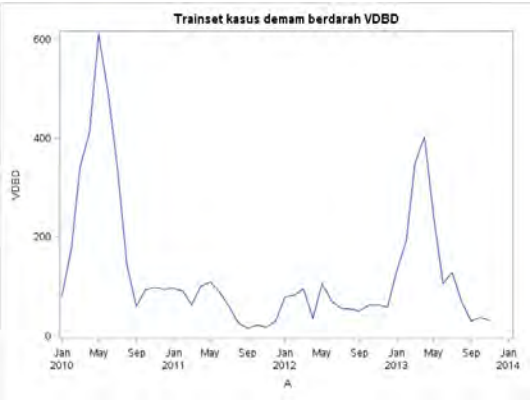
Syarat utama yang harus dipenuhi pada pemodelan ARIMA adalah data variabel dependen harus stasioner. Definisi data yang stasioner adalah data yang unsur statistiknya memiliki rata-rata, variasi dan auto korelasi yang konstan [36].

Uji stasioneritas dilakukan dengan melihat plot data dan hasil dari uji ADF. Pembuatan grafik dilakukan dalam SAS dengan menjalankan perintah yang ada pada SAS Code 5.1.

```
1 proc sgplot data=work.trainset;
2 series x=a y=vdbd;
3 title 'Trainset kasus demam berdarah VDBD';
4 run;
```

SAS Code 5.1 - Membuat Grafik dalam SAS

Gambar 5.1 merupakan plot dari trainset VDBD kasus demam berdarah. Secara visual data pada VDBD terlihat belum stasioner.



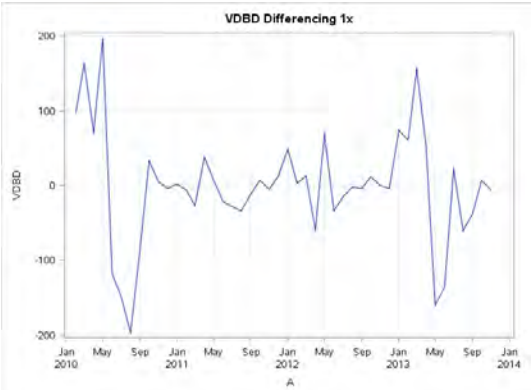
Gambar 5.1 - Plot VDBD

Uji ADF dilakukan dengan menggunakan tools R. Tabel 5.1 merupakan hasil dari uji stasioneritas untuk VDBD.

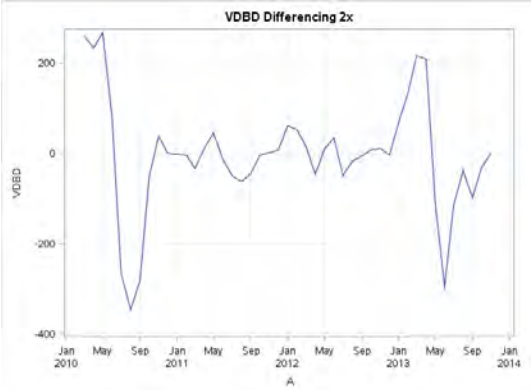
Tabel 5.1 - Hasil Uji Stasioneritas VDBD

ADF TEST – Data : vdbd_train_ts		
Dickey-Fuller	Lag Order	p-value
-2,8387	2	0,2398

Pada uji ADF ini, dilihat dari nilai p-value lebih besar dari 0,05, sehingga dari hasil uji ADF ini dapat disimpulkan bahwa data VDBD tidak stasioner, dan perlu dilakukan diferensiasi.



Gambar 5.2 - Plot VDBD setelah didiferensiasi 1 kali



Gambar 5.3 - Plot VDBD setelah didiferensiasi 2 kali

Gambar 5.2 dan Gambar 5.3 merupakan plot VDBD setelah dilakukan diferensiasi 1 kali dan 2 kali. Walaupun sudah dilakukan diferensiasi, secara visual VDBD masih belum terlihat stasioner. Cara lain untuk memperoleh stasioneritas adalah mentransformasikan data secara logaritma. Tujuan dari transformasi ini adalah untuk menstabilkan varians pada data [24].

```
1 data work.trainset;
2     set work.trainset;
3     vdbdlog2=log(log(vdbd));
4 run;
```

SAS Code 5.2 - Transformasi log pada variabel

Langkah pertama yang dilakukan untuk transformasi adalah membuat variabel VDBDLOG yang merupakan perubahan variabel VDBD dengan fungsi logaritma, VDBDLOG = log(VDBD). Pada penelitian ini dilakukan dua kali transformasi logaritma pada variabel dependen. Hal ini dikarenakan dirasa hasil transformasi logaritma pertama masih memiliki varians yang cukup besar. Rumus yang digunakan untuk melakukan transformasi adalah VDBDLOG2 = log(log(VDBD)). SAS Code 5.2 merupakan perintah yang dijalankan untuk membuat variabel baru VDBDLOG2. Gambar 5.4 merupakan hasil plot VDBDLOG2.

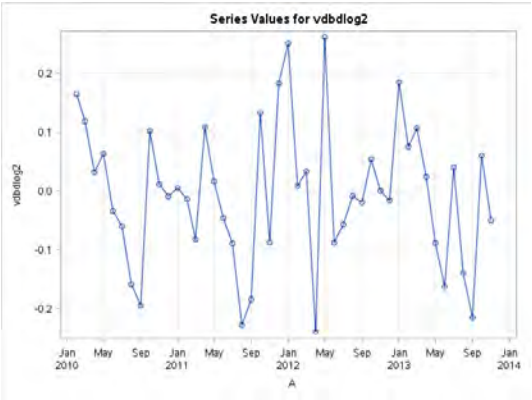
Gambar 5.4 - Plot VDBDLOG2

Uji ADF dilakukan untuk menentukan kestasioneran VDBDLOG2. Tabel 5.2 merupakan hasil uji stasioneritas dari VDBDLOG2.

Tabel 5.2 - Hasil Uji Stasioneritas VDBDLOG2

ADF TEST – Data : log(log(vdbd_train_ts))		
Dickey-Fuller	**Lag Order**	**p-value**
-3.6141	3	0,04204

Dari Tabel 5.2 dapat diketahui bahwa variabel VDBDLOG2 sudah stasioner karena nilai p-value sudah kurang dari 0,05. Namun secara visual data tersebut tidak terlihat stasioner, sehingga untuk itu diperlukan differensiasi satu kali (d=1). Gambar 5.5 merupakan hasil plot VDBDLOG2 (d=1).



Gambar 5.5 - Plot VDBDLOG2 setelah didiferensiasi 1 kali

Tabel 5.3 merupakan hasil uji ADF data VDBDLOG2 (d=1). Dari hasil uji ADF tersebut dapat diketahui bahwa VDBDLOG2(d=1) merupakan data yang stasioner.

Tabel 5.3 - Hasil Uji Stasioneritas VDBDLOG2 setelah didiferensiasi 1 kali

ADF TEST – Data : log(log(vdbd_train_ts))		
Dickey-Fuller	**Lag Order**	**p-value**
-4,1989	3	0,01

Estimasi nilai parameter d pada model ARIMA(p,d,q) untuk data historis demam berdarah adalah d=0 atau d=1.

5.1.1.2. Identifikasi parameter p dan q

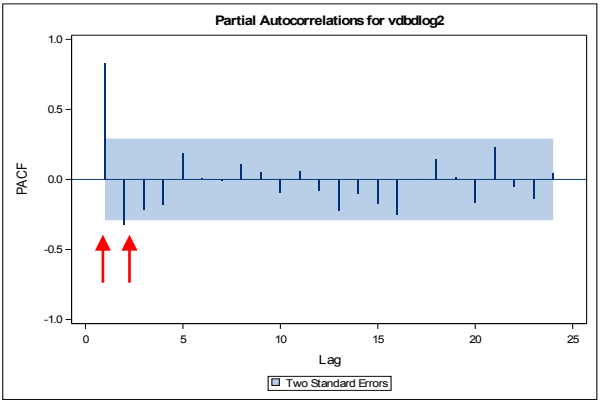
Dari uji stasioneritas sudah diketahui nilai parameter d model ARIMA(p,d,q) untuk kasus demam berdarah adalah 1. Identifikasi nilai parameter p dan q dilihat dari grafik PACF dan ACF untuk VDBDLOG2 d=0 dan VDBDLOG2 d=1. Untuk melihat grafik ACF dan PACF pada SAS menggunakan SAS Code 5.3.

```
1proc arima data=work.trainset;
2identify var=vdbdlog2;
3run;
```

SAS Code 5.3 - Melihat ACF dan PACF

1. Identifikasi Nilai Parameter p

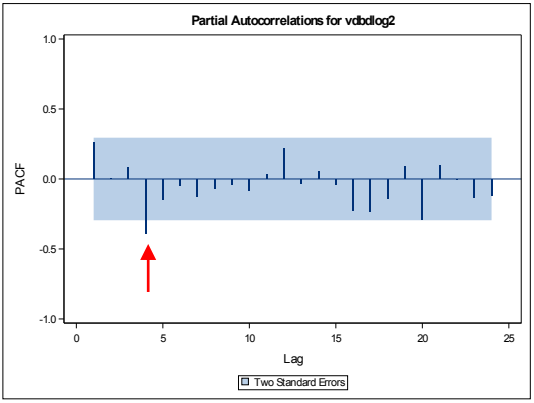
Identifikasi nilai parameter p dillihat dari grafik PACF dari variabel VDBDLOG2 d=0 dan VDBDLOG2 d=1. Pada grafik PACF yang dicari adalah lokasi lag yang nilainya naik/turun seara signifikan. Gambar 5.6 merupakan grafik PACF VDBDLOG2 d=0.



Gambar 5.6 - PACF VDBDLOG2 d=0

Pada Gambar 5.6 terdapat dua lag yang naik/turun secara signfikan yakni pada lag 1 dan lag 2, sehingga nilai identifikasi parameter p untuk d=0 adalah p=1 atau p=2.

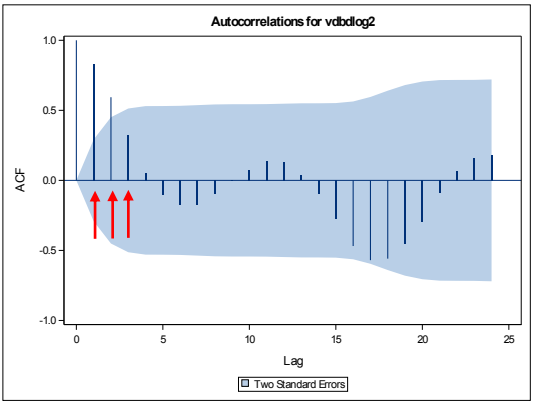
Gambar 5.7 merupakan grafik PACF untuk VDBDLOG2 d=1. Pada Gambar 5.7 terdapat satu lag yang naik secara signifikan yakni pada lag 4, sehingga nilai identifikasi parameter p untuk d=1 adalah p=4.



Gambar 5.7 - PACF VDBDLOG2 d=1

2. Identifikasi Parameter q

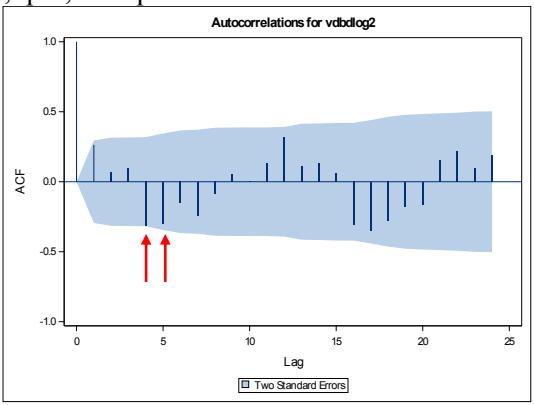
Identifikasi nilai parameter q dillihat dari grafik ACF dari variabel VDBDLOG2 d=0 dan VDBDLOG2 d=1. Pada grafik ACF yang dicari adalah lokasi lag yang nilainya naik/turun seara signifikan. Gambar 5.8 merupakan grafik ACF VDBDLOG2 d=0.



Gambar 5.8 - ACF VDBDLOG2 d=0

Gambar 5.8 menampilkan grafik ACF dari VDBDLOG2 d=0 yang bentuknya menyerupai grafik sinus, hal ini mengindikasikan bahwa model ini adalah model AR

(MA=0). Pada Gambar 5.8 jika modelnya ini bukanlah model AR, dilihat pada lag 1, 2 dan 3 terdapat nilai ACF yang signfikan. Sehingga estimasi nilai paremeter q untuk d=0 adalah q=0, q=1, q=2, dan q=3.



Gambar 5.9 - ACF VDBDLOG2 d=1

Gambar 5.9 merupakan grafik ACF untuk VDBDLOG2 d=1. Pada Gambar 5.9 terdapat dua lag yang naik secara signifikan yakni pada lag 4 dan lag 5, sehingga nilai identifikasi parameter q untuk d=1 adalah q=4 atau q=5.

3. Model hasil identifikasi

Tabel 5.4 merangkum hasil identifikasi nilai parameter p,d dan q untuk model ARIMA VDBDLOG2.

Tabel 5.4 - Identifikasi nilai parameter p,d, dan q ARIMA VDBDLOG2

Nilai estimasi p	Nilai estimasi d	Nilai estimasi q
1,2	0	1,2,3
4	1	4,5

Dengan beberapa identifikasi parameter tersebut menghasilkan beberapa kombinasi model ARIMA(p,d,q) yang dijelaskan pada Tabel 5.5.

Tabel 5.5 - Identifikasi Model ARIMA(p,d,q) VDBDLOG2

Identifikasi Model ARIMA(p,d,q) untuk VDBDLOG2		
ARIMA(1,0,0)	ARIMA(1,0,1)	ARIMA(1,0,2)
ARIMA(1,0,3)	ARIMA(2,0,0)	ARIMA(2,0,1)
ARIMA(2,0,2)	ARIMA(2,0,3)	ARIMA(0,0,1)
ARIMA(0,0,2)	ARIMA(0,0,3)	ARIMA(4,1,0)
ARIMA(4,1,4)	ARIMA(4,1,5)	ARIMA(0,1,4)
ARIMA(0,1,5)		

Dari kombinasi model tersebut akan diestimasi nilai parameter untuk model AR dan MA, dengan metode estimasi *maximum likelihood*. Model AR dan MA yang memiliki hasil estimasi yang tidak signifikan dieliminasi.

5.1.1.3. Estimasi nilai parameter

Dari 16 model pada Tabel 5.5 parameter model akan diestimasi. Model ARIMA dengan parameter AR dan MA yang tidak signifikan akan dieliminasi. Dari 16 model yang sudah diidentifikasi terdapat 5 model yang lolos. Persyaratan sebuah model untuk lolos uji estimasi parameter adalah nilai parameter AR dan MA keduanya harus signifikan (pvalue atau Pr>|t| memiliki nilai <0,05). Jika ada salah satunya yang tidak signifikan maka model itu harus dieliminasi [31]. Hasil estimasi parameter ditunjukkan pada Tabel 5.6 .

Tabel 5.6 - Estimasi parameter VDBDLOG2

No	Model	AR		MA		Keterangan
		Estimasi	Pr> |t|	Estimasi	Pr> |t|	
1	ARIMA(1,0,0)	0,84533	<,0001	-	-	Buang
2	**ARIMA(1,0,3)**	**0,56531**	**0,0034**	**-0,4436**	**0,0083**	**Lolos**
3	ARIMA(2,0,2)	-0,75226	<,0001	-0,19189	0,3185	Buang
4	ARIMA(0,0,2)	-	-	-0,40985	0,0049	Buang
5	ARIMA(4,1,4)	0,19314	0,7835	0,4605	0,9899	Buang
6	ARIMA(0,1,5)	-	-	0,45164	0,9761	Buang

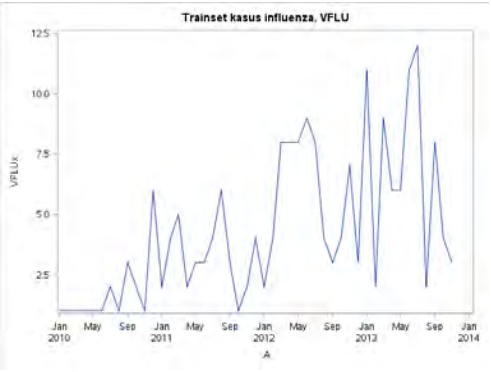
No	Model	AR		MA		Keterangan
		Estimasi	Pr> |t|	Estimasi	Pr> |t|	
7	**ARIMA(1,0,1)**	**0,74963**	**<,0001**	**-0,4036**	**0,0117**	**Lolos**
8	**ARIMA(2,0,0)**	**-0,3792**	**0,0064**	**-**	**-**	**Lolos**
9	ARIMA(2,0,3)	-0,28229	0,4009	-0,42782	0,0169	Buang
10	ARIMA(0,0,3)	-	-	-0,54444	<,0001	Buang
11	ARIMA(4,1,5)	-0,20075	0,5851	0,50374	0,9698	Buang
12	ARIMA(1,0,2)	0,74381	<,0001	-0,02272	0,9134	Buang
13	**ARIMA(2,0,1)**	**-0,7622**	**<,0001**	**0,53024**	**0,0328**	**Lolos**
14	ARIMA(0,0,1)	-	-	-0,81728	<,0001	Buang
15	**ARIMA(4,1,0)**	**-0,4357**	**0,0034**	**-**	**-**	**Lolos**
16	ARIMA(0,1,4)	-	-	0,16505	0,3151	Buang

Lima model dengan parameter AR dan MA yang signifikan adalah ARIMA(1,0,3), ARIMA(1,0,1) ARIMA(2,0,0), ARIMA(2,0,1), dan ARIMA(4,0,1). Model-model tersebut memiliki hasil estimasi parameter AR dan MA yang signifikan. Lima model yang lolos tersebut akan dilakukan uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

5.1.2. Model ARIMA Univariat untuk Kasus Influenza

Pada bagian ini akan dijelaskan pemodelan model ARIMA univariat untuk kasus demam berdarah.

