

SKRIPSI

**METODE *IMPROVED HOLT WINTERS* PADA PRAKIRAAN KEJADIAN
DEMAM BERDARAH DI KABUPATEN SLEMAN**

***IMPROVED HOLT WINTERS METHOD ON FORECAST INCIDENT OF
DENGUE FEVER IN SLEMAN DISTRICT***



MOHAMMAD IRFAN RADEN

11/316986/PA/14104

**PROGRAM STUDI S1 ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2018

SKRIPSI

**METODE *IMPROVED HOLT WINTERS* PADA PRAKIRAAN KEJADIAN
DEMAM BERDARAH DI KABUPATEN SLEMAN**

***IMPROVED HOLT WINTERS METHOD ON FORECAST INCIDENT OF
DENGUE FEVER IN SLEMAN DISTRICT***

Diajukan untuk memenuhi salah satu syarat memperoleh derajat
Sarjana Komputer



MOHAMMAD IRFAN RADEN

11/316986/PA/14104

**PROGRAM STUDI S1 ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2018

HALAMAN PENGESAHAN

SKRIPSI

**METODE *IMPROVED HOLT WINTERS* PADA PRAKIRAAN
KEJADIAN DEMAM BERDARAH DI KABUPATEN SLEMAN**

Telah disusun dan dipersiapkan oleh:

MOHAMMAD IRFAN RADEN

11/316986/PA/14104

Telah dipertahankan di depan Tim Penguji

pada tanggal 5 Juli 2018

Susunan Tim Penguji

Agus Sihabuddin, S.Si., M.Kom., Dr
Pembimbing

Anny Kartika Sari, S.Si., M.Sc., Ph.D
Ketua Penguji

Isna Alfi Bustoni, S.T., M.Eng
Anggota Penguji

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang sepengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 25 Juni 2018

Mohammad Irfan Raden

HALAMAN MOTTO DAN PERSEMBAHAN

Katakanlah: "Apabila hamba-hambaKu bertanya kepadamu (Wahai Muhammad) tentang Diriku, maka jawablah, bahwa Aku ini dekat. Aku mengabulkan permohonan orang yang berdoa apabila ia memohon kepadaKu, maka hendaknya mereka itu memenuhi perintahKu dan hendaklah mereka yakin kepadaKu, agar mereka selalu berada dalam kebenaran".

(Al-Baqarah : 186)

"Do'a adalah modal yang dimiliki semua orang untuk menjadi apapun dan mendapatkan apapun, tanpa memandang jabatan, status, kekayaan bahkan bentuk fisik." (Yusuf Mansur)

Karya ini penulis persembahkan untuk orang tua, kakak-kakak, adik-adik, dan seluruh keluarga besar yang selalu memberikan harapan bagi penulis, juga segenap rekan-rekan serta pembaca sekalian.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala limpahan rahmat, karunia, serta petunjuk-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "*Metode Improve Holt Winters Pada Prakiraan Kejadian Demam Berdarah di Kabupaten Sleman*".

Dalam penyusunan skripsi ini penulis telah banyak mendapatkan arahan, bantuan, serta dukungan dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Kedua orang tua, Bapak Mohammad Raden Budiman dan Ibu Nur Asma yang senantiasa memberikan semangat, motivasi, dukungan, kasih sayang dan doa sehingga penulis dapat menyelesaikan penulisan skripsi.
2. Prof. Dr. Triyono selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada.
3. Dr. Suprpto, M.Kom selaku Ketua Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada.
4. Bapak Sigit Priyanta, S.Si., M.Kom., Dr selaku Dosen Pembimbing Akademik atas bantuan, bimbingan, dan motivasi yang diberikan kepada penulis.
5. Bapak Agus Sihabuddin, S.Si., M.Kom., Dr selaku Dosen Pembimbing yang telah banyak memberikan arahan, motivasi dan waktu luangnya dalam penulisan skripsi dari awal hingga selesai.
6. Seluruh teman-teman Ilmu Komputer 2011 yang telah memberikan banyak kenangan dan dukungan selama menjalani kuliah di Universitas Gadjah Mada.
7. Segenap Dosen dan civitas akademik di lingkungan Program Studi Ilmu Komputer, Departemen Ilmu Komputer dan Elektronika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada.
8. Segenap keluarga besar Asrama Mahasiswa Aceh Meurapi Duwa yang selalu mendukung dalam menyelesaikan tugas skripsi.
9. Seluruh pihak yang telah membantu penulis dalam penyelesaian skripsi yang tidak dapat disebutkan satu persatu.

Akhir kata penulis berharap semoga skripsi ini dapat memberikan manfaat bagi kita semua, terutama bagi perkembangan ilmu pengetahuan serta perkembangan Ilmu Komputer dan Teknologi Informasi.

Yogyakarta, 07 Juni 2018

Penulis

DAFTAR ISI

SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN.....	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.1 Rumusan Masalah	3
1.2 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metode Penelitian.....	4
1.7 Sistematika Penulisan.....	5
BAB II INJAUAN PUSTAKA	8
BAB III DASAR TEORI	11
3.1 Prakiraan.....	11
3.2 Exponential Smoothing	12
3.3 <i>Additive Holt Winters</i>	13
3.4 <i>Multiplicative Holt Winters</i>	14
3.5 Metode <i>Improved Holt Winters</i>	14
3.6 Evaluasi Prakiraan	15
3.6.1 Mean Absolute Percentage Error	16
3.6.2 Mean Square Error	16
3.6.3 Root Mean Square Error	16
3.7 Data Stasioner.....	16
3.8 Autocorrelation Function (ACF)	17

3.9	Partial Autocorrelation Function (PACF)	18
3.10	Bobot nilai awal.....	18
3.11	Demam Berdarah <i>Dengue</i> (DBD)	19
3.12	Twitter	22
3.13	Deret berkala.....	22
BAB IV ANALISIS DAN PERANCANGAN SISTEM		24
4.1	Deskripsi Umum Metode Penelitian	24
4.2	Spesifikasi Data	24
4.2.1	Data Input	24
4.2.2	Data output.....	27
4.3	Proses Prakiraan	27
4.4	Analisis Pembuatan Sistem	28
4.5	Alur Pembuatan Sistem Dengan Metode Improved Holt Winters	28
4.6	Scraping dan preprocessing	30
4.7	Stasionerisasi data	32
BAB V IMPLEMENTASI.....		33
5.1	Deskripsi Implementasi	33
5.2	Implementasi Scraping Tweet Pada Twitter	33
5.3	Pemilihan Data Input.....	37
5.4	Implementasi pembuatan model.....	42
5.5	Implementasi stasionerisasi data	42
5.6	Implementasi menentukan nilai awal	43
5.7	Implementasi model <i>Improved Holt Winters</i>	43
BAB VI HASIL DAN PEMBAHASAN		46
6.1	Pengujian korelasi	46
6.2	Pengujian stasioneritas data.....	47
6.3	Pembagian Data <i>training</i> dan Data <i>Testing</i>	51
6.4	Penentuan nilai awal.....	52
6.5	Pengujian prakiraan demam berdarah	53
6.6	Pengujian prakiraan demam berdarah pada data jumlah tweet	54
6.7	Pengujian prakiraan demam berdarah pada data jumlah kejadian demam berdarah	55
BAB VII KESIMPULAN DAN SARAN		58
7.1	Kesimpulan.....	58

7.2 Saran	58
DAFTAR PUSTAKA	59

DAFTAR GAMBAR

Gambar 3. 1 Grafik jumlah penderita demam berdarah di Indonesia	20
Gambar 3. 2 Grafik Data penderita DBD tingkat Provinsi	21
Gambar 4. 1 Diagram alir proses prakiraan <i>Improved Holt Winter</i>	28
Gambar 4. 2 Diagram alir pembuaan sistem.....	29
Gambar 4. 3 Ttweet demam berdarah pada twitter	30
Gambar 4. 4 Proses scraping data tweet pada twitter.....	31
Gambar 5. 1 Inisialisasi library yang digunakan pada proses scraping	34
Gambar 5. 2 constructor untuk pencarian pada Url twitter	34
Gambar 5. 3 Constructor untuk ekstrak data.....	34
Gambar 5. 4 Constructor untuk ekstrak meta data	36
Gambar 5. 5 kode crawling pada terminal anaconda dengan fitur place	36
Gambar 5. 6 Crawling pada terminal dengan menggunakan nama kabupaten Sleman	36
Gambar 5. 7 Grafik Data asli jumlah kejadian demam berdarah, jumlah kejadian demam berdarah berdasarkan twitter, dan jumlah tweet mengenai demam berdarah	39
Gambar 5. 8 Constructor menggunakan matplotlib dan statsmodel	40
Gambar 5. 9 Kode membaca data demam berdarah	40
Gambar 5. 10 Grafik musiman dan <i>trend</i> pada jumlah data kejadian demam berdarah	40
Gambar 5.11 Gambar 5. 11 Kode differencing data non stasioner	42
Gambar 5. 12 kode program memasukkan data.....	43
Gambar 5. 13 kode program untuk membagi data <i>training</i> dan data <i>testing</i>	43
Gambar 5. 14 Kode program nilai awal α , β , γ	44
Gambar 5. 15 kode program algoritma model <i>improved holt winter</i> pada data <i>training</i>	44
Gambar 5. 16 kode program algoritma model <i>improved Holt Winters</i> pada data <i>testing</i>	44
Gambar 5. 17 kode program algoritma nilai mape, mse, rmse	45
Gambar 5. 18 kode program untuk membuat plot data prakiraan data <i>training</i> , data <i>testing</i> , dan hasil error (MAPE, MSE, RMSE).....	45

Gambar 6. 1 Tampilan grafik jumlah data tweet tentang demam berdarah	47
Gambar 6. 2 Tampilan grafik data jumlah kejadian demam berdarah	47
Gambar 6. 3 Tampilan correlogram data jumlah kejadian demam berdarah	48
Gambar 6. 4 Grafik correlogram data jumlah tweet tentang demam berdarah	49
Gambar 6. 5 Tampilan correlogram ACF differencing orde satu jumlah tweet mengenai demam berdarah	50
Gambar 6. 6 Tampilan correlogram ACF differencing orde satu jumlah kejadian demam berdarah	50
Gambar 6. 7 Grafik Jumlah kejadian demam berdarah di Kabupaten Sleman	57

DAFTAR TABEL

Tabel 2.1 Tinjauan daftar pustaka	9
Tabel 3.1 Data penderita DBD tingkat Provinsi	21
Tabel 4.1 Data jumlah kejadian demam berdarah kabupaten sleman	25
Tabel 4.2 Data twitter (jumlah kejadian demam berdarah, jumlah tweet)	26
Tabel 5.1 data asli jumlah kejadian demam berdarah, jumlah kejadian dbd berdasarkan tweet dan jumlah tweet mengenai demam berdarah	37
Tabel 6.1 Data training Jumlah tweet demam berdarah	51
Tabel 6.2 data testing jumlah tweet demam berdarah	52
Tabel 6.3 Trial dan error nilai α , β , γ pada data tweet	53

INTISARI

METODE *IMPROVED HOLT WINTERS* PADA PRAKIRAN KEJADIAN DEMAM BERDARAH DI KABUPATEN SLEMAN

Mohammad Irfan Raden

11/316986/PA/14104

Penyakit berbahaya demam berdarah terjadi setiap tahun di wilayah Kabupaten Sleman. Untuk membantu kesiapan masyarakat dalam menghadapi jumlah penderita demam berdarah dibutuhkan suatu prakiraan. Penelitian ini bertujuan untuk membuat prakiraan jumlah kejadian demam berdarah di Kabupaten Sleman dengan menggunakan data jumlah kejadian demam berdarah dan jumlah tweet mengenai demam berdarah. Proses prakiraan dengan menggunakan metode *Improve Holt Winters* yang merupakan pengembangan dari metode *Additive Holt Winters*. Hasilnya, data jumlah kejadian demam berdarah memiliki korelasi dengan jumlah tweet mengenai demam berdarah dengan nilai korelasi 0,335 lebih besar dari nilai kritis 0,284 dengan tingkat signifikansi 95%. Prakiraan dengan menggunakan metode *Improve Holt Winters* pada data jumlah kejadian demam berdarah menghasilkan nilai MAPE sebesar 5,94%, MSE sebesar 715,63 dan RMSE sebesar 26,75 dengan nilai awal α , β , γ terbaik (0,95; 0,05; 0,01). Dan pada data jumlah tweet menghasilkan nilai MAPE sebesar 4,01%, MSE 5,28 dan RMSE sebesar 2,3 dengan nilai awal α , β , γ terbaik (0,9; 0,02; 0,01).

Kata kunci: prakiraan, twitter, *improved Holt Winters*, demam berdarah, Holt Winters.

ABSTRACT

IMPROVED HOLT WINTERS METHOD ON FORECAST INCIDENT OF DENGUE FEVER IN SLEMAN DISTRICT

Mohammad Irfan Raden

11/316986/PA/14104

Danger of dengue fever occurs every year in the region of Sleman Regency. To assist readiness community in handling the number of DHF patients requires forecasts. This research aims to predict the number of dengue incidence in Sleman District by using data on the number of dengue fever and number of tweets about dengue fever. Forecasting process using Improve Holt Winters method which is the development of Additive Holt Winters method. As a result, data on the number of dengue fever cases has a correlation with the number of tweets about dengue fever with a correlation value of 0.335 greater than the critical value of 0.284 with a 95% significance level. Forecast by using Improve Holt Winters method on data of dengue fever number resulted MAPE value of 5.94%, MSE of 715.63 and RMSE of 26.75 with best initial value of α , β , γ (0.95, 0.05 ; 0.01). And on tweet data yields MAPE value of 4.01%, MSE 5,28 and RMSE 2,3 with best initial value α , β , γ (0,9; 0,02; 0,01)

Keyword: *forecast, improved Holt Winters, Holt Winters additive, twitter, dengue fever*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Demam berdarah (DBD) termasuk salah satu penyakit yang berbahaya di dunia yang dapat mengakibatkan kematian pada korbannya. Demam berdarah adalah virus penyakit yang ditularkan oleh nyamuk *Aedes Aegypti*, nyamuk yang paling cepat berkembang di dunia dan telah menyebabkan hampir 390 juta orang terinfeksi setiap tahunnya. WHO memperkirakan jumlah penderita demam berdarah berpotensi mencapai hingga 40% dari seluruh penduduk dunia. Berdasarkan pada data tahun 2004 – 2010, WHO melaporkan bahwa 75% penderita demam berdarah berasal dari Asia Pasifik (WHO, 2013). Indonesia tercatat sebagai negara endemis demam berdarah terbesar kedua diantara 30 negara endemis lainnya. Pada tahun 2014, jumlah penderita DBD di Indonesia sebanyak 100.347 kasus dan pada tahun 2015 mengalami peningkatan menjadi 129.650 kasus dengan jumlah kematian sebanyak 1.071 orang (Menkes, 2015).

Daerah Istimewa Yogyakarta (DIY) merupakan salah satu Provinsi di Indonesia yang dinyatakan sebagai wilayah endemis demam berdarah. Pada tahun 2011, jumlah penderita demam berdarah adalah sebanyak 982 kasus, lima diantaranya meninggal dunia. Pada tahun 2012, jumlah penderita demam berdarah adalah sebanyak 970 kasus. (Dinkes, DIY 2012). Provinsi Daerah Istimewa Yogyakarta berada pada peringkat keempat terbesar wilayah penderita DBD dengan angka 92,96 per 100.000 penduduk. Jumlah penderita DBD cenderung mengalami kenaikan pada setiap tahunnya (Menkes, 2015).

Saat ini masyarakat Indonesia telah sangat populer dengan penggunaan teknologi media sosial. Media sosial adalah salah satu aplikasi teknologi yang dapat melingkupi struktur sosial yang terdiri dari elemen individu dan elemen kelompok atau organisasi (Heaps, 2009). Media sosial dapat digunakan sebagai interaksi antar pengguna yang dapat saling berhubungan karena kesamaan sosialitas, dalam menuliskan fakta dan opini dengan jangkauan yang luas. Berdasarkan tingkat kepopuleran dalam penggunaan media sosial, maka salah satu aplikasi media sosial yang paling banyak digunakan oleh masyarakat Indonesia hingga saat ini adalah

media sosial twitter. Menurut hasil riset yang dilakukan memos pada desember 2016, Indonesia termasuk salah satu negara pengguna twitter terbanyak di dunia dengan jumlah 29 juta pengguna dan telah memposting 2,4% dari 10,6 miliar tweet diseluruh dunia.

Twitter merupakan media yang memungkinkan untuk melihat suatu data dan informasi secara mendalam dan real time (Eysenbach, 2009). Pada penelitian sebelumnya telah dilakukan analisis data mengenai penyebaran sebuah isu kesehatan menggunakan twitter (Chew, 2010). Dalam penelitian tersebut dilakukan analisis data untuk melihat atensi public, kepekaan perasaan, dan pengetahuan public. Pada penelitian lainnya telah dilakukan prakiraan jumlah kejadian *Influenza Like Illness* (ILI) dengan penggunaan data Twitter untuk meningkatkan keakuratan model dalam melakukan prakiraan.

Data jumlah kasus penderita demam berdarah disajikan dalam waktu harian, mingguan, bulanan, dan tahunan. Sehingga pada data tersebut dapat diurutkan berdasarkan urutan waktu kejadian deret berkala (*time series*). Data deret berkala berfungsi untuk menggambarkan suatu perkembangan data dari waktu ke waktu. Data deret berkala juga sering digunakan untuk melakukan proses prakiraan. Prakiraan didefinisikan sebagai suatu rangkaian proses kegiatan yang dilakukan untuk memperkirakan kejadian yang akan terjadi pada waktu berikutnya dengan mengamati dan menganalisa urutan waktu yang terjadi pada waktu sebelumnya.

Dalam melakukan suatu prakiraan deret berkala terdapat berbagai macam metode diantaranya metode Kalman Filter, ARIMA, Jenkins, dan *Exponential Smoothing* (Arsyad, 1994). Metode *exponential smoothing* adalah metode penghalusan eksponensial bersifat sederhana dan mudah dipahami khususnya bagi prakiraan pendek dari data deret waktu yang Panjang (Makridakis, 1999). Metode *exponential smoothing* dibagi lagi menjadi beberapa metode yaitu *Single Exponential Smoothing*, *holt*, *Holt Winters* dan lainnya. Perbedaan metode tersebut terletak pada penggunaan parameternya dimana *Single Exponential Smoothing* hanya menggunakan satu parameter, metode *holt* menggunakan dua parameter, dan metode *Holt Winters* menggunakan tiga parameter. Dalam proses prakiraan dengan menggunakan metode *Holt Winters* juga terdapat dua model yang sering digunakan

yaitu model *additive* dan model *multiplicative*. Kedua model ini juga masih dikembangkan guna mendapatkan nilai akurasi yang lebih baik.

Oleh karena itu dalam penelitian ini peneliti ingin melakukan implementasi prakiraan dengan menggunakan model *Improved Holt Winters* (IHW) pada data demam berdarah. Model IHW adalah pengembangan dari model *Additive Holt Winters* (AHW) dengan perbedaan terletak pada pemulusan keseluruhan (level). Pada metode IHW pemulusan nilai awal α hanya terjadi pada pengamatan nilai Y_t tidak pada faktor *seasonal* guna mengurangi fluktuasi faktor *seasonal* pada data (Tratar, 2014). Peneliti memilih untuk menggunakan metode *Improve Holt Winters* karena metode tersebut bisa digunakan pada data demam berdarah yang diestimasi membentuk pola data *trend* dan musiman.

Prakiraan dengan menggunakan metode *Improve Holt Winters* diharapkan akan memberikan nilai akurasi yang baik dalam mengatasi pola data dengan parameter level, *trend*, dan musiman. Dengan metode IHW diharapkan hasil akurasi dari prakiraan demam berdarah dapat mendekati nilai aslinya sehingga bisa digunakan sebagai salah satu upaya untuk mengetahui jumlah kasus demam berdarah pada waktu yang akan datang juga dapat membantu tindakan preventif dalam upaya pencegahan atau tindakan kuratif dalam upaya pengobatan sehingga kesiapan dalam penanganan kasus demam berdarah dapat lebih meningkat.

1.1 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan masalah dalam penelitian ini adalah “bagaimana hasil akurasi dari implementasi metode *Improved Holt Winters* pada prakiraan jumlah kejadian demam berdarah di Kabupaten Sleman”.

1.2 Batasan Masalah

Batasan permasalahan pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan yaitu data jumlah kejadian demam berdarah tahun 2010-2014 di wilayah kabupaten Sleman. Data diambil dari Dinas Kesehatan (Dinkes) DIY.

2. Data fasilitas kesehatan masyarakat dan nama lokasi di DIY diambil dari Sistem Informasi Tenaga dan Sarana Kesehatan pada alamat url <https://dinkes.slemankab.go.id/alamat-rumah-sakit-di-kab-sleman-2> digunakan untuk penyaringan konten tweet berdasarkan lokasi dalam pengumpulan data mengenai jumlah kejadian demam berdarah pada twitter.
3. Prakiraan yang dibuat pada penelitian ini menggunakan metode *Improved Holt Winters*.

1.4 Tujuan Penelitian

Tujuan dilakukan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode Improve Holt Winters pada prakiraan kejadian demam berdarah di Kabupaten sleman.
2. Mencari korelasi antara data jumlah kejadian demam berdarah dengan data jumlah tweet dan jumlah kejadian demam berdarah berdasarkan twitter.
3. Mengetahui nilai akurasi metode prakiraan Improve Holt Winters dengan pada data tweet demam berdarah dan jumlah kejadian demam berdarah.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah memberikan pengetahuan kepada akademisi mengenai prakiraan dengan menggunakan model *Improved Holt Winters* terhadap data penyakit demam berdarah. Dengan diketahuinya nilai akurasi tersebut maka akan membantu masyarakat, pemerintah daerah, dan Instansi lainnya dalam mempersiapkan tindakan pencegahan serta pengobatan menjadi lebih baik.

1.6 Metode Penelitian

Tahapan-tahapan dan metode yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Studi Pustaka

Studi pustaka merupakan tahapan metode penelitian paling awal yang dilakukan untuk mempelajari literature yang mendukung dalam penelitian. Literatur yang dipelajari berkaitan dengan demam berdarah, data *time series* pola musiman, analisis prakiraan, *Holt Winters* model,

analisis data twitter, dan bahasa pemrograman python. Sumber yang digunakan sebagai literature antara lain jurnal ilmiah, buku cetak, buku elektronik, dan laporan hasil penelitian, skripsi dan disertasi.

2. Pengumpulan data

Pengumpulan data dilakukan dengan dua tahap yaitu tahap pertama pengumpulan data dari Menteri Kesehatan berupa data jumlah kejadian demam berdarah dari tahun 2010 sampai tahun 2014 dan tahap kedua berupa data hasil proses scrapping preprocessing twitter.

3. Analisis kebutuhan model dan sistem

Analisis kebutuhan model dan sistem merupakan tahapan untuk menganalisa kebutuhan serta mendeskripsikan dari model dan sistem yang akan dibangun dalam penelitian.

4. Perancangan model dan sistem

Perancangan model dan sistem adalah tahapan dalam pembuatan rancangan yang menggambarkan alur kerja dengan menggunakan diagram blok dan flowchart.

5. Implementasi

Implementasi adalah suatu proses yang dilakukan untuk membangun model dan sistem berdasarkan alur kerja rancangan. Tahapan ini berupa penulisan kode bahasa pemrograman untuk pengambilan data twitter, pemodelan prakiraan, analisis data, dan perhitungan nilai akurasi.

6. Pengujian dan evaluasi

Pengujian model dilakukan setelah proses implementasi selesai. Hasil output dari implementasi tersebut kemudian dilakukan evaluasi dengan menganalisis nilai akurasi dan nilai erornya dengan jumlah data yang sebenarnya.

1.7 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini merupakan gambaran secara umum pada setiap bab yang dituliskan sesuai dengan tahapan penulisan skripsi, yaitu sebagai berikut:

BAB I : PENDAHULUAN

Bab ini menguraikan secara singkat hal-hal mengenai latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian metode penelitian dan sistematika penulisan skripsi yang dilakukan oleh peneliti.

BAB II : STUDI PUSTAKA

Bab ini menguraikan tentang studi literatur yang digunakan oleh penulis dalam keterkaitannya pada penelitian dengan melihat, mempelajari dan membandingkan dari penelitian yang pernah dilakukan oleh peneliti sebelumnya.

BAB III : LANDASAN TEORI

Bab ini menerangkan dasar – dasar teori yang digunakan dalam penelitian sehingga penelitian mengacu pada teori yang benar dalam ilmu pengetahuan serta menjadi dasar dalam memecahkan masalah pada penelitian ini.

BAB IV : ANALISIS DAN PERANCANGAN SISTEM

Bab ini menerangkan tentang analisis kebutuhan sistem pembuatan model *Improved Holt Winter*, gambaran data secara spesifik, serta perancangan sistem yang akan dibangun dalam penelitian sesuai hasil yang ingin dicapai pada penelitian. Sehingga dalam implementasi dapat berjalan dengan baik dan benar.

BAB V : IMPLEMENTASI

Bab ini menguraikan tentang proses dan pelaksanaan yang dilakukan selama penelitian, menampilkan data-data input, kode bahasa pemrograman maupun output data hasil.

BAB VI : PEMBAHASAN

Bab ini menerangkan pembahasan dari hasil implementasi. Nilai output prakiraan yang diperoleh dapat dibandingkan dengan data asli sehingga dapat diketahui nilai akurasi.

BAB VII : KESIMPULAN DAN SARAN

Bab ini merupakan kesimpulan yang ditarik dari hasil pembahasan penelitian. Selanjutnya kekurangan yang ada pada penelitian akan dituliskan pada saran sehingga dapat membantu pada penelitian berikutnya untuk pengembangan dan optimasinya.

BAB II

TINJAUAN PUSTAKA

Brian dkk (2013) melakukan penelitian mengenai analisis perbandingan *exponential smoothing*, *holt-winters*, dan *polynomial newton* dalam meramalkan data produksi ubi kayu di Provinsi Jawa Tengah. Hasil yang didapatkan dari penelitian tersebut adalah nilai akurasi dengan menggunakan metode *Holt Winters* menjadi yang paling baik daripada menggunakan metode *exponential smoothing* dan *polynomial newton*. Dapat disimpulkan bahwa pada penelitian prakiraan produksi ubi kayu lebih tepat apabila menggunakan metode *Holt Winters*.

Deossa dkk (2014) melakukan suatu penelitian untuk meramalkan kebutuhan energi pada pembangkit listrik. Tujuan penelitian ini adalah untuk menyesuaikan kapasitas energi dan menghemat biaya produksi. Penelitian prakiraan ini dilakukan dengan menggunakan metode *Additive Holt Winters* (AHW). Kemudian Rossi dkk (2015) melanjutkan prakiraan metode AHW pada penggunaan *resource* sebuah data *center* dengan tujuan untuk menghemat penggunaan energi. Hasil dari penelitian keduanya yaitu nilai perhitungan prakiraan mendekati dengan kondisi nyata.

Tratar, L.F (2014) dalam penelitiannya melakukan prakiraan dengan mengembangkan metode *Improved Holt Winters* (IHW) pada studi kasus kunjungan turis yang menginap di Negara Republik Slovenia. Data yang digunakan berasal dari data kantor statistic di Republik Slovenia yaitu berupa data kunjungan turis pada tahun 2000-2009. Penelitian tersebut melakukan perbandingan metode AHW, MHW dan IHW dengan tujuan untuk melihat nilai akurasi yang paling baik. Pada hasil perhitungan penelitiannya menghasilkan nilai MSE (*Mean Square Error*) AHW (770,745), MHW (547,267), IHW (535,270). Nilai MSE IHW adalah yang paling kecil diantara metode MHW dan AHW sehingga dapat disimpulkan bahwa metode IHW memiliki nilai akurasi paling tinggi pada kasus jumlah pengunjung turis di Republik Slovenia.

Unal dkk (2015) melakukan penelitian mengenai prakiraan turis dengan memanfaatkan fitur makro pada Microsoft excel. Penelitian prakiraan ini menggunakan metode *Holt Winters* dengan membandingkan metode *Additive Holt*

Winters (AHW) dan *Multiplicative Holt Winters* (MHW). Hasil dari penelitian tersebut adalah AHW memberikan hasil prakiraan yang lebih baik nilai akurasinya dibandingkan dengan metode MHW.

Hutomo (2016) membuat model prakiraan kebutuhan persediaan benih menggunakan metode *holt - winter exponential smoothing*. Pada penelitiannya menghasilkan nilai MAPE sebesar 7% untuk prakiraan produk P21 dan 9% untuk P27. Pada rantai pasokan produk benih P21 mengalami penurunan nilai *bullwhip* dari 1,21655 menjadi 1,04754 dan menghilangkan *bullwhip effect* pada rantai pasokan. Pada rantai pasokan benih P27 mengalami penurunan dari 1,15140 menjadi 1,04632.

Akbar (2018) melakukan penelitian mengenai prakiraan kejadian demam berdarah di Daerah Istimewa Yogyakarta berdasarkan data twitter dengan menggunakan metode MARIMA Fungsi Transfer. Pada penelitiannya menggunakan data jumlah kejadian demam berdarah dari Menteri Kesehatan Indonesia ditambah dengan data twitter. Hasil dari penelitiannya adalah jumlah kejadian demam berdarah memiliki korelasi dengan jumlah tweet pada twitter dengan nilai korelasi 0,284 lebih besar dari critical value dengan 95% tingkat kepercayaan. Hasil akurasi yang dipatkan dari menggunakan metode MARIMA masih lebih rendah daripada dengan menggunakan metode *Ekspponential Smoothing* (ES).

Tabel 2 1 Tinjauan daftar pustaka

No	Penulis	Metode	Tujuan	Hasil / Keterangan
1	Brian dkk (2013)	ES – HW - PN	Perbandingan metode prakiraan produksi ubi kayu jawa tengah	HW memberikan hasil akurasi yang lebih baik dari ES dan PN
2	Deossa dkk (2014)	<i>Holt Winters Additive</i>	Meramalkan kebutuhan energi pada pembangkit listrik	Pendekatan dengan metode AHW mendekati pada kondisi nyata kebutuhan energi pembangkit listrik
3	Rossi dkk (2015)	<i>Holt Winters Additive</i>	Meramalkan penghematan	Prakiraan mendekati pada kondisi nyata pada

			pembagkit listrik pada data center resource	penghematan energi listrik
4	Tratar (2014)	AHW, MHW, dan IHW	Mengembangkan metode IHW pada kunjungan turis di Republik Slovenia	IHW memiliki nilai MSE (<i>Mean Square Error</i>) paling kecil daripada AHW dan MHW
5	Unal dkk (2015)	Perbandingan <i>Holt Winters Additive</i> – <i>Holt Winters Multiplicative</i>	Perbandingan prakiraan kunjungan turis dengan memanfaatkan fitur makro pada Microsoft excel	Prakiraan menggunakan metode HW <i>Additive</i> lebih baik daripada HW <i>Multiplicative</i> dalam prakiraan kunjungan turis.
6	Hutomo (2016)	<i>Holt Winters</i> , Esponential smoothing	Mengembangkan model prakiraan kebutuhan persediaan benih dan mengurangi <i>bullwhip effect</i> pada rantai pasok	Kebutuhan persediaan sebagai patokan dalam melakukan pemesanan dapat mengurangi nilai <i>bullwhip effect</i> .
7	Akbar (2017)	MARIMA fungsi Transfer	Prakiraan pada kejadian demam berdarah di Daerah Istimewa Yogyakarta	Hasil akurasi prakiraan masih lebih rendah dari hasil prakiraan menggunakan metode eksponensial smoothing.

Pada tabel 2.1 menjelaskan secara singkat daftar tinjauan pustaka yang digunakan dalam penelitian pada rentang tahun 2013-2017 yang berkaitan dengan metode *Holt Winters*. Selanjutnya penulis akan melakukan penelitian mengenai prakiraan kejadian demam berdarah di Kabupaten Sleman dengan menggunakan metode *Improved Holt Winters* dimana metode tersebut melakukan perubahan pada persamaan *level* metode *Additive Holt Winters* dengan perhitungan nilai alpha hanya terjadi pada data aktual saja dan tidak terjadi pada faktor musiman guna untuk mengurangi fluktuatif pola data musiman.

BAB III

DASAR TEORI

3.1 Prakiraan

Prakiraan atau *forecasting* adalah serangkaian kegiatan untuk melakukan perkiraan sebelum peristiwa terjadi berdasarkan hasil pengamatan, analisis data, dan perhitungan data masa lalu (Sumayang, 2003). Prakiraan penting dalam berbagai bidang kehidupan baik kepada individu, perusahaan ataupun pemerintahan. Pada kegiatan individu prakiraan dapat digunakan untuk memperkirakan tingkat keberhasilan dalam pekerjaan, investasi dan lainnya. Sedangkan dalam kegiatan perusahaan prakiraan digunakan untuk memprediksi produksi, penjualan, pemasaran, dan lainnya. Pada pemerintahan prakiraan digunakan dalam bidang ekonomi, pembangunan infrastruktur kota, pendidikan, keamanan, kesehatan, penanggulangan bencana dan lainnya. Prakiraan merupakan dasar dari segala macam perencanaan untuk menghubungkan lingkungan yang tidak stabil (Makridakis, 1993). Prakiraan memiliki tujuan untuk mengkaji suatu kebijakan yang berlaku saat ini dan masa lampau untuk melihat dan menentukan sejauh mana kebijakan itu dapat berpengaruh. Dengan adanya prakiraan dapat meningkatkan efektivitas suatu perencanaan (Render dan Heizer, 2009).

Prakiraan yang baik memiliki beberapa kriteria yang penting antara lain tingginya nilai akurasi dan rendahnya nilai kesalahan. Akurasi dari suatu prakiraan dapat diukur pada tingkat kebisaan dan kekonsistensian dengan membandingkan nilai prakiraan hingga mendekati dengan nilai kenyataan. Nilai akurasi dan nilai kesalahan bergantung pada metode yang digunakan dan lamanya periode berlangsung. Berdasarkan lamanya periode, prakiraan dapat dikelompokkan dalam tiga bagian, yaitu prakiraan jangka pendek, jangka menengah dan jangka Panjang. Prakiraan jangka pendek yaitu untuk jangka waktu kurang dari 3 bulan. Prakiraan jangka menengah yaitu dalam waktu 3 hingga 18 bulan. Prakiraan jangka Panjang untuk jangka waktu lebih besar dari 18 bulan. (Herjanto, 2008)

Metode prakiraan dibagi menjadi dua kategori utama, yaitu metode kualitatif dan metode kuantitatif. Metode kualitatif adalah metode yang dilakukan apabila tidak tersedia data pada masa yang lalu. Metode kualitatif dilakukan dengan

mempertimbangkan pendapat-pendapat dari para ahli untuk dijadikan prakiraan. Metode kualitatif dibagi lagi menjadi dua, yaitu metode eksploratoris dan normatif. Metode kuantitatif yaitu metode yang dilakukan dengan menggunakan data pada masa yang lalu untuk dijadikan prakiraan pada waktu yang akan datang. metode kualitatif dibagi menjadi dua bagian yaitu metode deret berkala (*time series*) dan metode kausal. Prakiraan dengan metode kuantitatif dapat dilakukan apabila tersedia informasi tentang masa lalu, dapat dikuantitatifkan dalam bentuk data numerik, dan dapat diasumsikan beberapa aspek pola masa lalu akan terjadi terus berlanjut di waktu yang akan datang (Makridakis dkk., 1992).

3.2 Exponential Smoothing

Metode *exponential smoothing* atau penghalusan eksponensial adalah metode prakiraan yang dilakukan dengan menggunakan pem bobotan terhadap data masa lalu dengan cara eksponensial sehingga bobot pada data paling akhir menjadi lebih besar dalam rata-rata bergerak. Metode penghalusan eksponensial bersifat sederhana dan mudah dipahami khususnya bagi prakiraan pendek dari data *time series* yang panjang. Metode pemulusan eksponensial merupakan teknik prakiraan yang ekonomis dan tidak kompleks dengan hasil ramalan yang cukup baik dalam variasi aplikasi yang luas. Ada beberapa metode yang dapat dilakukan dalam metode penghalusan eksponensial, yaitu diantaranya:

1. Metode *Single Exponential Smoothing*, adalah metode pemulusan yang paling sederhana dimana hanya terdapat satu parameter yang perlu diestimasi.
2. Metode *Holt*, adalah metode pemulusan yang menggunakan dua parameter berbeda untuk mengakomodasi prakiraan pada data *trend*.
3. Metode Holt-Winters adalah metode pemulusan yang menggunakan tiga parameter diantaranya; parameter konstanta pemulusan, parameter untuk trend, dan parameter untuk musiman. Metode Holt-Winters dapat digunakan untuk data nonstasioner. Dalam metode *holt-winters* terdapat dua tipe metode pemulusan, yaitu metode *Additive Holt Winters* dan *Multiplicative Holt Winters* (Kalekar, 2004). Penentuan penggunaan model bergantung pada plot data yang akan diramalkan.