5.1.2.1. Uji Stasioneritas



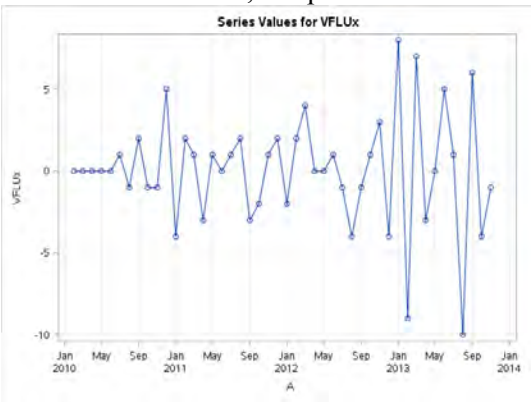
Gambar 5.10 - Plot VFLU

Gambar 5.10 merupakan hasil plot dari variabel kasus influenza (VFLU). Secara visual variabel VFLU masih belum stasioner. Tabel 5.7 merupakan hasil uji stasioneritas dari VFLU.

Tabel 5.7 - Uji Stasioneritas VFLU

ADF TEST – Data : vflu_train_ts		
Dickey-Fuller	Lag Order	p-value
-3,3911	3	0,06913

Dari hasil uji ADF, nilai p-value lebih besar dari 0,05, sehingga dari hasil uji ADF ini dapat disimpulkan bahwa variabel VFLU tidak stasioner, dan perlu dilakukan diferensiasi.



Gambar 5.11 - Plot VFLU setelah didiferensiasi 1 kali

Gambar 5.11 merupakan plot dari variabel VFLU yang sudah didiferensiasi (d=1). Secara visual datanya terlihat stasioner. Tabel 5.8 merupakan hasil dari uji stasioneritas untuk variabel VFLU (d=1).

Tabel 5.8 - Uji stasioneritas VFLU d=1

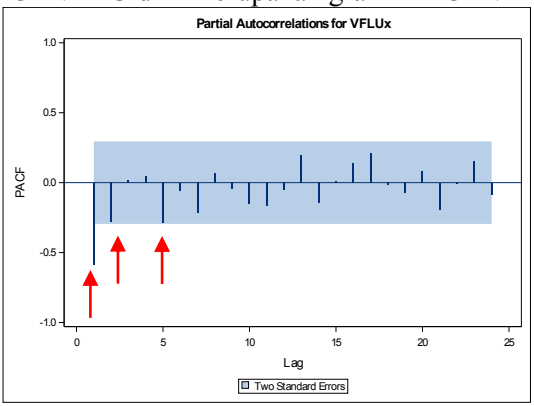
ADF TEST – Data : diff(vflu_train_ts)		
Dickey-Fuller	Lag Order	p-value
-3,4819	3	0,0500

Dari hasil uji stasioneritas, nilai p-value lebih kecil atau sama dengan dari 0,05, sehingga dari hasil uji ADF ini dapat disimpulkan bahwa variabel VFLU (d=1) sudah stasioner, dan bisa dilanjutkan ke analisis berikutnya.

5.1.2.2. Identifikasi Parameter p dan q

1. Identifikasi parameter p

Identifikasi nilai parameter p dillihat dari grafik PACF dari variabel VFLU d=1. Pada grafik PACF yang dicari adalah lokasi lag yang nilainya naik/turun seara signifikan. Gambar 5.12 - PACF VFLU d=1 merupakan grafik PACF VFLU d=1.

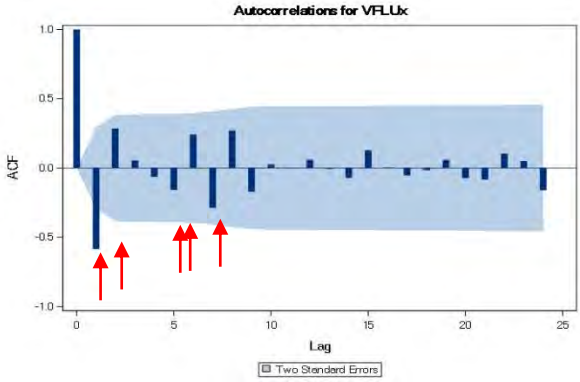


Gambar 5.12 - PACF VFLU d=1

Pada Gambar 5.12 terdapat tiga lag yang naik/turun secara signfikan yakni pada lag 1, lag 2 dan lag 5, sehingga nilai identifikasi parameter p untuk d=1 adalah p=1, p=2 atau p=5.

2. Identifikasi parameter q

Identifikasi nilai parameter q dillihat dari grafik ACF dari variabel VFLU d=1. Pada grafik ACF yang dicari adalah lokasi lag yang nilainya naik/turun seara signifikan. **Error! eference source not found.** merupakan grafik ACF VFLU d=1.



ACF VFLU d=1

Pada **Error! Reference source not found.** terdapat dua ag yang naik/turun secara signfikan yakni pada lag 1 dan setelah lag 2 nilai ACF berkurang secara drastic. Kemudian nilai yang naik pada lag 6,7, dan 8. Sehingga nilai identifikasi parameter q untuk d=1 adalah q=1, q=2, q=6, q=7 dan q=8.

3. Model hasil identifikasi

Tabel 5.9 merangkum hasil identifikasi nilai parameter p,d dan q untuk model ARIMA VFLU.

Tabel 5.9 - Identifikasi nilai parameter p,d, dan q ARIMA VFLU

Nilai identifikasi p	Nilai identifikasi d	Nilai identifikasi q
1,2	1	1,2,6,7,8

Dengan beberapa identifikasi parameter tersebut menghasilkan beberapa kombinasi model ARIMA(p,d,q) yang dijelaskan pada Tabel 5.10.

Tabel 5.10 - Identifikasi model ARIMA(p,d,q) VFLU

Identifikasi Model ARIMA(p,d,q) untuk VFLU		
ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(2,1,0)
ARIMA(2,1,1)	ARIMA(5,1,0)	ARIMA(5,1,1)
ARIMA(0,1,1)	ARIMA(0,1,2)	ARIMA(2,1,2)
ARIMA(1,1,2)	ARIMA(5,1,2)	ARIMA(0,1,8)
ARIMA(0,1,6)	ARIMA(0,1,7)	ARIMA(1,1,8)
ARIMA(1,1,6)	ARIMA(1,1,7)	ARIMA(2,1,8)
ARIMA(2,1,6)	ARIMA(2,1,7)	ARIMA(5,1,8)
ARIMA(5,1,6)	ARIMA(5,1,7)	

Dari kombinasi model tersebut akan diestimasi nilai parameter untuk model AR dan MA, dengan metode estimasi *maximum likelihood*. Model AR dan MA yang memiliki hasil estimasi yang tidak signifikan dieliminasi.

5.1.2.3. Estimasi nilai parameter

Dari 23 model pada Tabel 5.10 parameter model ARIMA untuk VFLU akan diestimasi. Model ARIMA dengan parameter AR dan MA yang tidak signifikan akan dieliminasi. Persyaratan sebuah model untuk lolos uji estimasi parameter adalah nilai parameter AR dan MA keduanya harus signifikan. Jika ada salah satunya yang tidak signifikan maka model itu harus dieliminasi [39]. Dari 23 model yang sudah diidentifikasi terdapat 3 model yang lolos. Hasil estimasi parameter ditunjukkan pada Tabel 5.11.

Tabel 5.11 - Estimasi parameter VFLU

No	Model	AR		MA		Keterangan
		Estimasi	Pr> |t|	Estimasi	Pr> |t|	
1	**ARIMA(1,1,0)**	**-0,57483**	**<,0001**	**-**	**-**	**Lolos**
2	ARIMA(2,1,1)	0,28524	0,1791	0,99947	0,9644	
3	ARIMA(0,1,1)	-	-	0,99972	0,9637	
4	ARIMA(1,1,2)	-0,46124	0,8639	0,41674	0,9992	

No	Model	AR		MA		Keterangan
		Estimasi	Pr> |t|	Estimasi	Pr> |t|	
5	ARIMA(1,1,1)	0,06002	0,7398	0,99994	0,9947	
6	**ARIMA(5,1,0)**	**-0,41649**	**0,0165**	**-**	**-**	**Lolos**
7	ARIMA(0,1,2)	-	-	0,05558	0,9984	
8	ARIMA(5,1,2)	-0,16039	0,601	0,94863	0,9935	
9	**ARIMA(2,1,0)**	**-0,29509**	**0,0486**	**-**	**-**	**Lolos**
10	ARIMA(5,1,1)	-0,41629	0,0196	0,99992	0,9949	
11	ARIMA(2,1,2)	-0,42065	0,226	-0,19021	0,5921	
12	ARIMA(0,1,6)	-	-	-0,19319	0,9877	
13	ARIMA(0,1,7)	-	-	0,22066	0,9907	
14	ARIMA(0,1,8)	-	-	-0,21162	0,8535	
15	ARIMA(1,1,6)	-0,94299	<,0001	-0,10222	0,9891	
16	ARIMA(1,1,7)	-0,87053	0,0125	0,07248	0,9878	
17	ARIMA(1,1,8)	-0,64963	0,3658	-0,10163	0,9895	
18	ARIMA(2,1,6)	-0,10112	0,9391	-0,13899	0,9863	
19	ARIMA(2,1,7)	-0,692	0,0004	-0,27798	0,9764	
20	ARIMA(2,1,8)	-0,57781	0,011	-0,80172	0,9906	
21	ARIMA(5,1,6)	0,35704	0,994	-0,05899	0,996	
22	ARIMA(5,1,7)	-0,05479	0,9863	0,19102	0,9938	
23	ARIMA(5,1,8)	-0,55872	0,8641	0,08738	0,993	

Tiga model dengan parameter AR dan MA yang signifikan adalah ARIMA(1,1,0), ARIMA(5,1,0) dan ARIMA(2,1,0). Model-model tersebut memiliki hasil estimasi parameter AR dan MA yang signifikan. Tiga model yang lolos tersebut akan dilakukan uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

5.1.3. Evaluasi dan Pemilihan Model Univariat

Dari hasil estimasi model ARIMA(p,d,q) untuk kasus demam berdarah dan kasus influenza tahap selanjutnya adalah menguji kelayakan model ARIMA dengan melihat keberadaan

white noise pada residualnya. Model yang residualnya tidak *white noise* (memiliki autokorelasi) akan dieliminasi. Dari uji kelayakan model ini diperoleh model yang layak.

Tahap selanjutnya adalah mengevaluasi hasil peramalan model-model yang layak dengan melihat nilai AIC dan MAPE dari hasil peramalan. Model yang akan dipilih adalah model yang memiliki nilai AIC dan MAPE lebih kecil.

5.1.3.1. Uji kelayakan model

Uji kelayakan model dilakukan dengan menganalisis *white noise* dari residual dengan uji Ljung Box. Hipotesa awal (H0) dari Ljung Box adalah *white noise*. Untuk mengetahui apakah residual ini *white noise* atau tidak melihat nilai signifikansi dari Chi-Square (Pr>ChiSq). Jika nilai signifikan (<0,05), maka H0 ditolak [31].

Keberadaan *white noise* pada residual model menunjukkan bahwa model tersebut sudah layak [24]. Namun jika residual model tersebut tidak *white noise*, atau masih memiliki autokorelasi, maka model tersebut masih bisa diperbaiki, dan harus dieliminasi.

Tabel 5.12 merupakan hasil uji kelayakan model ARIMA untuk kasus demam berdarah. Hasil dari uji kelayakan model menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMA untuk kasus demam berdarah sudah layak.

Tabel 5.12 - Uji kelayakan model ARIMA demam berdarah

Model	Chi-Square	Pr > ChiSq	Keterangan
ARIMA(1,0,3)	1,16	0,5612	White Noise
ARIMA(1,0,1)	7,95	0,0933	White Noise
ARIMA(2,0,0)	8,3	0,0811	White Noise
ARIMA(2,0,1)	6,91	0,0747	White Noise
ARIMA(4,1,0)	1,07	0,5844	White Noise

Tabel 5.13 merupakan hasil uji kelayakan model ARIMA untuk kasus influenza. Hasil dari uji kelayakan model menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMA untuk kasus influenza sudah layak.

Tabel 5.13 - Uji kelayakan model ARIMA kasus influenza

Model	Chi-Square	Pr > ChiSq	Keterangan
ARIMA(1,1,0)	9,51	0,0905	White Noise
ARIMA(5,1,0)	3,93	0,05	White Noise
ARIMA(2,1,0)	6,02	0,1974	White Noise

5.1.3.2. Pemilihan Model

Pemilihan model akan dilakukan terhadap model-model yang sudah dihasilkan. Pemilihan model dilakukan berdasarkan nilai MAPE. Model dengan nilai MAPE yang paling kecil yang dipilih.

Tabel 5.14 merupakan perbandingan hasil peramalan dari identifikasi model ARIMA(p,d,q) untuk kasus demam berdarah. Seluruh model ini disusun berdasarkan variabel dependen VDBDLOG2, dengan hasil peramalan telah ditransformasikan kembali ke kondisi normal.

```
data work.hasilperamalan;
    set work.hasilperamalan;
    x_hasil = exp(xlog);
    forecast_hasil = exp(forecast + std*std/2);
    l95_hasil = exp(l95);
    u95_hasil = exp(u95);
run;
```

SAS Code 5.4 - Transformasi dari log ke normal

VDBDLOG2 merupakan hasil transformasi logaritma dua kali pada variabel VDBD, sehingga untuk memperoleh nilai VDBD dari hasil peramalan dengan variabel dependen

VDBDLOG2 dilakukan dengan menjalankan perintah pada SAS Code 5.4.

Tabel 5.14 - Hasil peramalan ARIMA(p,d,q) univariat untuk kasus demam berdarah

No	Model ARIMA(p,d,q)	MAPE
1	ARIMA(2,0,1)	32,34%
2	ARIMA(1,0,3)	32,56%
3	ARIMA(2,0,0)	33,37%
4	ARIMA(4,1,0)	33,87%
5	ARIMA(1,0,1)	34,19%

Pada Tabel 5.14 ARIMA(2,0,1) memiliki performa paling baik dibandingkan hasil identifikasi model lainnya berdasarkan nilai MAPE. ARIMA(2,0,1) dengan variabel dependen ditransformasi logaritma dua kali dipilih menjadi model ARIMA univariat dengan performa paling baik untuk kasus demam berdarah. Tabel 5.15 merupakan perbandingan hasil peramalan dari identifikasi model ARIMA(p,d,q) untuk kasus influenza.

Tabel 5.15 - Hasil peramalan ARIMA(p,d,q) univariat untuk kasus influenza

No	Model ARIMA(p,d,q)	MAPE
1	ARIMA(5,1,0)	54,78%
2	ARIMA(2,1,0)	58,99%
3	ARIMA(1,1,0)	65,57%

Pada Tabel 5.15 ARIMA(5,1,0) performa paling baik dibandingkan hasil identifikasi model lainnya, dengan MAPE senilai 54,78%. Model ARIMA(5,1,0) dipilih sebagai dasar pemodelan multivariat kasus influenza.

5.2. Model Multivariat

Pada bagian berikut ini akan dijelaskan mengenai pemodelan multivariat dengan metode ARIMAX. Tahapan dalam penyusunan model multivariat ini berangkat dari model univariat dengan performa terbaik.

Dari model dengan performa terbaik tersebut, model multivariat diidentifikasi berdasarkan variabel eksternal yang memiliki korelasi signifikan dengan variabel dependen. Hasil identifikasi tersebut selanjutnya dibawa ke tahap estimasi parameter, yang tujuannya adalah untuk mencari nilai parameter AR, MA dan Variabel Eksternal yang signifikan (pvalue <0,05). Model yang hasil parameternya tidak signifikan akan dibuang atau dimodifikasi nilai order AR atau MA hingga mendapatkan hasil estimasi parameter AR, MA dan variabel eksternal yang signifikan.

Setelah ditemukan hasil estimasi parameter tersebut dilakukan uji kelayakan model, dengan melihat keberadaan *white noise* pada nilai residual. Model yang menunjukkan keberadaan *white noise* pada residual adalah model yang sudah layak.

5.2.1. Model ARIMA Multivariat (ARIMAX) untuk Kasus Demam Berdarah

Pemodelan ARIMAX untuk kasus demam berdarah dilakukan dengan menganalisis kandidat variabel eksternal atau disebut juga dengan variabel eksternal yang bisa menjadi bagian dari model. Nilai parameter p, d, q diambil dari model ARIMA univariat yang memiliki performa paling baik. Untuk kasus demam berdarah model univariat dengan performa paling baik adalah ARIMA(2,0,1) dengan variabel dependen kasus demam berdarah yang ditransformasi secara logaritma sebanyak dua kali, seperti yang dilakukan terhadap variabel dependen pada pemodelan univariat.