3.3 Additive Holt Winters

Model *Additive Holt Winters* (AHW) adalah model eksponensial dengan merepresentasikan *time series* sebagai berikut:

$$y_t = b_1 + b_2t + S_t + E_t \quad (3.1)$$

dimana:

b_1 = komponen permanen

b_2 = komponen kecenderungan

S_t = faktor musiman aditif

E_t = *random error* komponen

Metode AHW digunakan untuk variasi data musiman dari data runtun waktu yang konstan. Pada akhir periode ke- t , nilai ramalan (Y_{t+k}) untuk periode $(t+k)$ diperoleh dari persamaan:

$$Y_{t+k} = (L_t + kT_t) + S_{t+k-c} \quad (3.2)$$

Dengan bentuk pemulusan sebagai berikut:

a. Pemulusan Keseluruhan (*level*)

$$L_t = \alpha(Y_t - S_{t-c}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3.3)$$

b. Pemulusan Kecenderungan

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (3.4)$$

c. Pemulusan musiman

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-c} \quad (3.5)$$

Keterangan:

Y_t = nilai actual pada periode akhir t

α = parameter penghalusan untuk data ($0 < \alpha < 1$)

β = parameter penghalusan untuk kecenderungan ($0 < \beta < 1$)

γ = parameter penghalusan untuk musiman ($0 < \gamma < 1$)

T_t = pemulusan kecenderungan

L_t = pemulusan keseluruhan

S_t = pemulusan musiman

3.4 *Multiplicative Holt Winters*

Multiplicative Holt Winters (MHW) adalah model prakiraan *time series* yang mana amplitud dari *poola seasonal* adalah proporsional untuk rata-rata tingkatan deret. Deret waktu menunjukkan *seasonal Multiplicative*. Dalam model ini mengasumsikan *time series* dengan model:

$$Y_t = (b_1 + b_2 t) S_t + E_t \quad (3.6)$$

Dimana:

b_1 = komponen permanen

b_2 = komponen kecenderungan

S_t = faktor musiman

E_t = komponen *random error*

Metode MHW digunakan untuk variasi data musiman dari data runtun waktu yang mengalami peningkatan atau penurunan (fluktuasi). Nilai ramalan (Y_{t+k}) untuk periode ($t + k$) yang ditinjau pada akhir periode ke- t dari model ini adalah

$$Y_{t+k} = (L_t + kT_t) S_{t+k-c} \quad (3.7)$$

Dengan nilai pemulusan sebagai berikut:

a. Pemulusan Keseluruhan (*level*)

$$\alpha \frac{Y_t}{S_{t-c}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3.8)$$

b. Pemulusan Kecenderungan (*trend*)

$$T_t = \beta(L_t - L_{t-1})S + (1 - \beta)T_{t-1} \quad (3.9)$$

c. Pemulusan Musiman (*seasonal*)

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-c} \quad (3.10)$$

Dengan $0 \leq \alpha, \beta, \gamma \leq 1$, S_{t-c} nilai estimasi faktor musiman, c adalah panjang musiman dan $k = 1, 2, \dots, c$.

3.5 *Motode Improved Holt Winters*

Metode *Improved Holt Winters* (IHW) merupakan adaptasi dari metode *Holt Winters Additive* dengan perbedaan terletak pada pemulusan keseluruhannya (*level*), Algoritmanya adalah sebagai berikut:

Inisialisasi:

$$Ls = \frac{1}{s} \sum_{i=1}^s Y_i \quad (3.11)$$

$$Bs = \frac{1}{s} \left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{2s} - Y_s}{s} \right] \quad (3.12)$$

$$S_i = Y_i - Ls, i = 1, \dots, s \quad (3.13)$$

Dengan perhitungan:

- a. Pemulusan keseluruhan (*level*)

$$L_t = \alpha Y_t - S_{t-c} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3.14)$$

- b. Pemulusan kecenderungan (*trend*)

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (3.15)$$

- c. Pemulusan musiman (*seasonal*)

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-c} \quad (3.16)$$

- d. Prakiraan (*forecast*)

$$F_{t+1} = L_t + b_t + S_{t+1-s} \quad (3.17)$$

Keterangan:

Y_t = nilai actual pada periode akhir t

α = parameter penghalusan untuk data ($0 < \alpha < 1$)

β = parameter penghalusan untuk kecenderungan ($0 < \beta < 1$)

γ = parameter penghalusan untuk musiman ($0 < \gamma < 1$)

T_t = pemulusan kecenderungan

L_t = pemulusan keseluruhan

S_t = pemulusan musiman

Untuk metode IHW pemulusan parameter α hanya terjadi pada pengamatan data Y_t dan tidak pada faktor *seasonal* S_{t-c} . Hal ini dilakukan untuk mengeliminasi fluktuasi musimannya dari pemulusan nilai pada Y_t . Ketika $\alpha Y_t > S_{t-c}$ (pemulusan nilai dalam periode t adalah lebih besar daripada nilai rata-rata musimannya dalam periode $t-c$) *level*nya meningkat pada perbandingan dengan *level* pada periode sebelumnya; penyesuaian terjadi sebaliknya saat $\alpha Y_t < S_{t-c}$ nilai awal untuk indeks *level*, *trend* dan *seasonal* adalah identik dengan metode AHW.

3.6 Evaluasi Prakiraan

Evaluasi prakiraan dilakukan untuk mengetahui kesalahan pada nilai prakiraan. Perhitungan nilai kesalahan pramalan menggunakan MAPE, MSE dan RMSE.

3.6.1 Mean Absolute Presentage Error

Mean Absolute Presentage Error (MAPE) adalah presentase tingkat error perhitungan yang didapat dari persamaan berikut:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{At - Ft}{At} \right| \times 100 \quad (3.18)$$

Dimana n adalah banyaknya *Forecasting*, A adalah nilai aktual pada bulan tersebut dan F adalah *forecasting* pada bulan tersebut. Hasil dari perhitungan MAPE dibuat dalam persentase.

3.6.2 Mean Square Error

Mean Square Error (MSE) adalah metode lain untuk mengevaluasi metode prakiraan. Masing-masing kesalahan dikuadratkan kemudian dijumlahkan dan ditambahkan dengan jumlah observasi. Dengan pendekatan ini akan mengatur kesalahan prakiraan yang besar sehingga menghasilkan kesalahan yang sedang yang kemungkinan lebih baik untuk kesalahan kecil.

$$MSE = \frac{\sum e_i^2}{n} = \frac{\sum (X_i - F_i)^2}{n} \quad (3.19)$$

Keterangan:

X = Data Aktual

F = Data prakiraan

n = Jumlah data

3.6.3 Root Mean Square Error

Root Mean Square Error (RMSE) merupakan nilai rata-rata dari jumlah kuadrat kesalahan yang digunakan untuk menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. RMSE dapat mengukur tingkat akurasi hasil prakiraan suatu model prakiraan. Nilai RMSE semakin rendah menunjukkan bahwa nilai prakiraan semakin mendekati nilai aslinya.

3.7 Data Stasioner

Data stasioner merupakan data yang berada di sekitar nilai rata-rata yang konstan, tidak mengalami pertumbuhan atau penurunan, tidak tergantung pada waktu dan varian dari fluktuasi data (Makridakis, dkk, 1997). Proses stasioner adalah proses menyeimbangkan data menjadi konstan. Peluang keseimbangan yang

berdistribusi pada waktu (t_1, t_2, \dots, t_m) memiliki peluang distribusi yang sama dengan $(t_1 + k; t_2 + k, \dots, t_m + k)$ (Abraham and Johannes: 194, 2005). Stasioneritas merupakan hal yang utama dalam pembuatan model *time series* pada prakiraan (Bisgard dan Kulahci, 2011). Prakiraan dapat dilakukan apabila kondisi data sudah stasioner.

Terdapat tiga kemungkinan yang menyebabkan data menjadi tidak stasioner, yaitu tidak stasioner varian, tidak stasioner rata-rata, dan tidak stasioner varian & rata-rata. Data deret waktu (*time series*) dapat dikatakan memiliki stasioner yang kuat apabila telah memenuhi persamaan:

- (i) $\text{Var}(X_t) < \infty$ for all $t \in \mathbb{Z}$,
- (ii) $\mu_{X(t)} = \mu$ for all $t \in \mathbb{Z}$,
- (iii) $\gamma_{X(r, s)} = \gamma_{X(r + t, s + t)}$ for all $r, s, t \in \mathbb{Z}$

di mana :

$\text{Var}(X_t)$ = varian dari (X_t)

$\mu_{X(t)}$ = rata -rata dari (X_t)

$\gamma_{X(r, s)}$ = autokovarian dari $X(r, s)$

Suatu proses stasioner dapat bersifat kuat (strickly stationary), dimana nilai mean (μ), varians (σ^2) dan kovarians (τ_k) tidak terpengaruh oleh waktu pengamatan. Ketidakstasioneran dalam *time series* dibedakan menjadi dua (2), yaitu tidak stasioner dalam mean (disebabkan μ_t tidak konstan) dan tidak stasioner dalam varians (disebabkan σ_t^2 yang dependent terhadap deret waktu). Tidak stasioner dalam mean dapat di atasi dengan melakukan differencing (pembedaan) dan stabilizing varians (transformasi) untuk menstasionerkan varians (Mufidah, 2010)

3.8 Autocorrelation Function (ACF)

Autocorrelation function atau fungsi autokorelasi adalah hubungan atau korelasi antara data pengamatan satu dengan data pengamatan lainnya pada suatu data deret berkala. Nilai-nilai pada data deret berkala memiliki korelasi yang sama dengan selisih waktu (time lag) 0, 1, 2 periode atau lebih. Artinya, jika lag =2, maka nilai t akan dikorelasikan dengan $t+2$ (artinya data pertama akan dikorelasikan dengan data ketiga), jika menggunakan lag=4, maka kemudian t akan dikorelasikan

dengan $t+4$, begitu seterusnya. Hal ini bisaanya ditulis dengan rumus korelasi (X_t , X_{t+k}) dengan k adalah jumlah lag dan t adalah nilainya.

Fungsi Autokorelasi adalah:

$$\rho_k = \frac{\text{cov}(Z_t, Z_{t+k})}{\sqrt{\text{var}(Z_t)}\sqrt{\text{var}(Z_{t+k})}} \quad (3.20)$$

3.9 Partial Autocorrelation Function (PACF)

Partial Autocorrelation Function adalah suatu fungsi untuk mengukur tingkat hubungan antara Z_t dan Z_{t+k} apabila pengaruh dari time lag 1, 2, . . . , $k-1$ dianggap terpisah. Kondisi pengamatan dari proses (Z_t) yang stasioner ditunjukkan dengan nilai mean $E(Z_t) = \mu$, $\text{var}(Z_t) = E\{Z_t - \mu\}^2 = \sigma^2$ adalah proses yang konstan. Covarian (Z_t, Z_{t-k}) yang merupakan fungsi hanya dari perbedaan waktu $|t-(t-k)|$. (Abraham and Johannes: 194 - 209, 2005). Perbedaannya PACF dengan ACF adalah jika ACF mencari korelasi antara X_t dan X_{t+k} ($k = \text{lag}$), maka PACF juga menghitung korelasi antara X_t dan X_{t+k} namun sebelumnya menghilangkan data antara X_t dan X_{t+k} terlebih dahulu. Berarti data yang dihilangkan adalah data X_{t+1} sampai dengan X_{t+k-1}

3.10 Bobot nilai awal

Ada tiga parameter penghalusan yang digunakan dalam metode penghalusan eksponensial Holt-Winters, yakni:

1. Alpha (α) merupakan parameter yang mengontrol penghalusan relative pada pengamatan yang baru dilakukan. Jika alpha bernilai mendekati 1 maka hanya pengamatan terbaru yang digunakan secara eksklusif. Sebaliknya bila alpha mendekati 0 maka pengamatan yang lain dihitung dengan bobot sepadan dengan yang terbaru.
2. Beta (β) merupakan parameter yang mengontrol penghalusan relatif pada pengamatan yang baru dilakukan untuk mengestimasi kemunculan *trend* nilai beta berkisar dari 0 sampai 1.
3. Gamma (γ) merupakan parameter yang mengontrol penghalusan relative pada pengamatan yang baru dilakukan untuk mengestimasi kemunculan unsur musiman. Nilai gamma berkisar dari 0 sampai 1 (Mulyana, 2004).

Menurut metode pemulusan eksponensial dari *Holt Winters*, nilai awal juga dapat ditentukan nilainya secara sembarang dengan ketentuan yaitu:

$$S_{L+1} = X_{L+1} \quad (3.21)$$

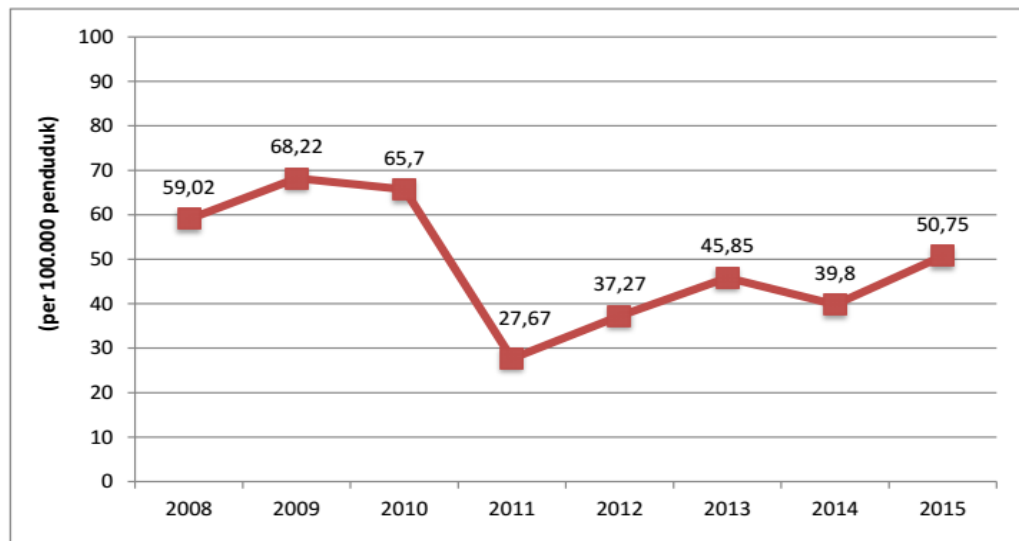
3.11 Demam Berdarah *Dengue* (DBD)

Penyakit Demam Berdarah *Dengue* (DBD) adalah penyakit yang disebabkan oleh virus *dengue* yang tergolong *Arthropod-Borne Virus*, genus *Flavivirus*, dan famili *Flaviviridae*. DBD ditularkan melalui gigitan nyamuk dari genus *Aedes*, terutama *Aedes aegypti* atau *Aedes albopictus*. Masa inkubasi ekstrinsik (di dalam tubuh nyamuk) berlangsung sekitar 8-10 hari, sedangkan inkubasi intrinsik (dalam tubuh manusia) berkisar antara 4-6 hari dan diikuti dengan respon imun. Penyakit DBD cenderung semakin luas penyebarannya sejalan dengan meningkatnya kepadatan dan mobilitas penduduk serta semakin meluasnya virus penyebab dan nyamuk penularnya (*aedes aegypti*) tersebar di rumah-rumah maupun tempat umum, kecuali wilayah dengan ketinggian lebih dari 1000 meter di atas permukaan laut. Penyakit DBD dapat muncul sepanjang tahun dan menyerang seluruh kalangan umur. (Menkes RI, 2011).

Virus *dengue* dilaporkan telah menjangkiti lebih dari 100 negara di dunia, terutama pada wilayah perkotaan yang memiliki jumlah penduduk padat dan ke wilayah pemukiman/pedesaan. Virus *dengue* ditemukan di sebagian besar wilayah yang memiliki iklim tropis dan subtropis, seperti pada wilayah Asia Tenggara, Amerika Tengah, Amerika Selatan dan Karibia (Kurane, 2007). Jumlah orang yang terinfeksi diperkirakan sekitar 50 sampai 100 juta orang dan mengakibatkan 22.000 kematian setiap tahun. Jumlah populasi yang berada pada wilayah endemis DBD sebanyak 40% dari seluruh penduduk dunia yang memungkinkan terinfeksi virus *dengue* melalui gigitan nyamuk setempat (Knewlton, dkk. 2009).

Jumlah kasus DBD cenderung terus meningkat dan banyak menimbulkan kematian terutama pada anak-anak, 90% di antaranya menyerang anak dengan umur di bawah 15 tahun. Di Indonesia, pada tahun 2015 jumlah penderita DBD yang dilaporkan adalah sebanyak 129.650 kasus dengan jumlah kematian sebanyak 1.071 orang (IR/Angka kesakitan= 50,75 per 100.000 penduduk dan CFR/angka kematian= 0,83%). Dibandingkan tahun 2014 dengan kasus sebanyak 100.347 serta

IR 39,80 terjadi peningkatan kasus pada tahun 2015. Berikut grafik angka kesakitan DBD selama kurun waktu 2008-2015 (Ditjen P2P, Kemenkes RI, 2016).

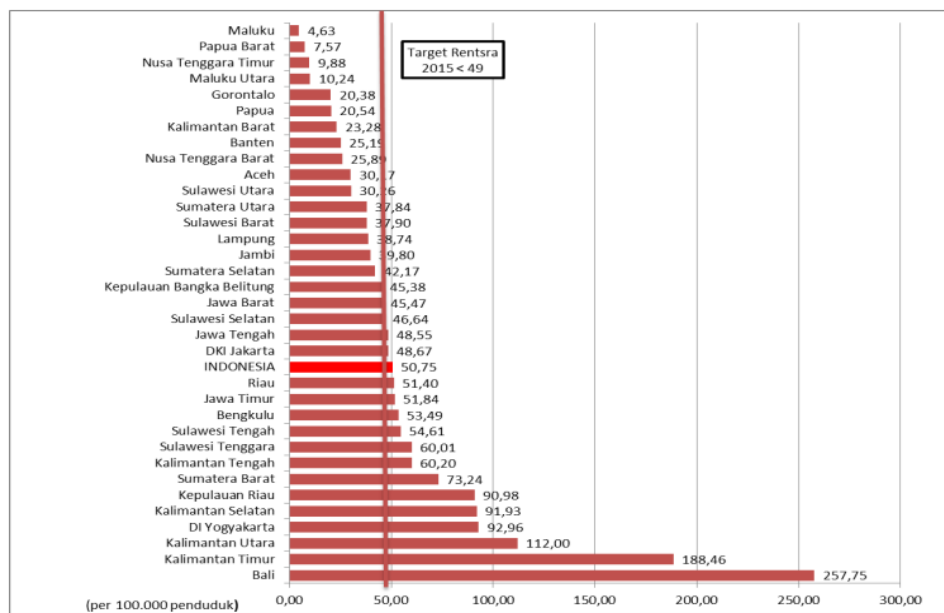


Gambar 3. 1 Grafik jumlah penderita demam berdarah di Indonesia

Pada tahun 2015 Provinsi Daerah Istimewa Yogyakarta merupakan salah satu wilayah dengan jumlah penderita DBD terbanyak ke-4 setelah Provinsi Bali, Kalimantan Timur dan Kalimantan Utara. Pada tahun 2015 jumlah kejadian Demam Berdarah telah tercatat sebanyak 92,96 per 100.000 penduduk. Berikut adalah data sejumlah kasus penderita demam berdarah pada tahun 2015 berdasarkan tingkat provinsi.

Tabel 3. 1 Data penderita DBD tingkat Provinsi

No	Provinsi	Jumlah Penduduk	Demam Berdarah Dengue			
			Jumlah Kasus	Incidence Rate per 100.000 Penduduk	Jumlah Kasus Meninggal	Case Fatality Rate (%)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Aceh	5.001.953	1.589	30,17	6	0,40
2	Sumatera Utara	13.937.797	5.274	37,84	30	0,57
3	Sumatera Barat	5.196.289	3.806	73,24	22	0,58
4	Riau	6.344.402	3.261	51,40	22	0,67
5	Jambi	3.402.052	1.354	39,80	8	0,59
6	Sumatera Selatan	8.052.315	3.396	42,17	16	0,47
7	Bengkulu	1.874.944	1.003	53,49	20	1,99
8	Lampung	8.117.268	3.145	38,74	32	1,02
9	Kepulauan Bangka Belitung	1.372.813	623	45,38	6	0,96
10	Kepulauan Riau	1.973.043	1.795	90,98	16	0,89
11	DKI Jakarta	10.177.924	4.954	48,67	11	0,22
12	Jawa Barat	46.709.569	21.237	45,47	14	0,07
13	Jawa Tengah	33.774.141	16.398	48,55	255	1,56
14	DI Yogyakarta	3.679.176	3.420	92,96	35	1,02
15	Jawa Timur	38.847.561	20.138	51,84	283	1,41
16	Banten	11.955.243	3.011	25,19	43	1,43
17	Bali	4.152.833	10.704	257,75	28	0,26
18	Nusa Tenggara Barat	4.835.577	1.252	25,89	5	0,40
19	Nusa Tenggara Timur	5.120.061	506	9,88	1	0,20
20	Kalimantan Barat	4.789.574	1.115	23,28	15	1,35
21	Kalimantan Tengah	2.495.035	1.502	60,20	10	0,67
22	Kalimantan Selatan	3.989.793	3.668	91,93	40	1,09
23	Kalimantan Timur	3.426.638	6.458	188,46	65	1,01
24	Kalimantan Utara	641.936	719	112,00	2	0,28
25	Sulawesi Utara	2.412.118	730	30,26	17	2,33
26	Sulawesi Tengah	2.876.689	1.571	54,61	13	0,83
27	Sulawesi Selatan	8.520.304	3.974	46,64	19	0,48
28	Sulawesi Tenggara	2.499.540	1.500	60,01	9	0,60
29	Gorontalo	1.133.237	231	20,38	14	6,06
30	Sulawesi Barat	1.282.162	486	37,90	0	0,00
31	Maluku	1.686.469	78	4,63	6	7,69
32	Maluku Utara	1.162.345	119	10,24	2	1,68
33	Papua Barat	871.510	66	7,57	3	4,55
34	Papua	3.149.375	647	20,54	3	0,46
Indonesia		253.461.686	129.650	50,75	1.071	0,83



Gambar 3. 2 Grafik Data penderita DBD tingkat Provinsi

Penyakit DBD dapat terjadi dan berkembang biak secara meluas dikarenakan oleh beberapa faktor diantaranya; agent (virus dengue), host / manusia yang rentan terjangkit virus serta lingkungan yang memungkinkan tumbuh dan berkembang biaknya nyamuk *Aedes spp.* Selain itu, juga dipengaruhi oleh faktor lainnya seperti kepadatan dan mobilitas penduduk, kualitas perumahan, jarak antar

rumah, pendidikan, pekerjaan, sikap hidup, curah hujan, golongan umur, iklim dan lainnya.

3.12 Twitter

Twitter adalah sebuah situs web mikroblog yang didirikan oleh Jack Dorsey dan diresmikan pada tahun 2006 dimana menawarkan penggunaannya untuk dapat mengirim dan menerima konten atau pesan tweet (Twitter, 2006). Mikroblog merupakan alat komunikasi online yang memungkinkan penggunaannya untuk dapat berbagi pikiran dan ide tentang suatu objek atau fenomena tertentu dalam ukuran teks yang sedikit. Tweet adalah konten atau teks tulisan yang dibatasi hingga 140 karakter dimana batasan teks tersebut membuat pesan menjadi lebih ringkas untuk disampaikan kepada pengguna. Pengguna dapat saling berinteraksi dengan menuliskan simbol “@” pada tweet untuk terhubung dengan referensi pengguna lainnya dan menggunakan *hashtags* (#) untuk membuat topik tweet sehingga mudah untuk melihat kembali percakapan tweet. Twitter bukan hanya alat yang berguna untuk terhubung dengan teman-teman untuk mengetahui perkembangan kegiatan sehari-hari. Kekuatan sebenarnya dari twitter adalah penyebaran informasi secara viral. Pengguna dapat mengikuti organisasi berita, selebriti, film dan informasi lainnya (Twitter, 2006)

Tingginya tingkat popularitas Twitter menyebabkan jejaring sosial dapat digunakan sebagai indikator adanya suatu kejadian yang menjadi perhatian publik. Hal ini dapat dilihat dari jumlah tweet yang dilakukan masyarakat.

3.13 Deret berkala

Deret berkala (*Time series*) merupakan sekumpulan data statistik yang dikumpulkan secara berkala sesuai dengan urutan waktu. Metode dari analisis deret waktu adalah yang memprakarsai untuk proses stokastik pada umumnya dan rantai markov. Tujuan dari analisis deret waktu adalah untuk menggambarkan dan meringkas data deret berkala agar menjadi model dimensi yang sederhana dan untuk membuat prakiraan dengan metode kuantitatif. Deret waktu dibagi menjadi dua menurut jenis datanya, yaitu data kontinu dan data diskret. Data kontinu adalah data yang memiliki deret waktu kontinu dengan selang waktu tertentu. Sedangkan

data diskret adalah data yang diambil pada waktu-waktu tertentu dan bisaanya berjarak waktu yang sama (Wei, 2006). Salah satu metode sederhana untuk menggambarkan kecenderungan *time series* adalah dengan menguraikan rangkaian data menjadi empat elemen, yaitu:

1. Data kecenderungan (*trend*)

Data kecenderungan atau *trend* sekuler adalah kecenderungan gerakan data dalam jangka panjang rata-rata. Kecenderungan data dapat bergerak menurun, meningkat atau tetap. Cara sederhana untuk mendetek *trend* dalam data musiman adalah dengan mengambil rata-rata selama jangka waktu tertentu. Jika nilai rata-rata berubah seiring dengan waktu, maka dapat dikatakan bahwa ada bukti dari sebuah *trend* dalam data.

2. Data efek musiman

Data efek musiman adalah data yang mengalami perubahan siklus naik-turun berhubungan dengan kalender, misalnya harian, mingguan, bulanan, tahunan.

3. Data siklis

Data siklis adalah data yang terdapat di sekitar garis pola dan mengalami perubahan siklus, misalnya seperti pada perubahan siklus bisnis. Pola siklis hampir mirip dengan data yang memiliki pola musiman. Akan tetapi pola siklis memiliki rentang waktu yang lebih panjang daripada pola musiman.

4. Data residual / irregular

Data residual adalah data yang perubahannya terjadi secara random atau sistematis yang disebabkan oleh faktor-faktor yang tidak terduga dan bisaanya muncul dalam jangka pendek yang mengakibatkan terjadinya variasi.

BAB IV

ANALISIS DAN PERANCANGAN SISTEM

4.1 Deskripsi Umum Metode Penelitian

Penelitian ini mengimplementasikan metode *Improved Holt Winters* (IHW) ke dalam suatu sistem. Sehingga sistem dapat melakukan prakiraan jumlah kasus kejadian demam berdarah di Kabupaten Sleman, Daerah Istimewa Yogyakarta (DIY) berdasarkan jumlah data kejadian demam berdarah.

Penelitian ini dibagi menjadi dua tahapan yaitu tahapan pengumpulan data demam berdarah di Kabupaten Sleman dan tahapan implementasi prakiraan dengan model *Improved Holt Winters*. Pada tahapan pertama adalah pengumpulan data demam berdarah di Kabupaten Sleman - DIY. Data yang dikumpulkan adalah data asli jumlah kejadian demam berdarah, data jumlah kejadian demam berdarah berdasarkan tweet dan data jumlah tweet tentang demam berdarah. Semua data diambil dari rentang waktu tahun 2010 hingga tahun 2014. Data asli jumlah kejadian demam berdarah diperoleh berasal dari Dinas Kesehatan Yogyakarta. Data jumlah kejadian demam berdarah berdasarkan tweet dan jumlah tweet mengenai demam berdarah diperoleh berasal dari hasil scraping pada twitter.

Data yang telah diperoleh kemudian akan diolah ke dalam database pada Microsoft excel untuk selanjutnya dilakukan prakiraan. Pada tahap kedua yaitu tahapan implementasi prakiraan dengan menggunakan model *Improved Holt Winters*. Pada tahapan ini yaitu pembuatan program untuk mengimplementasikan model *Improved Holt Winters* terhadap kumpulan data dengan menggunakan Bahasa pemrograman python. program yang telah dibuat nantinya mampu untuk melakukan prakiraan demam berdarah. Hasil prakiraan akan dianalisis sehingga dapat diketahui nilai akurasi nya.

4.2 Spesifikasi Data

4.2.1 Data Input

Data Input adalah data yang digunakan sebagai masukan untuk membangun suatu sistem. Data tersebut diproses sehingga dapat menghasilkan suatu output yaitu berupa hasil prakiraan. Data input yang digunakan adalah:

1. Data jumlah kejadian demam berdarah di wilayah Kabupaten Sleman.

Data jumlah kejadian penderita demam berdarah diperoleh dari Dinas Kesehatan Daerah Istimewa Yogyakarta (Dinkes DIY) tahun 2010-2014. Data tersebut dikelompokkan ke dalam jumlah data demam berdarah per bulan pada Kabupaten Sleman.

Tabel 4. 1 Data jumlah kejadian demam berdarah kabupaten sleman

Bulan	Jumlah kejadian demam berdarah	Sep-11	3	May-13	77
		Bulan	Jumlah kejadian demam berdarah	Bulan	Jumlah kejadian demam berdarah
Jan-10	148				
Feb-10	120	Oct-11	2	Jun-13	32
Mar-10	83	Nov-11	9	Jul-13	78
Apr-10	71	Dec-11	5	Sep-13	33
May-10	22	Jan-12	17	Oct-13	13
Jun-10	24	Feb-12	4	Nov-13	28
Jul-10	39	Mar-12	16	Dec-13	49
Aug-10	23	Apr-12	14	Jan-14	75
Sep-10	6	May-12	20	Feb-14	62
Oct-10	7	Jun-12	19	Mar-14	60
Nov-10	21	Jul-12	24	Apr-14	46
Dec-10	39	Aug-12	11	May-14	43
Jan-11	22	Sep-12	15	Jun-14	39
Feb-11	41	Oct-12	15	Jul-14	23
Mar-11	28	Nov-12	39	Aug-14	42
Apr-11	15	Dec-12	42	Sep-14	38
May-11	18	Jan-13	133	Oct-14	22
Jun-11	11	Feb-13	110	Nov-14	43
Jul-11	7	Mar-13	77	Dec-14	45
Aug-11	5	Apr-13	81		

Data tweet berbahasa Indonesia Data twitter diperoleh dari hasil scrapping terhadap halaman situs www.mobile.twitter.com. Scrapping dilakukan pada halaman hasil search dengan keyword berupa kata dan frasa yang berkaitan dengan demam berdarah (“demam berdarah”, “demam dengue”, “dbd”, ”sakit demam berdarah, “dengue fever”, “dhf”, “dengue”, “dengue hemorrhagic fever”). Data hasil scrapping akan disimpan dalam database. Data yang sudah terkumpul akan dihitung jumlah nya berdasarkan konten dan berdasarkan jumlah tweet mengenai demam berdarah.

Tabel 4. 2 Data twitter (jumlah kejadian demam berdarah, jumlah tweet)

Bulan	Jumlah kejadian demam berdarah berdasarkan tweet	Jumlah tweet
Jan-10	0	0
Feb-10	0	0
Mar-10	0	0
Apr-10	0	0
May-10	0	0
Jun-10	0	0
Jul-10	0	0
Aug-10	0	0
Sep-10	1	2
Oct-10	6	5
Nov-10	0	0
Dec-10	2	3
Jan-11	5	7
Feb-11	2	4
Mar-11	3	6
Apr-11	2	3
May-11	0	0
Jun-11	1	3
Jul-11	0	1
Aug-11	2	6
Sep-11	1	1
Oct-11	0	1
Nov-11	1	1
Dec-11	2	4
Jan-12	1	2
Feb-12	1	3
Mar-12	1	3
Apr-12	5	6
May-12	4	7
Jun-12	1	3
Jul-12	3	6
Aug-12	0	0
Sep-12	1	1
Oct-12	0	2
Nov-12	1	3
Dec-12	5	10
Jan-13	4	10

Feb-13	9	22	Feb-14	4	12
Mar-13	12	27	Mar-14	2	13
Apr-13	13	21	Apr-14	4	20
May-13	11	25	May-14	9	16
Jun-13	5	16	Jun-14	7	15
Jul-13	6	8	Jul-14	4	9
Aug-13	4	9	Aug-14	11	11
Sep-13	17	14	Sep-14	9	18
Oct-13	16	15	Oct-14	9	13
Nov-13	8	12	Nov-14	3	10
Dec-13	7	22	Dec-14	3	11
Jan-14	7	22			

4.2.2 Data output

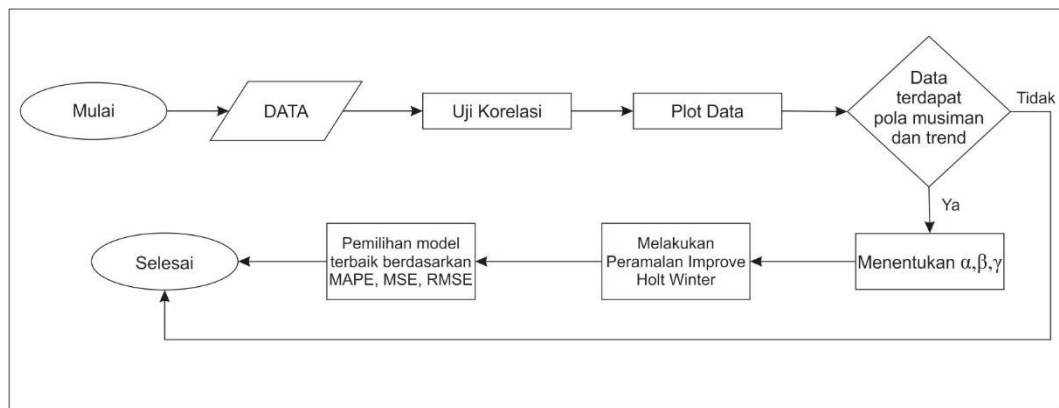
Data output adalah data yang dihasilkan setelah data input diproses dengan menggunakan metode permalan *Improved Holt Winter*. Data output ini berupa hasil prakiraan jumlah kejadian demam berdarah dengan analisis menggunakan metode *Improved Holt Winters* dan nilai akurasinya.

4.3 Proses Prakiraan

Data yang telah dikumpulkan pada database kemudian dilakukan proses prakiraan. Dari ketiga data tersebut kemudian dilakukan pengujian korelasi dengan tingkat signifikansi 95% untuk melihat korelasinya antara data asli jumlah kejadian demam berdarah dengan jumlah kejadian demam berdarah berdasarkan twitter dan data asli jumlah kejadian demam berdarah dengan jumlah tweet mengenai demam berdarah. Apabila pada data tersebut memiliki saling korelasi maka selanjutnya dilakukan plotting data. Plotting data dibuat selama kurun waktu yang diteliti untuk melihat apakah terdapat *trend* dan musiman pada data. Selanjutnya akan dilakukan uji kointegrasi pada data tersebut untuk mengetahui kestasioneran data yang diperoleh. Uji kointegrasi selesai maka selanjutnya akan dilakukan proses prakiraan dengan menggunakan metode IHW. Dengan nilai-nilai parameter α , β , dan

γ dapat ditentukan melalui *trial* dan *error* dengan tujuan untuk meminimumkan MAPE.

Setelah beberapa percobaan nilai awal prakiraan diperoleh, maka selanjutnya dilakukan perbandingan untuk memilih nilai awal yang paling baik. Perbandingan dilakukan dengan melihat hasil pengukuran tingkat kesalahan. Pengukuran kesalahan penelitian ini menggunakan MAPE, yang diharapkan nilainya sangat kecil dan dapat merepresentasikan data.