5.2.1.1. Kandidat Variabel Eksternal untuk Kasus Demam Berdarah

Tabel 5.16 merupakan deskripsi kandidat variabel eksternal yang akan digunakan untuk pembentukan model ARIMAX kasus demam berdarah. Variabel eksternal diambil dari trainset dan sudah ditransformasi secara logaritma sebanayak dua kali, menyesuaikan dengan transformasi yang dilakukan terhadap variabel dependen (VDBDLOG2). Masing-masing variabel eksternal ini digeser dengan operasi lag 1 sampai dengan 10. Sehingga total ada 77 variabel eksternal yang perlu dianalisis.

Tabel 5.16 - Kandidat variabel eksternal ARIMAX demam berdarah

No	Variabel Eksternal	Sumber	Keterangan
1	dbd_log2	Google Trends	Search index bulanan dari pencarian kata-kata "dbd"
2	demam_log2	Google Trends	Search index bulanan dari pencarian kata-kata "demam"
3	demamberdarah_log2	Google Trends	Search index bulanan dari pencarian kata-kata "demam berdarah"
4	dengue_log2	Google Trends	Search index bulanan dari pencarian kata-kata "dengue"
5	HMD_log2	BMKG	Kelembaban relatif per bulan
6	SUHU_log2	BMKG	Suhu rata-rata per bulan
7	CH_log2	BMKG	Curah hujan rata-rata per bulan

Dari 77 kandidat variabel eksternal tersebut dilakukan uji korelasi terhadap VDBDLOG2, dengan signifikansi p-value < 0,05. Variabel eksternal yang memiliki nilai p-value lebih dari 0,05 akan dieliminasi. Tabel 5.17 merupakan rangkuman hasil uji korelasi 77 kandidat variabel eksternal. Keterangan uji korelasi pada Tabel 5.17 adalah sebagai berikut :

- Tanda * menandakan bahwa variabel tersebut berkorelasi signifikan dengan variabel dependen VDBDLOG2
- Angka pada variabel yang berkorelasi signifikan menandakan kekuatan korelasi berdasarkan interpretasi hasil dari nilai Spearman Coefficient.
  1. *(1)* menandakan korelasi yang ***SANGAT KUAT***
  2. *(2)* menandakan korelasi yang ***KUAT***
  3. *(3)* menandakan korelasi yang ***SEDANG***
  4. *(4)* menandakan korelasi yang ***LEMAH*** , sehingga tidak perlu dihiraukan.

Tabel 5.17 - uji korelasi kandidat variabel eksternal ARIMAX demam berdarah

Variabel Eksternal	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag6	Lag7	Lag 8	Lag 9	Lag10
SUHU_log2	-0,039	0,188	0,288	0,399* *(3)*	0,319* *(3)*	0,314* *(3)*	0,304	0,206	0,192	0,110	-0,027
CH_log2	0,292* *(4)*	0,449* *(3)*	0,463* *(3)*	0,361* *(3)*	0,284	0,019	-0,213	-0,452* *(3)*	-0,639* *(2)*	-0,466* *(3)*	-0,272
HMD_log2	0,621* *(2)*	0,744* *(1)*	0,695* *(2)*	0,554* *(2)*	0,343* *(3)*	0,090	-0,136	-0,351* *(3)*	-0,455* *(3)*	-0,472* *(3)*	-0,431* *(3)*
dengue_log2	0,373* *(3)*	0,555* *(2)*	0,541* *(2)*	0,265	0,278	-0,018	-0,058	-0,172	-0,318* *(3)*	-0,1582	-0,059
demamberdarah_log2	0,0942	0,202	0,307* *(3)*	0,166	0,089	-0,076	-0,227	-0,386* *(3)*	-0,331* *(3)*	-0,221	-0,0594
demam_log2	0,098	0,264	0,350* *(3)*	0,283	0,227	0,0287	-0,172	-0,187	-0,321* *(3)*	-0,217	-0,0281
dbd_log2	-0,0203	0,182	0,423* *(3)*	0,407* *(3)*	0,263	0,031	-0,216	-0,432* *(3)*	-0,535* *(2)*	-0,474* *(3)*	-0,320

Dari keseluruhan 77 kandidat variabel eksternal terdapat 33 variabel eksternal yang menunjukkan korelasi yang signifikan terhadap VDBDLOG2.

Dari 33 variabel eksternal yang lolos, disaring kembali dengan membuang variabel eksternal dengan korelasi LEMAH terhadap VDBDLOG2. Tabel 5.18 merupakan penjelasan detil variabel yang memiliki korelasi signifikan terhadap VDBDLOG2 dan digunakan dalam pembuatan model.

Tabel 5.18 - Variabel dengan korelasi signifikan terhadap VDBDLOG2

No	Nama Variabel Eksternal	Lag	Spearman Coeff.	pvalue	Interpretasi korelasi
1	CH_log2_1	1	0,44876	0,0018	Sedang
2	CH_log2_2	2	0,46295	0,0014	Sedang
3	CH_log2_3	3	0,36119	0,016	Sedang
4	CH_log2_7	7	-0,45243	0,0034	Sedang
5	CH_log2_8	8	-0,63889	<,0001	kuat
6	CH_log2_9	9	-0,46621	0,0032	Sedang
7	dbd_log2_2	2	0,42303	0,0038	Sedang
8	dbd_log2_3	3	0,40723	0,0061	Sedang
9	dbd_log2_7	7	-0,43214	0,0054	Sedang
10	dbd_log2_8	8	-0,5353	0,0004	kuat
11	dbd_log2_9	9	-0,47363	0,0027	Sedang
12	demam_log2_2	2	0,34957	0,0186	Sedang
13	demam_log2_8	8	-0,3213	0,0461	Sedang
14	demamberdarah_log2_2	2	0,30744	0,0399	Sedang
15	demamberdarah_log2_7	7	-0,38607	0,0139	Sedang
16	demamberdarah_log2_8	8	-0,33048	0,0399	Sedang

No	Nama Variabel Eksternal	Lag	Spearman Coeff.	pvalue	Interpretasi korelasi
17	dengue_log2	0	0,3731	0,0098	Sedang
18	dengue_log2_1	1	0,55471	<,0001	kuat
19	dengue_log2_2	2	0,54058	0,0001	kuat
20	dengue_log2_8	8	-0,31774	0,0487	Sedang
21	HMD_log2	0	0,62135	<,0001	kuat
22	HMD_log2_1	1	0,74382	<,0001	sangat kuat
23	HMD_log2_2	2	0,69506	<,0001	kuat
24	HMD_log2_3	3	0,55362	<,0001	kuat
25	HMD_log2_4	4	0,34261	0,0245	Sedang
26	HMD_log2_7	7	-0,35099	0,0264	Sedang
27	HMD_log2_8	8	-0,45533	0,0036	Sedang
28	HMD_log2_9	9	-0,47155	0,0028	Sedang
29	HMD_log2_10	10	-0,43118	0,0077	Sedang
30	SUHU_log2_3	3	0,39861	0,0074	Sedang
31	SUHU_log2_4	4	0,31914	0,037	Sedang
32	SUHU_log2_5	5	0,31409	0,0428	Sedang

Variabel eksternal pada Tabel 5.18 akan digunakan untuk menyusun model ARIMAX untuk kasus demam berdarah.

5.2.1.2. Identifikasi model ARIMAX kasus demam berdarah

Pada bagian berikut dijelaskan mengenai pengidentifikasian model ARIMAX untuk kasus demam berdarah. Dasar parameter p, d, dan q untuk model ARIMAX kasus demam berdarah adalah ARIMAX(2,0,1). Nilai parameter tersebut dipilih dari model dengan performa terbaik pada model univariat kasus demam berdarah.

5.2.1.2.1. Identifikasi model ARIMAX dengan variabel eksternal iklim

Pemodelan ARIMAX dengan variabel eksternal iklim melibatkan seluruh variabel eksternal iklim yang memiliki korelasi signifikan terhadap variabel dependen VDBDLOG2. Variabel eksternal iklim yang digunakan dijelaskan pada Tabel 5.19.

Tabel 5.19 - Variabel eksternal iklim demam berdarah

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	CH_log2_1	1	Sedang
2	CH_log2_2	2	Sedang
3	CH_log2_3	3	Sedang
4	CH_log2_7	7	Sedang
5	CH_log2_8	8	kuat
6	CH_log2_9	9	Sedang
7	HMD_log2_1	1	sangat kuat
8	HMD_log2_2	2	kuat
9	HMD_log2_3	3	kuat
10	HMD_log2_4	4	Sedang
11	HMD_log2_7	7	Sedang
12	HMD_log2_8	8	Sedang
13	HMD_log2_9	9	Sedang
14	HMD_log2_10	10	Sedang
15	SUHU_log2_3	3	Sedang
16	SUHU_log2_4	4	Sedang

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
17	SUHU_log2_5	5	Sedang
18	HMD_log2	0	kuat

Selain dilakukan pemodelan dengan menggunakan 1 variabel eksternal iklim, dibuat juga kombinasi variabel eksternal iklim yang terdiri dari variabel Curah Hujan (CH), kelembaban relatif (HMD), dan suhu udara rata-rata (SUHU). Contoh kombinasi variabel iklim dijelaskan pada Tabel 5.20.

Tabel 5.20 - Contoh kombinasi variabel iklim

Kombinasi Iklim	Variabel eksternal
Iklim 1	SUHU_LOG2_3
	CH_LOG2_1
	HMD_LOG2_0

Terdapat total 198 kombinasi variabel eksternal iklim yang dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(2,0,1). Model yang dipilih adalah model yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (pvalue < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.1.2.2. Identifikasi model ARIMAX dengan variabel eksternal Google Trends

Pemodelan ARIMAX dengan variabel eksternal Google Trends melibatkan seluruh variabel eksternal Google Trends yang memiliki korelasi signifikan terhadap variabel dependen VDBDLOG2. Variabel eksternal Google Trends yang digunakan, dijelaskan pada Tabel 5.21.

Tabel 5.21 - Variabel eksternal Google Trends demam berdarah

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	dbd_log2_2	2	Sedang
2	dbd_log2_3	3	Sedang
3	dbd_log2_7	7	Sedang
4	dbd_log2_8	8	kuat
5	dbd_log2_9	9	Sedang
6	demam_log2_2	2	Sedang
7	demam_log2_8	8	Sedang
8	demamberdarah_log2_2	2	Sedang
9	demamberdarah_log2_7	7	Sedang
10	demamberdarah_log2_8	8	Sedang
11	dengue_log2	0	Sedang
12	dengue_log2_1	1	kuat
13	dengue_log2_2	2	kuat
14	dengue_log2_8	8	Sedang

Selain dilakukan pemodelan dengan menggunakan 1 variabel eksternal iklim, dibuat juga kombinasi variabel eksternal Google Trends yang terdiri dari variabel hasil pencarian kata “dbd” (dbd), “demam” (demam), “demam berdarah” (demamberdarah) dan “dengue” (dengue). Contoh kombinasi variabel Google Trends dijelaskan pada Tabel 5.22.

Tabel 5.22 - Contoh kombinasi variabel iklim

Kombinasi Iklim	Variabel eksternal
GT1	dengue_log2_0
	demamberdarah_log2_2
	demam_log2_2
	dbd_log2_2

Terdapat total 120 kombinasi variabel eksternal Google Trends yang dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(2,0,1). Model yang dipilih adalah model yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (pvalue < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.1.2.3. Identifikasi model ARIMAX dengan variabel eksternal Google Trends dan Iklim

Pemodelan ARIMAX dengan variabel eksternal Google Trends dan Iklim melibatkan kombinasi variabel eksternal Google Trends dan Iklim yang memiliki korelasi signifikan KUAT dan SANGAT KUAT terhadap variabel dependen VDBDLOG2. Hal ini bertujuan untuk membatasi jumlah kombinasi variabel Google Trends dan Iklim yang dihasilkan. Tabel 5.23 menjelaskan variabel eksternal Google Trends dan iklim yang digunakan.

Tabel 5.23 - Variabel eksternal Google Trends dan Iklim kasus demam berdarah

No	Jenis Variabel	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	Google Trends	dbd_log2_8	8	kuat
2		dengue_log2_1	1	kuat
3		dengue_log2_2	2	kuat
4	Iklim	CH_log2_8	8	kuat
5		HMD_log2_1	1	sangat kuat
6		HMD_log2_2	2	kuat
7		HMD_log2_3	3	kuat
8		HMD_log2	0	kuat

Dari 8 variabel eksternal tersebut dihasilkan 15 kombinasi variabel Google Trends dan Iklim. Tabel 5.24 menjelaskan kombinasi variabel Google Trends dan iklim yang akan dianalisis untuk model ARIMAX kasus demam berdarah.

Tabel 5.24 - Kombinasi Google Trends dan Iklim kasus demam berdarah

Kombinasi Variabel	Variabel Eksternal	Kombinasi Variabel	Variabel Eksternal
GTIKLIM 1	CH_log2_8	GTIKLIM 9	HMD_log2_1
	dbd_log2_8		dengue_log2_2
GTIKLIM 2	CH_log2_8	GTIKLIM 10	HMD_log2_2
	dengue_log2_1		dbd_log2_8
GTIKLIM 3	CH_log2_8	GTIKLIM 11	HMD_log2_2
	dengue_log2_2		dengue_log2_1
GTIKLIM 4	HMD_log2	GTIKLIM 12	HMD_log2_2
	dbd_log2_8		dengue_log2_2
GTIKLIM 5	HMD_log2	GTIKLIM 13	HMD_log2_3
	dengue_log2_1		dbd_log2_8
GTIKLIM 6	HMD_log2	GTIKLIM 14	HMD_log2_3
	dengue_log2_2		dengue_log2_1
GTIKLIM 7	HMD_log2_1	GTIKLIM 15	HMD_log2_3
	dbd_log2_8		dengue_log2_2
GTIKLIM 8	HMD_log2_1		
	dengue_log2_1		

Seluruh kombinasi variabel eksternal Google Trends dan Iklim pada Tabel 5.24 akan dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(2,0,1). Model yang dipilih adalah model yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (p-value < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.1.3. Estimasi parameter model ARIMAX kasus demam berdarah

Dari model yang sudah diidentifikasi tahapan selanjutnya adalah mengestimasi nilai parameter AR, MA dan Variabel eksternal. Tahap estimasi parameter, yang bertujuan untuk mencari nilai parameter AR, MA dan Variabel Eksternal yang signifikan (pvalue atau Pr>|t| memiliki nilai <0,05). Model yang hasil parameternya tidak signifikan akan dibuang atau dimodifikasi nilai order AR atau MA hingga mendapatkan hasil estimasi parameter AR, MA dan variabel eksternal yang signifikan. Pada bagian ini dijelaskan model-model yang sudah memiliki nilai parameter AR, MA dan Variabel eksternal yang signifikan.

5.2.1.3.1. Estimasi parameter model ARIMAX dengan variabel eksternal iklim

Estimasi parameter model ARIMAX dengan variabel eksternal iklim dilakukan terhadap 216 model ARIMAX, 18 model dengan satu variabel eksternal dan 198 model dengan kombinasi variabel. Dari 216 model tersebut ditemukan 9 model dengan nilai estimasi parameter AR, MA dan Variabel Eksternal yang signifikan. Sembilan model tersebut di jelaskan pada Tabel 5.25.

Tabel 5.25 - Hasil Estimasi ARIMAX variabel iklim kasus demam berdarah

No	Model	Variabel Eksternal			AR		MA	
		Var	Estimasi	Pr> |t|	Estimasi	Pr> |t|	Estimasi	Pr> |t|
1	ARIMAX(1,0,0)	CH_log2_8	-0,089	0,0009	0,78634	<,0001		
2	ARIMAX(0,0,3)	HMD_LOG2_1	5,874	<,0001			-0,51954	0,0003
3	ARIMAX(0,0,1)	HMD_LOG2_1	8,052	<,0001			-0,7166	<,0001
4	ARIMAX(0,0,2)	HMD_LOG2_1	7,690	<,0001			-0,30798	0,0416
5	ARIMAX(0,0,1)	CH_LOG2_8	-0,089	0,0007			-0,71106	<,0001
		HMD_LOG2_2	4,620	<,0001				
6	ARIMAX(1,0,0)	CH_LOG2_8	-0,070	0,0027	0,61915	<,0001		
		HMD_LOG2_1	5,686	<,0001				
7	ARIMAX(1,0,0)	CH_LOG2_8	-0,076	0,0043	0,61039	0,0002		
		HMD_LOG2_2	4,761	0,0038				
8	ARIMAX(0,0,1)	CH_LOG2_8	-0,067	0,0078	-0,49893	0,0009		
		HMD_LOG2_1	6,688	<,0001				
9	ARIMAX(0,0,2)	CH_LOG2_8	-0,089	0,0009	-0,36916	0,0482		
		HMD_LOG2_2	5,182	0,0002				

Sembilan model pada Tabel 5.25 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

5.2.1.3.2. Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends

Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dilakukan terhadap 134 model ARIMAX, 14 model dengan satu variabel eksternal dan 120 model dengan kombinasi variabel Google Trends. Dari 134 model tersebut ditemukan 6 model dengan nilai estimasi parameter AR, MA dan Variabel Eksternal yang signifikan. Delapan model tersebut di jelaskan pada Tabel 5.26. Enam model pada Tabel 5.26 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

Tabel 5.26 - Hasil Estimasi ARIMAX variabel Google Trends kasus demam berdarah

No	Model	Variabel Eksternal			AR		MA	
		Vars	Estimasi	Pr> |t|	Estimasi	Pr> |t|	Estimasi	Pr> |t|
1	ARIMAX(2,0,1)	dengue_log2_1	0,35677	0,033	-0,74428	<,0001	0,54451	0,0498
2	ARIMAX(1,0,3)	dengue_log2_1	0,32721	0,0486	0,552	0,005	-0,42622	0,0162
3	ARIMAX(1,0,3)	demamberdarah_log2_2	0,16776	0,0003	0,63735	<,0001	-0,62746	<,0001
4	ARIMAX(1,0,0)	demamberdarah_log2_2	0,17203	0,0005	0,90258	<,0001		
		dbd_log2_2	0,04358	0,0481				
5	ARIMAX(1,0,0)	dengue_log2_1	0,52375	0,0021	0,89091	<,0001		
		demamberdarah_log2_2	0,17694	0,0002				
6	ARIMAX(2,0,0)	dengue_log2_1	0,40698	0,0036	-0,33857	0,0239		
		demamberdarah_log2_2	0,1574	<,0001				

5.2.1.3.3. Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dan Iklim

Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dan iklim dilakukan terhadap 15 model ARIMAX dengan kombinasi variabel Google Trends dan iklim. Dari 15 model tersebut, terdapat 2 model yang memiliki hasil estimasi parameter yang signifikan. Tabel 5.27 menjelaskan model yang memiliki hasil estimasi parameter signifikan untuk model ARIMAX dengan variabel eksternal Google Trends dan Iklim.