Gambar 4. 1 Diagram alir proses prakiraan Improved Holt Winter

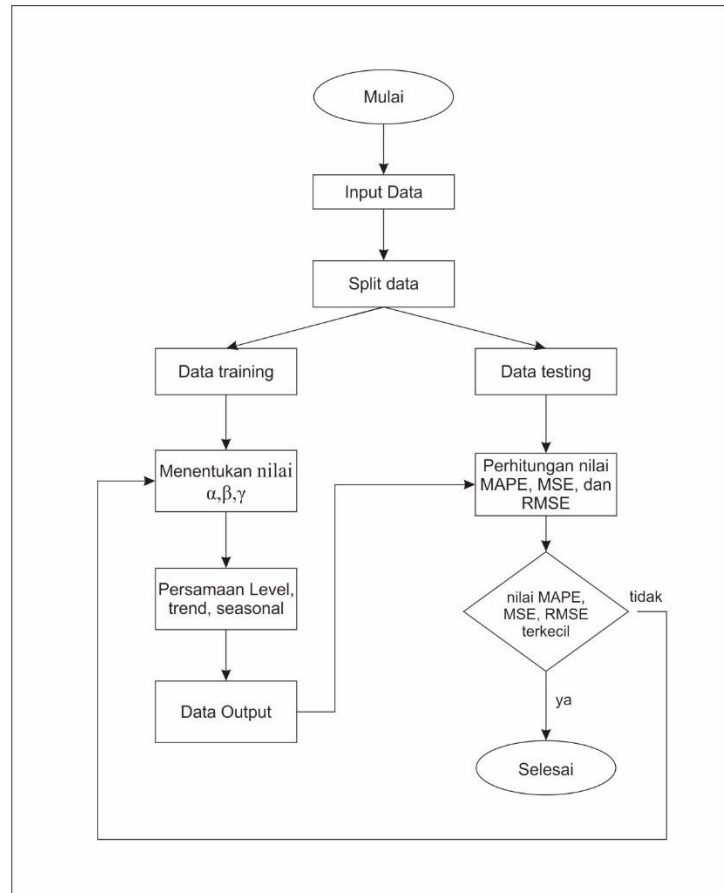
4.4 Analisis Pembuatan Sistem

Sistem yang akan dibuat dapat memberikan hasil prakiraan dengan metode *Improved Holt Winter*. Sistem memiliki beberapa spesifikasi fungsional yaitu diantaranya:

1. Sistem dapat menampilkan persamaan model *Improved Holt Winter*
2. Sistem dapat melakukan prakiraan jumlah kejadian demam berdarah dengan menggunakan metode *Improved Holt Winter*
3. Sistem mampu menampilkan hasil prakiraan dan nilai percentage error dari data *testing* dan data *training* beserta dengan nilai akurasinya.

4.5 Alur Pembuatan Sistem Dengan Metode Improved Holt Winters

Secara umum proses alur kerja dari pembuatan sistem dengan menggunakan improve holt winters dapat digambarkan pada gambar 4.2



Gambar 4. 2 Diagram alir pembuatan sistem

Berdasarkan gambar 4.2 pembuatan sistem dimulai dengan memasukkan data yang akan dilakukan prakiraannya berupa data jumlah kejadian demam berdarah dan data jumlah tweet mengenai demam berdarah. Pada masing-masing data tersebut akan dibagi menjadi dua kelompok data yaitu data training dengan rentang waktu januari 2010-desember 2013 dan data testing dengan rentang waktu januari 2014-desember 2014. Selanjutnya adalah menentukan nilai awal alpha, beta, dan gamma dengan cara trial dan error. Pada nilai awal tersebut akan dilakukan perhitungan dengan menggunakan persamaan level, trend dan seasonal sehingga menghasilkan data output berupa nilai prakiraan pada januari 2014 – desember 2014. Data output yang dihasilkan kemudian dilakukan perhitungan nilai error MAPE, MSE, dan RMSE dengan bantuan data testing. hasil dari nilai eror akan dibandingkan dengan nilai eror selanjutnya. Apabila nilai error belum menemukan yang terkecil maka nilai alpha, beta dan gamma akan diubah nilainya. Apabila sudah mendapatkan nilai error terkecil maka sistem proses prakiraan dengan metode improve holt winters sudah selesai.

4.6 Scrapping dan preprocessing

Scrapping dan preprocessing twitter dilakukan dengan tujuan untuk mendapatkan data tweet yang berkaitan dengan demam berdarah. Scrapping dilakukan pada halaman pencarian website www.mobile.twitter.com. Kemudian dilakukan preprocessing dengan menghitung jumlah tweet bulanan dan filtering tweet untuk memilih tweet yang berasal dari wilayah DIY. Kemudian hasil dari preprocessing ini dapat digunakan sebagai masukan untuk melakukan proses prakiraan. Data tweet ini juga digabungkan dengan data kejadian penderita demam berdarah dan dijadikan menjadi dua jenis kelompok data berupa data *training* dan kelompok data *testing*.

Proses *scrapping* dilakukan pada halaman url www.twitter.com/search-advanced dengan melakukan pada tiga kategori pencarian yaitu pencarian dengan *query keyword* demam berdarah ditambah fitur lokasi Sleman, *keyword* demam berdarah AND Sleman, dan *keyword* demam berdarah saja. *Query* tersebut akan dimasukkan pada *field search* twitter dan hasilnya dapat ditampilkan pada gambar berikut:

Gambar 4. 3 Ttweet demam berdarah pada twitter

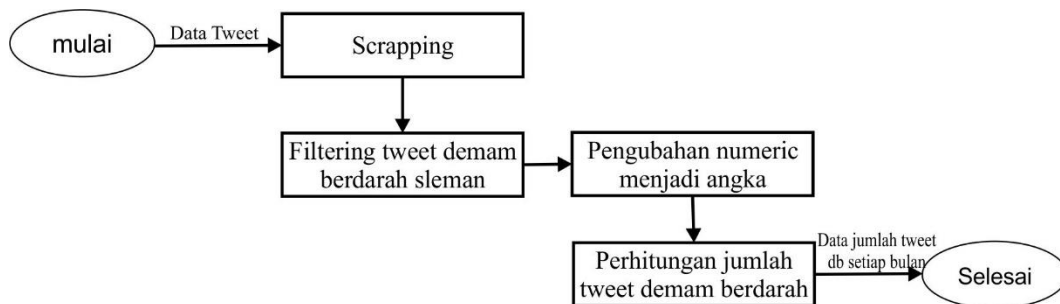


Pada Tweet pertama menunjukkan tampilan dari hasil pencarian berdasarkan query keyword demam berdarah dengan fitur lokasi Sleman dan sekitarnya dengan radius 15 mil. Pada tweet kedua adalah tampilan dari hasil pencarian dengan menggunakan keyword demam berdarah dan Sleman. Setelah

didapatkan data postingan tweet dalam kedua kategori selanjutnya dilakukan perhitungan. Aturan perhitungan dilakukan dua macam cara, yaitu:

1. Perhitungan jumlah kejadian demam berdarah berdasarkan isi tulisan tweet
2. Perhitungan jumlah tweet mengenai demam berdarah.

Berdasarkan hasil tweet maka dapat dilihat pada postingan @_RobertoSaurus dan @IniSleman jika menggunakan proses perhitungan pertama jumlah kejadian dihitung berdasarkan isi teks nya adalah satu orang kejadian dan 234 orang kejadian demam berdarah. Sedangkan jika menggunakan proses perhitungan kedua maka hasilnya berdasarkan jumlah tweet yaitu terdapat 2 tweet. Kedua proses perhitungan ini dilakukan agar mendapatkan korelasi terbaik antara jumlah kejadian demam berdarah dengan tweet mengenai demam berdarah.



Gambar 4. 4 Diagram alir scraping data tweet pada twitter

Pada gambar 4.3 dapat diuraikan proses dalam melakukan scraping dilakukan untuk mendapatkan data jumlah tweet demam berdarah setiap bulannya. Tahapan yang dilakukan adalah melakukan scrapping pada website twitter untuk menggali data teks yang umumnya bertipe HTML atau XHTML. Selanjutnya melakukan Filtering data teks pada twitter untuk menyaring data yang dibutuhkan yang berkaitan dengan tweet demam berdarah. Hasil dari filtering yang berupa data teks diubah menjadi data numeric dan dilakukan perhitungan jumlah tweet berdasarkan aturan perhitungan konten tweet dan jumlah tweet tentang demam berdarah pada tiap bulannya. Data ini kemudian akan disimpan pada file excel untuk dilakukan perhitungan koefisien korelasi antara jumlah data perhitungan tweet dan jumlah data kejadian demam berdarah sebenarnya. Setelah itu dilakukan uji signifikasnsi dengan menggunakan critical value dari nilai koefisien korelasi. Deret data yang memiliki korelasi akan digunakan dalam model prakiraan *improved*

Holt Winters. Dua buah deret data yang memiliki korelasi silang memungkinkan suatu deret data dapat diprakirakan berdasarkan deret data yang lain (Bechtel, 2005; Dewey, 2007). Deret data kemudian dibagi menjadi dua kelompok, yaitu kelompok data *training* sebesar 80% dan data *testing* sebesar 20% (Hydman, 2013). Kelompok data *training* digunakan untuk melakukan pembuatan model dan kelompok data *testing* digunakan untuk pengujian model.

4.7 Stasionerisasi data

Pada data tweet dan jumlah kejadian demam berdarah dilakukan uji stasioneritas dengan mengamati dan menganalisis nilai Autocorrelation Function (ACF) yang telah diubah ke dalam bentuk grafik corellogram. Data yang tidak memiliki stasioner akan dilakukan proses transformasi dan differencing dengan persamaan pada bab sebelumnya untuk mengubah data menjadi stasioner.

BAB V

IMPLEMENTASI

5.1 Deskripsi Implementasi

Berdasarkan analisis dan perancangan pada bab sebelumnya system dan model prakiraan yang dibangun pada penelitian ini adalah berbasis pada web. Sedangkan proses pembuatan model prakiraan *improved Holt Winters* diimplementasikan dalam bentuk class yang berisi kumpulan fungsi dengan menggunakan Bahasa pemrograman python. Bahasa pemrograman python memiliki library yang luas dan siap pakai untuk keperluan pemrosesan data. Perangkat keras dan perangkat lunak yang digunakan oleh peneliti dalam implementasi yaitu sebagai berikut:

1. Perangkat keras
 - a. laptop dengan spesifikasi *processor* Intel Core i5 1,8GHz (4CPUs)
 - b. RAM 4GB
 - c. kapasitas *hardisk* 500GB.
2. Perangkat lunak
 - a. sistem operasi Windows 10 Pro,
 - b. Anaconda dengan Python 3.6

5.2 Implementasi Scraping Tweet Pada Twitter

Seperti yang telah dijelaskan pada bab sebelumnya proses scraping dilakukan untuk mendapatkan jumlah data tweet demam berdarah dengan lokasi Sleman. Proses scrapping menggunakan bantuan framework scrapy (<https://scrapy.org/>) pada halaman hasil pencarian tweet yang diakses pada halaman hasil pencarian tweet [http://mobile.twitter.com/search?q=\[query\]](http://mobile.twitter.com/search?q=[query]), dimana query tersebut adalah kata kunci (keyword) yang digunakan oleh pengguna dengan ketentuan keywordnya adalah demam berdarah dan Sleman, keyword demam berdarah dan fitur place Sleman. Hasil dari keyword tersebut memiliki url yang berbeda untuk diinisiasi ke dalam constructor dari class spider. Class spider

merupakan turunan dari class scrapy spider. Inisiasi library yang digunakan dalam proses scraping dapat dilihat pada gambar 5.1.

```
1  from scrapy.spiders import CrawlSpider, Rule
2  from scrapy.selector import Selector
3  from scrapy.conf import settings
4  from scrapy import http
5  from scrapy.shell import inspect_response
6  import re
7  import json
8  import time
9  import logging
10 try:
11     from urllib import quote
12 except ImportError:
13     from urllib.parse import quote
14
15 from datetime import datetime
16
17 from TweetScraper.items import Tweet, User
```

Gambar 5. 1 Inisialisasi library yang digunakan pada proses scraping

Untuk dapat menseleksi data tweet program menggunakan selector. Data yang diimport dapat berbentuk quote text file html, php, json. Data yang ditampilkan berupa nama pengguna, tweet, dan waktu.

Kemudian pada constructor untuk membuat dan menyimpan hasil pencarian dapat dilihat pada gambar.

```
20
21 class TweetScraper(CrawlSpider):
22     name = 'TweetScraper'
23     allowed_domains = ['twitter.com']
24
25     def __init__(self, query='', lang='', crawl_user=False, top_tweet=False):
26
27         self.query = query
28         self.url = "https://twitter.com/i/search/timeline?l={}".format(lang)
29
30         if not top_tweet:
31             self.url = self.url + "&f=tweets"
32
33         self.url = self.url + "&q=%s&src=typed&max_position=%s"
34
35         self.crawl_user = crawl_user
36
37     def start_requests(self):
38         url = self.url % (quote(self.query), '')
39         yield http.Request(url, callback=self.parse_page)
40
41     def parse_page(self, response):
42         # inspect_response(response, self)
43         # handle current page
44         data = json.loads(response.body.decode("utf-8"))
45         for item in self.parse tweets block(data['items html']):
```

Gambar 5. 2 constructor untuk pencarian pada Url twitter

```

52
53     def parse_tweets_block(self, html_page):
54         page = Selector(text=html_page)
55
56         items = page.xpath('//li[@data-item-type="tweet"]/div')
57         for item in self.parse_tweet_item(items):
58             yield item
59
60     def parse_tweet_item(self, items):
61         for item in items:
62             try:
63                 tweet = Tweet()
64                 tweet['usernameTweet'] =
                    item.xpath('.//span[@class="username u-dir
                    u-textTruncate"]/b/text()').extract()[0]
65                 ID = item.xpath('.//@data-tweet-id').extract()
66                 if not ID:
67                     continue
68                 tweet['ID'] = ID[0]
69                 tweet['text'] = ' '.join(
                    item.xpath('.//div[@class="js-tweet-text-container"]/p//
                    text()').extract()).replace(' # ',
70
                    '#').replace(
71
                    ' @ ', '@')
72                 if tweet['text'] == '':
73
74                     if tweet['text'] == '':
75                         # If there is not text, we ignore the tweet
76                         continue
77                     tweet['url'] =
78                         item.xpath('.//@data-permalink-path').extract()[0]
79                     nbr_retweet =
80                         item.css('span.ProfileTweet-action--retweet >
81                         span.ProfileTweet-actionCount').xpath(
82                             '@data-tweet-stat-count').extract()
83                     if nbr_retweet:
84                         tweet['nbr_retweet'] = int(nbr_retweet[0])
85                     else:
86                         tweet['nbr_retweet'] = 0
87                     nbr_favorite =
88                         item.css('span.ProfileTweet-action--favorite >
89                         span.ProfileTweet-actionCount').xpath(
90                             '@data-tweet-stat-count').extract()
91                     if nbr_favorite:
92                         tweet['nbr_favorite'] = int(nbr_favorite[0])
93                     else:
94                         tweet['nbr_favorite'] = 0
95                     nbr_reply = item.css('span.ProfileTweet-action--reply
96                     > span.ProfileTweet-actionCount').xpath(
97                         '@data-tweet-stat-count').extract()
98                     if nbr_reply:
99                         tweet['nbr_reply'] = int(nbr_reply[0])
100                     else:
101                         tweet['nbr_reply'] = 0

```

Gambar 5. 3 Constructor untuk ekstrak data

```

92         tweet['nbr_reply'] = int(nbr_reply[0])
93     else:
94         tweet['nbr_reply'] = 0
95
96         tweet['datetime'] = datetime.fromtimestamp(int(
97
98             item.xpath('..//div[@class="stream-item-header"]/small
99                 l[@class="time"]/a/span/@data-time').extract()[
100                     0])).strftime('%Y-%m-%d %H:%M:%S')
101
102         has_cards = item.xpath('..//@data-card-type').extract()
103         if has_cards and has_cards[0] == 'photo':
104             tweet['has_image'] = True
105             tweet['images'] =
106                 item.xpath('..//*/div/@data-image-url').extract()
107         elif has_cards:
108             logger.debug('Not handle "data-card-type":\n%s' %
109                 item.xpath('.').extract()[0])
110
111         has_cards = item.xpath('..//@data-card2-type').extract()
112         if has_cards:
113             if has_cards[0] == 'animated_gif':
114                 tweet['has_video'] = True
115                 tweet['videos'] =
116                     item.xpath('..//*/source/@video-src').extract()

```

Gambar 5. 4 Constructor untuk ekstrak meta data

Gambar 5.3, 5.4 merupakan implementasi constructor yang berfungsi untuk melakukan scraping terhadap halaman pencarian. Untuk mendapatkan hasil ekstrak datanya maka perlu dilakukan crawling pada scrapy dengan menggunakan terminal anaconda.

```

1 scrapy crawl TweetScraper -a query="demam berdarah OR dbd OR dengue
  fever OR dhf OR dengue OR dengue hemorrhagic since:2010-01-01
  until:2014-12-31 near:sleman"

```

Gambar 5. 5 kode crawling pada terminal anaconda dengan fitur place

Pada gambar 5.6 data yang diambil berdasarkan keyword dengan fitur lokasi dimana tweet tersebut dibuat di wilayah sleman dengan urutan waktu pada bulan januari tahun 2010 - desember 2014. Berbeda dengan gambar 5.7 data yang diambil hanya berdasarkan keyword mengenai demam berdarah dan sleman.

```

1 scrapy crawl TweetScraper -a query="sleman AND demam berdarah OR dbd OR
  dengue fever OR dhf OR dengue OR dengue hemorrhagic since:2010-01-01
  until:2014-12-31"

```

Gambar 5. 6 Crawling pada terminal dengan menggunakan nama kabupaten Sleman

Hasil scraping dari tersebut disimpan ke dalam database untuk dilakukan perhitungan jumlah tweet per bulan. Data tweet tersebut diproses lebih lanjut untuk

menentukan tweet yang didalamnya terdapat nama lokasi dan fasilitas kesehatan masyarakat di kabupaten Sleman. Perhitungan dilakukan terhadap jumlah kejadian demam berdarah yang ada di dalam konten tweet dan jumlah tweet tentang demam berdarah.

5.3 Pemilihan Data Input

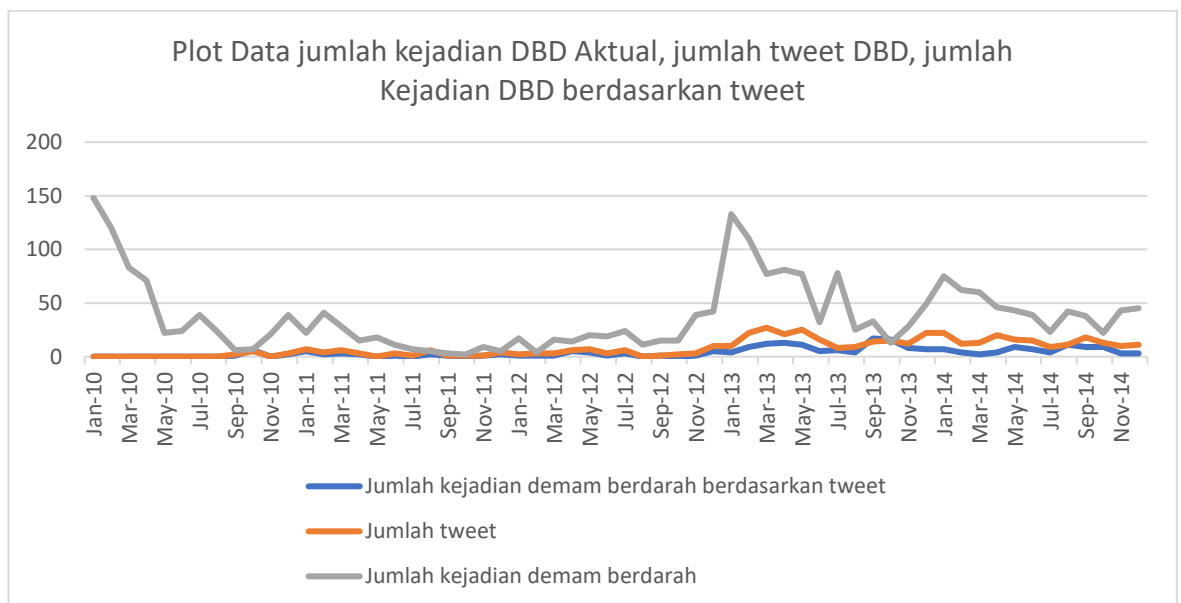
Proses scraping dan preprocessing menghasilkan dua jenis data tweet, yaitu data jumlah kejadian demam berdarah berdasarkan tweet dan data jumlah tweet tentang demam berdarah. Kedua data tersebut kemudian dihitung nilai korelasinya dengan data kejadian demam berdarah untuk menentukan data mana yang akan digunakan dalam pembuatan model. Data jumlah kejadian demam berdarah berdasarkan tweet berdasarkan tweet, data jumlah tweet dan data jumlah kejadian demam berdarah dari tahun 2010 hingga 2014 dapat dilihat pada tabel berikut.