Tabel 5.27 - Hasil estimasi parameter ARIMAX dengan Variabel Google Trends dan Iklim kasus demam berdarah

No	Model	Variabel Eksternal			AR	
		Vars	Estimasi	Pr> |t|	Estimasi	Pr> |t|
1	ARIMAX (1,0,0)	HMD_log2_1	6,0427	<,0001	0,77804	<,0001
		dengue_log2_1	0,39142	0,0257		
2	ARIMAX (1,0,0)	HMD_log2_2	4,9874	0,0016	0,79652	<,0001
		dengue_log2_1	0,48619	0,0086		

Model pada Tabel 5.27 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

5.2.1.4. Uji kelayakan model ARIMAX kasus demam berdarah

Uji kelayakan model dilakukan dengan menganalisis *white noise* dari residual dengan uji Ljung Box. Hipotesa awal (H0) dari Ljung Box adalah *white noise*. Untuk mengetahui apakah residual ini *white noise* atau tidak melihat nilai signifikansi dari Chi-Square (Pr>ChiSq). Jika nilai signifikan (<0,05), maka H0 ditolak [32].

Keberadaan *white noise* pada residual model menunjukkan bahwa model tersebut sudah layak [25]. Namun jika residual model tersebut tidak *white noise*, atau masih memiliki autokorelasi, maka model tersebut masih bisa diperbaiki, dan harus dieliminasi.

Tabel 5.28 merupakan hasil uji *white noise* residual untuk model ARIMAX dengan variabel eksternal Google Trends. Hasil dari uji *white noise* pada Tabel 5.28 menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMAX dengan variabel eksternal Google Trends kasus demam berdarah sudah layak.

Tabel 5.28 - Uji white noise ARIMAX variabel eksternal Google Trends kasus demam berdarah

No	Model	Variabel eksternal	Chi-Square	Pr>ChiSq	Ket.
1	ARIMAX(2,0,1)	dengue_log2_1	4,4	0,2215	White Noise
2	ARIMAX(1,0,3)	dengue_log2_1	0,53	0,7657	White Noise
3	ARIMAX(1,0,3)	demamberdarah_log2_2	3,24	0,1982	White Noise
4	ARIMAX(1,0,0)	demamberdarah_log2_2 dbd_log2_2	12,9	0,2996	White Noise
5	ARIMAX(1,0,0)	dengue_log2_1 demamberdarah_log2_2	18,09	0,0795	White Noise
6	ARIMAX(2,0,0)	dengue_log2_1 demamberdarah_log2_2	14,89	0,1361	White Noise

Tabel 5.29 merupakan hasil uji *white noise* residual untuk model ARIMAX dengan variabel eksternal iklim. Hasil dari uji *white noise* pada Tabel 5.29 menunjukkan bahwa

seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMAX dengan variabel eksternal iklim kasus demam berdarah sudah layak.

Tabel 5.29 - Uji white noise ARIMAX variabel eksternal iklim kasus demam berdarah

No	Model	Variabel eksternal	Chi-Square	Pr>ChiSq	Ket
1	ARIMAX(1,0,0)	CH_LOG2_8	14,11	0,227	White Noise
2	ARIMAX(0,0,3)	HMD_LOG2_1	6,39	0,6998	White Noise
3	ARIMAX(0,0,1)	HMD_LOG2_1	13,46	0,2643	White Noise
4	ARIMAX(0,0,2)	HMD_LOG2_1	10,45	0,402	White Noise
5	ARIMAX(0,0,1)	CH_LOG2_8 HMD_LOG2_2	10,34	0,5003	White Noise
6	ARIMAX(1,0,0)	CH_LOG2_8 HMD_LOG2_1	10,37	0,4973	White Noise
7	ARIMAX(1,0,0)	CH_LOG2_8 HMD_LOG2_2	7,23	0,7802	White Noise
8	ARIMAX(0,0,1)	CH_LOG2_8 HMD_LOG2_1	8,82	0,6386	White Noise
9	ARIMAX(0,0,2)	CH_LOG2_8 HMD_LOG2_2	6,88	0,7366	White Noise

Tabel 5.29 merupakan hasil uji *white noise* residual untuk model ARIMAX dengan variabel eksternal Google

Trends dan iklim. Hasil dari uji *white noise* pada Tabel 5.29 menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMAX dengan variabel eksternal Google Trends dan iklim kasus demam berdarah sudah layak.

Tabel 5.30 - Uji white noise ARIMAX variabel eksternal Google Trends dan iklim kasus demam berdarah

No	Model	Variabel eksternal	Chi-Square	Pr>ChiSq	Ket
1	ARIMAX(1,0,0)	HMD_log2_1 dengue_log2_1	12,69	0,3142	White Noise
2	ARIMAX(1,0,0)	HMD_log2_2 dengue_log2_1	10,61	0,4765	White Noise

5.2.2. Model ARIMA Multivariat (ARIMAX) untuk Kasus Influenza

Pemodelan ARIMAX untuk kasus influenza dilakukan dengan menganalisis kandidat variabel independen atau disebut juga dengan variabel eksternal yang bisa menjadi bagian dari model. Nilai parameter p, d, q diambil dari model ARIMA univariat yang memiliki performa paling baik. Untuk kasus influenza model univariat dengan performa paling baik adalah ARIMA(5,1,0) dengan variabel dependen kasus influenza (VFLU).

5.2.2.1. Kandidat Variabel Eksternal untuk Kasus Influenza

Tabel 5.31 merupakan deskripsi kandidat variabel eksternal yang akan digunakan untuk pembentukan model ARIMAX kasus influenza. Variabel eksternal diambil dari trainset menyesuaikan dengan transformasi yang dilakukan terhadap variabel dependen (VFLU). Variabel eksternal pada Tabel 5.31 tidak memerlukan transformasi, karena variabel dependen VFLU tidak melalui transformasi apapun. Masing-masing variabel eksternal ini digeser dengan operasi lag 1 sampai dengan 10. Sehingga total ada 99 variabel eksternal yang perlu dianalisis.

Tabel 5.31 - Kandidat variabel eksternal ARIMAX influenza

No	Variabel Eksternal	Sumber	Keterangan
1	batuk	Google Trends	Search index bulanan dari pencarian kata-kata "batuk"
2	batukpilek	Google Trends	Search index bulanan dari pencarian kata-kata "batuk pilek"
3	bersin	Google Trends	Search index bulanan dari pencarian kata-kata "bersin"

No	Variabel Eksternal	Sumber	Keterangan
4	flu	Google Trends	Search index bulanan dari pencarian kata-kata "flu"
5	influenza	Google Trends	Search index bulanan dari pencarian kata-kata "influenza"
6	pilek	Google Trends	Search index bulanan dari pencarian kata-kata "pilek"
7	HMD	BMKG	Kelembaban relatif per bulan
8	SUHU	BMKG	Suhu rata-rata per bulan
9	CH	BMKG	Curah hujan rata-rata per bulan

Dari 99 kandidat variabel eksternal tersebut dilakukan uji korelasi terhadap VFLU, dengan signifikansi p-value < 0,05. Variabel eksternal yang memiliki nilai p-value lebih dari 0,05 akan dieliminasi. Tabel 5.32 merupakan rangkuman hasil uji korelasi 99 kandidat variabel eksternal.

Keterangan hasil uji korelasi pada Tabel 5.31 adalah sebagai berikut :

- Tanda * menandakan bahwa variabel tersebut berkorelasi signifikan dengan variabel dependen VFLU
- Angka pada variabel yang berkorelasi signifikan menandakan kekuatan korelasi berdasarkan interpretasi hasil dari nilai Spearman Coefficient.
  1. *(1)* menandakan korelasi yang ***SANGAT KUAT***
  2. *(2)* menandakan korelasi yang ***KUAT***
  3. *(3)* menandakan korelasi yang ***SEDANG***
  4. *(4)* menandakan korelasi yang ***LEMAH*** , sehingga tidak perlu dihiraukan.

Tabel 5.32 - uji korelasi kandidat variabel eksternal ARIMAX influenza

Variabel Eksternal	Lag0	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	Lag9	Lag10
batuk	0,633* *(2)*	0,694* *(2)*	0,593* *(2)*	0,531* *(2)*	0,502* *(2)*	0,478* *(3)*	0,380* *(3)*	0,258	0,148	0,220	0,218
batukpilek	0,602* *(2)*	0,568* *(2)*	0,536* *(2)*	0,492* *(3)*	0,463* *(3)*	0,298	0,123	0,096	0,053	0,132	0,272
bersin	-0,026	0,078	0,090	-0,236	-0,146	-0,040	-0,139	-0,242	-0,114	0,093	-0,129
flu	0,410* *(3)*	0,457* *(3)*	0,540* *(2)*	0,517* *(2)*	0,399* *(3)*	0,397* *(3)*	0,428* *(3)*	0,263	0,141	0,0757	0,125
influenza	-0,351* *(3)*	-0,513* *(2)*	-0,383* *(3)*	-0,244	-0,084	-0,097	0,043	0,025	0,071	-0,016	-0,099
pilek	0,218	0,230	0,135	0,223	0,259	0,069	-0,070	-0,004	-0,060	0,009	0,056
HMD	-0,028	-0,0487	0,055	0,128	0,104	0,0195	-0,126	-0,237	-0,426* *(3)*	-0,546* *(2)*	-0,476* *(3)*
SUHU	-0,307* *(3)*	-0,127	-0,168	-0,120	-0,096	-0,088	0,029	0,146	0,032	-0,0203	-0,325* *(3)*
CH	-0,0913	-0,107	0,024	0,246	0,232	0,101	0,004	-0,339* *(3)*	-0,331* *(3)*	-0,540* *(2)*	-0,4997* *(3)*

Dari keseluruhan 99 kandidat variabel eksternal terdapat 31 variabel eksternal yang menunjukkan korelasi yang signifikan terhadap VFLU.

Dari 31 variabel eksternal yang lolos, disaring kembali dengan membuang variabel eksternal dengan korelasi LEMAH terhadap VFLU. Tabel 5.33 merupakan penjelasan detil variabel yang memiliki korelasi signifikan terhadap VFLU dan digunakan dalam pembuatan model.

.

Tabel 5.33 - Variabel dengan korelasi signifikan terhadap VFLU

No	Nama Variabel Eksternal	Lag	Spearman Coeff.	pvalue	Interpretasi korelasi
1	batuk	0	0,6326	<,0001	kuat
2	batuk_1	1	0,69415	<,0001	kuat
3	batuk_2	2	0,59279	<,0001	kuat
4	batuk_3	3	0,53052	0,0002	kuat
5	batuk_4	4	0,5017	0,0006	kuat
6	batuk_5	5	0,47785	0,0014	Sedang
7	batuk_6	6	0,37973	0,0143	Sedang
8	batukpilek	0	0,60188	<,0001	kuat
9	batukpilek_1	1	0,56802	<,0001	kuat
10	batukpilek_2	2	0,53627	0,0001	kuat
11	batukpilek_3	3	0,49165	0,0007	Sedang
12	batukpilek_4	4	0,46264	0,0018	Sedang
13	flu	0	0,41015	0,0042	Sedang
14	flu_1	1	0,45665	0,0014	Sedang
15	flu_2	2	0,53966	0,0001	kuat
16	flu_3	3	0,51736	0,0003	kuat
17	flu_4	4	0,3988	0,0081	Sedang

No	Nama Variabel Eksternal	Lag	Spearman Coeff.	pvalue	Interpretasi korelasi
18	flu_5	5	0,39701	0,0092	Sedang
19	flu_6	6	0,42839	0,0052	Sedang
20	influenza	0	-0,35099	0,0156	Sedang
21	influenza_1	1	-0,51301	0,0003	kuat
22	influenza_2	2	-0,38256	0,0095	Sedang
23	CH_7	7	-0,33921	0,0323	Sedang
24	CH_8	8	-0,33083	0,0397	Sedang
25	CH_9	9	-0,53973	0,0005	kuat
26	CH_10	10	-0,49972	0,0016	Sedang
27	SUHU	0	-0,30691	0,0359	Sedang
28	SUHU_10	10	-0,32491	0,0497	Sedang
29	HMD_8	8	-0,42642	0,0068	Sedang
30	HMD_9	9	-0,54639	0,0004	kuat
31	HMD_10	10	-0,47546	0,0029	Sedang

Variabel eksternal pada Tabel 5.33 akan digunakan untuk menyusun model ARIMAX untuk kasus influenza.

5.2.2.2. Identifikasi Model ARIMAX kasus influenza

Pada bagian berikut dijelaskan mengenai pengidentifikasian model ARIMAX untuk kasus influenza. Dasar parameter p, d, dan q untuk model ARIMAX kasus demam berdarah adalah ARIMAX(5,1,0). Nilai parameter tersebut dipilih dari model dengan performa terbaik pada model univariat kasus demam berdarah.

5.2.2.2.1. Identifikasi Model ARIMAX dengan variabel eksternal iklim

Pemodelan ARIMAX dengan variabel eksternal iklim melibatkan seluruh variabel eksternal iklim yang memiliki korelasi signifikan terhadap variabel dependen VFLU. Variabel eksternal iklim yang digunakan dijelaskan pada Tabel 5.34.

Tabel 5.34 - Variabel eksternal iklim influenza

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	CH_7	7	Sedang
2	CH_8	8	Sedang
3	CH_9	9	kuat
4	CH_10	10	Sedang
5	SUHU	0	Sedang
6	SUHU_10	10	Sedang
7	HMD_8	8	Sedang
8	HMD_9	9	kuat
9	HMD_10	10	Sedang

Selain dilakukan pemodelan dengan menggunakan 1 variabel eksternal iklim, dibuat juga kombinasi variabel eksternal iklim yang terdiri dari variabel Curah Hujan (CH), kelembaban relatif (HMD), dan suhu udara rata-rata (SUHU). Contoh kombinasi variabel iklim dijelaskan pada Tabel 5.35.

Tabel 5.35 - Contoh kombinasi variabel iklim kasus influenza

Kombinasi Iklim	Variabel eksternal
Iklim_flu1	SUHU
	CH_7
	HMD_8

Terdapat total 24 kombinasi variabel eksternal iklim yang dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(5,1,0). Model yang dipilih adalah model

yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (pvalue < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.2.2.2. Identifikasi Model ARIMAX dengan variabel eksternal Google Trends

Pemodelan ARIMAX dengan variabel eksternal Google Trends melibatkan seluruh variabel eksternal Google Trends yang memiliki korelasi signifikan terhadap variabel dependen VFLU. Variabel eksternal Google Trends yang digunakan, dijelaskan pada Tabel 5.36.

Tabel 5.36 - Variabel eksternal Google Trends influenza

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	batuk	0	kuat
2	batuk_1	1	kuat
3	batuk_2	2	kuat
4	batuk_3	3	kuat
5	batuk_4	4	kuat
6	batuk_5	5	Sedang
7	batuk_6	6	Sedang
8	batukpilek	0	kuat
9	batukpilek_1	1	kuat
10	batukpilek_2	2	kuat
11	batukpilek_3	3	Sedang
12	batukpilek_4	4	Sedang
13	flu	0	Sedang
14	flu_1	1	Sedang
15	flu_2	2	kuat
16	flu_3	3	kuat
17	flu_4	4	Sedang

No	Nama Variabel Eksternal	Lag	Interpretasi korelasi
18	flu_5	5	Sedang
19	flu_6	6	Sedang
20	influenza	0	Sedang
21	influenza_1	1	kuat
22	influenza_2	2	Sedang

Selain dilakukan pemodelan dengan menggunakan 1 variabel eksternal iklim, dibuat juga kombinasi variabel eksternal Google Trends yang terdiri dari variabel hasil pencarian kata "dbd" (dbd), "demam" (demam), "demam berdarah" (demamberdarah) dan "dengue" (dengue). Contoh kombinasi variabel Google Trends dijelaskan pada Tabel 5.37.

Tabel 5.37 - Contoh kombinasi variabel Google Trends kasus influenza

Kombinasi Iklim	Variabel eksternal
GT1	influenza
	Flu
	batukpilek
	batuk

Terdapat total 245 kombinasi variabel eksternal Google Trends yang dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(5,1,0). Model yang dipilih adalah model yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (pvalue < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.2.2.3. Identifikasi Model ARIMAX dengan variabel eksternal Google Trends dan Iklim

Pemodelan ARIMAX dengan variabel eksternal Google Trends dan Iklim melibatkan kombinasi variabel

eksternal Google Trends dan Iklim yang memiliki korelasi signifikan KUAT dan SANGAT KUAT terhadap variabel dependen VFLU. Hal ini bertujuan untuk membatasi jumlah kombinasi variabel Google Trends dan Iklim yang dihasilkan. Tabel 5.38 menjelaskan variabel eksternal Google Trends dan iklim kasus influenza yang digunakan.

Tabel 5.38 - Variabel eksternal Google Trends dan Iklim kasus influenza

No	Jenis Variabel	Nama Variabel Eksternal	Lag	Interpretasi korelasi
1	Google Trends	batuk	0	kuat
2		batuk_1	1	kuat
3		batuk_2	2	kuat
4		batuk_3	3	kuat
5		batuk_4	4	kuat
6		batukpilek_1	1	kuat
7		batukpilek_2	2	kuat
8		flu_2	2	kuat
9		flu_3	3	kuat
10		influenza_1	1	kuat
11	Iklim	CH_9	9	kuat
12		HMD_9	8	kuat

Dari 12 variabel eksternal tersebut dihasilkan 22 kombinasi variabel Google Trends dan Iklim. Tabel 5.39 menjelaskan kombinasi variabel Google Trends dan iklim yang akan dianalisis untuk model ARIMAX kasus influenza.

Tabel 5.39 - Kombinasi Google Trends dan Iklim kasus influenza

Kombinasi Variabel	Variabel Eksternal	Kombinasi Variabel	Variabel Eksternal
GTIKLIM_FLU 1	CH_9	GTIKLIM_FLU 12	HMD_9
	batuk		batuk
GTIKLIM_FLU 2	CH_9	GTIKLIM_FLU 13	HMD_9
	batuk_1		batuk_1

Kombinasi Variabel	Variabel Eksternal	Kombinasi Variabel	Variabel Eksternal
GTIKLIM_ FLU 3	CH_9	GTIKLIM_ FLU 14	HMD_9
	batuk_2		batuk_2
GTIKLIM_ FLU 4	CH_9	GTIKLIM_ FLU 15	HMD_9
	batuk_3		batuk_3
GTIKLIM_ FLU 5	CH_9	GTIKLIM_ FLU 16	HMD_9
	batuk_4		batuk_4
GTIKLIM_ FLU 6	CH_9	GTIKLIM_ FLU 17	HMD_9
	batukpilek		batukpilek
GTIKLIM_ FLU 7	CH_9	GTIKLIM_ FLU 18	HMD_9
	batukpilek_1		batukpilek_1
GTIKLIM_ FLU 8	CH_9	GTIKLIM_ FLU 19	HMD_9
	batukpilek_2		batukpilek_2
GTIKLIM_ FLU 9	CH_9	GTIKLIM_ FLU 20	HMD_9
	flu_2		flu_2
GTIKLIM_ FLU 10	CH_9	GTIKLIM_ FLU 21	HMD_9
	flu_3		flu_3
GTIKLIM_ FLU 11	CH_9	GTIKLIM_ FLU 22	HMD_9
	influenza_1		influenza_1

Seluruh kombinasi variabel eksternal Google Trends dan Iklim pada Tabel 5.39 akan dianalisis dengan melihat nilai hasil nilai estimasi parameter ARIMAX(2,0,1). Model yang dipilih adalah model yang memiliki nilai estimasi parameter AR, MA, dan variabel eksternal yang signifikan (pvalue < 0,05). Model dengan nilai estimasi parameter AR, MA dan variabel eksternal yang tidak signifikan akan dibuang, atau dimodifikasi hingga diperoleh model dengan nilai estimasi parameter yang signifikan.

5.2.2.3. Estimasi parameter model ARIMAX kasus influenza

Dari model yang sudah diidentifikasi tahapan selanjutnya adalah mengestimasi nilai parameter AR, MA dan Variabel eksternal. Tahap estimasi parameter, yang bertujuan untuk mencari nilai parameter AR, MA dan Variabel Eksternal yang signifikan (pvalue atau Pr>|t| memiliki nilai <0,05). Model yang hasil parameternya tidak signifikan akan dibuang atau dimodifikasi nilai order AR atau MA hingga mendapatkan hasil estimasi parameter AR, MA dan variabel eksternal yang signifikan. Pada bagian ini dijelaskan model-model yang sudah memiliki nilai parameter AR, MA dan Variabel eksternal yang signifikan.

5.2.2.3.1. Estimasi parameter model ARIMAX dengan variabel eksternal iklim

Estimasi parameter model ARIMAX dengan variabel eksternal iklim dilakukan terhadap 33 model ARIMAX, 9 model dengan satu variabel eksternal dan 24 model dengan kombinasi variabel. Dari 33 model tersebut ditemukan 8 model dengan nilai estimasi parameter AR, MA dan Variabel Eksternal yang signifikan. Delapan model tersebut di jelaskan pada Tabel 5.40. Delapan model pada Tabel 5.40 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

Tabel 5.40 - Hasil Estimasi ARIMAX variabel iklim kasus influenza

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
1	ARIMAX (5,1,0)	CH_9	-0,3245	<,0001	-0,4226	0,0307
2	ARIMAX (5,1,0)	CH_10	-0,2634	0,0005	-0,519	0,0062
3	ARIMAX (5,1,0)	HMD_8	-0,2982	0,0015	-0,4439	0,0335
4	ARIMAX (5,1,0)	HMD_9	-0,3787	<,0001	-0,6017	0,0004
5	ARIMAX (5,1,0)	HMD_10	-0,2474	0,0267	-0,4958	0,0094

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
6	ARIMAX (5,1,0)	SUHU	-1,095	0,0216	-0,4725	0,0185
		HMD_8	-0,2761	0,0005		
7	ARIMAX (5,1,0)	CH_10	-0,2118	0,0001	-0,6153	0,0004
		HMD_8	-0,2169	0,0005		
8	ARIMAX (5,1,0)	SUHU_10	-1,471	0,0227	-0,453	0,0146
		CH_10	-0,2612	0,0001		

5.2.2.3.2. Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends

Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dilakukan terhadap 267 model ARIMAX, 22 model dengan satu variabel eksternal dan 245 model dengan kombinasi variabel Google Trends. Dari 267 model tersebut ditemukan 10 model dengan nilai estimasi parameter AR, MA dan Variabel Eksternal yang signifikan. Sepuluh model tersebut di jelaskan pada Tabel 5.41. Sepuluh model pada Tabel 5.41 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

Tabel 5.41 - Hasil Estimasi ARIMAX variabel Google Trends kasus influenza

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
1	ARIMAX (5,1,0)	influenza	-0,0455	0,037	-0,4373	0,0126
2	ARIMAX (5,1,0)	influenza_1	0,0615	0,0065	-0,3791	0,0322
3	ARIMAX (5,1,0)	flu_2	0,104	<,0001	-0,5926	0,0005
4	ARIMAX (5,1,0)	flu_3	0,0823	0,0084	-0,5143	0,0046
5	ARIMAX (5,1,0)	batukpilek	0,0752	0,0241	-0,5600	0,0008

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
6	ARIMAX (5,1,0)	batukpilek_3	0,1008	0,0045	-0,4145	0,0188
7	ARIMAX (5,1,0)	batuk_1	0,1619	<,0001	-0,6220	<,0001
8	ARIMAX (5,1,0)	batuk_2	0,0883	0,0411	-0,5018	0,0051
9	ARIMAX (5,1,0)	flu_3	0,0621	0,0437	-0,4802	0,0072
		influenza_1	0,0485	0,0306		
10	ARIMAX (5,1,0)	batuk_1	0,1824	0,0002	-0,4561	0,0021
		batukpilek	-0,1279	0,0016		
		influenza_1	-0,06909	0,0001		

5.2.2.3.3. Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dan iklim

Estimasi parameter model ARIMAX dengan variabel eksternal Google Trends dan iklim dilakukan terhadap 22 model ARIMAX dengan kombinasi variabel Google Trends dan iklim. Dari 22 model tersebut, terdapat 4 model yang memiliki hasil estimasi parameter yang signifikan. Tabel 5.42 menjelaskan model yang memiliki hasil estimasi parameter signifikan untuk model ARIMAX dengan variabel eksternal Google Trends dan Iklim untuk kasus influenza.

Tabel 5.42 - Hasil estimasi parameter ARIMAX dengan Variabel Google Trends dan Iklim kasus influenza

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
1	ARIMAX (2,1,0)	CH_9	-0.3560	<.0001	-0.6855	<.0001
		batukpilek	-0.0604	0.0395		
2	ARIMAX (5,1,0)	HMD_9	-0.2383	0.0012	-0.7303	<.0001
		batuk_1	0.10049	0.0039		

No	Model	External variables			AR	
		Vars	Est.	Pr>|t|	Est.	Pr> |t|
3	ARIMAX (5,1,0)	HMD_9	-0.3105	<.0001	-0.5731	0.0010
		influenza_1	-0.0456	0.0198		
4	ARIMAX (2,1,0)	HMD_9	-0.289	0.0022	-0.5128	0.0008
		influenza_1	-0.0579	0.0090		

Model pada Tabel 5.42 akan melewati uji kelayakan model dengan melihat keberadaan *white noise* pada residual.

5.2.2.4. Uji kelayakan model ARIMAX kasus influenza

Uji kelayakan model dilakukan dengan menganalisis *white noise* dari residual dengan uji Ljung Box. Hipotesa awal (H0) dari Ljung Box adalah *white noise*. Untuk mengetahui apakah residual ini *white noise* atau tidak melihat nilai signifikansi dari Chi-Square (Pr>ChiSq). Jika nilai signifikan (<0,05), maka H0 ditolak [31].

Keberadaan *white noise* pada residual model menunjukkan bahwa model tersebut sudah layak [24]. Namun jika residual model tersebut tidak *white noise*, atau masih memiliki autokorelasi, maka model tersebut masih bisa diperbaiki, dan harus dieliminasi.

Tabel 5.43 merupakan hasil uji *white noise* residual untuk model ARIMAX dengan variabel eksternal iklim. Hasil dari uji *white noise* pada Tabel 5.43 menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMAX dengan variabel eksternal iklim kasus influenza sudah layak.

Tabel 5.43 - Uji white noise ARIMAX variabel eksternal iklim kasus influenza

No	Model	Variabel eksternal	WHITE NOISE		Ket.
			Chi-Square	Pr>ChiSq	
1	ARIMA X(5,1,0)	CH_9	2,98	0,8872	White Noise
2	ARIMA X(5,1,0)	CH_10	6,6	0,4712	White Noise
3	ARIMA X(5,1,0)	HMD_8	12,67	0,0805	White Noise
4	ARIMA X(5,1,0)	HMD_9	4,36	0,738	White Noise
5	ARIMA X(5,1,0)	HMD_10	8,59	0,2837	White Noise
6	ARIMA X(5,1,0)	SUHU	14,11	0,0493	White Noise
		HMD_8			White Noise
7	ARIMA X(5,1,0)	CH_10	10,3	0,1722	White Noise
		HMD_8			White Noise
8	ARIMA X(5,1,0)	SUHU_10	5,56	0,5918	White Noise
		CH_10			White Noise

Tabel 5.44 merupakan hasil uji *white noise* residual untuk model ARIMAX kasus influenza dengan variabel eksternal Google Trends. Hasil dari uji *white noise* pada Tabel 5.44 menunjukkan bahwa seluruh model hasil estimasi memiliki residual *white noise*. Sehingga bisa disimpulkan bahwa model ARIMAX dengan variabel eksternal Google Trends kasus influenza sudah layak.

Tabel 5.44 - Uji white noise ARIMAX variabel eksternal Google Trends kasus influenza

No	Model	Variabel eksternal	WHITE NOISE		Ket.
			Chi-Square	Pr>Chi Sq	
1	ARIMA X(5,1,0)	influenza	5,58	0,5898	White Noise
2	ARIMA X(5,1,0)	influenza_1	5,71	0,5738	White Noise
3	ARIMA X(5,1,0)	flu_2	5,34	0,619	White Noise
4	ARIMA X(5,1,0)	flu_3	8,94	0,2573	White Noise
5	ARIMA X(5,1,0)	batukpilek	9,71	0,2059	White Noise
6	ARIMA X(5,1,0)	batukpilek_3	8,64	0,2797	White Noise
7	ARIMA X(5,1,0)	batuk_1	7,78	0,352	White Noise
8	ARIMA X(5,1,0)	batuk_2	7,36	0,3927	White Noise
9	ARIMA X(5,1,0)	flu_3 influenza_1	9,99	0,1889	White Noise
10	ARIMA X(5,1,0)	batuk_1 batukpilek influenza_1	15,84	0,1044	White Noise

Tabel 5.45 merupakan hasil uji *white noise* residual untuk model ARIMAX dengan variabel eksternal Google Trends dan iklim. Hasil dari uji *white noise* pada Tabel 5.45 menunjukkan bahwa terdapat satu model yang memiliki residual tidak *white noise*, yakni model ARIMAX(2,1,0) dengan variabel eksternal HMD_9 dan influenza_1. Selain model tersebut, 3 model ARIMAX dengan variabel eksternal Google Trends dan iklim kasus influenza sudah layak.

Tabel 5.45 - Uji white noise ARIMAX variabel eksternal Google Trends dan iklim kasus influenza

No	Model	Var. eksternal	WHITE NOISE		Ket.
			Chi-Square	Pr>ChiSq	
1	ARIMAX (2,1,0)	CH_9 batukpilek	8,88	0,5437	White Noise
2	ARIMAX (5,1,0)	HMD_9 batuk_1	4,67	0,7006	White Noise
3	ARIMAX (5,1,0)	HMD_9 influenza_1	9,39	0,226	White Noise
4	ARIMAX (2,1,0)	HMD_9 influenza_1	9,87	0,0428	Tidak White Noise

5.2.3. Evaluasi Model Multivariat

Pada bagian ini seluruh model ARIMAX yang sudah layak baik untuk kasus demam berdarah dan kasus influenza, dievaluasi dengan melihat nilai MAPE untuk memilih model dengan performa terbaik.

5.2.3.1. Kasus Demam Berdarah

Pada kasus demam berdarah akan dipilih model ARIMAX dengan variabel eksternal iklim, model ARIMAX dengan variabel eksternal Google Trends, dan model ARIMAX dengan kombinasi variabel Google Trends dan iklim. Seluruh model ini disusun berdasarkan variabel dependen VDBDLOG2, beserta variabel eksternal yang transformasinya menyesuaikan variabel dependen VDBDLOG2. Hasil akhir peramalan telah ditransformasikan kembali ke kondisi normal

5.2.3.1.1. Model ARIMAX dengan variabel eksternal iklim

Tabel 5.46 merupakan performa 9 model ARIMAX dengan variabel eksternal iklim. Model ARIMAX kasus demam berdarah dengan variabel eksternal iklim memiliki performa

keseluruhan rata-rata MAPE 29,3%. Performa paling baik terdapat pada model ARIMAX(0,0,1) variabel eksternal CH_LOG2_8 dan HMD_LOG2_1 dengan nilai MAPE 24,78%.

Tabel 5.46 - Performa model ARIMAX variabel iklim kasus demam berdarah

No	Model	Variabel eksternal	MAPE
			Trainset
1	ARIMAX(1,0,0)	CH_log2_8	29,26%
2	ARIMAX(0,0,3)	HMD_LOG2_1	31,69%
3	ARIMAX(0,0,1)	HMD_LOG2_1	35,94%
4	ARIMAX(0,0,2)	HMD_LOG2_1	33,93%
5	ARIMAX(0,0,1)	CH_LOG2_8	28,60%
		HMD_LOG2_2	
6	ARIMAX(1,0,0)	CH_LOG2_8	24,89%
		HMD_LOG2_1	
7	ARIMAX(1,0,0)	CH_LOG2_8	26,91%
		HMD_LOG2_2	
8	*ARIMAX(0,0,1)*	*CH_LOG2_8*	*24,78%*
		HMD_LOG2_1	
9	ARIMAX(0,0,2)	CH_LOG2_8	27,31%
		HMD_LOG2_2	

5.2.3.1.2. Model ARIMAX dengan kombinasi variabel Google Trends

Tabel 5.47 merupakan performa 6 model ARIMAX dengan variabel eksternal Google Trends. Model ARIMAX kasus demam berdarah dengan variabel eksternal Google Trends memiliki performa keseluruhan rata-rata MAPE 31,28%. Performa paling baik terdapat pada model

ARIMAX(2,0,0) variabel eksternal dengue_log2_1 dan demamberdarah_log2_2 dengan nilai MAPE 29,11%.