Tabel 5. 1 data asli jumlah kejadian demam berdarah, jumlah kejadian dbd berdasarkan tweet dan jumlah tweet mengenai demam berdarah

Bulan	Jumlah Kejadian Demam berdarah	Data Hasil scraping tweet	
		Jumlah kejadian DBD berdasarkan tweet	Jumlah tweet tentang DBD
JAN 2010	148	0	0
FEB 2010	120	0	0
MAR 2010	83	0	0
APR 2010	71	0	0
MEI 2010	22	0	0
JUN 2010	24	0	0
JUL 2010	39	0	0
AGU 2010	23	0	0
SEP 2010	6	1	2
OKT 2010	7	6	5
NOV 2010	21	0	0
DES 2010	39	2	7
JAN 2011	22	5	7

FEB 2011	41	2	4
MAR 2011	28	3	6
APR 2011	15	2	3
MEI 2011	18	0	0
JUN 2011	11	1	3
JUL 2011	7	0	1
AGU 2011	5	2	6
SEP 2011	3	1	1
OKT 2011	2	0	1
NOV 2011	9	1	1
DES 2011	5	2	4
JAN 2012	17	1	2
FEB 2012	4	1	3
MAR 2012	16	1	3
APR 2012	14	5	6
MEI 2012	20	4	7
JUN 2012	19	1	3
JUL 2012	24	3	6
AGU 2012	11	0	0
SEP 2012	15	1	1
OKT 2012	15	0	2
NOV 2012	39	1	3
DES 2012	42	5	10
JAN 2013	133	4	10
FEB 2013	110	9	22
MAR 2013	77	12	27
APR 2013	81	13	21
MEI 2013	77	11	25
JUN 2013	32	5	16
JUL 2013	78	6	8
AGU 2013	25	4	9
SEP 2013	33	17	14
OKT 2013	13	16	15
NOV 2013	28	8	12

DES 2013	49	7	22
JAN 2014	75	7	22
FEB 2014	62	4	12
MAR 2014	60	2	13
APR 2014	46	4	20
MEI 2014	43	9	16
JUN 2014	39	7	15
JUL 2014	23	4	9
AGU 2014	42	11	11
SEP 2014	38	9	18
OKT 2014	22	9	13
NOV 2014	43	3	10
DES 2014	45	3	11



Gambar 5. 7 Grafik Data asli jumlah kejadian demam berdarah, jumlah kejadian demam berdarah berdasarkan twitter, dan jumlah tweet mengenai demam berdarah

Berdasarkan gambar grafik di atas, dapat diketahui bahwa jumlah kejadian demam berdarah di Kabupaten Sleman pada tahun 2010 hingga tahun 2014 fluktuatif. Kenaikan dan penurunan jumlah kejadian demam berdarah hampir sama

setiap tahunnya. Dari grafik tersebut dapat terdeteksi musiman serta pola *trend* sehingga dapat diterapkan analisis Improve *Holt Winters*.

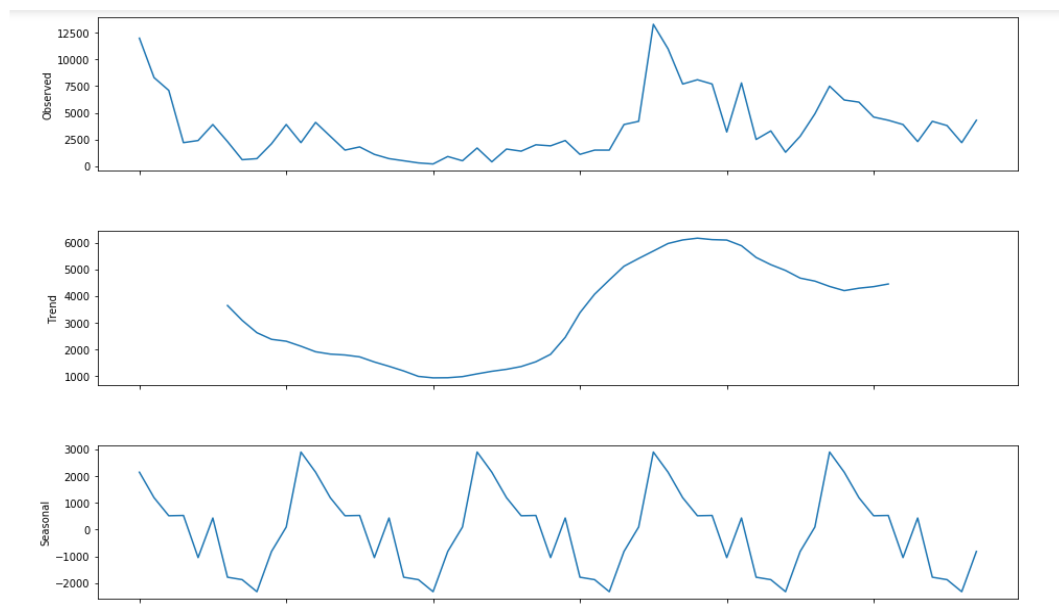
Untuk mendeteksi adanya musiman ataupun *trend* dapat diimplementasikan dengan menggunakan library statsmodel dan matplotlib.

```
10 import matplotlib.pyplot as plt
11 from statsmodels.tsa.stattools import acf
12 from statsmodels.tsa.stattools import pacf
13 from statsmodels.tsa.seasonal import seasonal_decompose
14
```

Gambar 5. 8 Constructor menggunakan matplotlib dan statsmodel

```
15 df = pd.read_csv('C:/users/irfan/desktop/forecast/JDbd.csv', index_col=0)
16 df.index.name=None
17 df.reset_index(inplace=True)
18 df.drop(df.index[58], inplace=True)
19
```

Gambar 5. 9 Kode membaca data demam berdarah



Gambar 5. 10 Grafik musiman dan trend pada jumlah data kejadian demam berdarah

Pada gambar 5.10 didapatkan grafik *trend* dan *seasonal* dengan menggunakan modul statsmodel dan library fungsi *decompose_seasonal*. pada fungsi *decompose_seasonal* data trend dan seasonal ditentukan pada seperti pada gambar berikut 5.11.

```
if extrapolate_trend == 'freq':
    extrapolate_trend = freq - 1

if extrapolate_trend > 0:
    trend = _extrapolate_trend(trend, extrapolate_trend + 1)

if model.startswith('m'):
    detrended = x / trend
else:
    detrended = x - trend

period_averages = seasonal_mean(detrended, freq)

if model.startswith('m'):
    period_averages /= np.mean(period_averages, axis=0)
else:
    period_averages -= np.mean(period_averages, axis=0)

seasonal = np.tile(period_averages.T, nobs // freq + 1).T[:nobs]
```

Gambar 5.11 Kode trend dan seasonal pada `decompose_seasonal`

Fungsi trend diperoleh dari jumlah frekuensi pada data rentang waktu keseluruhan. Apabila frekuensi menunjukkan pertumbuhan ($t + 1$) pada data sebelumnya (t) maka terjadi trend peningkatan yang digambarkan dengan plot extrapolarisasi. Sebaliknya terjadi penurunan trend apabila frekuensi mengalami penurunan dalam rentang waktu t . sedangkan seasonal diperoleh dari frekuensi pada rata-rata periodenya.

Selanjutnya untuk proses implementasinya maka data dibagi menjadi dua kelompok, yaitu data *training* dan kelompok data *testing*. Data *training* sebesar 80% dari total data digunakan untuk proses pembuatan model dan data *testing* sebesar 20% dari total digunakan untuk proses pengujian model. Dengan demikian proses pembuatan model menggunakan data *training* dari tahun 2010 hingga 2013, sedangkan proses pengujian menggunakan data *testing* tahun 2014.

Pemilihan data tweet yang diduga memiliki korelasi hubungan dengan jumlah kejadian demam berdarah yang terjadi di Kabupaten Sleman dilakukan dengan cara menghitung nilai koefisien korelasi antara kedua data. Perhitungan nilai koefisien korelasi dilakukan diantara data jumlah kejadian demam berdarah serta data jumlah kejadian demam berdasarkan tweet dengan jumlah kejadian demam berdarah serta jumlah tweet tentang demam berdarah dengan jumlah kejadian demam berdarah.

Dari data selama bulan januari 2010 hingga desember 2013, koefisien korelasi antara jumlah kejadian demam berdarah berdasarkan tweet dengan jumlah

kejadian demam berdarah sebesar 0,2069. Sedangkan koefisien korelasi antara jumlah tweet tentang demam berdarah dengan jumlah kejadian demam berdarah sebesar 0,335.

Nilai critical value dari koefisien dari data *training* (Data sebanyak 48 pengamatan) dengan 95% tingkat kepercayaan adalah sebesar 0,284. Berdasarkan critical value tersebut, maka terdapat nilai korelasi yang signifikan antara data jumlah tweet tentang demam berdarah dengan data jumlah kejadian demam berdarah Kabupaten Sleman pada tahun 2010 hingga 2013. Selanjutnya maka data jumlah tweet dapat digunakan pada proses pembuatan model *Improved Holt Winter*.

5.4 Implementasi pembuatan model

Pada penelitian ini, pembuatan model *Holt Winters* diimplementasikan ke dalam class forecast dan beberapa modul dieksekusi melalui command line interface dan jupyter. Jupyter merupakan suatu aplikasi web yang dapat digunakan untuk melakukan eksekusi kode program secara relative (live code).

5.5 Implementasi stasionerisasi data

Setelah data input ditentukan, maka dilakukan proses pengecekan stasioneritas data memanfaatkan grafik correlogram dari nilai ACF. Setelah nilai ACF didapat nilai ACF tersebut diubah menjadi grafik correlogram. Implementasi nilai ACF dan grafik correlogram dapat diolah dengan menggunakan correlogram. Data jumlah demam berdarah dan jumlah tweet mengenai demam berdarah dapat diuji stasionerisasinya pada tingkat level. Untuk mengubah data non stasioner menjadi stasioner dilakukan differencing.

```
19 differenced = series.diff(12)
20 differenced = differenced[12:]
21 differenced.to_csv('C:/users/irfan/desktop/series/jumlahtweetstasioner.csv')
22 differenced.plot()
23 plot_acf(series)
24 pyplot.show()
```

Gambar 5. 11 Kode differencing data non stasioner

Data yang telah diubah menjadi stasioner disimpan ke database untuk selanjutnya dibuat correlogram atau plot ACF pada data.

5.6 Implementasi menentukan nilai awal

Setelah data dalam keadaan stasioner proses selanjutnya adalah menentukan nilai awal. Sama halnya dengan metode penghalusan eksponensial lainnya, dibutuhkan nilai awal komponen untuk memulai perhitungan. Untuk menginisialisasi metode prakiraan Holt-Winters, diperlukan nilai awal untuk penghalusan *trend* dan indeks musiman. Untuk mendapatkan estimasi nilai awal dari indeks musiman, diperlukan setidaknya data lengkap selama satu musim. Dengan demikian nilai *trend* dan penghalusan diinisialisasi pada periode S. Nilai awal konstanta penghalusan didapatkan dengan menggunakan nilai rata-rata musim pertama. Metode yang digunakan untuk mengestimasi parameter model adalah dengan cara simulasi, yaitu mensimulasikan kisaran nilai α , β , γ pada interval (0,1) dengan cara trial dan error.

5.7 Implementasi model *Improved Holt Winters*

Model persamaan IHW diimplementasikan dalam bahasa pemrograman python. Data yang diambil untuk dilakukan prakiraan adalah data jumlah tweet mengenai demam berdarah.

```
series = Series.from_csv('C:/users/irfan/desktop/series/jumlahtweet.csv', header=0)
mon_base = Series.from_csv('C:/users/irfan/desktop/series/jumlahtweet.csv')
mon_base.columns = ['bulan', 'jumlahtweet']
mon_base = mon_base.reset_index()
```

Gambar 5. 12 kode program memasukkan data

Data jumlah tweet mengenai demam berdarah pada file jumlahtweet.csv dimasukkan ke dalam program dalam bentuk file csv dengan dibuat index column berupa bulan dan jumlah tweet.

```
25 split = int(0.8 * len(mon_base))
26 train = mon_base[:split]
27 test = mon_base[split:]
28
```

Gambar 5. 13 kode program untuk membagi data training dan data testing

Selanjutnya data dibagi menjadi dua kelompok yaitu data *training* dan data *testing* dengan pembagian data *training* sebesar 80% (sebanyak 48 data) dan data

testing 20% (sebanyak 12 data). Langkah awal dalam membuat model *improved* holt winter ini yaitu memasukkan nilai awal dengan kisaran nilai α , β , γ pada interval (0,1) dengan cara trial dan error.

```
29 alpha = 0.9
30 beta = 0.02
31 gamma = 0.01
32
```

Gambar 5. 14 Kode program nilai awal α , β , γ

Pada gambar 5.14 adalah hasil dari penentuan nilai awal setelah melakukan trial error yang merupakan nilai awal terbaik untuk menghasilkan nilai RMSE, MAPE, MSE. Selanjutnya adalah membuat algoritma pemodelan *improved* holt winter dengan memasukkan nilai awal pada data *training* dan data *testing*.

```
45 y_hat = np.array([(a_hat + b_hat) + F_fit[0]])
46 for i in range(1, len(y)):
47     a_hat = np.append(a_hat,
48                       alpha * y[i] - F_fit[i - p] + (1 - alpha) *
49                       (a_hat[i - 1] + b_hat[i - 1]))
50     b_hat = np.append(b_hat,
51                       beta * (a_hat[i] - a_hat[i - 1]) +
52                       b_hat[i - 1] * (1 - beta))
53     F_fit = np.append(F_fit,
54                      gamma * (y[i] - a_hat[i]) + F_fit[i - p] * (1 - gamma))
55     y_hat = np.append(y_hat,
56                      (a_hat[i] + b_hat[i]) + F_fit[i])
57
```

Gambar 5. 15 kode program algoritma model *improved* holt winter pada data *training*

```
62 y_test = test.jumlahtweet.values.reshape(-1, 1)
63
64 F = F_fit[-p:]
65 a_hat_test = np.array([a_hat[-1]])
66 b_hat_test = np.array([b_hat[-1]])
67 y_hat_test = np.array([(a_hat_test + b_hat_test) + F[0]])
68
69 for i in range(1, len(y_test)):
70     a_hat_test = np.append(a_hat_test,
71                           alpha * y_test[i] - F[i - p] + (1 - alpha) *
72                           (a_hat_test[i - 1] + b_hat_test[i - 1]))
73     b_hat_test = np.append(b_hat_test,
74                           beta * (a_hat_test[i] - a_hat_test[i - 1]) +
75                           b_hat_test[i - 1] * (1 - beta))
76     F = np.append(F,
77                  gamma * (y_test[i] - a_hat_test[i]) + F[i - p] * (1 - gamma))
78     y_hat_test = np.append(y_hat_test,
79                            (a_hat_test[i] + b_hat_test[i]) - F[i])
80
```

Gambar 5. 16 kode program algoritma model improved Holt Winters pada data testing

```
81 y_hat_test = y_hat_test.reshape(-1, 1)
82 mape = np.sum(np.abs(y_hat_test - y_test)) / np.sum(y_test) * 100
83 mse = np.mean(np.sum((y_hat_test - y_test)**2))
84 rmse = np.sqrt(mse)
85
```

Gambar 5. 17 kode program algoritma nilai mape, mse, rmse

```
86 plt.figure(figsize=(12,8))
87 plt.plot(test['bulan'], y_hat_test, "--", label="Forecast")
88 plt.plot(train['bulan'], y_hat, "--", label="Training Forecast")
89 plt.plot(test['bulan'], test['jumlhtweet'], label="Actual")
90 plt.plot(train['bulan'], train['jumlhtweet'], label="Training")
91 plt.title("Holt Winters Forecast Test: MAPE = %.2f\nMSE: %.2f\nRMSE: %.2f" %(mape, mse, rmse))
92 plt.legend(loc="best")
93 plt.show()
94
```

Gambar 5. 18 kode program untuk membuat plot data prakiraan data training, data testing, dan hasil error (MAPE, MSE, RMSE)

BAB VI

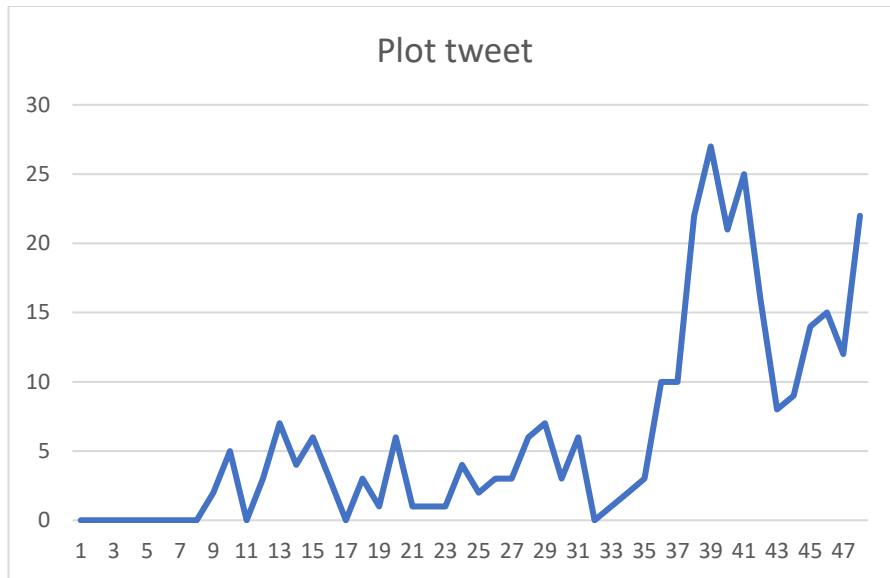
HASIL DAN PEMBAHASAN

6.1 Pengujian korelasi

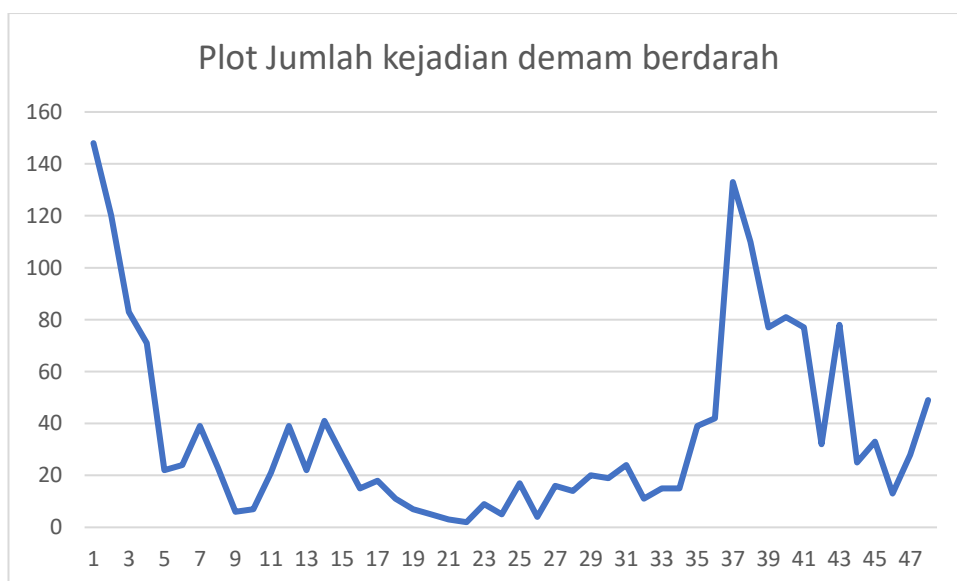
Pengujian korelasi menggunakan data jumlah kejadian demam berdarah dengan jumlah tweet tentang demam berdarah dan jumlah kejadian demam berdarah berdasarkan konten twitter. Data yang digunakan merupakan deret data dari bulan janari tahun 2010 hingga desember tahun 2013. Pemilihan data ini berdasarkan uji signifikansi terhadap nilai koefisien korelasi antara kedua deret data. Koefisien korelasi antara jumlah kejadian demam berdarah dan jumlah tweet mengenai demam berdarah adalah sebesar 0,335 lebih besar dari nilai kritis sebesar 0,284 dengan tingkat signifikansi 95% yang menandakan bahwa pertumbuhan kedua deret data tersebut saling berkaitan secara positif, yaitu apabila salah satu data mengalami peningkatan jumlah, maka data yang berkaitan juga akan mengalami peningkatan.

Korelasi antara jumlah kejadian demam berdarah dengan jumlah kejadian demam berdarah berdasarkan twitter menghasilkan nilai korelasi sebesar 0,206 lebih kecil dari nilai kritis 0,284 yang menandakan bahwa kedua deret data tersebut memiliki korelasi yang negative atau tidak memiliki korelasi. Hal ini disebabkan oleh beberapa tweet berisikan konten berita yang menyertakan jumlah kejadian demam berdarah dalam rentang waktu tertentu dan retweet dari tweet berisi konten berita. Sehingga pada data jumlah kejadian demam berdarah berdasarkan twitter tidak bisa dilakukan prakiraan.

Untuk menggambarkan kedua data yang memiliki korelasi maka dibuat plot antara jumlah kejadian demam berdarah dengan jumlah tweet mengenai demam berdarah.



Gambar 6. 1 Tampilan grafik jumlah data tweet tentang demam berdarah

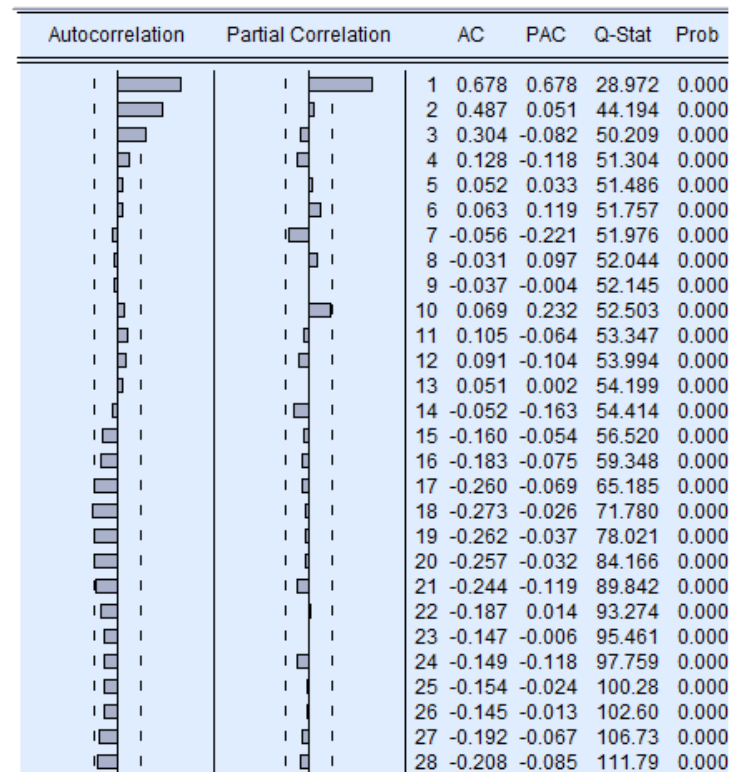


Gambar 6. 2 Tampilan grafik data jumlah kejadian demam berdarah

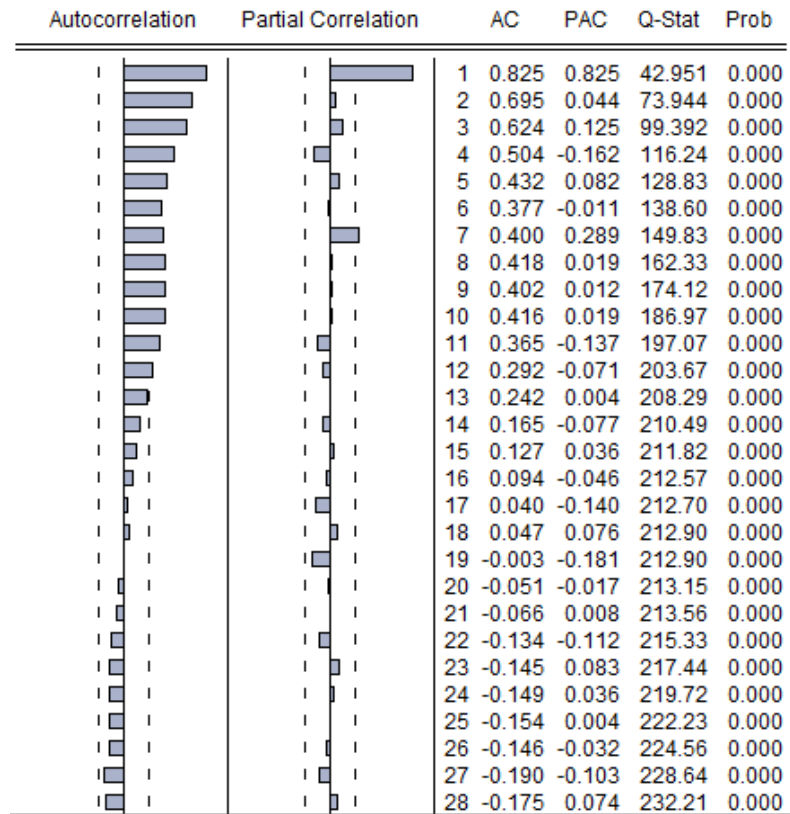
6.2 Pengujian stasioneritas data

Metode *improved Holt Winters* dapat dilakukan pada data stasioner ataupun non stasioner. Pada data jumlah kejadian demam berdarah dan jumlah tweet dilakukan pengecekan stasioneritas data dengan membuat diagram correlogram. Pengecekan stasioneritas dilakukan hanya untuk mengetahui apakah data dalam

keadaan stasioner atau non stasioner. Hasil dari correlogram dapat dilihat pada gambar 6.3 dan 6.4

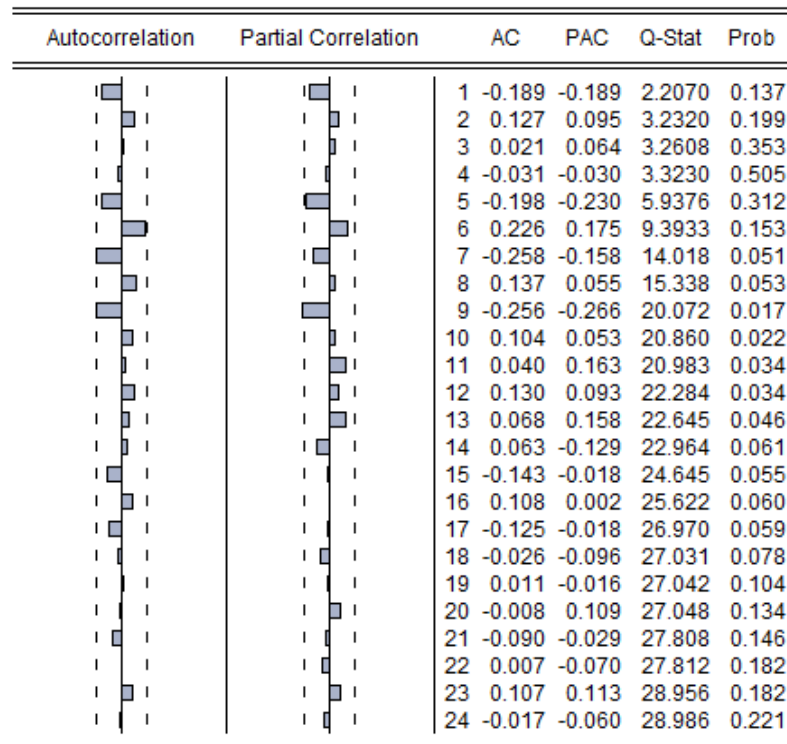


Gambar 6. 3 Tampilan correlogram data jumlah kejadian demam berdarah

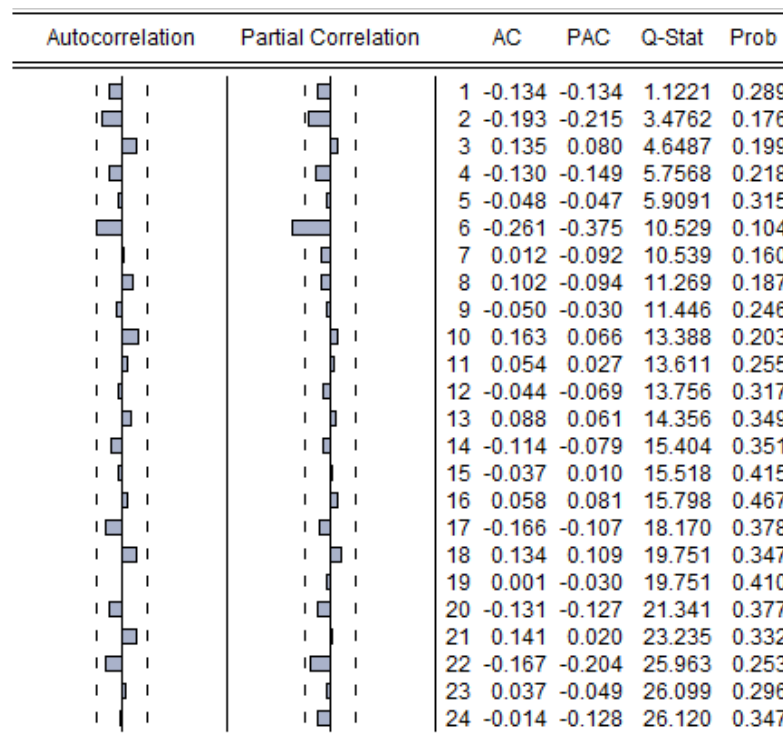


Gambar 6. 4 Grafik correlogram data jumlah tweet tentang demam berdarah

Berdasarkan gambar 6.3 dan 6.4, grafik correlogram pada jumlah kejadian demam berdarah dan jumlah tweet menghasilkan nilai ACF yang meluruh menuju nol secara eksponensial, hal ini mengindikasikan bahwa data jumlah kejadian demam berdarah dan data jumlah tweet tentang demam berdarah dalam keadaan yang tidak stasioner. Kedua data tersebut dapat diubah kedalam bentuk yang stasioner dengan cara melakukan differencing. Kemudian dilakukan pengecekan stasioner kembali.



Gambar 6. 5 Tampilan correlogram ACF differencing orde satu jumlah tweet mengenai demam berdarah



Gambar 6. 6 Tampilan correlogram ACF differencing orde satu jumlah kejadian demam berdarah

Setelah melakukan differencing maka berdasarkan gambar 6.5 dan 6.6 grafik correlogram ACF menuju nol secara cepat setelah beberapa lag. Hal ini menunjukkan bahwa data jumlah tweet tentang demam berdarah telah stasioner pada differencing orde pertama. Pada data keduanya juga tidak terdapat perbedaan nilai yang sangat signifikan dengan nol.

6.3 Pembagian Data *training* dan Data *Testing*

Setelah data dibagi menjadi dua kelompok berupa yaitu data *training* dan data *testing* hasilnya adalah sebagai berikut:

Tabel 6. 1 Data training Jumlah tweet demam berdarah

Index	bulan	umlahtwee
0	2010-01-01 00:00:00	0
1	2010-02-01 00:00:00	0
2	2010-03-01 00:00:00	0
3	2010-04-01 00:00:00	0
4	2010-05-01 00:00:00	0
5	2010-06-01 00:00:00	0
6	2010-07-01 00:00:00	0
7	2010-08-01 00:00:00	0
8	2010-09-01 00:00:00	2
9	2010-10-01 00:00:00	5
10	2010-11-01 00:00:00	0
11	2010-12-01 00:00:00	3
12	2011-01-01 00:00:00	7
13	2011-02-01 00:00:00	4
14	2011-03-01 00:00:00	6
15	2011-04-01 00:00:00	3
16	2011-05-01 00:00:00	0
17	2011-06-01 00:00:00	3
18	2011-07-01 00:00:00	1
19	2011-08-01 00:00:00	6
20	2011-09-01 00:00:00	1
21	2011-10-01 00:00:00	1

22	2011-11-01 00:00:00	1
23	2011-12-01 00:00:00	4
24	2012-01-01 00:00:00	2
25	2012-02-01 00:00:00	3
26	2012-03-01 00:00:00	3
27	2012-04-01 00:00:00	6
28	2012-05-01 00:00:00	7
29	2012-06-01 00:00:00	3
30	2012-07-01 00:00:00	6
31	2012-08-01 00:00:00	0
32	2012-09-01 00:00:00	1
33	2012-10-01 00:00:00	2
34	2012-11-01 00:00:00	3
35	2012-12-01 00:00:00	10
36	2013-01-01 00:00:00	10
37	2013-02-01 00:00:00	22
38	2013-03-01 00:00:00	27
39	2013-04-01 00:00:00	21
40	2013-05-01 00:00:00	25
41	2013-06-01 00:00:00	16
42	2013-07-01 00:00:00	8
43	2013-08-01 00:00:00	9

44	2013-09-01 00:00:00	14
45	2013-10-01 00:00:00	15
46	2013-11-01 00:00:00	12
47	2013-12-01 00:00:00	22

Pada data *training* sebesar 80% dari seluruh data terdapat 48 data jumlah tweet mengenai demam berdarah pada bulan januari tahun 2010 hingga bulan desember tahun 2013. Data *training* digunakan untuk membuat hasil prakiraan dengan model *improved Holt Winters*. Sedangkan pada data *testing* sebesar 20% dari seluruh data terdapat 12 data jumlah tweet mengenai demam berdarah. Data *testing* ini nantinya akan digunakan untuk perhitungan nilai mape, mse, dan rmse.

Tabel 6. 2 data testing jumlah tweet demam berdarah

Index	bulan	jumlahtweet
48	2014-01-01 00:00:00	22
49	2014-02-01 00:00:00	12
50	2014-03-01 00:00:00	13
51	2014-04-01 00:00:00	20
52	2014-05-01 00:00:00	16
53	2014-06-01 00:00:00	15
54	2014-07-01 00:00:00	9
55	2014-08-01 00:00:00	11
56	2014-09-01 00:00:00	18
57	2014-10-01 00:00:00	13
58	2014-11-01 00:00:00	10
59	2014-12-01 00:00:00	11

6.4 Penentuan nilai awal

Penentan nilai awal menggunakan alpha, beta dan gamma dengan cara *trial - error* terbaik untuk menghasilkan nilai MAPE, MSE dan RMSE. Hasil dari trial-error ditampilkan pada tabel 6.3, 6.4.

Tabel 6. 3 Trial dan error nilai α , β , γ pada data tweet

No	Alpha	Beta	gamma	MAPE	MSE	RMSE
1	0,1	0,1	0,1	1083,51	312632,36	559,14
2	0,5	0,1	0,1	51	912,53	30,21
3	0,6	0,1	0,1	39,52	548,11	23,41
4	0,8	0,1	0,1	22,73	163,97	12,81
5	0,9	0,1	0,1	16,66	78,02	8,83
6	0,9	0,2	0,1	19,07	103,56	10,14
7	0,9	0,08	0,1	15,19	66,31	8,14
8	0,9	0,02	0,1	6,61	18,10	4,25
9	0,9	0,01	0,1	8,76	25,65	5,06
10	0,9	0,02	0,05	4,72	8,20	2,86
11	0,9	0,02	0,01	4,01	5,28	2,30

Berdasarkan tabel 6.3, pada data jumlah tweet mengenai demam berdarah dengan menggunakan nilai awal alpha 0.9, beta 0.02, dan gamma 0.01 menghasilkan nilai MAPE, MSE, dan RMSE terkecil yaitu sebesar MAPE 4.01, MSE 5.28, RMSE 2,30.