Tabel 5.47 - Performa model ARIMAX variabel Google Trends kasus demam berdarah

No	Model	Variabel eksternal	MAPE
			Trainset
1	ARIMAX(2,0,1)	dengue_log2_1	30,95%
2	ARIMAX(1,0,3)	dengue_log2_1	30,29%
3	ARIMAX(1,0,3)	demamberdarah_log2_2	31,45%
4	ARIMAX(1,0,0)	demamberdarah_log2_2	34,88%
		dbd_log2_2	
5	ARIMAX(1,0,0)	dengue_log2_1	31,00%
		demamberdarah_log2_2	
6	*ARIMAX(2,0,0)*	*dengue_log2_1*	*29,11%*
		demamberdarah_log2_2	

5.2.3.1.3. Model ARIMAX dengan kombinasi variabel Google Trends dan iklim

Tabel 5.48 merupakan performa 2 model ARIMAX dengan variabel eksternal Google Trends dan iklim. Model ARIMAX kasus demam berdarah dengan variabel eksternal Google Trends dan iklim memiliki performa keseluruhan rata-rata MAPE 31%. Performa paling baik terdapat pada model ARIMAX(1,0,0) variabel eksternal HMD_log2_1 dan dengue_log2_1 dengan nilai MAPE 28,26%.

Tabel 5.48 - Performa model ARIMAX variabel Google Trends dan iklim kasus demam berdarah

No	Model	Variabel eksternal	MAPE
			Trainset
1	*ARIMAX(1,0,0)*	*HMD_log2_1*	*28,26%*
		dengue_log2_1	

No	Model	Variabel eksternal	MAPE Trainset
2	ARIMAX(1,0,0)	HMD_log2_2 dengue_log2_1	33,83%

5.2.3.2. Kasus Influenza

Pada kasus influenza akan dipilih model ARIMAX dengan kombinasi variabel eksternal iklim, model ARIMAX dengan kombinasi variabel eksternal Google Trends, dan model ARIMAX dengan kombinasi variabel Google Trends dan iklim. Seluruh model ini disusun berdasarkan variabel dependen VFLU.

5.2.3.2.1. Model ARIMAX dengan kombinasi variabel iklim

Tabel 5.49 merupakan performa 8 model ARIMAX dengan variabel eksternal iklim. Model ARIMAX kasus influenza dengan variabel eksternal iklim memiliki performa keseluruhan rata-rata MAPE 50%. Performa paling baik terdapat pada model ARIMAX(5,1,0) variabel eksternal HMD_9 dengan nilai MAPE 42,2%.

Tabel 5.49 - Performa model ARIMAX variabel iklim kasus influenza

No	Model	Variabel Eksternal	MAPE Trainset
1	ARIMAX(5,1,0)	CH_9	54,3%
2	ARIMAX(5,1,0)	CH_10	53,8%
3	ARIMAX(5,1,0)	HMD_8	52,6%
4	ARIMAX(5,1,0)	HMD_9	42,2%
5	ARIMAX(5,1,0)	HMD_10	55,2%
6	ARIMAX(5,1,0)	SUHU HMD_8	43,6%
7	ARIMAX(5,1,0)	CH_10 HMD_8	44,6%
8	ARIMAX(5,1,0)	SUHU_10 CH_10	53,6%

5.2.3.2.2. Model ARIMAX dengan kombinasi variabel Google Trends

Tabel 5.50 merupakan performa 10 model ARIMAX dengan variabel eksternal Google Trends. Model ARIMAX kasus influenza dengan variabel eksternal iklim memiliki performa keseluruhan rata-rata MAPE 54,03%. Performa paling baik terdapat pada model ARIMAX(5,1,0) variabel eksternal batuk_1 dengan nilai MAPE 46,2%.

Tabel 5.50 - Performa model ARIMAX variabel Google Trends kasus influenza

No	Model	Variabel eksternal	MAPE
			Trainset
1	ARIMAX(5,1,0)	influenza	60,7%
2	ARIMAX(5,1,0)	influenza_1	55,7%
3	ARIMAX(5,1,0)	flu_2	51,2%
4	ARIMAX(5,1,0)	flu_3	51,9%
5	ARIMAX(5,1,0)	batukpilek	56,9%
6	ARIMAX(5,1,0)	batukpilek_3	59,5%
7	*ARIMAX(5,1,0)*	*batuk_1*	*46,2%*
8	ARIMAX(5,1,0)	batuk_2	57,8%
9	ARIMAX(5,1,0)	flu_3	53,0%
		influenza_1	
10	ARIMAX(5,1,0)	batuk_1	47,3%
		batukpilek	
		influenza_1	

5.2.3.2.3. Model ARIMAX dengan kombinasi variabel Google Trends dan iklim

Tabel 5.51 merupakan performa 4 model ARIMAX dengan variabel eksternal Google Trends dan iklim. Model ARIMAX kasus influenza dengan variabel eksternal Google Trends iklim memiliki performa keseluruhan rata-rata MAPE

47,4%. Performa paling baik terdapat pada model ARIMAX(5,1,0) variabel eksternal HMD_9 dan influenza_1 dengan nilai MAPE 35,4%

Tabel 5.51 - Performa model ARIMAX variabel Google Trends dan iklim kasus influenza

No	Model	Variabel eksternal	MAPE
			Trainset
1	ARIMAX(2,1,0)	CH_9	61,8%
		batukpilek	
2	*ARIMAX(5,1,0)*	*HMD_9*	*35,4%*
		batuk_1	
3	ARIMAX(5,1,0)	HMD_9	41,1%
		influenza_1	
4	ARIMAX(2,1,0)	HMD_9	51,4%
		influenza_1	

5.3. Rangkuman dan persamaan hasil pemodelan

Bagian ini merangkum hasil pemodelan univariat dan multivariat yang dilakukan pada bagian sebelumnya. Persamaan model juga dijelaskan pada bagian ini.

5.3.1. Kasus demam berdarah

Rangkuman hasil pemodelan kasus demam berdarah ditunjukkan pada Tabel 5.52.

Tabel 5.52 - Rangkuman pemodelan kasus demam berdarah

No	Model	Variabel Eksternal	MAPE
1	ARIMA(2,0,1)	-	32,34%
2	ARIMAX(0,0,1)	CH_LOG2_8	24,78%
		HMD_LOG2_1	
3	ARIMAX(2,0,0)	dengue_log2_1	29,11%
		demamberdarah_log2_2	

No	Model	Variabel Eksternal	MAPE
4	ARIMAX(1,0,0)	HMD_LOG2_1	28,26%
		dengue_log2_1	

Model-model yang dihasilkan berupa persamaan. Nilai parameter pada persamaan ini diperoleh dari hasil estimasi pada SAS. Berikut adalah persamaan dari model yang dipilih untuk kasus demam berdarah.

1. Model univariat

Model univariat untuk kasus demam berdarah yang dipilih adalah model ARIMA(2,0,1). Persamaan 5.1 merupakan persamaan model ARIMA(2,0,1) untuk kasus demam berdarah.

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

$$dengan\ nilai\ c = \mu(1 - \phi_1 - \phi_2)$$

$$c = 1{,}46965(1 - 1{,}61305 - 0{,}76222) = 0{,}21923$$

Persamaan 5.1 - Persamaan model univariat kasus demam berdarah

y_t adalah nilai kasus demam berdarah dalam transformasi logaritmis (VDBDLOG2) pada periode t dan ε_t adalah nilai *random error* pada periode t. Dengan c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR.

2. Model multivariat dengan variabel eksternal iklim

Model multivariat dengan variabel eksternal iklim untuk kasus demam berdarah yang dipilih adalah model ARIMAX(0,0,1) dengan variabel eksternal curah hujan pada lag 8 dan variabel eksternal kelembaban pada lag 1. Persamaan 5.2 merupakan persamaan model ARIMAX(0,0,1) dengan variabel eksternal iklim.

$$y_t = c + \varepsilon_t - 0{,}49893\varepsilon_{t-1} - 0{,}06692CH_{t-8} + 6{,}68784HMD_{t-1}$$
$$dengan\ nilai\ c = -8{,}32902$$

Persamaan 5.2 - Persamaan model multivariat variabel eksternal iklim kasus demam berdarah

y_t adalah nilai kasus demam berdarah dalam transformasi logaritmis (VDBDLOG2) pada periode t dan ε_t adalah nilai *random error* pada periode t. CH_t adalah variabel eksternal curah hujan yang sudah ditransformasi logaritmis (CH_log2) pada periode t, karena terdapat operasi lag 8, maka pada persamaan menjadi CH_{t-8}. HMD_t adalah variabel eksternal kelembaban yang sudah ditransformasi logaritmis (HMD_log2) pada periode t, karena terdapat operasi lag 1 maka pada persamaan menjadi HMD_{t-1}. Nilai c merupakan konstanta yang diperoleh dari nilai rata-rata, karena model ini model MA murni, sehingga tidak perlu melibatkan nilai estimasi parameter AR untuk memperoleh nilai c.

3. Model multivariat dengan variabel eksternal Google Trends

Model multivariat dengan variabel eksternal Google Trends untuk kasus demam berdarah yang dipilih adalah model ARIMAX(2,0,0) dengan variabel eksternal search index query "dengue" pada lag 1 dan variabel eksternal search query "demam berdarah" pada lag 2. Persamaan 5.3 merupakan persamaan model ARIMAX(2,0,0) dengan variabel eksternal Google Trends.

$$y_t = c + \varepsilon_t + 1{,}18464y_{t-1} - 0{,}33857y_{t-2} + 0{,}40698x1_{t-1} + 0{,}1574x2_{t-2}$$
$$dengan\ nilai\ c = \mu(1 - \phi_1 - \phi_2)$$
$$c = 0{,}78128(1 - 1{,}8464 + 0{,}33857) = 0{,}120258$$

Persamaan 5.3 - Persamaan model multivariat variabel eksternal Google Trends kasus demam berdarah

y_t adalah nilai kasus demam berdarah dalam transformasi logaritmis (VDBDLOG2) pada periode t. $x1_t$ adalah variabel eksternal search index query "dengue" yang sudah ditransformasi logaritmis (dengue_log2) pada periode t, karena terdapat operasi lag 1, maka pada persamaan menjadi $x1_{t-1}$. $x2_t$ adalah variabel eksternal search index query "demam berdarah" yang sudah ditransformasi logaritmis (demamberdarah_log2) pada periode t, karena terdapat operasi lag 2 maka pada persamaan menjadi $x2_{t-2}$. Dengan c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR.

4. Model multivariat dengan variabel eksternal iklim dan Google Trends

Model multivariat dengan variabel eksternal iklim dan Google Trends untuk kasus demam berdarah yang dipilih adalah model ARIMAX(1,0,0) dengan variabel eksternal search index query "dengue" pada lag 1 dan variabel eksternal iklim kelembaban pada lag 1. Persamaan 5.4 merupakan persamaan model ARIMAX(1,0,0) dengan variabel eksternal iklim dan Google Trends kasus influenza.

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

$$dengan\ nilai\ c = \mu(1 - \phi_1)$$

$$c = -0{,}790183(1 - 0{,}77804) = -1{,}7539$$

Persamaan 5.4 - Persamaan model multivariat variabel eksternal iklim dan Google Trends kasus demam berdarah

y_t adalah nilai kasus demam berdarah dalam transformasi logaritmis (VDBDLOG2) pada periode t. $x1_t$ adalah variabel eksternal search index query "dengue" yang sudah ditransformasi logaritmis (dengue_log2) pada periode t, karena terdapat operasi lag 1, maka pada persamaan menjadi $x1_{t-1}$. HMD_t adalah variabel

eksternal iklim kelembaban yang sudah ditransformasi logaritmis (HMD_log2) pada periode t, karena terdapat operasi lag 1 maka pada persamaan menjadi HMD_{t-1}. Dengan c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR.

5.3.2. Kasus influenza

Rangkuman hasil pemodelan kasus influenza ditunjukkan pada Tabel 5.53.

Tabel 5.53 - Rangkuman pemodelan kasus influenza

No	Model	Variabel Eksternal	MAPE
1	ARIMA(5,1,0)	-	54,78%
2	ARIMAX(5,1,0)	HMD_9	42,20%
3	ARIMAX(5,1,0)	batuk_1	46,24%
4	ARIMAX(5,1,0)	HMD_9	35,40%
		batuk_1	

Model-model yang dihasilkan berupa persamaan. Berikut adalah persamaan dari model yang dipilih untuk kasus influenza.

1. Model univariat

Model univariat untuk kasus influenza yang dipilih adalah model ARIMA(5,1,0). Persamaan 5.5 merupakan persamaan untuk model univariat kasus influenza.

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

$$dengan\ nilai\ c = \mu(1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)$$

$$c = 0{,}115261(1+0{,}76151+0{,}31349+0{,}12029+0{,}33234+0{,}41649$$

$$c = 0{,}339344$$

Persamaan 5.5 - Persamaan model univariat kasus influenza(1)

y_t adalah nilai kasus influenza (VFLU) pada periode t. Variabel c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR. Model kasus influenza memiliki nilai parameter d=1, yang artinya model ini melalui proses differencing sebanyak 1 kali. Sehingga variabel responnya berupa *backshift operator (B)*. Persamaan 5.6 merupakan bentuk lain persamaan dari Persamaan 5.5 tanpa *backshift operator*.

$$By_t = y_{t-1}$$

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

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

Persamaan 5.6 - Persamaan model univariat kasus influenza(2)

2. Model multivariat dengan variabel eksternal iklim

Model multivariat dengan variabel eksternal iklim untuk kasus influenza yang dipilih adalah model ARIMAX(5,1,0) dengan variabel eksternal iklim kelembaban pada lag 9. Persamaan 5.7 merupakan persamaan untuk model multivariat dengan variabel eksternal iklim kasus influenza.

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

$$dengan\ nilai\ c,$$

$$c = \mu(1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)$$

$$c = 0{,}115261(1 + 0{,}76151 + 0{,}31349 + 0{,}12029 + 0{,}33234 + 0{,}41649$$

$$c = 0{,}339344$$

Persamaan 5.7 - Persamaan model multivariat variabel eksternal iklim kasus influenza(1)

y_t adalah nilai kasus influenza (VFLU) pada periode t. Variabel c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR. HMD_t adalah variabel eksternal iklim kelembaban pada periode t, karena terdapat operasi lag 9 maka pada persamaan menjadi HMD_{t-9}.

Model kasus influenza memiliki nilai parameter d=1, yang artinya model ini melalui proses differencing sebanyak 1 kali. Sehingga variabel responnya berupa *backshift operator (B)*. Persamaan 5.8 merupakan bentuk lain persamaan dari Persamaan 5.7 tanpa *backshift operator*.

$$By_t = y_{t-1}$$

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

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

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

Persamaan 5.8 - Persamaan model multivariat variabel eksternal iklim kasus influenza(2)

3. Model multivariat dengan variabel eksternal Google Trends

Model multivariat dengan variabel eksternal Google Trends untuk kasus influenza yang dipilih adalah model ARIMAX(5,1,0) dengan variabel eksternal search index query "batuk" pada lag 1. Persamaan 5.9 merupakan persamaan untuk model multivariat dengan variabel eksternal Google Trends kasus influenza.

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

$dengan\ nilai\ c,$

$$c = \mu(1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)$$

$$c = 0{,}041714(1 + 0{,}90297 + 0{,}63271 + 0{,}57546 + 0{,}78566 + 0{,}62196)$$

$$c = 0{,}0092312$$

Persamaan 5.9 - Persamaan model multivariat variabel eksternal Google Trends kasus influenza(1)

y_t adalah nilai kasus influenza (VFLU) pada periode t. Variabel c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR. $x3_t$ adalah variabel eksternal search index query "batuk", karena terdapat operasi lag 1 maka pada persamaan menjadi $x3_{t-1}$.

Model kasus influenza memiliki nilai parameter d=1, yang artinya model ini melalui proses differencing sebanyak 1 kali. Sehingga variabel responnya berupa *backshift operator (B)*. Persamaan 5.10 merupakan bentuk lain persamaan dari Persamaan 5.9 tanpa *backshift operator*.

$$By_t = y_{t-1}$$

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

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

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

Persamaan 5.10 - Persamaan model multivariat variabel eksternal Google Trends kasus influenza(2)

4. Model multivariat dengan variabel eksternal iklim dan Google Trends

Model multivariat dengan variabel eksternal Google Trends untuk kasus influenza yang dipilih adalah model ARIMAX(5,1,0) dengan variabel eksternal search index query "batuk" pada lag 1 dan variabel eksternal iklim kelembaban pada lag 9. Persamaan 5.11 merupakan persamaan untuk model multivariat dengan variabel eksternal iklim dan Google Trends kasus influenza.

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

$$dengan\ nilai\ c,$$

$$c = \mu(1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)$$

$$c = 0{,}03530(1+1{,}03457+0{,}73255+0{,}64346+0{,}90542+0{,}73025)$$

$$c = 0{,}178118$$

Persamaan 5.11 -Persamaan model multivariat variabel eksternal iklim dan Google Trends kasus influenza(1)

y_t adalah nilai kasus influenza (VFLU) pada periode t. Variabel c merupakan konstanta yang diperoleh dari nilai rata-rata dan nilai seluruh estimasi parameter AR. $x3_t$ adalah variabel eksternal search index query "batuk" pada periode t, karena terdapat operasi lag 1 maka pada persamaan menjadi $x3_{t-1}$. HMD_t adalah variabel eksternal iklim kelembaban pada periode t, karena terdapat operasi lag 9 maka pada persamaan menjadi HMD_{t-9}.