Tabel 6. 4 Trial dan error nilai α , β , γ pada data jumlah kejadian demam berdarah

No	alpha	beta	gamma	MAPE	MSE	RMSE
1	0,9	0,02	0,01	7,2	805,41	28,38
2	0,9	0,05	0,01	6,95	796,11	28,22
3	0,95	0,05	0,01	5,94	715,63	26,75

Berdasarkan table 6.4 pada data jumlah kejadian demam berdarah dengan menggunakan nilai awal alpha 0.95, beta 0.05, dan gamma 0.01 menghasilkan nilai MAPE, MSE, dan RMSE terkecil yaitu sebesar MAPE 5.94, MSE 715,63, dan RMSE 26,75.

6.5 Pengujian prakiraan demam berdarah

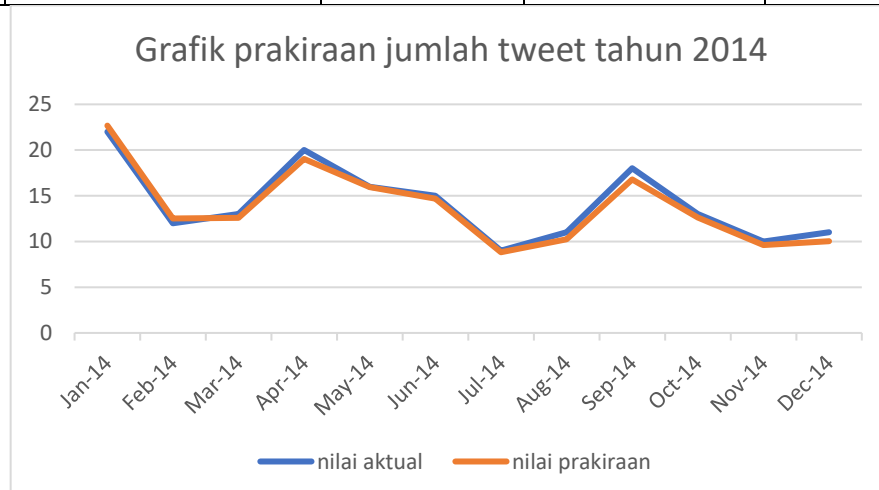
Pengujian ini menggunakan data training Januari 2010 hingga desember 2013 untuk melakukan prakiraan jumlah kejadian demam berdarah pada tahun 2014 dan kemudian dibandingkan dengan data testing untuk dicari nilai MAPE, MSE, RMSE.

6.6 Pengujian prakiraan demam berdarah pada data jumlah tweet

Hasil perhitungan prakiraan jumlah tweet mengenai demam berdarah tahun 2014 dengan nilai awal alpha 0.9, nilai beta 0.02 dan nilai gamma 0.01 seperti terlihat pada table 6.5. Hasil grafik prakiraannya ditampilkan pada gambar 6.7

Tabel 6. 5 Hasil pengujian data tweet demam berdarah tahun 2014

No	Bulan	Data Aktual jumlah tweet	Data prakiraan jumlah tweet	APE %
1	JAN 2014	22	23	4,55
2	FEB 2014	12	13	8,33
3	MAR 2014	13	13	0,00
4	APRIL 2014	20	19	5,00
5	MEI 2014	16	16	0,00
6	JUNI 2014	15	15	0,00
7	JULI 2014	9	9	0,00
8	AGUSTUS 2014	11	10	9,09
9	SEPTEMBER 2014	18	17	5,56
10	OKTOBER 2014	13	13	0,00
11	NOVEMBER 2014	10	10	0,00
12	DESEMBER 2014	11	10	9,09



Gambar 6.7 Grafik nilai error pada jumlah tweet tahun 2014

Berdasarkan tabel 6.5 nilai error tertinggi terjadi pada bulan desember 2014 dengan nilai APE sebesar 8,92%. Sedangkan untuk nilai error terendah terjadi pada

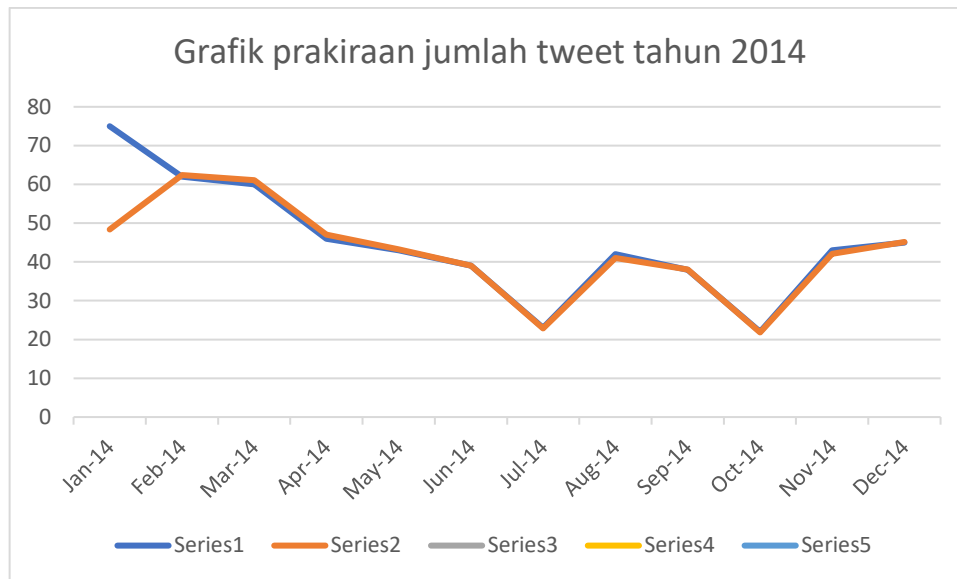
bulan mei 2014 dengan nilai APE sebesar 0,36%. Untuk penghitungan mape yang dihasilkan dari bulan januari 2014 hingga desember 2014 adalah sebesar 4,01 %.

6.7 Pengujian prakiraan demam berdarah pada data jumlah kejadian demam berdarah

Hasil perhitungan prakiraan pada data jumlah kejadian demam berdarah tahun 2014 dengan nilai awal alpha 0.95, nilai beta 0.05 dan nilai gamma 0.01 seperti terlihat pada table 6.6. Hasil grafik prakiraannya ditampilkan pada gambar 6.7

Tabel 6. 5 Hasil pengujian data tweet demam berdarah tahun 2014

No	Bulan	Data Aktual jumlah tweet	Data prakiraan jumlah demam berdarah	APE (%)
1	JAN 2014	75	48.3312	36,00
2	FEB 2014	62	62.4085	0,00
3	MAR 2014	60	61.0708	1,67
4	APRIL 2014	46	47.0401	2,17
5	MEI 2014	43	43.288	0,00
6	JUNI 2014	39	39.0543	0,00
7	JULI 2014	23	22.8551	0,00
8	AGUSTUS 2014	42	41.012	2,38
9	SEPTEMBER 2014	38	38.0358	0,00
10	OKTOBER 2014	22	21.829	0,00
11	NOVEMBER 2014	43	42.0674	2,33
12	DESEMBER 2014	45	45.1597	0,00

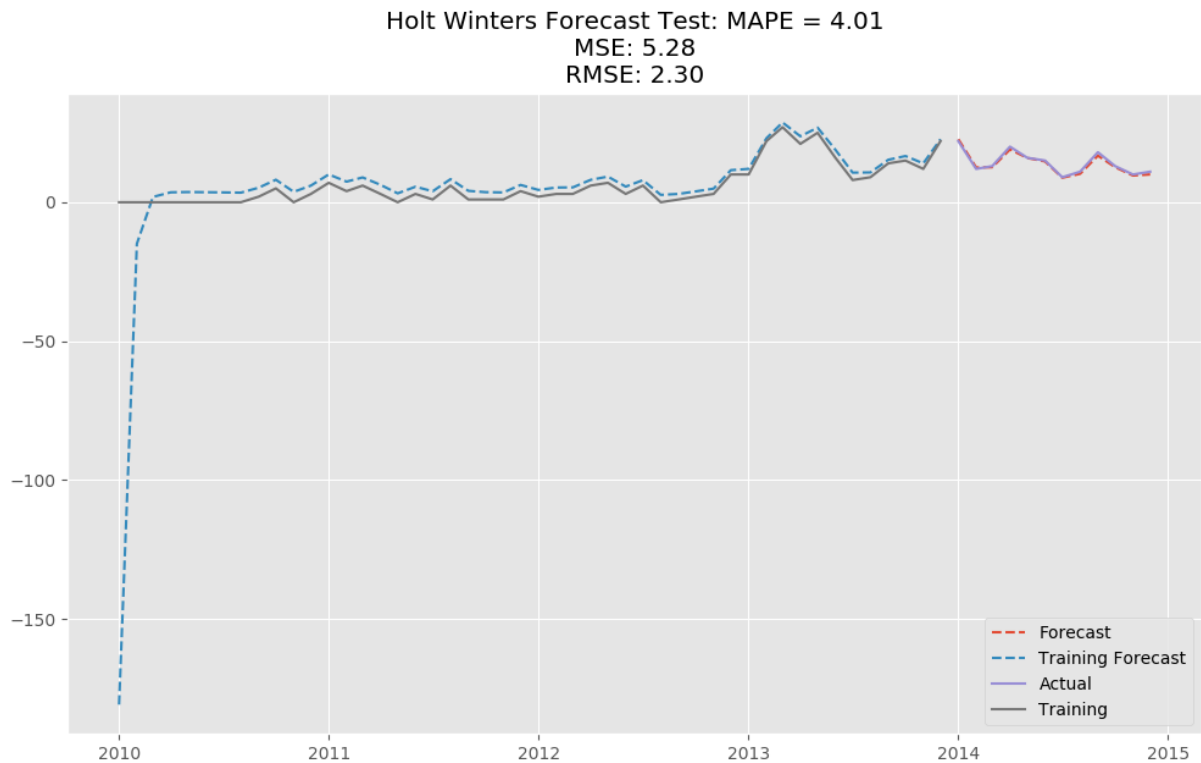


Gambar 6. 8 Grafik nilai error pada data tweet tahun 2014

Berdasarkan tabel 6.5 nilai error tertinggi terjadi pada bulan januari 2014 dengan nilai APE sebesar 36%. Sedangkan untuk nilai error terendah terjadi pada bulan mei, juni, juli, September, oktober 2014 dengan nilai APE sebesar 0,0%. Untuk penghitungan mape yang dihasilkan dari bulan januari 2014 hingga desember 2014 adalah sebesar 5,94 %.

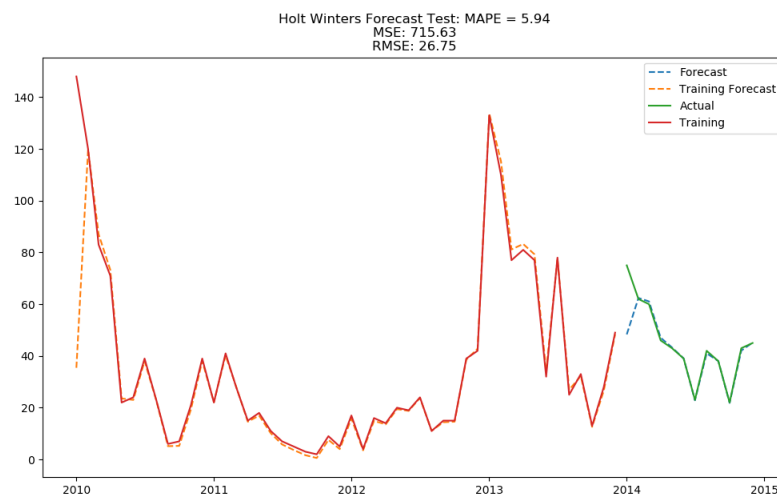
6.8 Analisis prakiraan Improve Holt Winters

Hasil perhitungan prakiraan jumlah tweet mengenai demam berdarah dengan menggunakan nilai α 0.9, β 0,02, γ 0,01 yang menghasilkan nilai MAPE 4,01 % untuk prakiraan mulai januari 2014 hingga desember 2014 sehingga persamaan pemulusan level adalah $L_t = 0,9Y_t - S_{t-c} + (1 - \alpha)(L_{t-1} + T_{t-1})$ konstanta pemulusan untuk pola musiman $\beta = 0,02$ sehingga persamaan pemulusan pola musiman adalah $T_t = 0,2(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$, konstanta pemulusan untuk pola *trend* $\gamma = 0,1$ sehingga persamaan pemulusan pola *trend* adalah $S_t = 0,1(Y_t - L_t) + (1 - \gamma)S_{t-c}$ dan m oleh prakiraan m periode ke depan adalah $F_{t+1} = L_t + b_t + S_{t+1-s}$ Hasil prakiraan jumlah tweet mengenai demam berdarah sudah mendekati dengan nilai aslinya. Grafik hasil prakiraan jumlah tweet mengenai demam berdarah pada gambar 6.8.



Gambar 6. 9 Grafik hasil prakiraan jumlah tweet mengenai demam berdarah

Selanjutnya untuk hasil perhitungan prakiraan jumlah jumlah kejadian demam berdarah dengan menggunakan nilai α 0.95, β 0,05, γ 0,01 yang menghasilkan nilai MAPE 5,94 % untuk prakiraan mulai januari 2014 hingga desember 2014 sebesar. Hasil prakiraan jumlah tweet mengenai demam berdarah sudah mendekati dengan nilai aslinya. Grafik hasil prakiraan jumlah kejadian demam berdarah pada gambar 6.9.



Gambar 6. 7 Grafik Jumlah kejadian demam berdarah di Kabupaten Sleman

BAB VII

KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan hasil penelitian, diperoleh bahwa prakiraan menggunakan metode *Improved Holt-Winters* dapat diambil kesimpulan sebagai berikut:

1. Jumlah tweet mengenai demam berdarah di Kabupaten Sleman memiliki korelasi dengan jumlah kejadian demam berdarah di Kabupaten Sleman. Akan tetapi untuk jumlah kejadian demam berdarah berdasarkan konten data tweet tidak memiliki korelasi dengan jumlah kejadian demam berdarah.
2. Prakiraan dengan menggunakan metode *Improved Holt Winters* pada data jumlah tweet mengenai demam berdarah menghasilkan konstanta data asli alpha 0,9, beta 0,02, gamma 0,01. Prakiraan pada data jumlah kejadian demam berdarah menghasilkan konstanta data asli alpha 0,95, beta 0,05, gamma 0,01.
3. Prakiraan jumlah tweet mengenai kejadian demam berdarah dengan metode Improve Holt-Winters menghasilkan nilai Mean Square Error (MSE) yaitu 5,28 ,nilai Mean Absolute Percentage Error (MAPE) yaitu 4,01 %, dan nilai RMSE yaitu 2,30. Prakiraan jumlah tweet mengenai kejadian demam berdarah dengan metode Improve Holt-Winters menghasilkan nilai Mean Square Error (MSE) yaitu 715,63 ,nilai Mean Absolute Percentage Error (MAPE) yaitu 5,94%, dan nilai RMSE yaitu 26,75.
4. Nilai akurasi dengan menggunakan metode *Improved Holt Winters* pada data tweet mengenai demam berdarah dan data jumlah kejadian demam berdarah sudah baik, dimana data hasil prakiraan mendekati nilai sebenarnya.

7.2 Saran

Pada penelitian selanjutnya, diharapkan deret data yang lebih panjang guna menganalisis lebih lanjut korelasi antara kedua deret data. Selain itu, dapat menggunakan metode lainnya seperti aditif holt winter dan multiplicative holt winter untuk melihat perbandingan nilai akurasi yang lebih baik.

DAFTAR PUSTAKA

- Akbar, V.H., 2018, Prakiraan Kejadian Demam Berdarah Berdasarkan Data Twitter Dengan Menggunakan Metode Marima Fungsi Transfer, Tesis, Universitas Gadjah Mada
- Chatfield, C., 1978. The Holt-Winters Forecasting Procedure. *Journal of the Royal Statistical Society. Series C* 27, 264-279.
- Deossa, P., Marquez, A., dan Espinosa, J., 2014., Integration of Economic MPC, energy load and price with Holt Winters Model, *Proceedings 2014 IEEE PES Transmission & Distribution Conference and Exposition - Latin America*, 10-13 September 2014, 1-5
- Djumaty, B.L., Tanamah, A.R., Wowor, A.D., 2013, Analisis Perbandingan Metode Holt-Winters, Single Exponential Smoothing Dan Polinomial Newton Dalam Meramalkan Data Produksi Ubi Kayu, *Prosiding Matematika, Sains, dan Teknologi Informasi (MSTI)*, 14 Juni 2013, 27-31
- Heizer, J., Render, B., 2009. *Manajemen Operasi*, Buku 1 Edisi 9. Jakarta: Salemba Empat.
- Herjanto, Eddy. 2008. *Manajemen Operasi*, Edisi Ketiga. Jakarta: Grasindo.
- Hutomo, P.D., 2016, Model Prakiraan Kebutuhan Persediaan Benih Menggunakan Metode Holt Winter – ES, Tesis, Universitas Gadjah Mada
- Kalekar, P.S., 2004. *Time series Forecasting using Holt-Winter Exponential Smoothing*. India: Kanwal Rekhi School of Information Technology.
- Knowlton K, Solomon G, Rotkin-Ellman M, Pitch F. 2009. *Mosquito-Borne Dengue Fever Threat Spreading in the Americas*. New York.
- Kusriastuti R. 2005. *Kebijaksanaan Penanggulangan Demam Berdarah Dengue di Indonesia*. Jakarta. Depkes R.I.
- Makridakis, S., S.C. Wheelwright & V.E. McGee. Terjemahan U.S. Andriyanto & A. Basith. 1995. *Metode dan Aplikasi Prakiraan*, Jilid 1, Edisi kedua. Penerbit Erlangga. Jakarta. 532 pp.
- Sari CIN. 2005. *Pengaruh Lingkungan Terhadap Perkembangan Penyakit Malaria Dan Demam Berdarah Dengue*. Bogor: IPB;
- Sumayang, Lalu. 2003. *Dasar-Dasar Manajemen Produksi dan Operasi*. Salemba Empat, Jakarta.
- Tratar, L.B, 2014. *Improved Holt-Winters method of overnight stays of tourists in republic of slovenia*. *Economic and Business* 16, 5-17.



UNIVERSITAS
GADJAH MADA

**METODE IMPROVED HOLT WINTERS PADA PRAKIRAAN KEJADIAN DEMAM BERDARAH DI
KABUPATEN SLEMAN**

MOHAMMAD IRFAN RADEN, Agus Sihabuddin, S.Si., M.Kom., Dr

Universitas Gadjah Mada, 2018 | Diunduh dari <http://etd.repository.ugm.ac.id/>

Utami, Tiani. W., Darsyah, Moh. Yamin. 2015. *Prakiraan Data Saham dengan
Metode Winter's*. Jurnal Statistika, Vol. 3, No. 2.