Model kasus influenza memiliki nilai parameter d=1, yang artinya model ini melalui proses differencing sebanyak 1 kali. Sehingga variabel responnya berupa *backshift operator (B)*. Persamaan 5.12 merupakan bentuk

lain persamaan dari Persamaan 5.11 tanpa *backshift operator*.

$$By_t = y_{t-1}$$

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

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

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

Persamaan 5.12 - Persamaan model multivariat variabel eksternal Google Trends kasus influenza(2)

BAB VI
HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai hasil peramalan dan analisis hasil pemodelan.

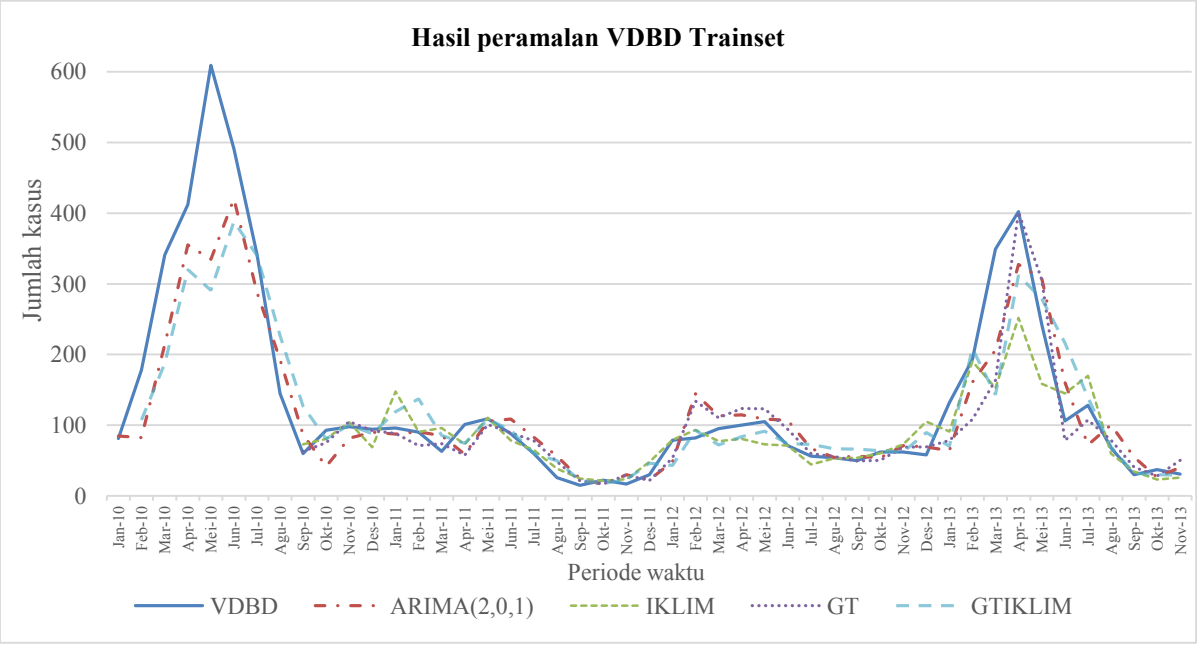
6.1. Hasil peramalan

Dari model-model yang sudah dipilih, digunakan untuk meramal dengan menggunakan data testset untuk masing-masing kasus penyakit. Tabel 6.1 merupakan performa hasil peramalan kasus demam berdarah terhadap data testset.

Tabel 6.1 - Performa hasil peramalan kasus demam berdarah

No	Model	Variabel Eksternal	MAPE	
			Trainset	Testset
1	ARIMA(2,0,1)	-	32,34%	32,34%
2	ARIMAX(0,0,1)	CH_LOG2_8 HMD_LOG2_1	24,78%	26%
3	ARIMAX(2,0,0)	dengue_log2_1 demamberdarah_log2_2	29,11%	19%
4	ARIMAX(1,0,0)	HMD_LOG2_1 dengue_log2_1	28,26%	30%

Terlihat pada Tabel 6.1 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 6.1 merupakan plot perbandingan hasil peramalan menggunakan data trainset.



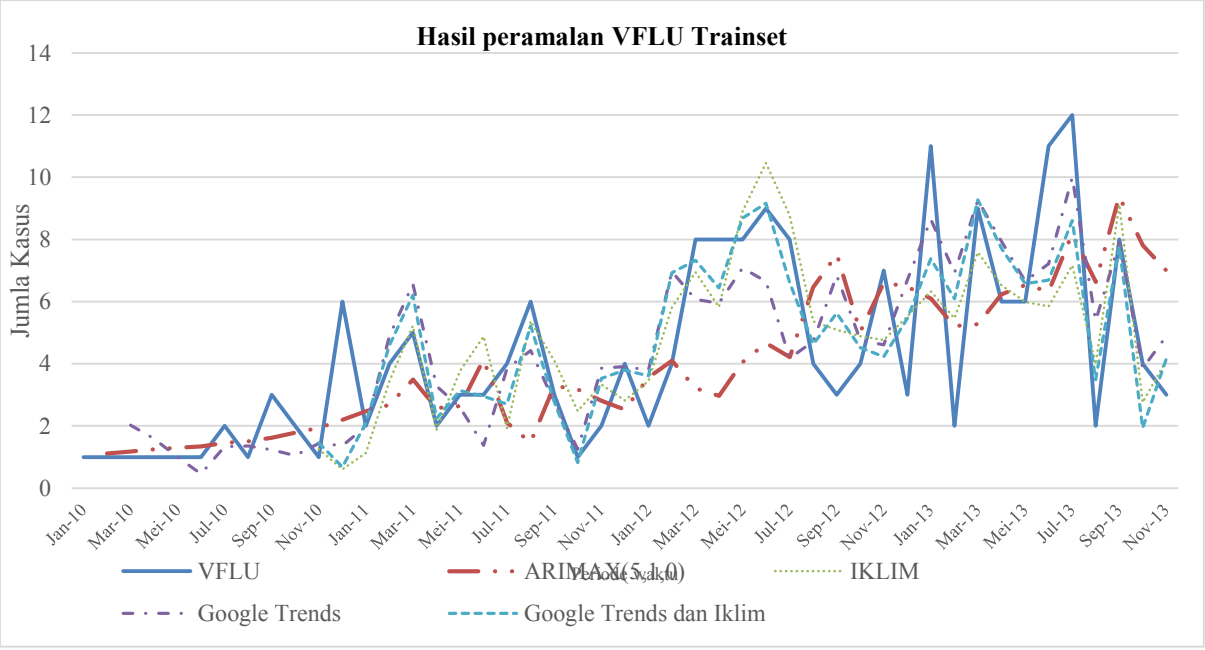
Gambar 6.1 - Hasil peramalan VDBD trainset

Tabel 6.2 merupakan performa hasil peramalan kasus influenza terhadap testset. Terlihat pada Tabel 6.2 performa hasil peramalan kasus influenza dengan menggunakan data trainset mengalami peningkatan seiring dengan penambahan variabel eksternal.

Tabel 6.2 - Performa hasil peramalan kasus influenza

No	Model	Variabel Eksternal	MAPE	
			Trainset	Testset
1	ARIMA(5,1,0)		54,78%	102%
2	ARIMAX(5,1,0)	HMD_9	42,20%	72,40%
3	ARIMAX(5,1,0)	batuk_1	46,24%	84,50%
4	ARIMAX(5,1,0)	HMD_9 batuk_1	35,40%	72%

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 6.2 merupakan plot perbandingan hasil peramalan menggunakan data trainset. Hasil MAPE pada testset naik drastis dikarennakan pola data trainset dan testset variabel dependen VFLU yang sangat berbeda.



Gambar 6.2 - Hasil peramalan VFLU trainset

6.2. Analisis hasil dan diskusi

Kasus demam berdarah dan kasus influenza menunjukkan korelasi yang siginfikan 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. Tabel 6.3 menunjukkan beberapa model ARIMAX yang disusun dengan variabel eksternal dengan kekuatan korelasi sedang.

Tabel 6.3 - Model ARIMAX dengan variabel eksternal kekuatan korelasi sedang

Kasus	Model	Variabel eksternal	Kekuatan korelasi
Influenza	ARIMAX(5,1,0)	CH_10	sedang
	ARIMAX(5,1,0)	HMD_8	sedang
	ARIMAX(5,1,0)	HMD_10	sedang
	ARIMAX(5,1,0)	SUHU	sedang
		HMD_8	sedang
	ARIMAX(5,1,0)	CH_10	sedang
		HMD_8	sedang
	ARIMAX(5,1,0)	SUHU_10	sedang
		CH_10	sedang
Demam berdarah	ARIMAX(1,0,3)	demamberdarah_log2_2	sedang
	ARIMAX(1,0,0)	demamberdarah_log2_2	sedang

Kasus	Model	Variabel eksternal	Kekuatan korelasi
		dbd_log2_2	sedang
	ARIMAX(1,0,0)	dengue_log2_1	kuat
		demamberdarah _log2_2	sedang
	ARIMAX(2,0,0)	dengue_log2_1	kuat
		demamberdarah _log2_2	sedang

Pada penelitian ini terlihat bahwa model ARIMAX dengan performa terbaik kebanyakan menggunakan variabel eksternal dengan kekuatan korelasi kuat atau sangat kuat. Tabel 6.4 menunjukkan model-model dengan performa terbaik dan keterangan mengenai kekuatan korelasi dari variabel eksternal masing-masing model.

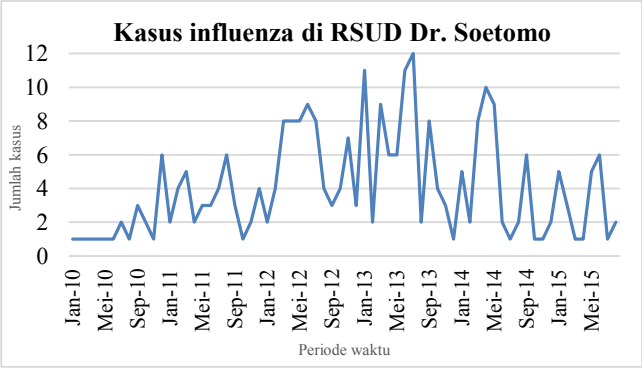
Tabel 6.4 - Model dengan performa terbaik dan keterangan korelasi

Kasus	Model	Variabel Eksternal	Keteran gan korelasi
Demam berdarah	ARIMAX(0,0,1)	CH_LOG2_8	kuat
		HMD_LOG2_1	sangat kuat
	ARIMAX(2,0,0)	dengue_log2_1	kuat
		demamberdarah_log2 _2	sedang
	ARIMAX(1,0,0)	HMD_LOG2_1	sangat kuat
		dengue_log2_1	sedang
Influenza	ARIMAX(5,1,0)	HMD_9	kuat
	ARIMAX(5,1,0)	batuk_1	kuat
	ARIMAX(5,1,0)	HMD_9	kuat
		batuk_1	kuat

Performa hasil peramalan kasus demam berdarah sudah layak, walaupun hasilnya masih belum baik. Model ARIMA

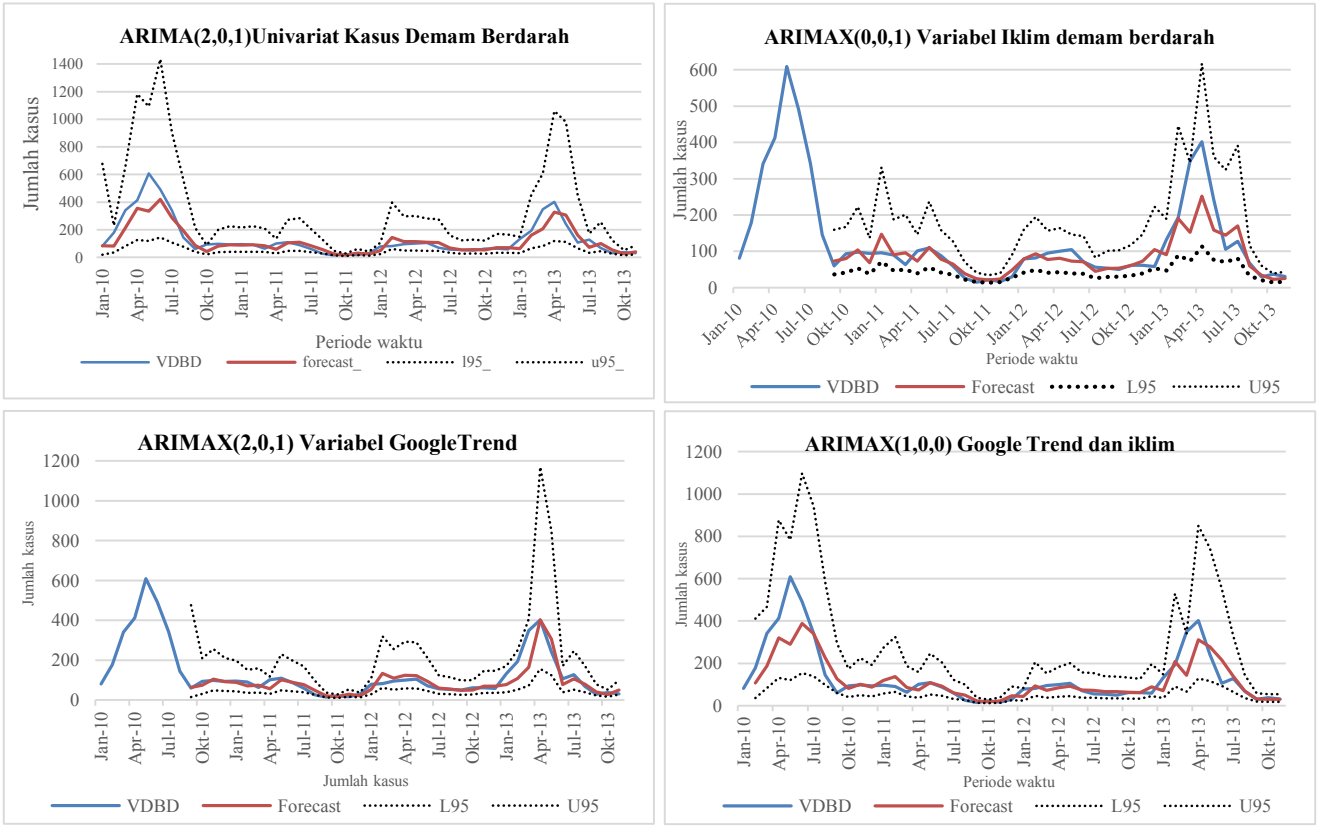
univariat dan multivariat untuk kasus demam berdarah sudah layak karena residual dari seluruh modelnya sudah *white noise*, dan data aktual kasus demam berdarah berada di dalam *range* 95% *confidence interval* dari seluruh model. Hal ini ditampilkan pada Gambar 6.4.

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 (Gambar 6.3), dengan naik dan turun yang tajam. Model ARIMA akan memiliki performa yang baik selama pergerakan datanya bertahap, dan tidak fluktuatif [37]. Alternatif metode peramalan untuk data dengan tingkat fluktuasi yang sangat tinggi bisa menggunakan metode peramalan *artifical neural network* (ANN). Menurut Sharda dan Patil (1990) pada artikel oleh Zhang (1998) metode ANN menunjukkan performa yang lebih baik dibandingkan dengan metode ARIMA dengan data yang polanya tidak reguler, dengan fluktuasi yang tinggi [38].



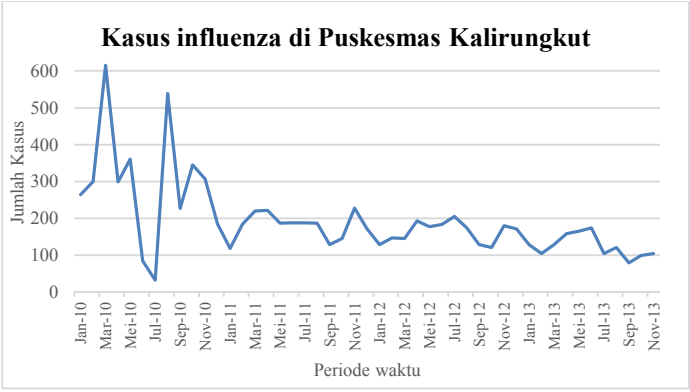
Gambar 6.3 - Kasus influenza di RSUD Dr. Soetomo

Data kejadian penyakit influenza dari RSUD Dr. Soetomo pada penelitian ini memiliki kejanggalan. Rata-rata kasus influenza yang terjadi di RSUD Dr. Soetemo ada 5 kasus



Gambar 6.4 - Confidence interval 95% model ARIMA kasus demam berdarah

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 6.5.



Gambar 6.5 - Kasus influenza di Puskemas 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 6.5 merupakan rata-rata MAPE model ARIMAX yang ditemukan pada penelitian ini.

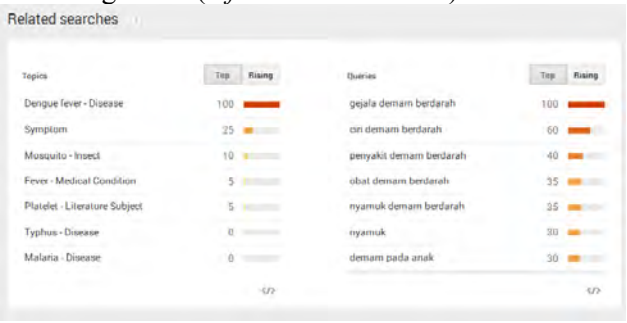
Tabel 6.5 – Perbandingan nilai MAPE model univariat dengan rata-rata MAPE model ARIMAX performa terbaik

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.

Penentuan *search query* pada pencarian Google Trends untuk menjadi variabel eksternal pada model peramalan sangat penting. Pada penelitian ini *search query* terkait penyakit demam berdarah dan influenza ditentukan berdasarkan pada *related search suggestion* yang muncul pada Google Trends. Contoh ditunjukkan pada Gambar 6.6. Walaupun hasil *related search* ini menunjukkan adanya korelasi yang signifikan terhadap jumlah kasus penyakit, pemilihan *query* menggunakan hasil *related search* memiliki kelemahan. Kelemahannya adalah hasil *related search* yang muncul ditentukan berdasarkan korelasi frekuensi *query* pencarian sebelumnya pada periode tertentu [40]. Sehingga *related search* yang muncul bisa berubah seiring waktu berjalan. Menjadikan metode ini kurang tepat dalam menentukan *search query* terkait penyakit demam berdarah dan influenza. Saran untuk penelitian ke depan dalam menentukan *query* pada Google Trends bisa dilakukan dengan menggunakan kuesioner yang menanyakan "Kata-kata apa yang anda pikirkan saat mencari informasi

meengenai demam berdarah/influenza?". Sungjin et.al(2013) menggunakan kuesioner dalam menentukan *search query* yang berkaitan dengan ILI (*influenza like-illness*).



Gambar 6.6 - Related search suggestion pada Google Trends

Pada penelitian-penelitian sebelumnya, hasil pencarian di Google Trends menunjukkan potensi untuk meramalkan kejadian kasus influenza [8] [16]. 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.

Halaman ini sengaja dikosongkan

LAMPIRAN

1. Data variabel dependen jumlah kasus penyakit demam berdarah perbulan

Bulan	Tahun					
	2010	2011	2012	2013	2014	2015
Januari	81	96	79	132	30	110
Februari	178	90	82	193	27	188
Maret	341	63	95	349	42	196
April	412	101	100	402	60	204
Mei	609	109	105	242	91	135
Juni	491	88	71	106	85	86
Juli	342	60	56	128	43	57
Agustus	145	26	54	68	37	29
September	60	15	50	30	36	
Oktober	93	22	62	37	29	
November	98	17	62	31	24	
Desember	94	30	58	20	47	

2. Data variabel dependen jumlah kasus penyakit influenza perbulan

Bulan	Tahun					
	2010	2011	2012	2013	2014	2015
Januari	1	2	2	11	5	5
Februari	1	4	4	2	2	3
Maret	1	5	8	9	8	1
April	1	2	8	6	10	1
Mei	1	3	8	6	9	5
Juni	1	3	9	11	2	6
Juli	2	4	8	12	1	1

Bulan	Tahun					
	2010	2011	2012	2013	2014	2015
Agustus	1	6	4	2	2	2
September	3	3	3	8	6	
Oktober	2	1	4	4	1	
November	1	2	7	3	1	
Desember	6	4	3	1	2	

3. Data variabel eksternal iklim temperature (SUHU), kelembaban udara (HMD), dan curah hujan (CH) perbulan

Periode	SUHU	HMD	CH
Jan-10	28,99	76,10	11,77
Feb-10	28,85	78,36	8,82
Mar-10	28,46	80,13	5,72
Apr-10	28,46	79,37	6,13
Mei-10	28,69	78,87	2,70
Jun-10	28,72	77,40	6,50
Jul-10	28,82	77,87	5,78
Agu-10	28,65	77,94	8,93
Sep-10	28,43	79,60	8,22
Okt-10	28,35	80,06	6,34
Nov-10	28,87	76,53	1,81
Des-10	27,86	80,81	10,47
Jan-11	27,51	81,87	6,94
Feb-11	27,89	79,96	5,81
Mar-11	28,13	79,97	16,47
Apr-11	28,32	80,27	9,82
Mei-11	28,57	77,84	2,26
Jun-11	27,74	71,37	0,99
Jul-11	27,71	70,58	0,00

Periode	SUHU	HMD	CH
Agu-11	27,59	66,42	0,00
Sep-11	28,49	65,23	0,00
Okt-11	30,00	64,35	0,81
Nov-11	29,24	75,40	7,60
Des-11	28,59	78,42	13,19
Jan-12	27,47	77,26	12,91
Feb-12	27,98	76,09	9,44
Mar-12	27,73	74,93	8,50
Apr-12	28,93	73,76	2,01
Mei-12	28,60	72,60	2,28
Jun-12	28,00	71,43	0,10
Jul-12	27,24	72,90	0,79
Agu-12	27,19	74,37	1,48
Sep-12	28,54	75,84	2,17
Okt-12	30,23	77,31	2,87
Nov-12	30,11	78,77	1,33
Des-12	28,66	80,24	6,66
Jan-13	27,96	81,71	23,17
Feb-13	28,25	80,57	8,22
Mar-13	28,42	80,77	14,84
Apr-13	28,72	80,43	5,64
Mei-13	28,67	80,58	6,41
Jun-13	28,29	81,50	8,69
Jul-13	27,43	76,13	3,23
Agu-13	27,60	68,90	0,10
Sep-13	28,55	64,27	0,00
Okt-13	30,18	62,77	0,19
Nov-13	29,21	72,53	3,49
Des-13	27,85	81,35	13,74

Periode	SUHU	HMD	CH
Jan-14	27,37	81,74	15,09
Feb-14	27,54	82,79	11,08
Mar-14	28,56	79,45	12,21
Apr-14	28,76	79,17	8,85
Mei-14	29,48	75,26	1,76
Jun-14	28,90	75,23	2,87
Jul-14	28,10	72,16	0,23
Agu-14	27,81	68,61	0,17
Sep-14	28,24	63,07	0,00
Okt-14	29,94	60,90	0,25
Nov-14	30,14	68,10	2,53
Des-14	28,30	80,32	13,47
Jan-15	28,00	79,77	22,88
Feb-15	27,82	80,83	7,40
Mar-15	28,36	80,07	14,07
Apr-15	28,16	82,10	9,35
Mei-15	28,31	76,90	6,85
Jun-15	27,93	73,06	0,00
Jul-15	27,11	73,25	0,00
Agu-15	27,61	70,19	0,04

4. Data variabel eksternal search index Google Trends untuk kasus demam berdarah per bulan

Periode	"demam berdarah"	"dbd"	"demam"	"dengue"
Jan-10	10,56	2,79	43,93	78,27
Feb-10	10,96	4,01	43,92	70,40
Mar-10	9,04	2,81	38,83	69,57
Apr-10	9,85	2,89	43,53	65,72

Periode	"demam berdarah"	"dbd"	"demam"	"dengue”
Mei-10	8,52	1,97	34,49	62,24
Jun-10	6,28	0,99	33,39	77,07
Jul-10	3,01	0,97	28,89	51,24
Agu-10	6,44	1,02	28,95	44,40
Sep-10	7,20	1,06	35,38	51,24
Okt-10	6,48	2,50	33,48	49,15
Nov-10	8,72	3,75	34,64	65,00
Des-10	9,16	3,97	37,74	47,48
Jan-11	14,24	5,18	33,80	74,15
Feb-11	14,67	7,20	49,50	66,68
Mar-11	12,01	4,91	49,25	65,88
Apr-11	13,01	4,92	10,75	62,21
Mei-11	11,17	3,27	9,20	58,90
Jun-11	8,19	1,61	36,78	72,93
Jul-11	3,90	1,55	31,81	48,47
Agu-11	8,30	1,61	31,85	41,99
Sep-11	9,23	1,64	24,50	48,45
Okt-11	8,27	3,81	37,00	46,45
Nov-11	11,07	5,62	35,25	61,42
Des-11	8,00	5,87	42,25	51,00
Jan-12	17,20	7,56	47,20	78,00
Feb-12	17,75	3,50	44,50	58,00
Mar-12	18,75	7,01	48,00	54,00
Apr-12	22,20	6,96	55,40	60,00
Mei-12	17,75	4,58	37,25	39,00
Jun-12	9,75	2,24	38,00	56,00
Jul-12	3,40	2,13	34,40	59,00
Agu-12	19,50	2,19	32,50	62,00
Sep-12	12,40	2,22	44,60	65,00

Periode	"demam berdarah"	"dbd"	"demam"	"dengue”
Okt-12	3,50	5,11	37,50	39,00
Nov-12	18,00	7,49	40,00	58,00
Des-12	24,80	7,78	41,60	49,00
Jan-13	32,75	9,95	65,25	100,00
Feb-13	25,00	13,60	69,25	63,00
Mar-13	23,00	8,80	65,20	72,00
Apr-13	25,00	9,50	71,00	58,00
Mei-13	18,00	6,25	57,00	58,00
Jun-13	16,00	2,00	63,20	78,00
Jul-13	12,50	2,25	51,25	44,00
Agu-13	11,75	2,93	53,25	24,00
Sep-13	12,20	3,60	56,20	53,00
Okt-13	15,00	9,25	53,00	41,00
Nov-13	14,25	9,75	52,25	58,00
Des-13	14,40	9,40	52,00	33,00
Jan-14	14,25	10,00	51,50	29,00
Feb-14	16,50	10,00	63,00	48,00
Mar-14	13,20	9,75	46,00	46,00
Apr-14	15,00	9,50	50,75	36,00
Mei-14	14,50	8,75	46,00	37,00
Jun-14	11,40	4,80	41,60	50,00
Jul-14	2,75	3,95	37,50	47,00
Agu-14	7,00	3,10	36,60	26,00
Sep-14	13,75	2,25	43,50	26,00
Okt-14	16,50	2,25	46,00	43,00
Nov-14	15,60	9,80	47,20	51,00
Des-14	16,75	11,00	50,00	34,00
Jan-15	33,25	16,75	75,25	54,00

Periode	"demam berdarah"	"dbd"	"demam"	"dengue”
Feb-15	45,50	28,00	89,50	64,00
Mar-15	26,40	16,20	66,00	59,00
Apr-15	22,25	13,00	59,00	63,00
Mei-15	21,60	4,00	61,80	72,00
Jun-15	18,25	4,75	53,00	72,00
Jul-15	10,00	4,18	40,00	20,00
Agu-15	11,20	3,60	48,20	33,00

5. Data variabel eksternal search index Google Trends untuk kasus influenza perbulan

Periode	"bersin"	"influenza"	"flu"	"batuk"	"pilek"	"batuk pilek"
Jan-10	65,4	45,7	40,9	18,5	31,3	24,2
Feb-10	62,4	43,6	47,2	24,8	40,3	32,9
Mar-10	54,9	49,2	44,5	25,3	41,6	39,6
Apr-10	63,8	47,0	47,3	23,0	42,1	38,8
Mei-10	63,7	36,1	43,2	20,6	35,0	34,2
Jun-10	61,2	32,3	39,8	22,7	33,0	27,2
Jul-10	55,2	38,7	41,0	20,7	46,1	28,6
Agu-10	47,6	54,2	38,2	23,6	48,7	31,6
Sep-10	56,7	83,3	41,3	22,9	45,4	28,4
Okt-10	77,9	73,0	39,6	22,5	37,7	27,9
Nov-10	76,6	58,0	34,6	20,5	38,3	24,5
Des-10	34,0	45,6	42,0	23,2	39,9	28,7
Jan-11	53,8	42,3	48,2	35,2	40,1	30,3
Feb-11	67,0	40,4	43,5	44,8	51,3	40,9
Mar-11	54,7	45,5	28,7	37,5	52,7	49,1
Apr-11	37,4	43,5	31,1	27,8	53,1	48,0
Mei-11	33,5	33,4	29,3	6,0	43,9	42,2
Jun-11	49,7	29,8	27,8	31,8	41,2	33,3
Jul-11	89,0	35,7	29,0	28,4	57,4	34,9
Agu-11	47,0	64,0	25,3	28,0	60,4	38,6
Sep-11	47,0	58,0	52,3	26,3	56,1	34,6
Okt-11	47,0	58,0	49,6	29,8	46,4	33,8
Nov-11	47,0	41,0	35,5	26,8	46,9	29,6
Des-11	48,0	43,0	51,8	30,3	48,7	34,5
Jan-12	49	37	59,8	38	48,84	36

Periode	"bersin"	"influenza"	"flu"	"batuk"	"pilek"	"batuk pilek"
Feb-12	64	28	80,5	49	63,75	46
Mar-12	46	28	82,5	56	67,75	64
Apr-12	81	31	86,4	52,2	59,80	61
Mei-12	52	31	68,25	44,25	14,50	46
Jun-12	60	31	51	39	14,50	40
Jul-12	63	36	51,4	31,8	44,00	40
Agu-12	60	42	45,75	39,5	38,27	40
Sep-12	61	100	55	39,4	32,37	40
Okt-12	62	55	46,5	37,75	26,88	40
Nov-12	65	47	48,5	35,5	29,36	43
Des-12	53	39	65,6	44,4	32,33	45
Jan-13	53	39	59,75	49,25	30	47
Feb-13	46	41	64,75	71,75	62	68
Mar-13	61	43	56,2	63,6	67,4	82
Apr-13	84	40	58,5	54,25	52,5	73
Mei-13	83	21	50,75	53	51	55
Jun-13	37	20	53	54,2	49	44
Jul-13	59	35	61	48	63,25	44
Agu-13	74	26	62,25	56,5	66,25	61
Sep-13	61	51	53,8	55,2	55,8	56
Okt-13	42	41	56,25	49,25	50,5	52
Nov-13	38	36	56,25	42,5	46,25	39
Des-13	57	26	52,8	46,2	48,4	43
Jan-14	62	22	70,25	51,5	52,5	42
Feb-14	60	16	71,75	57	56,75	65
Mar-14	75	25	87,2	61,6	50	55
Apr-14	61	23	88,25	64,25	62,25	70
Mei-14	50	23	83,75	67,25	58,5	70

Periode	"bersin"	"influenza"	"flu"	"batuk"	"pilek"	"batuk pilek"
Jun-14	76	19	58,4	57,8	59,4	51
Jul-14	75	19	48,5	44	53	46
Agu-14	59	19	65	58,6	63,8	45
Sep-14	64	37	58,75	55,5	62,25	48
Okt-14	76	53	59,75	55,25	49	49
Nov-14	79	46	49,8	50	52,2	45
Des-14	45	24	64,25	53,5	54,75	55
Jan-15	87	22	66,25	59,25	58,5	60
Feb-15	85	32	78	81,75	76,75	73
Mar-15	61	37	64,6	84,6	80,8	100
Apr-15	87	34	65	72,75	66,75	83
Mei-15	83	24	58	70	73,2	77
Jun-15	100	19	67,75	73	66,75	60
Jul-15	83	15	69,5	76	81,25	74
Agu-15	73	35	81,4	74,6	80,4	83

BAB VII
KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran terkait pengerjaan tugas akhir ini. Bagian kesimpulan berisi kesimpulan dari hasil pengerjaan tugas akhir ini. Sedangkan pada bagian saran, berisi saran dan masukan untuk perbaikan, tindak lanjut, atau pengembangan penelitian tugas akhir ini.

7.1 Kesimpulan

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.

7.2 Saran

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. Kramberck, 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," Kementrian Kesehatan Republik Indonesia, Jakarta, 2014.

[5] M. M. Sintorini, "Pengaruh Iklim terhadap Kasus Demam Bedarah 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] Kementrian Kesehatan, *Rencana Strategis Kementrian Kesehatan Tahun 2015-2019,* Jakarta: Kementrian 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/denguetrends. [Accessed 29 September 2015].

[15] Google, "Google Flu Trends," Google, 2015. [Online]. Available: http://www.google.org/flutrends. [Accessed 29 September 2015].

[16] O. M. Araz, D. Bentley and R. L. Muellman, "Using Google Flu Trends data in forecasting influenza-like-illness relatde 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.

BIODATA PENULIS

Penulis lahir di Bekasi, 28 Maret 1995. Penulis telah menempuh pendidikan formal di SD Putradarma Bekasi, SMP Hang Tuah 1 Surabaya, SMA Negeri 15 Surabaya, dan akhirnya penulis menempuh pendidikan tinggi di Jurusan Sistem Informasi pada tahun 2012.

Selama masa perkuliahan penulis aktif mengikuti organisasi luar kampus AIESEC Surabaya selama 2 tahun. Selain organisasi luar kampus, penulis juga aktif dalam kepanitiaan Information Systems Expo di tahun 2014. Di masa kuliah penulis berkesempatan untuk mengikuti *volunteering programme* di kota Hohhot, Provinsi Inner Mongolia, Cina selama 2 bulan. Selama masa kuliah penulis memiliki ketertarikan di bidang *IT Consulting* dan *open data*. Hal ini dibuktikan dengan penulis menjuarai kompetisi *IT Business Case* dan *data scraping* di tahun 2015. Di penghujung masa kuliah penulis berkesempatan untuk mengikuti *Global Korean Scholarship Science and Engineering Students 2016* di Chonbuk National University, Korea Selatan. Penulis bercita-cita untuk menjadi *expert* dalam bidang *data analytics*. Penulis dapat dihubungi melalui email : larasarist@gmail.com.

Halaman ini sengaja dikosongkan